An International ICT R&D Journal Sponsored by ZTE Corporation

ISSN 1673-5188
CN 34-1294/ TN

ZTE COMMUNICATIONS

中兴通讯技术(英文版)

September 2021, Vol. 19 No. 3

Special Topic: Wireless Intelligence for Behavior Sensing and Recognition

The 9th Editorial Board of ZTE Communications

CONTENTS

ZTE COMMUNICATIONS September 2021 Vol. 19 No. 3 (Issue 75)

Special Topic

Wireless Intelligence for Behavior Sensing and Recognition

CONTENTS

ZTE COMMUNICATIONS September 2021 Vol. 19 No. 3 (Issue 75)

Serial parameters:CN 34-1294/TN*2003*q*16*94*en*P*¥20.00*2000*11*2021-09

MA Jianhua, GUO Bin

Editorial: Special Topic on Wireless Intelligence for Behavior Sensing and Recognition

Guest Editor

MA Jianhua is a professor with Faculty of Computer and Information Sciences, Hosei University, Japan. He served as the Chair of Digital Media Department of Hosei University in 2011 – 2012. His research interests include networking, pervasive computing, social computing, wearable technology, IoT, smart things, etc. He first proposed Ubiquitous Intelligence (UI) towards Smart World (SW), which he envisioned in 2004 and was featured in the *European ID People Magazine* in 2005. He is a founder of three IEEE technical committees (Cybermatics, Smart World and Hyper Intelligence).

Guest Editor

GUO Bin is a professor with Northwestern Polytechnical University, China. He received his Ph. D. degree in computer science from Keio University, Japan in 2009. His current research interests include ubiquitous computing, mobile crowd sensing, and urban computing. He has served as an associate editor of *IEEE Communications Magazine*, *IEEE Transactions on Human-Machine-Systems*, *ACM IMWUT*, and so on. He received the support of the National Science Fund for Distinguished Young Scholars in 2020. He is a senior member of IEEE and CCF.

Behavior sensing and recognition is the core technology that enables a wide variety of applications, e. g., healthcare, smart homes and public safety. Traditional approaches mainly use cameras and mobile/wearable sensors. However, all these approaches have certain disadvantages. For example, camera-based approaches have the limitations of requiring line of sight with enough lighting and breaching human privacy potentially. Mobile/wearable sensor-based approaches are inconvenient sometimes as users have to always wear certain sensor-rich devices. Recently, intelligence-based wireless techniques (e. g., Wi-Fi, ultra-wideband (UWB), mmWave, 5G, acoustic and light communications) are emerging as novel approaches to behavior sensing and recognition. Compared with traditional approaches, wireless signal-based approaches have a set of advantages; for example, they do not require lighting, provide better coverage as they can operate through certain barriers, preserve user privacy, and do not require users to carry any devices as they rely on the wireless signals reflected by humans, and even animals. As a result, the recognition of quite a number of behaviors, which is difficult based on traditional approaches, has now become possible, e. g., the recognition of fine-grained movements (such as gesture and lip language), keystrokes, drawings and gait patterns, behavior-based authentication, intrusion detection, presence monitoring, vital signals (such as breathing rate and heart rate), etc.

This special issue of *ZTE Communications* provides the opportunity for technical researchers and product developers to review and discuss the state-of-the-art and trends of wireless intelligence for behavior sensing and recognition techniques and systems. We have six invited papers covering different topics of wireless intelligence, including key techniques, applications and the literature review.

The first article, "HiddenTag: Enabling Person Identification Without Privacy Exposure", by QIU et al. summarizes the research of person identification and introduces a novel approach without privacy exposure by leveraging acoustic and Wi-Fi channel state information (CSI) sensing. The authors utilize smartphones to identify different users via profiling indoor activities by inaudible sound. The proposed approach generates specific sound signals and acoustic features for constructing the human body signatures. The authors also use CSI sensing to trigger the system of acoustic sensing. Based upon the implemented prototype and evaluation results, this introduced approach can distinguish multiple people in 10 – 15 s with 95.1% accuracy and maintain 84% – 90% online accuracy.

The second article, "Device-Free In-Air Gesture Recognition Based on RFID Tag Array", by WU et al. introduces a novel system to recognize both dynamic gestures and static gestures via a device-free approach based on an RFID tag array, which can be used for human computer interaction. Particularly, the authors carefully extract the temporal-spatial feature of received signals to distinguish dynamic gestures and static gestures. Moreover, they design a convolutional neural network and Long Short-Term Memory (CNN-LSTM) combined framework to recognize the gestures by considering the temporal and spatial features together. The authors have implemented a system prototype and conducted the extensive experiments that show over 90% accuracy in the gesture recognition.

The third article, "Indoor Environment and Human Sensing via Millimeter Wave Radio: A Review", by LIU et al. is a systemic review of mmWave sensing. The authors summarize the

DOI: 10.12142/ZTECOM.202103001
J. H. Ma and B. Guo, "Editorial: special topic on wireless intelligence for behavior sensing and recognition," *ZTE Communications*, vol. 19, no. 3, pp. 01 – 02, Sept. 2021. doi: 10.12142/ZTECOM.202103001.

prior works with a focus on two typical mmWave sensing tasks: environment reconstruction and human behavior recognition. Through these works, the unique advantages of mmWave, i.e., fine-grained sensing resolution and robust sensing for multiple co-existing objects, are highlighted. Finally, they use a table to further classify the works and discuss the potential directions for future work, which may inspire new ideas in the mmWave sensing field.

The fourth article, "Using UAV to Detect Truth for Clean Data Collection in Sensor-Cloud Systems", by LI et al. presents the research about a flexible and convenient data collection method in sensor-cloud systems and classifies the methodologies into two categories: the mobile vehicle (MV) based and mobile edge users (MEUs) based. In view of the latter, the authors further elaborate its pros and cons. As a conclusion, the authors point out that the important issue in such applications is the security of data collection which mainly comes from the security of MEUs and sensing devices, and put forward a scheme that uses unmanned aerial vehicles to detect the truth of sensing devices and MEUs (UAV-DT) to ensure the security of the data source and transmission in the sensor-cloud systems.

The fifth article, "Artificial Intelligence Rehabilitation Evaluation and Training System for Degeneration of Joint Disease", by LIU et al. studies the degeneration of joint disease using AI technology. To find effective, convenient and inexpensive therapies, the authors magnificently combine Traditional Chinese Medicine (Daoyin) with AI to design a rehabilitation assessment system. Several technologies are incorporated, including key-point detection, posture estimation, heart rate detection and deriving respiration from electrocardiogram (ECG) signals. The system is embedded into a portable wireless device for beneficial use in daily life. Their experiments show the effectiveness of the proposed system.

The last article, "A Survey of Intelligent Sensing Technologies in Autonomous Driving", by SHAO et al. presents a thorough review of intelligent sensing technologies for autonomous driving. Autonomous driving is a promising technique for the future of vehicles. The authors investigate the impact of sensing technologies on car architecture shaping and explore the generic intelligent sensing algorithms in autonomous driving. They also discuss the future trends of intelligent sensing in autonomous driving, such as end-to-end policy decision models.

We are grateful to all the authors for their valuable contributions and to all the reviewers for their valuable and constructive feedback. We hope that this special issue is interesting and useful to all readers.

HiddenTag: Enabling Person Identification Without Privacy Exposure

QIU Chen[1], DAI Tao[2], GUO Bin[1], YU Zhiwen[1], LIU Sicong[1]

(1. Northwestern Polytechnical University, Xi'an 710072, China;
2. Chang'an University, Xi'an 710064, China)

Abstract: Person identification is the key to enable personalized services in smart homes, including the smart voice assistant, augmented reality, and targeted advertisement. Although research in the past decades in person identification has brought technologies with high accuracy, existing solutions either require explicit user interaction or rely on images and video processing, and thus suffer from cost and privacy limitations. In this paper, we introduce a device-free personal identification system – HiddenTag, which utilizes smartphones to identify different users via profiling indoor activities with inaudible sound and channel state information (CSI). HiddenTag sends inaudible sound and senses its diffraction and multi-path reflection using smartphones. Based upon the multi-path effects and human body absorption, we design suitable sound signals and acoustic features for constructing the human body signatures. In addition, we use CSI to trigger the system of acoustic sensing. Extensive experiments indicate that HiddenTag can distinguish multi-person in 10 – 15 s with 95.1% accuracy. We implement a prototype of HiddenTag with an online system by Android smartphones and maintain 84% – 90% online accuracy.

Keywords: person identification; acoustic sensing; CSI; smart home

DOI: 10.12142/ZTECOM.202103002

https://kns.cnki.net/kcms/detail/34.1294.TN.20210819.1211.001.html, published online August 19, 2021

Manuscript received: 2021-06-15

Citation (IEEE Format): C. Qiu, T. Dai, B. Guo, et al., "HiddenTag: enabling person identification without privacy exposure," *ZTE Communications*, vol. 19, no. 3, pp. 03 – 12, Sept. 2021. doi: 10.12142/ZTECOM.202103002.

1 Introduction

Numerous applications are enabled with the realization of smart living environments and Internet of Things (IoT). Person identification is essential for smart home services, such as real-time recommendations on TV and human-machine interactions in video games[1 – 2]. Based on the personalized applications, users can obtain desirable services pervasively[3 – 5]. Therefore, an accurate, light-training and real-time person identification approach is needed.

Existing person identification mechanisms have many limitations that prevent them from being adopted pervasively. One of the biggest limitations is that they are often intrusive to users' privacy. Camera and computer vision based solutions can recognize different persons effectively, but unfortunately users' faces, gestures and other information may be exposed to others[6 – 8]. For example, monitoring a person's face when she/he is sitting on the sofa and walking in the hallway may cause privacy concerns.

Moreover, many person identification methods need a user to do extra work to help recognize the user. For example, smart speakers such as Amazon Echo and Google Home can identify users by their voiceprints. This approach requires users to speak to trigger recognition[9], which is a reactive solution. We therefore ask the question: can we identify users with-

out asking them to do any additional work and preserve their privacy?

To this end, we introduce HiddenTag, a new device-free person recognition system without pre-installed infrastructure or additional sensors. Only with the built-in smartphone, as shown in Fig. 1, the acoustic sender provides the high frequency sound (18 – 21 kHz) from which people cannot hear. When a user enters the smart environment, the user can keep the normal activities, such as walking, standing, and other types of human activities, and all these can be profiled by an acoustic receiver on off-the-shelf mobile devices. Based on the multi-path effects and bodies absorption in experimental scenarios, HiddenTag constructs the acoustic signatures for different persons. We design high frequency based features and enrich these features by utilizing sweeping and multi-tone techniques. Besides, we explore channel state information (CSI) to detect the human body and trigger the person identification approach. By leveraging machine learning models, our system recognizes different users efficiently in smart home environments. Case studies show the online identification can reach 90.2% accuracy and the corresponding offline group achieves 96.0% accuracy.

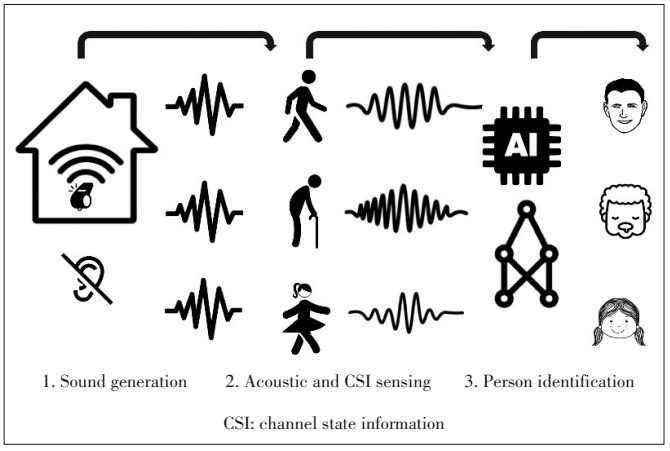
In addition, we pre-trained some common types of noises in the learning model and made HiddenTag more robust to noises. According to the collected historical data and temporal correlation feature, our system further calibrates some errors by using the proposed prediction model. Related SmartThings such as smart LED bulbs and media players are able to provide personalized services based on classification results. This paper makes the following contributions: 1) To the best of our knowledge, HiddenTag is the first high frequency (18 – 21 kHz) acoustic sensing solution for person identification; 2) HiddenTag has introduced sweeping and multi-tone techniques to enrich feature spaces. Adding common types of noises makes HiddenTag robust to real environment noises; 3) HiddenTag is implemented both online and offline. The proposed offline system achieves 96.2% accuracy with four users and the corresponding online system reaches 90.2% accuracy.

The rest of the paper is organized as follows. Section 2 introduces the system design. Experiments and simulations are shown in Section 3. Section 4 further discusses the evaluations. We provide related work and comparison in Section 5. Conclusions and future work are in Section 6.

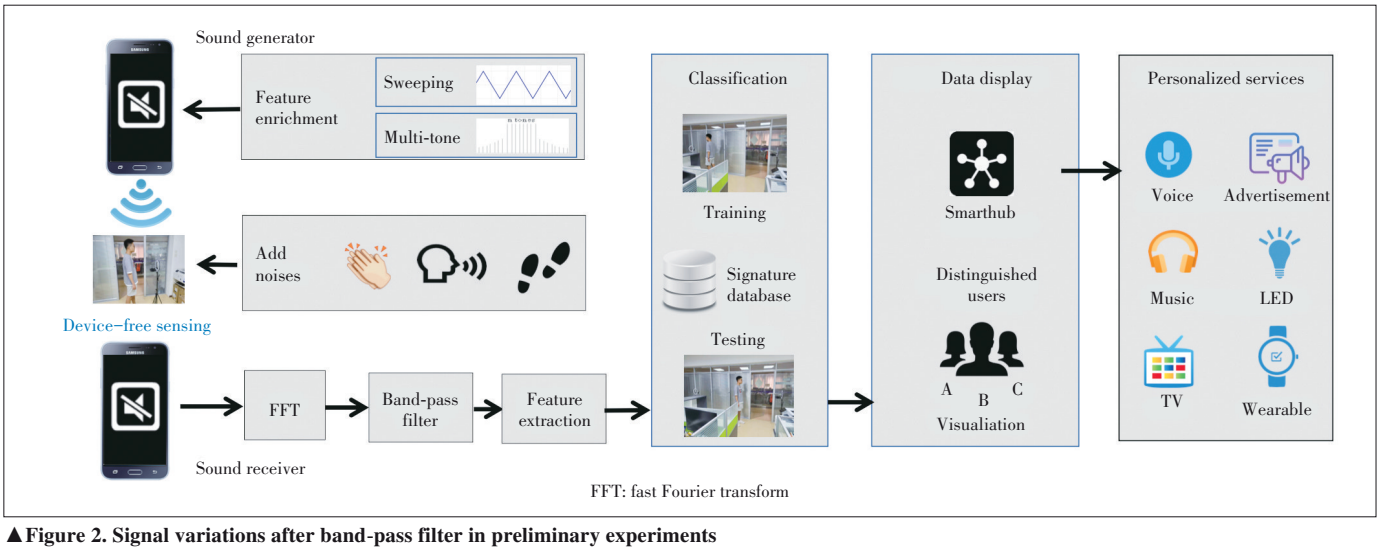
2 System Design

2.1 Overview of Our Approach

• HiddenTag employs existing smartphones without complex hardware modifications. The procedure of HiddenTag is illustrated in Fig. 2, where HiddenTag is a device-free system based on acoustic sensing. Off-the-shelf smartphones send high frequency (18 – 21 kHz) sound signals via speakers. The sound emitter can select one from the following three models: single-tone model, multi-tone model, and sweeping model. Af-



▲Figure 1. Concept view of HiddenTag



▲Figure 2. Signal variations after band-pass filter in preliminary experiments

fected by the user's indoor activities, the acoustic signals are changed in the propagation channels. Receivers of HiddenTag sense the varied acoustic information by microphones. By leveraging a band-pass filter, our system only processes the sound in the frequency range of 18 kHz and 21 kHz, which cannot be heard by human beings.

• In the training phase, based on feature engineering, the system trains different users and labels corresponding data. In the testing phase, HiddenTag adopts a band-pass filter to reduce noises. Classifiers based on the machine learning model are built to identify different users in smart home environments. Further, HiddenTag implements various personalized services (smart LED, music, smart TV, etc.) relying on the results of person identification.

2.2 Preliminaries

The fundamental idea of HiddenTag is that users can be recognized by their acoustic signatures. When the recorder receives acoustic signals, different degradations occur at different frequencies due to frequency selective fading. Additionally, multi-path effects, diffraction, and reflection also impact the acoustic signals. Once users walk in an indoor environment, walking activities and human bodies cause unique multi-path effects and body absorption. Fig. 3 illustrates the causes of such attenuation.

To verify this perspective, we conduct preliminary experiments in an empty room, the size of which is 5 m×5 m. We employ two Huawei Mate 30 smartphones as the acoustic sender and the receiver. The heights of the sender and receiver are 75 cm. The acoustic sender generates sounds frequencies from 18 kHz to 21 kHz. The sampling frequency is 48 kHz. In sweeping mode, the sweeping period is 0.02 s.

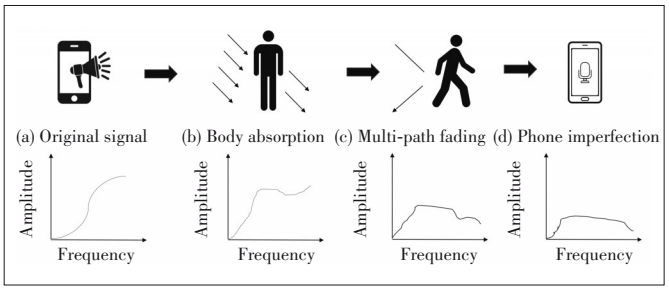
We record acoustic data for three control groups. In the beginning, the experimental room is empty. In the following two groups, User 1 and User 2 enter the room and walk around the sender and receiver.

As shown in Fig. 4, the x-axis indicates the time of the experiment and the y axis refers to the range of sound frequency. We conclude that for each control group, the power distributions on the different spectrums are different, and less power is distributed on the spectrum when there are users compared with the empty group.

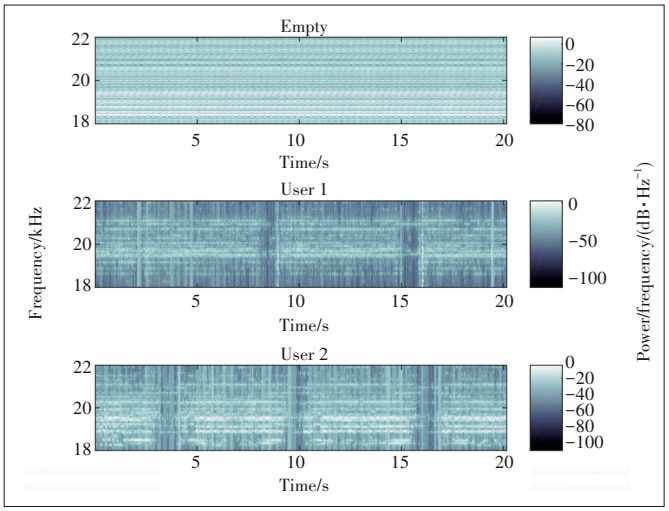
Therefore we design an approach that leverages the different signatures to identify users in the following.

2.3 Sending Inaudible Sound

As we explore the inaudible sound that can be generated from built-in smartphones, choosing parameters for sound generation is a challenge. According to our experimental results and literature, only generating single-tone acoustic signals between 18 kHz and 21 kHz is difficult to support accurate person identification because of the limited information. The feature space is constrained by a fixed sending frequency.



▲ **Figure 3. Signal variations after band-pass filter in preliminary experiments**



▲ **Figure 4. Time-spectral comparison for different ambient mediums**

$$S(t) = A \cdot \sin(2\pi f_0(t) \cdot t) . \quad (1)$$

As shown in Eq. (1), $S(t)$ is the amplitude value of the sin wave, and f_0 is the frequency of the sound wave that we send. If f_0 is a fixed value, the value of $S(t)$ can only reflect the wave at a certain frequency, which means that we do not utilize the inaudible sound on smartphones efficiently. Therefore, we introduce two other models, namely the sweeping model and the multi-tone model, to improve the identification accuracy by enriching feature space.

1) Sweeping model: We propose periodic frequency sweeping from 18 kHz to 21 kHz and set sampling frequency as 48 kHz. Consequently, the frequencies change quickly and cover all the frequencies from 18 kHz to 21 kHz in a short time period. This selection makes the generated sound inaudible for most people, but enriching the feature space for acoustic sensing.

$$f_0(t) = f_l + (f_u - f_l) \times \Delta t / t_d . \quad (2)$$

Different from Eq. (1) where f_0 is a fixed value, the value of f_0 is determined by Eq. (2) in the sweeping model. f_u and f_l indicate the upper bound and lower bound of the sweeping range. t_d is the duration of each sweeping period. Δt is the increment of the current time. As a result, the feature space of sweeping mode includes the data information from different

frequencies.

2) Multi-tone model: Even though the sweeping model includes different sound frequencies in a certain time period, for a specific time point, it can only emit a fixed frequency. In this subsection, we propose a multi-tone model. The sender provides more than one sound wave at the same time. The sender emits inaudible sound waves composed of multiple frequencies. Each component of the synthetic sound represents one sound wave at the designed frequency. Consequently, the multi-tone model enables the opportunity to cover more frequencies simultaneously. However, if HiddenTag emits sound at different frequencies, the distributed power on each frequency will decrease. We will apply the three models and compare the results in the section of performance evaluation.

In general, the multi-tone model can enrich feature space by increasing the number of tones. However, the increasing number of tones will reduce the power assigned to each tone. If the power distributed on each tone is too low, the identification results will decrease when we apply support vector machine (SVM) classification. Fig. 5 shows the result of FFT for the 3 sound generation models.

2.4 Receiving Sound

The process of receiving sounds is introduced as follows.

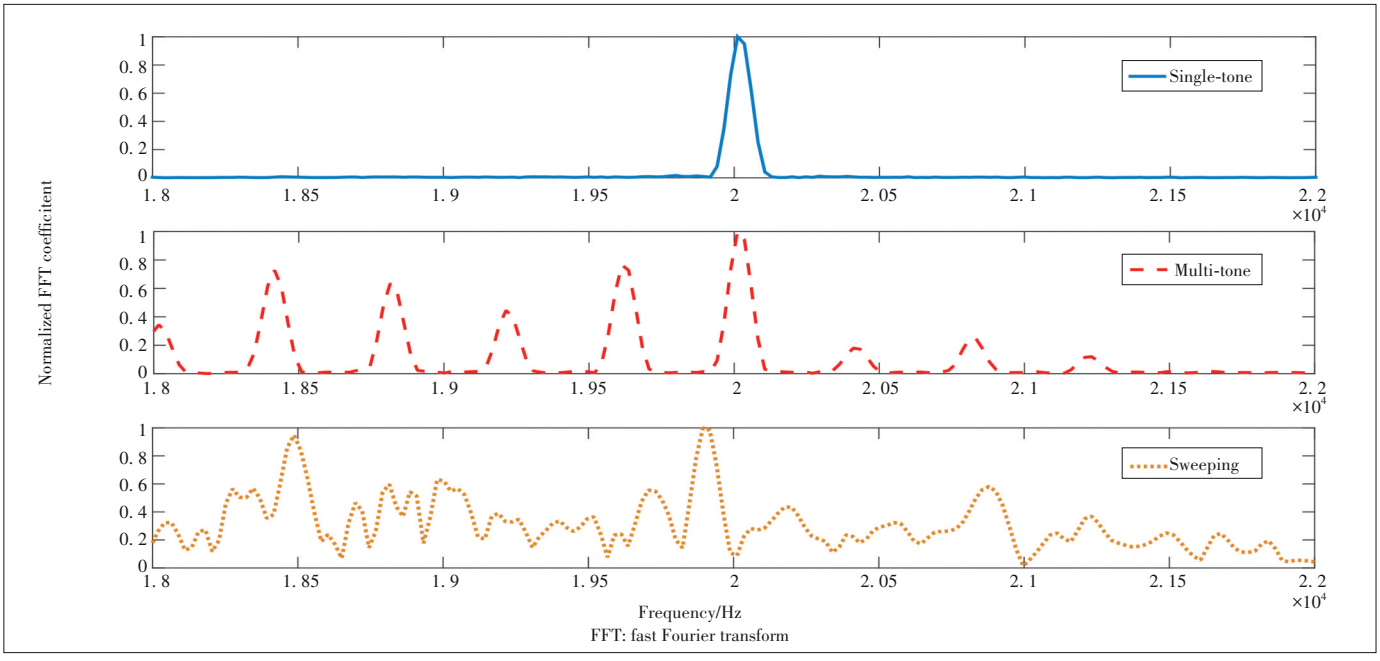
1) Sensing trigger: In HiddenTag, a sensing trigger is needed for person identification. Sensing trigger in our system detects users in a certain area rather than the whole home space. That is, HiddenTag should not recognize users everywhere except for the targeted sensing areas in the smart home. When a user enters the targeted area, HiddenTag will be turned on to collect acoustic data. Otherwise, the HiddenTag remains inactive. This switch can save the energy of smartphones and avoid high frequency acoustic signals when they are unnecessary. In our system, we adopt WiFi CSI signals[10 - 11], which are accurate and pervasive RF signals in smart homes, as the sensing trigger. Once our system detects CSI variations between wireless routers and receivers, HiddenTag will start acoustic sensing in the experimental area where the receiver locates.

2) Fast Fourier transform (FFT): Modern smartphones are able to generate sound waves with frequencies from 20 kHz to 22 kHz. There is an interesting phenomenon: most people cannot hear the sound between 18 kHz and 22 kHz. Considering that the users in the smart home do not suffer from the hearable noises, we can leverage such sound to identify different users. We use two smartphones in which the FFT converts time domain signals into representation in the frequency domain. That is, the FFT takes a block of time-domain data and returns the frequency spectrum of the data. Based on applying FFT and inverse fast Fourier transform (IFFT), we obtain data from both the time domain and frequency domain.

3) Band-pass filtering: In order to reduce the noises from the background and focus on the high acoustic frequency range, we adopt a band-pass filter. A band-pass filter passes signals with frequencies in a certain range and attenuates signals with frequencies out of that range. We keep the sound signals in the frequency range between 18 kHz and 21 kHz. The order of the band-pass filter is 9.

2.5 Launch Machine Learning Engine

1) Constructing acoustic features: Designing suitable feature space is important and challenging for high frequency



▲Figure 5. Three models of sound generation

sound. Different from most speech recognition works, classical features such Mel frequency cepstral coefficient (MFCC) [12] and AFTE[13] do not work well in our system. In HiddenTag, we explore classical features in statistics and extract them from both the time domain and frequency domain.

The features are calculated for a time window, the size of which can be adjusted based on the system's recommendations. In each time window, Table 1 shows the main features adopted in our system.

Specifically, we introduce power spectral entropy and crest factor in detail. In specific, entropy is a common measurement of disorder within a macroscopic system. In HiddenTag, spectral entropy is defined as following steps. First, we compute the spectrum $X(\omega_i)$ of the received signal. Next, we calculate the power spectral density (PSD) of the received signal via squaring its amplitude and normalizing it by the number of bins.

$$P\left(\omega_i\right) = \frac{1}{N}|X_{w_i}|^2 . \tag{3}$$

Then, we normalize the calculated PSD so that it can be viewed as a probability density function (PDF).

$$p_i = \frac{P(\omega_i)}{\sum_i P(\omega_i)} . \tag{4}$$

The power spectral entropy can be now calculated using a standard formula for an entropy calculation.

$$\mathrm{PSE} = -\sum_{i=1}^{n} p_i ln p_i . \tag{5}$$

Crest factor is a feature indicating the ratio of peak values to the effective value for a waveform. For example, crest factor 1 indicates no peaks and higher crest factors indicate peaks. In our system, as shown in Eq. (6), the crest factor refers to the peak amplitude (x_{peak}) of the waveform divided by the root mean square (RMS) value (x_{RMS}) of the waveform. Let C_{dB} denote the crest factor and RMS denote the square root of mean square (the arithmetic mean of the squares of a set of numbers), we have:

$$C_{dB} = 20\log_{10}\frac{x_{\mathrm{peak}}}{x_{\mathrm{RMS}}} . \tag{6}$$

2) Handling noises: Although we have used band-pass filters to reduce the noises which are not in the target range, there are other noises distributed on the frequency area from 18 – 21 kHz. These noise samples may reduce the classification accuracy of HiddenTag. Considering common noises in smart home environments include speaking, clapping, and some background noises, our system can add to or remove four types of noises (background, clapping, speaking and door knocking) from the dataset automatically when we train classification models. Besides, we can assign different ingredients to each type of noise. Once the noises occur in the testing phase, since the training model includes common noises, our system is confident in handling such a problem.

3) Classification: HiddenTag leverages SVM as the classification algorithm. Before implementing SVM in the proposed system, we should consider two problems. Which type of kernel shall be adopted? How to set the value of the penalty parameter? In our datasets, since the number of features is larger than that of observations, according to characterizations of common kernels, we select linear kernels for our SVM approach. Additionally, a low-value penalty parameter in SVM tends to make the decision surface smooth, while a high penalty parameter tries to train all samples correctly by giving the model freedom to select more samples as support vectors. We need to select the penalty parameter in SVM to achieve optimal results. HiddenTag adopts grid search to choose the penalty parameter. Besides, since our system aims to identify users in a short time period, the observation samples are limited. According to the features of common kernels used in SVM (linear kernel, radial basis function (RBF) kernel, etc.), we adopt linear kernels to obtain the optimized classification results.

4) Calibrate exceptions by prediction: Even if HiddenTag is able to identify different users, there still exists the probability of recognizing users incorrectly. Based on our observations, if the proposed system identifies users successfully for most cases, when some exceptions happen, we can calibrate the errors by historic information. In our system, as shown in Algorithm 1, we introduce an approach to avoid exceptions by leveraging the historical information. In each round, when we identify a user, our system not only counts the classification result from SVM in the current round, but also adds the previous results with a certain proportion (α). The parameter α can be adjusted according to the feedback of test cases.

▼**Table 1. Main features extracted in HiddenTag**

Features	Explanation
Crest factor	The value indicates how extreme the peaks are in a waveform
Energy	The energy of the signal in the time domain
Entropy of energy	The entropy of energy in the time domain
Spectral centroid	The center of the gravity of the frequency domain spectra
Spectral spread	The average spread of the spectrum in relation to its centroid
Spectral roll-off	The frequency below 90% of the magnitude distribution of the spectrum is concentrated
Spectral flux	The squared difference between two successive spectral frames
Spectral entropy	The entropy of the spectral energies
Spectral flatness	The ratio of the geometric mean to the arithmetic means of a power spectrum
Zero crossing rate	The rate of sign-changes along with a signal

Algorithm 1. Calibration algorithm for exceptions in HiddenTag

Require: α – between 0 and 1; $P_i'(j)$ – classification result

of user i in time period j before calibration
Ensure: U_{max} – the user with maximal prediction probability (identification result); n is the number of users, m is the number of time rounds
for int $i = 0; i < n; i$++ **do**
  for int $j = 0; j < m; j$++ **do**
    predict user by current round result and historic data
    $P_i(j) = P_i(j-1) \times \alpha + P_i'(j)$
  end for
end for
U_{max} is the user with maximal prediction probability
    select the user with the highest confidence in SVM
Return U_{max}

2.6 Applications of Personal Identification at Smart Home

Because HiddenTag can distinguish users in a smart home with convincing accuracy, we implement more applications via SmartHome Hub to provide personalized services. Our system integrates smart LED and speakers to show the identification results. For the installed smart LED, it will be assigned with different colors to different users. Once the user is identified, the corresponding color will be shown on the bulbs. The speaker can play personalized music for different users. If the user's web account is associated with HiddenTag, the preference music will be played on the smart speaker once the user is recognized. The system does not require explicit user interactions, such as login to an account, recognizing, recalling, or executing users' preferences. More applications can be integrated through SmartHome Hub based on the results of person identification.

3 Evaluation

3.1 Experiment Setup

HiddenTag includes an Android application and a module-view-control (MVC) based website to process acoustic data and recognize users. All the devices are deployed in a smart home environment. We use a Huawei Mate 30 smartphone as the controller, sender, and receiver. A proposed mobile application plays inaudible sound (18 – 21 kHz) on senders. It supports three models: single-tone, multi-tone, and sweeping frequency. Initially, we choose a multi-tone model for our experiments. Our system has generated 15 tones which are distributed from 18 – 21 kHz uniformly. The speaker's volume is set to 100%. The distance between the sender and receiver is 3 m. The area between the sender and receiver is empty. The sender and receiver are placed 75 cm above the floor. After receiving the varied acoustic signals by human activities, received acoustic data will be transmitted to the Dell T3640 server via WiFi. Based on the Python Scikit-Learn library, HiddenTag classifies different users via SVM. The c (penalty parameter) value is selected by grid search.

In our evaluation, we seek to answer three questions: Does HiddenTag identify different users successfully? Since there are often more than 3 family members living in a home environment, how many distinct users that our system can identify? What factors can affect the experimental results?

3.2 Evaluation Metrics

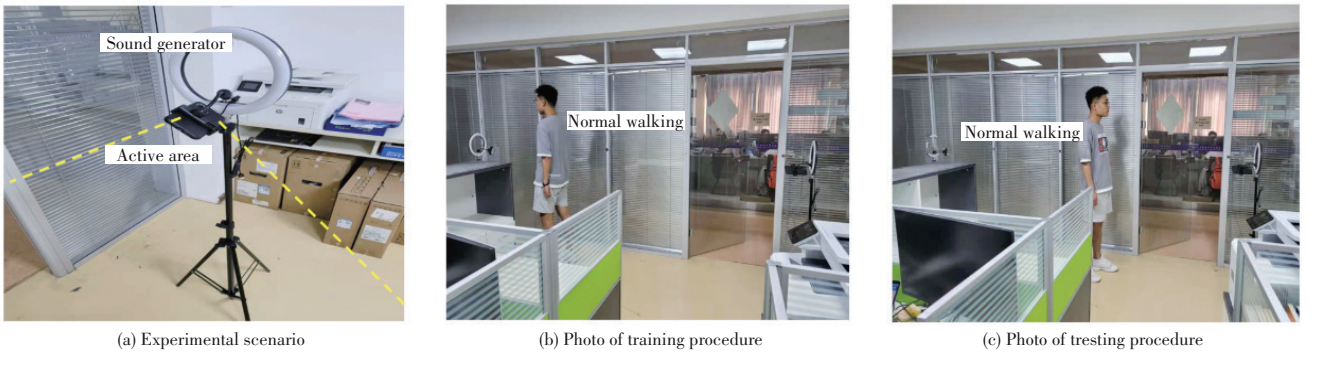
In the offline analysis, we use accuracy in a confusion matrix to describe person identification results. For online test results, we define that accuracy is the success rate for our recognition.

3.3 Case Study

We divide the case study into two phases: the training phase and the testing phase.

In the training phase, when each user enters the experimental environment, our system will detect user activities and start to profile the user. A user walks normally between the sender and the receiver. The user can also turn around and stand shortly. This training procedure lasts for 60 s. After the training procedure, the user leaves the experimental room.

When the user returns to the room, once she/he walks into the same experimental area, HiddenTag starts to recognize the user and shows the confidence of user recognition. This step is the testing phase. Fig. 6 illustrates the experimental environ-



(a) Experimental scenario (b) Photo of training procedure (c) Photo of tresting procedure

▲**Figure 6. Experimental scenario and case study**

ment and corresponding case study.

In this subsection, we observe the group with four users as shown in Table 2. Four users participated in the experiment. Each user was trained and tested separately. Table 3 is the confusion matrix for the classification. As shown in Fig. 7 (a), testing accuracy can achieve 96.1%. We thus conclude that the time length of training influences the accuracy. The longer time of training obtains better results. However, considering our application scenario should limit training procedure to a certain time length, we choose 60 s in our implementation.

Then, we focus on the number of users in our case study. We extend our experiments from 4 users to 10 users. After changing the number of users, based on Fig. 7 (b), we notice our system still achieves an accuracy of more than 90%. Although the system performs better when the system includes fewer users, HiddenTag can still process 10 users with acceptable accuracy.

▼Table 2. Information of four volunteers

User	A	B	C	D
Height/cm	176	177	174	163
Weight/kg	65	80	70	55
Age	25	31	33	40
Gender	M	M	F	F

▼Table 3. Confusion matrix of four-volunteer experiment

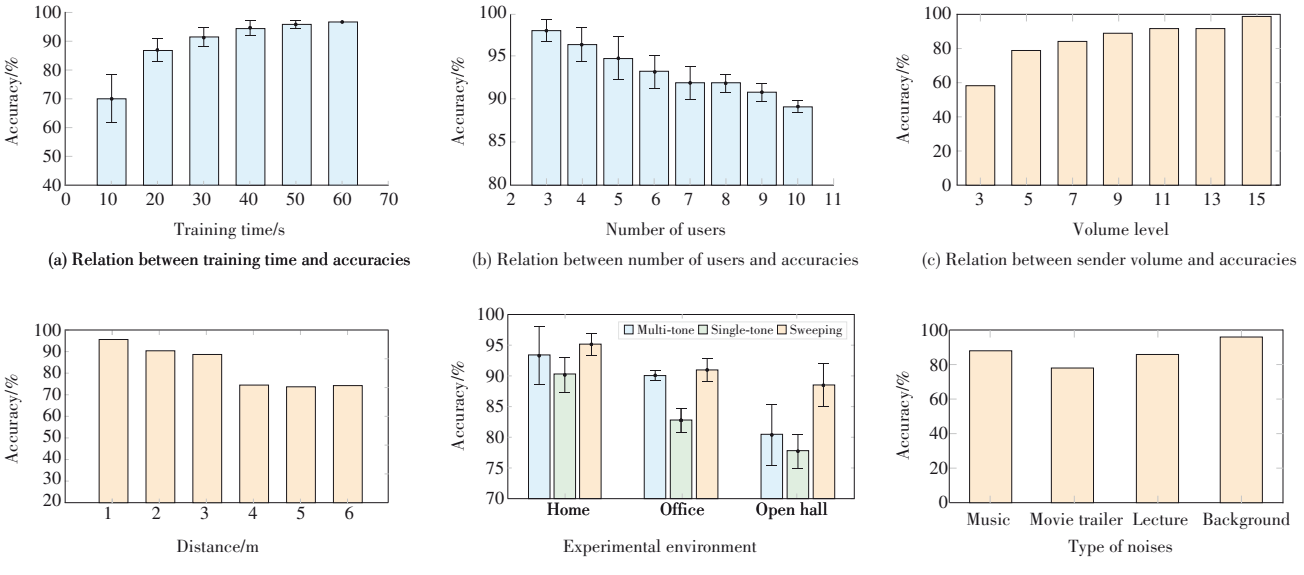
Actual/Classified	A	B	C	D
A	93.0%	1.0%	0.0%	6.0%
B	1.0%	98.0%	0.0%	1.0%
C	0.0%	0.0%	96.0%	4.0%
D	5.0%	16.0%	0.0%	79.0%

Different volumes of the sender sound will change the sound signal strength and identification accuracy. We did the control group experiments to detect which percentage is the best volume for our experiment. Fig. 7 (c) shows the improvement with increasing volume.

Additionally, the distance between the sender and receiver is another factor that affects recognition results. According to existing experimental settings, we only adjust the distance between the sender and receiver. Fig. 7 (d) shows that with closer distance, the group achieves better accuracies. Only when the distance is too short to profile walking activities (within 1 m), the accuracy will decrease.

Then, we compare three sound generation models and discuss which one is the best model for the proposed system. In Fig. 7 (e), we conduct other two control groups by using a single-tone model and a sweeping model. For the single-tone model, we set the frequency of the sound to 20 kHz. For the sweeping model, we sweep frequency from 18 kHz to 21 kHz once per second. We compare the three techniques in different scenarios (smart homes, offices and open halls), and come to the conclusion that sweeping and multi-tone models outperform single-tone models. Because multi-tone and sweeping models increase accuracies by enriching feature space. The multi-tone model is subjected to power decrease and thus needs a power amplifier to improve performance.

Additionally, based on our observations, the errors of the proposed system mainly occur in the first or second frame. Within the time increasing, the errors will decrease sharply. This phenomenon is caused by two reasons. First, the acoustic signature of each user cannot form in a very short time period.



(a) Relation between training time and accuracies
(b) Relation between number of users and accuracies
(c) Relation between sender volume and accuracies
(d) Relation between distance and accuracies
(e) Experiments by different acoustic models
(f) Experiments with different types of noises

▲Figure 7. Experimental results of evaluations

Once a user has walked 3 – 5 gait cycles, our system can recognize the user based on the acoustic signature. Second, as illustrated in Algorithm 1, the results will be calibrated by historical data. The beginning frames do not have the capability of enhancing accuracies by counting the results in previous rounds.

4 Diving into Depth

In this section, we further analyze HiddenTag based on these factors: online performance, experimental environment, and noise handling.

4.1 Pushing Offline to Online

In order to deploy HiddenTag in a real platform, we develop an online system to show the real-time identification results. HiddenTag adopts Node.js and Python Flask to display the real-time accuracies. The time delay of classification results can be controlled from 2 s to 5 s. Table 4 illustrates the comparison between online and offline results in the same experimental scenario. As shown in Eq. (7), for a certain user, the online accuracy is the ratio that times of successful identifications (N_s) divided by the total times of identifications (N_t).

$$Acc_0 = \frac{N_s}{N_t} \times 100\% . \tag{7}$$

Although online accuracy is lower than offline accuracy, it still reaches 85.0% – 90.0%. Next, we extend the online experiments from a room to other scenarios with different layouts and materials. We test HiddenTag in a conference room and a coffee room. The two experiments have achieved online accuracies of 87.5% and 90.5%. The case studies have proved HiddenTag can work normally in different environments.

There are two reasons that the accuracy is lower in the online system. First, in offline classification, SVM is able to choose optimal parameters by brute-force searching. However, it is difficult to optimize all the parameters in a short time period due to computational limitations in an online system. Second, the experimental environment is changing between online training and online testing. We discuss this issue in the following subsection.

4.2 Environment Changing

Our experiments are conducted in the same environment. Unfortunately, the same experiment scenario is always changing due to variations of environmental factors, such as temperature and humidity. To assess its impact, a user walked in the experimental scenario by a five-minute interval. Table 5 shows that the same user is classified by HiddenTag for four times in different time slots. Even if one user enters the room for four times, each event can be identified as different users with 62.25% accuracy. To eliminate the environmental changing, the system needs to continuously collect long-term training data implicitly. Relying on a larger dataset that includes more environments variations, we can identify people even if the environment changes sharply.

▼Table 4. Comparison between online and offline results

	Sweeping	Single-Tone	Multi-Tone
Offline	95.2%	91.0%	93.1%
Online	90.0%	80.0%	85.0%

▼ Table 5. Confusion matrix of identifying the same user in different time periods

		Classified			
		1st	2nd	3rd	4th
Actual	1st	48.0%	6.0%	3.0%	43.0%
	2nd	11.0%	74.0%	8.0%	7.0%
	3rd	1.0%	5.0%	74.0%	20.0%
	4th	40.0%	0.0%	7.0%	53.0%

4.3 Noise Handling

We simulate typical noises when conducting HiddenTag in an experimental scenario. In this experiment, we use a Huawei Mate 30 smartphone to play 3 audio files including a song named "Amazing Grace", the trailer of Game of Throne 8, and a lecture talk of a machine learning class on Coursera. The playing smartphone is close to the receiver (30 cm). By adopting the proposed noise handling method introduced in Section 2.5, Fig. 7 (f) shows that our system still reaches acceptable accuracies even if it encounters different types of noises. HiddenTag can work normally under some types of noises, but the accuracies decrease when it encounters some noises, such as the noises made by the working elevator and starting of the heater.

5 Related Work and Comparison

Existing person identification approaches broadly rely on computer vision and image techniques. By analyzing users' faces and fingerprints, researchers and engineers have provided numerous solutions to user recognition. As a classical face recognition approach, Turk and Pentlend leverage Eigenfaces to define the face space and identify people[14]. Recent researchers use the deep network to enhance recognition accuracy[6, 15 – 16]. DeepID3[6] designs a high-performance deep convolution network and adds supervision to early convolutional layers, and it represents the state-of-the-art technology on YouTube Faces benchmarks. Voice recognition is another type of common approach to identify a user. MUDA et al.[17] explore MFCC and dynamic time warping (DTW) techniques to recognize users. In addition, biometrics techniques, such as fingerprint and retina, are other common types of person identification[18 – 21]. Unfortunately, all of these methods face privacy concerns. Although these approaches can recognize users by biometric information, the key personal and private information has to be exposed.

Recently, some alternative methods have been proposed to

identify persons. Researchers adopt wireless sensing to identify persons, gestures, and even micro-activities[22 - 23]. By classifying variations of WiFi signals, WiWho[23] leverages CSI to describe the user's walking behaviors and identify users in WiFi environments. However, wireless sensing methods often need specific devices, such as the emitter and the receiver with CSI drivers, which are not common in smart home environments.

Acoustic sensing has been a hot topic recently, and lots of corresponding applications, such as speech recognition, indoor localization are implemented in smart homes[24 - 27]. GEIGER et al.[28] presented a system for identifying humans by their walking sound, by leveraging MFCC and Hidden Markov Model, which has reached the offline identification rate of 65.5% for 155 subjects. This approach depends on the sounds of footsteps. Once the shoes and floors are changed, the system might not work normally. This method does not consider noise handling and online accuracies in a real environment. Actually, there is no solution for person identification area by acoustic sensing without human voice or step sound.

As shown in Table 6, different from the existing solutions, HiddenTag is a device-free and highly accurate person identification approach. By using built-in smartphones, we can recognize users only by profiling the common indoor activities at home and in office environments.

6 Conclusions and Future Work

HiddenTag represents the first device-free system that employs inaudible acoustic sensing to achieve accurate person identification. Through this process without any hardware modification, we gain important insights: 1) acoustic information with frequencies from 18 - 21 kHz can profile human indoor activities and recognize users in smart home environments; 2) sweeping frequency and multi-tone models can improve SVM classification for acoustic datasets by enriching features; 3) online and offline identification accuracy can reach more than 90% in simplified testing and training procedures which are close to normal activities in the environments similar to smart homes. We believe HiddenTag's salient advantages will enable a myriad of personalized services in smart homes, including smart voice assistants, augmented reality, energy saving, and various pervasive applications.

Moving forward, we are aiming to further improve the identification accuracies by leveraging other machine learning techniques such as recurrent neural networks and generative adversarial networks and enrich the acoustic features by leveraging transfer learning. In addition, we aim to extend single person identification to multi-person with more walking patterns.

References

[1] CLIFFORD B R, BULL R. The psychology of person identification [M]. London, UK: Routledge, 2017. DOI: 10.4324/9781315533537
[2] BADRINARAYANANN V A , SIERRA J J , MARTIN K M . A dual identification framework of online multiplayer video games: The case of massively multiplayer online role playing games (MMORPGs) [J]. Journal of business research, 2015, 68(5): 1045-1052
[3] FANG B Y, CO J, ZHANG M. DeepASL: enabling ubiquitous and non-intrusive word and sentence-level sign language translation [C]//The 15th ACM Conference on Embedded Network Sensor Systems. Delft, Netherlands: ACM, 2017: 1 - 13. DOI: 10.1145/3131672.3131693
[4] ALI K , LIU A X , WEI W , et al. Keystroke Recognition Using WiFi Signals [C]// ACM MobiCom. Paris, France: ACM, 2015
[5] YI J, LEE Y. Heimdall: mobile GPU coordination platform for augmented reality applications [C]//The 26th Annual International Conference on Mobile Computing and Networking. London, United Kingdom: ACM, 2020: 1 - 14. DOI: 10.1145/3372224.3419192
[6] LSUN Y, LIANG D, WANG X G, et al. DeepID3: face recognition with very deep neural networks [J]. Computer Science, 2015
[7] WANG M, DENG W H. Deep face recognition: a survey [EB/OL]. [2021-03-20]. https://export.arxiv.org/pdf/1804.06655
[8] ZHANG Z Y. Microsoft Kinect sensor and its effect [J]. IEEE multimedia, 19(2): 4 - 10, 2012
[9] LÓPEZ G, QUESADA L, GUERRERO L A. Alexa vs. Siri vs. Cortana vs. Google assistant: a comparison of speech-based natural user interfaces [C]//The AHFE 2019 International Conference on Human Factors and Systems Interaction. Washington, USA: AHFE, 2019. DOI: 10.1007/978-3-319-60366-7_23
[10] HALPERIN D, HU W J, SHETH A, et al. Tool release [J]. ACM SIGCOMM computer communication review, 2011, 41(1): 53. DOI: 10.1145/1925861.1925870
[11] WANG X Y, GAO L J, MAO S W. CSI phase fingerprinting for indoor localization with a deep learning approach [J]. IEEE Internet of Things journal, 2016, 3(6): 1113 - 1123. DOI: 10.1109/JIOT.2016.2558659
[12] LOGAN B. Mel frequency cepstral coefficients for music modeling [EB/OL]. [2021-03-20]. http://citeseerx. ist. psu. edu/viewdoc/summary? doi=10.1.1.11.9216
[13] SMCKINNEY M, BREEBAART J. Features for audio and music classification [EB/OL]. [2021-04-02]. https://ismir2003. ismir. net/presentations/McKinney. pdf
[14] TURK M, PENTLAND A. Eigenfaces for recognition [J]. Journal of cognitive neuroscience, 1991, 3(1): 71 - 86. DOI: 10.1162/jocn.1991.3.1.71
[15] XIAO T, LI H S, OUYANG W L, et al. Learning deep feature representations with domain guided dropout for person re-identification [C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016:

▼**Table 6. Comparison between HiddenTag and other classical approaches**

	Information Type	Training Cost	Hardwares Required	Privacy Level	Accuracy
DeepID3	Image	High	Cameras	Low	92% offline
WiWho	CSI	Normal (100 s)	Special CSI devices	High	80% - 92% offline
Step sound	Normal sound	Low (3 s)	Built-in smartphones	High	65% offline
HiddenTag	High frequency sound (18 - 21 kHz)	Normal (60 s)	Built-in smartphones	High	96% offline, 85% - 90% online

CSI: channel state information

1249 – 1258. DOI: 10.1109/CVPR.2016.140
[16] YU H X, ZHENG W S. Weakly supervised discriminative feature learning with state information for person identification [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 5527 – 5537. DOI: 10.1109/CVPR42600.2020.00557
[17] MUDA L, BEGAM M, ELAMVAZUTHI I. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques [EB/OL]. [2021-04-02]. https://arxiv.org/abs/1003.4083
[18] BRUNELLI R, FALAVIGNA D. Person identification using multiple cues [J]. IEEE transactions on pattern analysis and machine intelligence, 1995, 17(10): 955 – 966. DOI: 10.1109/34.464560
[19] PERALTA D, TRIGUERO I, SANCHEZ-REILLO R, et al. Fast fingerprint identification for large databases [J]. Pattern recognition, 47: 588 – 602, 2014. 10.1016/j.patcog.2013.08.002
[20] RAO G S, NAGARAJU C, REDDY L, et al. A novel fingerprints identification system based on the edge detection [J]. International journal of computer science and network security, 8: 394 – 397, 2008
[21] TROJE N F, WESTHOFF C, LAVROV M. Person identification from biological motion: effects of structural and kinematic cues [J]. Perception & psychophysics, 2005, 67(4): 667 – 675. DOI: 10.3758/BF03193523
[22] WANG Z, YU Z W, LOU X Y, et al. Gesture-radar: a dual Doppler radar based system for robust recognition and quantitative profiling of human gestures [J]. IEEE transactions on human-machine systems, 2021, 51(1): 32 – 43. DOI: 10.1109/THMS.2020.3036637
[23] ZENG Y Z, PATHAK P H, MOHAPATRA P. WiWho: WiFi-based person identification in smart spaces [C]//15th ACM/IEEE International Conference on Information Processing in Sensor Networks. Vienna, Austria: IEEE, 2016: 1-12. DOI: 10.1109/IPSN.2016.7460727
[24] MOON Y, KIM K J, SHIN D H. Voices of the Internet of Things: an exploration of multiple voice effects in smart homes [M]//Distributed, ambient and pervasive interactions. Cham: Springer International Publishing, 2016: 270 – 278. DOI: 10.1007/978-3-319-39862-4_25
[25] TUNG Y C, SHIN K G. EchoTag: accurate infrastructure-free indoor location tagging with smartphones [C]//The 21st Annual International Conference on Mobile Computing and Networking. Paris, France: ACM, 2015: 525 – 536. DOI: 10.1145/2789168.2790102
[26] YANG Z J, WEI Y L, SHEN S, et al. Ear-AR: Indoor acoustic augmented reality on earphones [C]//The 26th Annual International Conference on Mobile Computing and Networking. London, United Kingdom: ACM, 2020: 1 – 14. DOI: 10.1145/3372224.3419213
[27] ZHOU B, ELBADRY M, GAO R P, et al. BatMapper: acoustic sensing based indoor floor plan construction using smartphones [C]//The 15th Annual International Conference on Mobile Systems, Applications, and Services. Niagara Falls, USA: ACM, 2017: 42 – 55. DOI: 10.1145/3081333.3081363
[28] GEIGER J T, KNEIßL M, SCHULLER B W, et al. Acoustic gait-based person identification using hidden Markov models [C]//The 2014 Workshop on Mapping Personality Traits Challenge and Workshop. Istanbul, Turkey: ACM, 2014: 25 – 30. DOI: 10.1145/2668024.2668027

Biographies

QIU Chen (qiuchen@nwpu.edu.cn) received the Ph.D. degree in computer science from Michigan State University, USA in 2017. He is currently an associate professor with Northwestern Polytechnical University, China. His research interests include pervasive computing, mobile computing, and applied machine learning. He is a member of the IEEE.

DAI Tao received his B.S., M.S. and Ph.D. degrees in software engineering from Xi'an Jiaotong University, China in 2008, 2011 and 2020, respectively. He is currently a lecturer at the School of Economics and Management, Chang'an University, China. He was a visiting student at the School of Computer Science, Carnegie Mellon University, USA from September 2018 to September 2019. His main research interests include natural language processing, information retrieval, and machine learning.

GUO Bin received the Ph.D. degree in computer science from Keio University, Japan in 2009 and then went to the French National Institute of Telecommunications for postdoctoral research. He is a professor with Northwestern Polytechnical University, China. His research interests include ubiquitous computing, mobile crowd sensing, and HCI. He is a senior member of the IEEE.

YU Zhiwen received the Ph.D. degree from Northwestern Polytechnical University, China. He is currently a professor and Dean with the School of Computer Science, Northwestern Polytechnical University. His research interests include pervasive computing and human-computer interaction. He is a senior member of the IEEE.

LIU Sicong received the B.S., M.S. and Ph.D. degrees from Xidian University, China in 2013, 2016, and 2020 respectively. From 2017 to 2018, she was a visiting scholar at Rice University, USA. She is currently an associate professor with Northwestern Polytechnical University, China. Her research interests include mobile computing system, mobile and embedded deep learning design, and automated deep model optimization.

Device-Free In-Air Gesture Recognition Based on RFID Tag Array

WU Jiaying, WANG Chuyu, XIE Lei
(State Key Laboratory for Novel Software Technology, Nanjing 210023, China)

Abstract: Due to the function of gestures to convey information, gesture recognition plays a more and more important part in human-computer interaction. Traditional methods to recognize gestures are mostly device-based, which means users need to contact the devices. To overcome the inconvenience of the device-based methods, studies on device-free gesture recognition have been conducted. However, computer vision methods bring privacy issues and light interference problems. Therefore, we turn to wireless technology. In this paper, we propose a device-free in-air gesture recognition method based on radio frequency identification (RFID) tag array. By capturing the signals reflected by gestures, we can extract the gesture features. For dynamic gestures, both temporal and spatial features need to be considered. For static gestures, spatial feature is the key, for which a neural network is adopted to recognize the gestures. Experiments show that the accuracy of dynamic gesture recognition on the test set is 92.17%, while the accuracy of static ones is 91.67%.

Keywords: gesture recognition; RFID tag array; neural network

DOI: 10.12142/ZTECOM.202103003

https://kns.cnki.net/kcms/detail/34.1294.TN.20210816.1023.002.html, published online August 16, 2021

Manuscript received: 2021-06-10

Citation (IEEE Format): J. Y. Wu, C. Y. Wang, and L. Xie, "Device-free in-air gesture recognition based on RFID tag array," *ZTE Communications*, vol. 19, no. 3, pp. 13 - 21, Sept. 2021. doi: 10.12142/ZTECOM.202103003.

1 Introduction

Recent years have seen the rapid growth of human-computer interaction and its applications range from somatosensory games to smart screens. In these applications, gestures are an important way to convey information. How to perceive gestures through the device, so as to realize accurate recognition and natural human-computer interaction, is a research hotspot.

Methods of gesture recognition include device-based ones and device-free ones. The device-based methods need users to wear or touch the device, so as to perceive human action[1-2]. However, wearing a device is not natural for users and the battery is a big headache. On the contrary, the device-free ones do not need users to touch the devices. They use computer vision or wireless technologies instead[3-4]. Though accurate, computer vision brings privacy concerns and is sensitive to light interference. Therefore, we turn to radio frequency identification (RFID), which is a kind of wireless technology. Based on RFID, users only need to perform

This work was supported by National Natural Science Foundation of China under Grant Nos. 61902175, 61872174 and 61832008 and Natural Science Foundation of China under Grant No. BK20190293.

gestures in the air naturally and do no need to care about the privacy issues.

Here comes the question: how to recognize gestures based on RFID? With regard to hand gestures, through a passive RFID tag, we can get the signal reflected by the hand. However, one tag is far from enough to accurately recognize gestures. Therefore, we use tag array to sense gestures. When performing a gesture, the hand has different influences on different parts of the tag array, which provides more diverse information for recognition. Features extracted from the signal along time can be regarded as an image sequence, which contains both spatial features and temporal features. For dynamic gestures, we should take both the spatial and temporal features into consideration. Here, a combined convolutional neural network and Long Short-Term Memory (CNN-LSTM) system is proposed to process spatial features from the tag array by the CNN and process temporal features in the image sequence by the LSTM. For static gestures, we can get the final feature image from the image sequence as the snapshot of the gesture, and use CNN to recognize it.

There exist some challenges, however. First, it is important to improve the robustness of recognition. When different users perform gestures at different speeds, the system need to be robust enough to recognize them. Second, for dynamic gestures, both spatial and temporal features need to be considered, which should be dealt with carefully. Third, for static gestures, we need to extract their features from dynamic signals. How to decide the final features for recognition is the key.

In this paper, our contributions are shown as follows:

1) We propose a device-free in-air gesture recognition method based on RFID tag array, which can recognize dynamic gestures and static gestures.

2) For dynamic gestures, we take both spatial and temporal features into account and discuss several structures for recognition. CNN and LSTM are combined to get better performance and adjustment is made to improve the robustness.

3) We implement a gesture recognition system based on our method. Experiments show that the accuracy of dynamic gesture recognition on the test set is 92.17%, and the accuracy of static ones is 91.67%.

2 Related Work

When it comes to human-computer interaction, there is no denying that action recognition is an important part. Through action, users convey order or information and interact with computer. From the perspective of the body part to be recognized, action recognition can be divided into three kinds[5]: gesture recognition, head and facial action recognition, and overall body action recognition. In this paper, we focus on gesture recognition, including dynamic gesture recognition and static gesture recognition. According to whether the user needs to wear or touch the device, action recognition can also be divided into device-based and device-free.

2.1 Device-Based Methods

Device-based methods need users to wear or touch and input the device. These methods include mechanical, tactile, ultrasonic, inertial and magnetic methods[6]. Wearable devices usually contain sensors such as accelerometers and gyroscopes, and based on these, they capture and recognize action[2, 7]. The signal returned by a sensor changes along with its moving, making it possible for us to extract action-related features. Sometimes, RFID tags can be attached to gloves or bracelets, acting as sensors[8–9]. Electromyography (EMG) and force myography (FMG) are also used to recognize action. JIANG et al.[1] propose a novel co-located approach (EMG and FMG) for capturing both sensing modalities, simultaneously, at the same location, so as to better recognize the gesture.

2.2 Device-Free Methods

Device-based methods can accurately perceive the gesture, but they are not natural for users. Besides, the battery problem is a big headache, making it more inconvenient. Therefore, researchers turn to device-free methods. Computer vision is a typical device-free method[3, 4, 10]. Structure, color and even depth information can be provided through ordinary cameras or depth cameras. However, as people pay more and more attention to privacy issues, they tend to refuse computer vision. With regard to wireless technologies, privacy is no longer a problem. These kinds of methods use wireless signals, usually electromagnetic or acoustic, to capture the action and recognize it. For examples, the Low-Latency Acoustic Phase (LLAP)[11] scheme uses ultrasonic signals to recognize character gestures. WiGest[12] uses WiFi signals. MHomeGes[13] recognizes arm gesture based on mmWave signals. As for RFID, RFIPad[14] makes use of the phase changes and received signal strength to recognize stroke movements and detect direction through, so as to recognize basic character gestures. Using hierarchical recognition, image processing and polynomial fitting, TagSheet[15] enables the recognition of sleeping postures.

3 Preliminaries

Originating from radar technology, RFID use radio frequency signals to sense and identify targets. A typical RFID system usually consists of readers, antennas and tags[16]. RFID tags include active tags, semi-active tags and passive tags. Active tags are battery-assisted, while passive ones need a reader to power them. Once powered by the signal from the reader, an RFID tag will transmit back the signal with its own information.

When a hand is put in front of a tag S, the signal returned by the tag includes the signal directly transmitted to the

tag S_{tag} and the signal reflected by the hand S_{hand}:

$$S = S_{tag} + S_{hand}. \tag{1}$$

Readings returned by a tag S include phase θ (rad) and Received Signal Strength Indication (RSSI) R (dBm). Therefore, with the hand in front of the tag, we can calculate the signal S[17]:

$$S = \sqrt{10^{\frac{R}{10}-3}}\, e^{j\theta} = \sqrt{10^{\frac{R}{10}-3}}\cos\theta + j\sqrt{10^{\frac{R}{10}-3}}\sin\theta. \tag{2}$$

Without the hand in front of the tag, S_{tag} can be calculated in the same way. Therefore, by subtracting S_{tag}, the signal reflected by the hand S_{hand} is obtained according to Eq. (1). Based on this, the actual power P_{actl} of S_{hand} and its theoretical power P_{theor} can be calculated:

$$P_{actl} = \left| S_{hand} \right|^2, \tag{3}$$

$$P_{theor} = \frac{\mathrm{C}}{d^4}, \tag{4}$$

where C is a constant and d is the distance from the hand to the tag.

According to the algorithm in Ref. [17], an actual power map and certain a theoretical power map can be got from a tag array. Based on these two maps, a possibility map can be created. A series of possibility maps along time form the feature image sequence we want.

4 Gesture Recognition System

Our device-free in-air gesture recognition system includes a feature extraction module and a gesture recognition module.

4.1 Feature Extraction

The feature image sequence can be extracted based on the preliminaries mentioned above. The extraction process includes preprocessing, gesture region segmentation and sequence generation.

1) Data preprocessing. A sliding window is used to preprocess the data. In this phase, several consecutive readings are combined into one, while the vacant readings of some tags are interpolation. The moving average filter is used to smooth the signal.

2) Gesture region segmentation. When there is a hand in front of the tag array, the signal we get will be far different from the one without a hand. Therefore, by calculating the distance to the signal without a hand, we can segment the gesture region whose distance is larger than the threshold.

3) Feature image sequence generation. Based on the former steps, an actual power map and certain a theoretical power map can be obtained. The feature image sequence will then be calculated by the algorithm in Ref. [17]. For a 5×7 RFID tag array, we can get a 15×21 cm^2 feature image and each pixel of the image measures the possibility that the hand is in front of the pixel grid. Fig. 1 shows such a feature image. The brighter a pixel grid on the map, the more likely it is that the hand will be in front of the corresponding pixel grid.

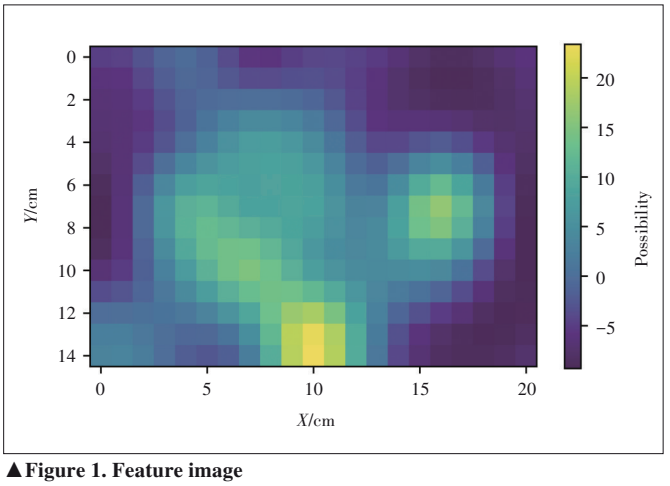
4.2 Dynamic Gesture Recognition

After a feature image sequence of a dynamic gesture is extracted, the problem of dynamic gesture recognition is transformed into a feature image sequence recognition problem. CNN has advantages in extracting spatial features from an image, while LSTM can handle the temporal feature in sequence well. In order to focus both spatial and temporal features in a feature image sequence, we can combine them as CNN-LSTM.

4.2.1 Network Structure

For efficiency, feature images in the long image sequence will be divided evenly into five groups. Images in the same group will be superimposed into one frame of the feature image. In this way, the original long image sequence is converted into a shorter sequence, which contains five frames of the feature image. The size of each image is 15×21 cm^2. Moreover, totally six kinds of gestures are needed to be recognized, so the label of each sample is coded as a six-dimensional one-hot code. The type corresponding to the highest value of the output vector is the predicted gesture type. Here, we adopt cross entropy as the loss function. To recognize the dynamic gesture from the feature sequence, we start with the CNN structure, then move to the LSTM structure, and finally discuss CNN-LSTM, the combined one.

1) CNN structure. To make it possible for CNN to learn from the data, images in the sequence are vertically stitched according to their temporal relationship. Then, the 75×21 cm^2 image is fed into the network as a sample input. The CNN structure is: the input layer, convolutional layer-1, pooling lay-



▲Figure 1. Feature image

er-1, convolutional layer-2, pooling layer-2, fully connected layer, and output layer. Activation function of the convolutional layer and fully connected layer is Rectified Linear Unit (ReLU), and one of the output layer is softmax. The kernel sizes of convolutional layer-1 and layer-2 are 3×3×*n_conv*1, 3×3×*n_conv*2, respectively. Besides, the fully connected layer maps the extracted features into an *n_fc*-dimensional vector. Here, (*n_conv*1, *n_conv*2, *n_fc*) forms the hyperparameters of the CNN structure.

2) LSTM structure. LSTM takes sequence data as input, with each element in sequence being a vector. Therefore, we flatten each image in the feature image sequence into a vector and feed them into LSTM along time. With all feature vectors fed, LSTM advances five steps in time dimension. Here, we take the output of last time dimension and map it into the output vector. As for hyperparameters, the number of hidden units *n_hd* is our target, which determines the ability of LSTM to extract temporal information along time sequence.

3) CNN-LSTM combined structure. For a dynamic gesture and its feature image sequence, CNN focuses on the spatial characteristic in image, which carry information of gesture impacts on different parts of the tag array, while LSTM focuses on the temporal characteristic in sequence, which carry information of the changes of the dynamic gesture along time. We combine them as CNN-LSTM and both spatial and temporal features are focused. Table 1 and Fig. 2 show the CNN-LSTM structure. The CNN part takes each image in feature image sequence as input and outputs a summary vector for the LSTM part. The LSTM part takes the summary vector sequence as input and extracts its temporal feature. In the last time step, the prediction vector is obtained through mapping. Here, (*n_conv*1, *n_conv*2, *n_fc*, *n_hd*) forms the hyperparameters.

4.2.2 Overfitting and Adjustment

In the process of learning, we should be vigilant against overfitting. In training, the loss on a training set declines, but the loss on the validation set tends to be flat or even starts to rise. It means that the model is still learning from the training set, but the features it learns tend to include irrelevant features, which is harmful to the generalization ability of the model.

To deal with the overfitting, we add a dropout layer between the CNN part and LSTM part. In training, it randomly drops neurons at this layer. All neurons will be used for prediction. Fig. 3 shows the loss curve with and without the dropout layer. As we can see, the validation loss begins to arise after about 130 epochs without dropout, indicating that overfitting occurs. On the contrary, this phenomenon is well-suppressed with dropout. With ten-fold cross validation, we set the dropout rate as 20%.

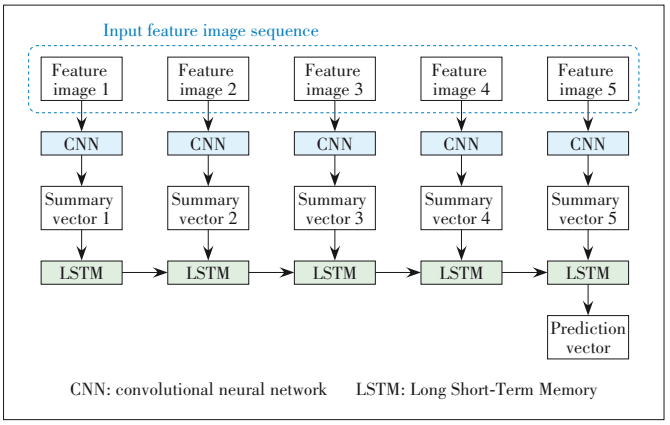
For other methods, one may think of regularization, such as L2 regularization, which adds a penalty term to loss function. L2 regularization tends to suppress excessive weight parameters. But for CNN, it doesn't make much sense. For LSTM, limitation to weight parameters will lead to rapid disappearance of the learned temporal information along time[18]. Therefore, we don't use regularization to suppress overfitting of CNN-LSTM.

Apart from the dropout layer, we adopt early stopping strategy in training to avoid serious overfitting. The red line in Fig. 3 is an example of early stopping line. If the loss reduction of the training set is less than a certain threshold or even the loss begins to rise, the training is considered to have made no progress in this epoch. If the model keeps making

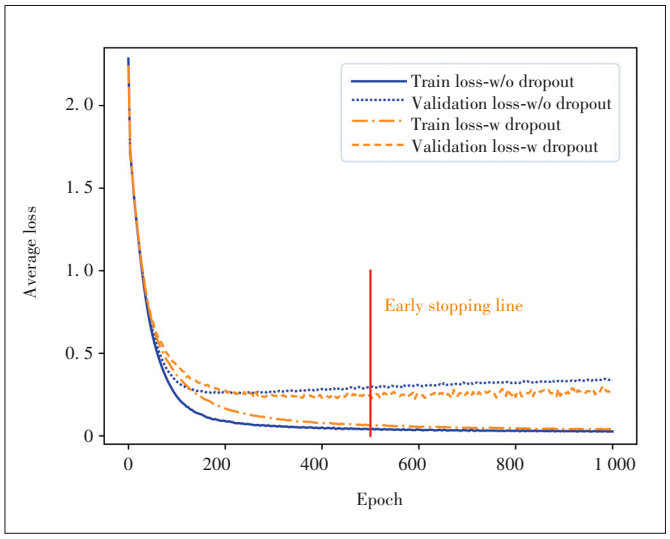
▼**Table 1. CNN-LSTM structure**

	Layer	Description
CNN part	Input layer	Input: feature image sequence Length of sequence: 5 Image size: 15×21
	Convolution layer-1	Extract *n_conv*1 features from the image Kernel size: 3×3×*n_conv*1 Step size: 1 Activation function: ReLU
	Pooling layer-1	Downsample the extracted features Pooling type: max pooling Template size: 2×2 Step size: 2
	Convolution layer-2	Extract *n_conv*2 features from the image Kernel size: 3×3×*n_conv*2 Step size: 1 Activation function: ReLU
	Pooling layer-2	Downsample the extracted features Pooling type: max pooling Template size: 2×2 Step size: 2
	Fully connected layer	Fully connect the features to *n_fc*-dimensional summary vector
LSTM part	LSTM layer	Extract features from summary vector sequence Time steps: 5 Number of hidden units: *n_hd*
	Fully connected layer	Fully connect the features to six-dimensional prediction vector

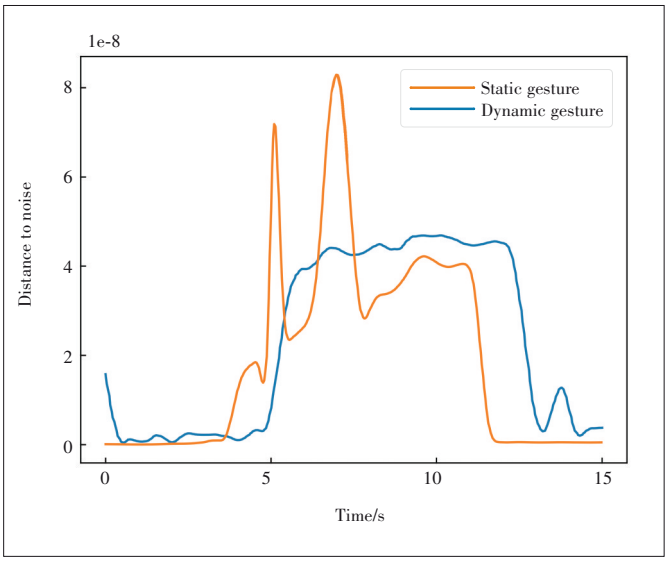
CNN: convolutional neural network LSTM: Long Short-Term Memory
ReLU: Rectified Linear Unit



▲**Figure 2. CNN-LSTM structure**



▲Figure 3. Loss curves for dynamic gesture recognition with and without dropout



▲Figure 4. Static gesture recognition by the distance to noise

no progress for certain epochs, we can stop the training in time. This prevents the model from continuing to learn even though it tends to overfit.

4.3 Static Gesture Recognition

For static gestures, when we put our hands in front of the tag array, the signal we received is still different from the signal without hands, which is defined as noise floor. Through the feature extraction module, we can still extract its feature. However, it is still difficult to distinguish between dynamic gestures and static gestures. We use the distance to noise floor to solve this problem. For a single tag, we calculate the difference between its current signal and the floor noise. And then, a module operation and a square operation are performed on this difference to get the distance to noise floor for a single tag. The sum of distances of all tags forms the distance of the tag array. As shown in Fig. 4, when a user is performing a dynamic gesture, the distance to noise floor changes a lot due to the movement of the hand. On the contrary, for a static gesture, the change is relatively small. Therefore, we can determine whether the signal is from a static gesture through the variance of the signal. If the variance is smaller than a threshold, the corresponding signal is regarded as being from a static gesture.

4.3.1 Feature Decision

With the feature image sequence extracted, how can we decide the final feature and deal with it? If we use the CNN-LSTM structure above, it makes no sense for the LSTM part to extract temporal features because a static gesture keeps unchanged when acquiring the signal, and there is almost no temporal change relation between the two adjacent feature images. Therefore, for a static gesture, the key is still the spatial characteristic—how the gesture affects different parts of the RFID tag array? Given this, there are two strategies for getting the final feature.

One is compression. The whole feature image sequence is compressed into one feature image. In other words, the final feature is the superimposition of the images in sequence. In this way, additive noise may be suppressed. It can be regarded as a smoothing method, which smooths the possible noise, but smooth the feature to a certain extent in the meantime.

The other is extracting or directly picking one feature image as the final feature. This can be regarded as a snapshot of the static gesture or an instant feature of it. Without smoothing, instant feature will remain. But at the same time, possible noise may also remain.

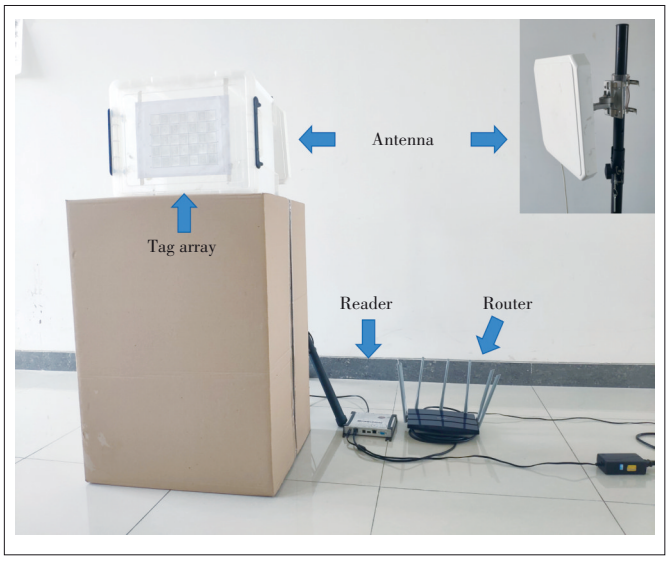
4.3.2 Network Structure

With the final feature image decided, the static gesture recognition problem is transformed into a feature image recognition problem. Here, we can adopt the CNN structure mention in Section 4.2.1 (the input layer, convolutional layer-1, pooling layer-1, convolutional layer-2, pooling layer-2, fully connected layer and output layer). The convolutional layers extract spatial features in the static gesture feature image, and the pooling layers down sample the features and reduce training overhead. The output layer outputs the final prediction vector.

5 Evaluation

5.1 Experiment Setup

Experiments are carried out in laboratory environment. As shown in Fig. 5, RF signal is transmitted through Impinj Speedway Revolution R420 UHF RFID Reader with laird s9028pcl RFID antenna. 5×7 AZ-9629 RFID tags are de-



▲Figure 5. Experiment deployment

ployed into a tag array and placed 0.5 m in front of the antenna. Users perform gestures in front of the array and PC gets signals from the reader through the router and then recognize them.

For dynamic gestures, six kinds of gestures are needed to be recognized: rotating left, rotating right, swiping left, swiping right, zooming in, and zooming out. We collect 3 600 samples, with 600 samples for each gesture. For static gestures, six kinds of gestures are needed to be recognized, including one-hand gestures (stop, victory, good, ok) and two-hand gestures (wrong, heart). We collect 600 samples, with 100 samples for each gesture. Fig. 6 shows the dynamic gestures and static gestures.

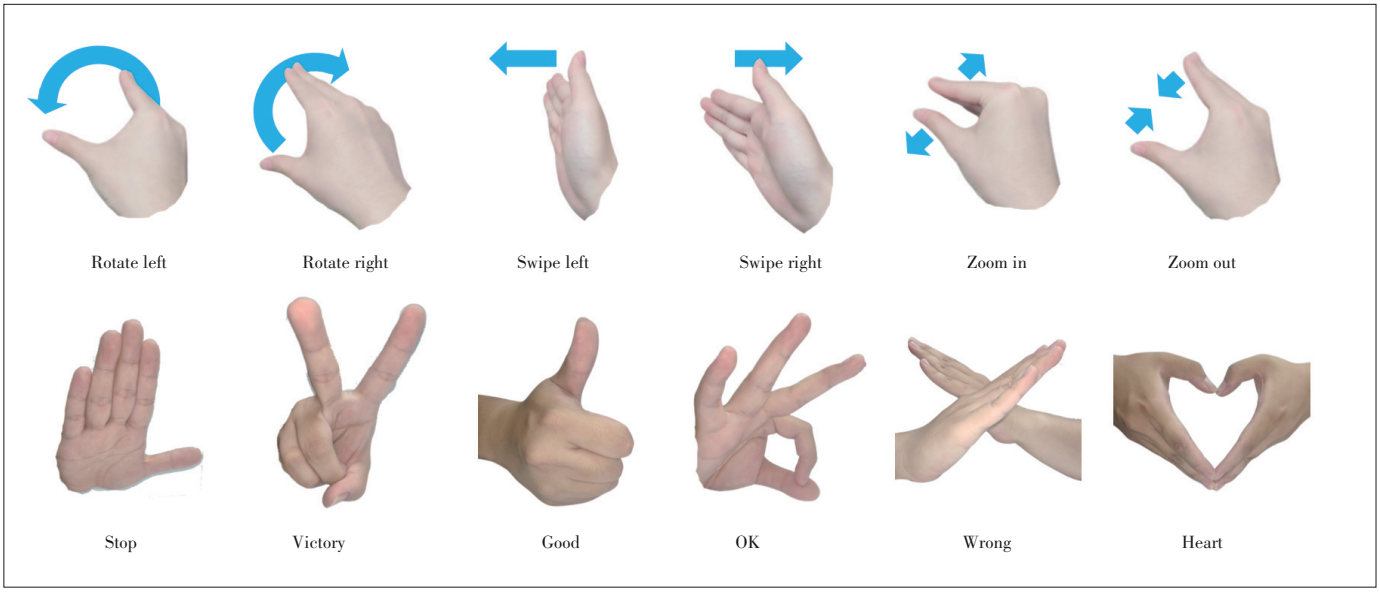
5.2 Evaluation on Model Structure

5.2.1 CNN Hyperparameter

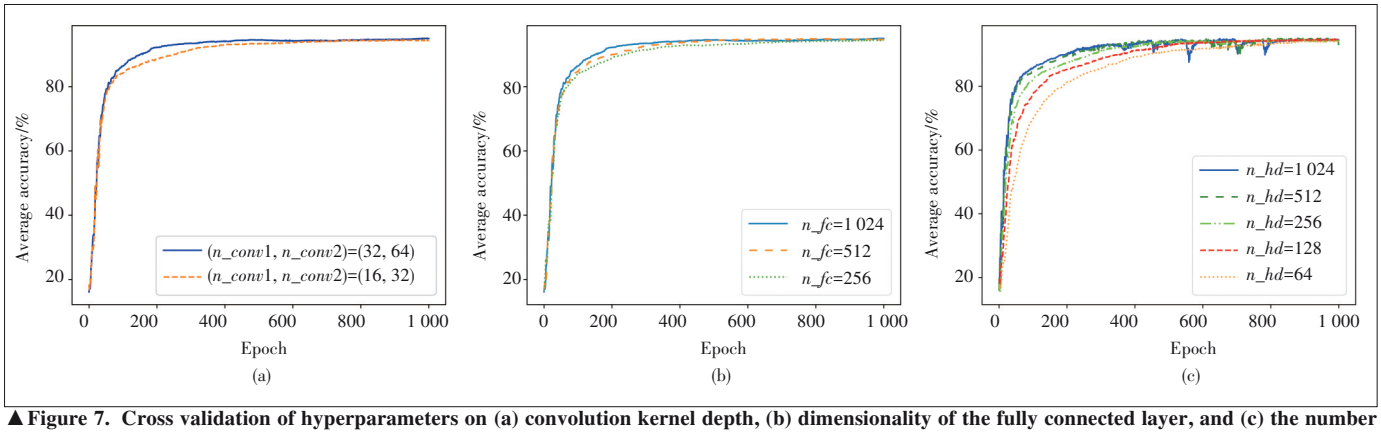
For the CNN structure mentioned in Section 4.2.1, we need to decide its hyperparameters(n_conv1, n_conv2, n_fc), that is, the kernel depths of the two convolutional layers and the feature dimensionality of the FC (fully connected) layer. To choose proper hyperparameters, we use ten-fold cross validation. The candidate values for hyperparameters (n_conv1, n_conv2) are $(32, 64)$ and $(16, 32)$. For a pure CNN structure, the average accuracy curve of different values on the validation set is shown in Fig. 7(a). Similarly, n_fc has candidate values as 1 024, 512 and 256 and the corresponding accuracy curve is shown in Fig. 7(b). The larger n_conv1 is, the less convergence time we need. Moreover, a small n_conv1 means that we can only extract a few features from the feature image sequences, limiting the capability of the model and even harming its performance. However, a too large hyperparameter also means the increased training costs and increased risk of overfitting. The same is true for the other two hyperparameters. Taking all these things into consideration, (n_conv1, n_conv2, n_fc) are set to $(32, 64, 1024)$. For CNN-LSTM structure, they are determined as $(32, 64, 256)$.

5.2.2 LSTM Hyperparameter

For the LSTM structure mentioned in Section 4.2.1, the hyperparameter is the number of hidden units n_hd. Ten-fold cross validation is used again. The average accuracy is shown in Fig. 7(c), with 1 024, 512, 256, 128 and 64 as candidate values. As we can see, if the number of hidden units is set too large, severe jitters will occur in the accuracy curve, indicating unstable performance of the model. But if it is set too small, it will take a long time to train and make the



▲Figure 6. Dynamic and static gestures used in our experiments



▲**Figure 7. Cross validation of hyperparameters on (a) convolution kernel depth, (b) dimensionality of the fully connected layer, and (c) the number of hidden units**

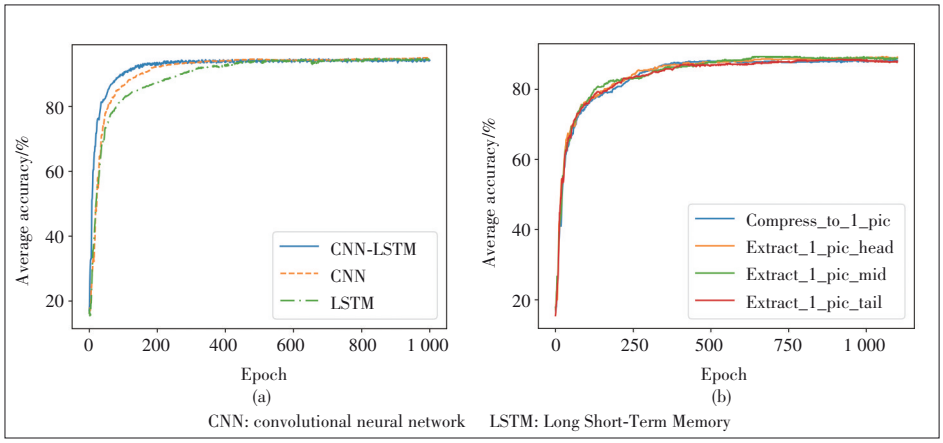
training expensive, limiting the ability of LSTM as well. Therefore, we choose 256 for *n_hd* in a pure LSTM structure and 128 in the CNN-LSTM structure.

5.2.3 Dynamic Gesture Model Selection

For dynamic gesture recognition, we also compare the three structures mentioned above through ten-fold cross validation. The average accuracy curve is shown in Fig. 8(a). In training, CNN focuses on spatial features and LSTM focuses on temporal features, while CNN-LSTM focuses on both of them. Therefore, CNN-LSTM facilitates learning more information in each epoch and costing fewer epochs to converge. That is why CNN-LSTM is chosen as the final network structure for dynamic gesture recognition.

5.2.4 Static Gesture Feature Decision

For static gesture recognition, we test the two strategies to get the final state through ten-fold cross validation. The first group use compression strategy to compress the whole feature image sequence into one picture. The second group use extracting strategy to pick feature images from the head, middle and tail of sequence respectively. From the test results shown in Fig. 8(b), four accuracy curves are close to each other, which means that both of the compression strategy and extracting strategy are feasible for recognition. Actually, these two strategies are tradeoff between smoothing noise and smoothing feature.



▲**Figure 8. Cross validation on (a) model selection for dynamic gesture and (b) feature decision for static gesture**

5.3 Evaluation of Dynamic Gesture Recognition

5.3.1 Overall Performance

With the 3 600 dynamic gesture samples, we take 600 of them as the test set and the rest as the training set. The accuracy on the test set is 92.17%, with the confusion matrix shown in Fig. 9(a). The accuracy of each gesture is above 89% and reaches up to 94%. According to the experiments, the system has accurate results in recognizing various dynamic gestures.

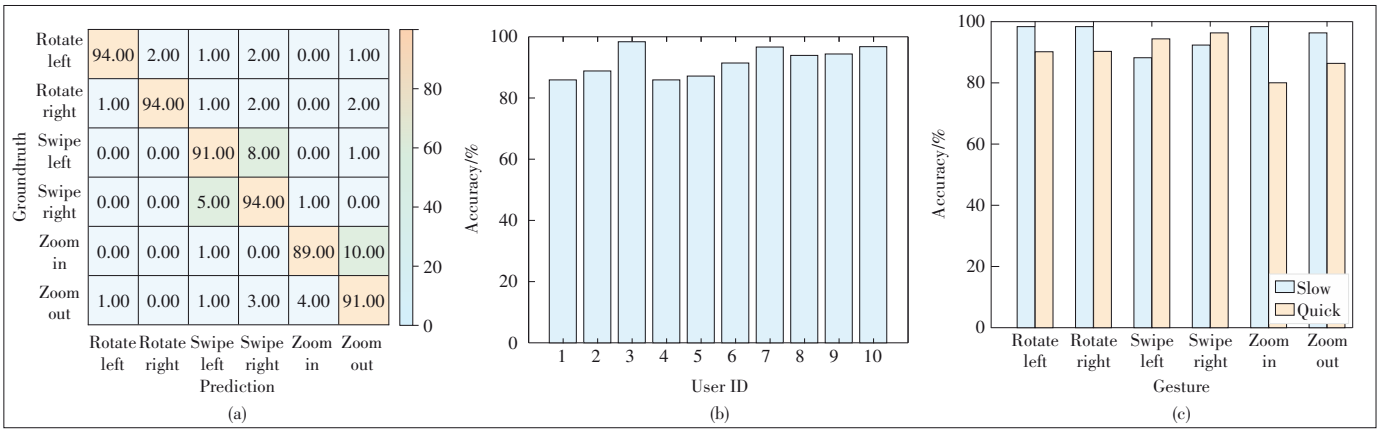
5.3.2 Evaluation on Different Users

For dynamic gestures, there are 10 users participating in the data collection of six gestures. As shown in Fig. 9(b), the accuracy of each user is above 85% and reaches up to 98.28%.

5.3.3 Evaluation at Different Speeds

Out of the 600 samples of each gesture, there are 300 fast gesture samples and 300 slow ones. The results on the test set are shown in Fig. 9(c). It can be seen that some types of gestures have higher accuracy at high speed but lower accuracy at low speed, and the other types have opposite situation. The accuracy of each gesture at different speeds is above 80% and reaches up to 96%.

The experiments show that the device-free in-air gesture recognition system based on RFID tag array can recognize several different types of



▲Figure 9. Evaluation of dynamic gesture recognition: (a) Confusion matrix, (b) accuracy of different users, and (c) accuracy at different speeds

dynamic gestures, and have high accuracy and robustness across different users at different speeds.

5.4 Evaluation of Static Gesture Recognition

5.4.1 Overall Performance

For the 600 static gesture samples, we take 60 of them as the test set and the rest as the training set. The accuracy on the test set is 91.67%. It can be seen from the confusion matrix shown in Fig. 10(a) that the accuracy of each gesture is above 80% and reaches up to 100%. The experiments show that the system has accurate results in recognizing various static gestures.

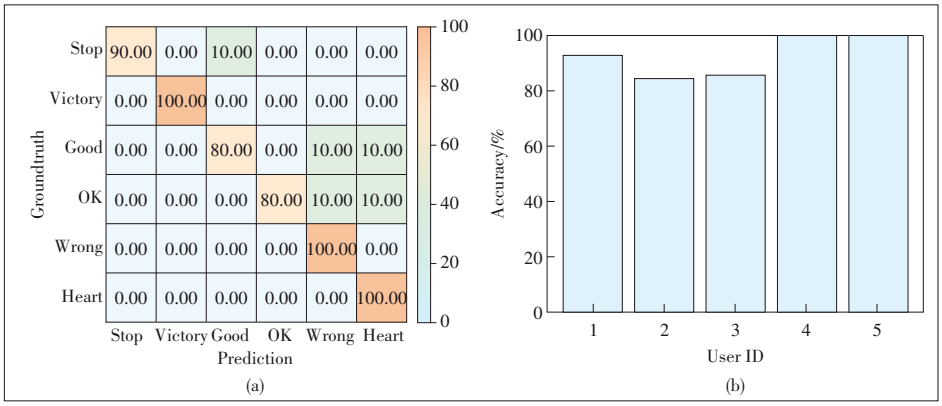
5.4.2 Evaluation on Different Users

There are five users participated in the data collection of six static gestures. The accuracy of different users is shown in Fig. 10(b). The accuracy of each user is above 84.62% and reaches up to 100%.

6 Discussion

1) Sensing distance. When performing a gesture, the recommended distance between the tag array and the hand is from 5 cm to 15 cm. If the hand is too far away from the tag array, the power of the reflected signal from the hand will be too small for us to extract the features, therefore harming the performance of the system. In this case, we need to increase the transmitting power of the reader, or even adopt beamforming technology to increase the power of the reflected signal. Besides, if the hand is too close to the tag array, it is likely to touch the tag array when performing gestures. This will change the input impedance of the tag and dramatically affect the received signal, reducing the accuracy of recognition. In this case, we can detect the touching action and require the user to repeat the gesture if we detect it. Another scheme is to build a more perfect reflection signal model that minimizes the impact of the touch action.

2) Hand size. The hand size will also affect the performance. When a user is performing gestures, signals reflected from the hand make different effects on different tags. The difference in hand sizes is not particularly significant in adults, so the effect on performance is not obvious. However, if the hand size is too small (such as a child's hand), the signal will be weak and affect the performance. To solve this problem, we may collect data of different hand sizes, thus making it possible for the model to extract size-independent features and improve the generalization ability.



▲Figure 10. Evaluation of static gesture recognition: (a) Confusion matrix and (b) accuracy at different speeds

7 Conclusions

This paper proposes a device-free method to recognize in-air dynamic and static gestures based on RFID tag array. The recognition system includes a feature extraction module and a gesture recognition module. Based on the feature image sequence extracted, we compare several structures and use CNN-LSTM to recognize dynamic gestures. For static gestures, the final feature is decid-

ed through two strategies and CNN is used for the recognition. Experiments show this method can recognize different gestures at different speeds across different users. The overall accuracy of the dynamic gesture test set is 92.17% and that of the static gesture test set is 91.67%.

References

[1] JIANG S, GAO Q, LIU H, et al. A novel, co-located EMG-FMG-sensing wearable armband for hand gesture recognition [J]. Sensors and actuators a: physical, 2020, 301: 111738. DOI: 10.1016/j.sna.2019.111738
[2] XIE R, CAO J. Accelerometer-based hand gesture recognition by neural network and similarity matching [J]. IEEE sensors journal, 2016, 16(11): 4537 – 4545. DOI: 10.1109/JSEN.2016.2546942
[3] SUN Y, WENG Y, LUO B, et al. Gesture recognition algorithm based on multi-scale feature fusion in RGB-D images [J]. IET image processing, 2020. DOI: 10.1049/iet-ipr.2020.0148
[4] WANG C, LIU Z, CHAN S C. Superpixel-based hand gesture recognition with kinect depth camera [J]. IEEE transactions on multimedia, 2014, 17(1): 29 – 39. DOI: 10.1109/TMM.2014.2374357
[5] MITRA S, ACHARYA T. Gesture recognition: A survey [J]. IEEE transactions on systems, man, and cybernetics, part C (applications and reviews), 2007, 37(3): 311 – 324. DOI: 10.1109/TSMCC.2007.893280
[6] RAUTARAY S S, AGRAWAL A. Vision based hand gesture recognition for human computer interaction: a survey [J]. Artificial intelligence review, 2015, 43(1): 1 – 54. DOI: 10.1007/s10462-012-9356-9
[7] BHASKARAN K A, NAIR A G, RAM K D, et al. Smart gloves for hand gesture recognition: sign language to speech conversion system [C]//International Conference on Robotics and Automation for Humanitarian Applications (RAHA). Coimbatore, India: IEEE, 2016: 1 – 6. DOI: 10.1109/RAHA.2016.7931887
[8] CHEN B, ZHANG Q, ZHAO R, et al. SGRS: a sequential gesture recognition system using COTS RFID [C]//IEEE Wireless Communications and Networking Conference (WCNC). Barcelona, Spain: IEEE, 2018: 1 – 6. DOI: 10.1109/WCNC.2018.8376998
[9] CHENG K, YE N, MALEKIAN R, et al. In-air gesture interaction: real time hand posture recognition using passive RFID tags [J]. IEEE access, 2019, 7: 94460 – 94472. DOI: 10.1109/ACCESS.2019.2928318
[10] HONGO H, OHYA M, YASUMOTO M, et al. Focus of attention for face and hand gesture recognition using multiple cameras [C]//Fourth IEEE International Conference on Automatic Face and Gesture Recognition Cat. No. PR00580. Grenoble, France: IEEE, 2000: 156 – 161. DOI: 10.1109/AFGR.2000.840627
[11] WANG W, LIU A X, SUN K. Device-free gesture tracking using acoustic signals [C]//22nd Annual International Conference on Mobile Computing and Networking. New York, USA: ACM, 2016: 82 – 94. DOI: 10.1145/2973750.2973764
[12] ABDELNASSER H, YOUSSEF M, HARRAS K A. Wigest: a ubiquitous wifi-based gesture recognition system [C]//IEEE Conference on Computer Communications (INFOCOM). Hong Kong, China: IEEE, 2015: 1472 – 1480. DOI: 10.1109/INFOCOM.2015.7218525
[13] LIU H, WANG Y, ZHOU A, et al. Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing [J]. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 2020, 4 (4): 1 – 28. DOI: 10.1145/3432235
[14] DING H, QIAN C, HAN J, et al. Rfipad: enabling cost-efficient and device-free in-air handwriting using passive tags [C]//37th International Conference on Distributed Computing Systems (ICDCS). Atlanta, USA: IEEE, 2017: 447 – 457. DOI: 10.1109/ICDCS.2017.141
[15] LIU J, CHEN X, CHEN S, et al. TagSheet: sleeping posture recognition with an unobtrusive passive tag matrix [C]//5th IEEE Conference on Computer Communications. Chengdu, China: IEEE, 2019: 874 – 882. DOI: 10.1109/INFOCOM.2019.8737599
[16] DOBKIN D M. The RF in RFID: passive UHF RFID in practice [M]. Oxford, UK: Newnes, 2007: 1 – 493. DOI: 10.1016/B978-0-7506-8209-1.X5001-3
[17] WANG C, LIU J, CHEN Y, et al. Multi-touch in the air: device-free finger tracking and gesture recognition via cots RFID [C]//IEEE Conference on Computer Communications. Honolulu, USA: IEEE, 2018: 1691 – 1699. DOI: 10.1109/INFOCOM.2018.8486346
[18] PASCANU R, MIKOLOV T, BENGIO Y. On the difficulty of training recurrent neural networks [C]//International Conference on Machine Learning. Atlanta, USA: PMLR, 2013: 1310 – 1318

Biographies

WU Jiaying is a Ph.D. student in the Department of Computer Science and Technology, Nanjing University, China, supervised by Prof. XIE Lei and WANG Chuyu. Her research interests include smart sensing and RFID.

WANG Chuyu (chuyu@nju.edu.cn) received his Ph.D. degree in computer science from Nanjing University, China in 2018. He is an assistant professor in the Department of Computer Science and Technology, Nanjing University, China. His research interests include RFID systems, software-defined radio, activity sensing, indoor localization, etc. WANG Chuyu is the corresponding author.

XIE Lei received his B.S. and Ph.D. degrees from Nanjing University, China in 2004 and 2010, respectively, all in computer science. He is a full professor in the Department of Computer Science and Technology, Nanjing University, China. He has published over 100 papers in *IEEE Transactions on Mobile Computing*, *IEEE Transactions on Parallel and Distributed Systems*, *ACM Transactions on Sensor Networks*, ACM MOBICOM, ACM UbiComp, ACM MobiHoc, IEEE INFOCOM, IEEE ICNP, IEEE ICDCS, etc.

Indoor Environment and Human Sensing via Millimeter Wave Radio: A Review

LIU Haipeng, ZHANG Xingyue, ZHOU Anfu, LIU Liang, MA Huadong
(Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: With the rapid development of 5G technology, more and more attention has been attracted to mmWave sensing. As an emerging sensing medium, mmWave has the advantages of both high sensitivity and precision. Different from its networking applications, the core method of mmWave sensing is to analyze the reflected signal changes containing the relevant information of different surrounding environments. In this paper, we conduct a systemic review for mmWave sensing. We first summarize the prior works on environmental sensing with different signal analysis methods. Then, we classify and discuss the work of sensing humans, including their behavior and gestures. Finally, we discuss and put forward more possibilities of mmWave human perception.

Keywords: millimeter wave radio; review; environment construction; human sensing; 5G

DOI: 10.12142/ZTECOM.202103004

https://kns.cnki.net/kcms/detail/34.1294.TN.20210806.1134.004.html, published online August 6, 2021

Manuscript received: 2021-06-10

Citation (IEEE Format): H. P. Liu, X. Y. Zhang, A. F. Zhou, et al., "Indoor environment and human sensing via millimeter wave radio: a review," *ZTE Communications*, vol. 19, no. 3, pp. 22 - 29, Sept. 2021. doi: 10.12142/ZTECOM.202103004.

1 Introduction

Millimeter wave (mmWave) communications are considered an essential component of 5G-and-beyond ultra-dense wireless networks, and the 5G breakout brings development opportunities to the study of mmWave sensing. Compared with traditional sound waves, ultrasonic waves and WiFi signals, mmWave sensing has great advantages, such as fine-grained resolution and the ability to detect subtle movements. Compared with camera-based sensors (such as visible light and infrared 3D structured light), mmWave can penetrate a few non-conductive objects including plastic, paper, glass, cloth, rain, and fog. Therefore, the unique sensing feature of mmWave has attracted more and more attention.

We, according to the different application scenarios, divide the existing methods into two categories.

1) Static indoor environment. The indoor structure being sensed is often utilized in robot vision and environment mapping, reacting upon indoor mmWave Wi-Fi networking. Depending on the capability of the equipment, we regroup the related work from a technology implementation perspective. We believe that such a classification is beneficial for researchers to quickly grasp not merely the cutting-edge works but also the technical details. In the beginning, this task is utilized for robot vision, i.e., help the robots discover obstacles in their routes[1]. Inspired by this application, researchers manage to "see" and build the surrounding construction through mmWave signals[2]. Utilizing the acquired surrounding spatial

The work is supported by the National Key R&D Program of China under Grant No. 2019YFB2102202, the A3 Foresight Program of NSFC (Grant No. 62061146002), the NSFC (61772084, 61720106007, 61832010), the Funds for Creative Research Groups of China (61921003), the Youth Top Talent Support Program, the 111 Project (B18008), and the Fundamental Research Funds for the Central Universities(2019XD-A13).

structure, researchers further facilitate the popularity of the mmWave Wi-Fi networking.

2) Dynamic human movements. Since mmWave signals have a short wavelength, it is sufficient to detect centimeter-level distance and sensitive to capture the millimeter-level motions around. Based on this knowledge, many researchers utilize mmWave signals to identify people through their gait[3], recognize several predefined hand gestures[4-5], and track someone's fingers[6]. More subtlely, the vital signs (such as respiratory and heart rates) can also be extracted by changes in reflected mmWave signals.

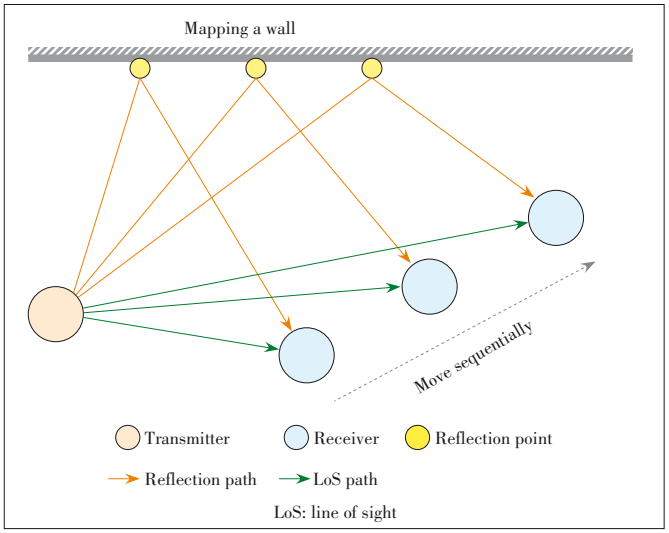
The core method of mmWave sensing is to analyze the signal changes reflected on the surrounding environment, so as to obtain its state. Therefore, in a practical operation process, we first need to use the transmitter module to transmit the signal and then receive the signal reflected and "modulated" by the surrounding environment.

The information of the target is contained in these signals, so the sensing systems then use different methods to "demodulate" the information of the surroundings, which will be processed and counted to get an overall message of the surrounding environment. However, there are two methods to perform the "demodulation": the traditional signal analysis and the AI-based learning. The former analyzes the angle and orientation of the reflection object by analyzing the changes in the communication information of the signal after reflection. The latter uses black-box machine learning methods to comprehend more complex human movements.

Interestingly, most studies on dynamic human movements rely on artificial intelligence (AI) technology because the features extracted from human movements are very complicated. These features include a great many reflection points with different distances, angles, speeds, and energy intensities. Therefore, the AI tool, originally designed for image learning, can discover the underlying relationship across such massive and complex feature information, so as to help the system sense human movements. Inversely, only a few studies on static indoor environments rely on AI. For the reader's convenience, we categorize these existing applications based on whether using AI and sensing scenarios. Based on this analysis, we explore each of the sensing scenarios in terms of breadth and depth and also put forward the prospect of mmWave sensing in future applications.

2 Sensing Indoor Environment

When networking in an indoor scenario, the available mmWave links often include multiple paths reflecting off the surrounding wall, in addition to the LoS path, as shown in Fig. 1. Different surrounding walls have different effects on the information of received signal strength (RSS) and phase of the mmWave link paths. Then by analyzing these effects, researchers can extract relevant information about the paths,



▲ **Figure 1. Indoor environment is inferred by concatenating the reflection points**

such as the time of flight, transmission length, or reflection angle. Using such information of paths, researchers construct the surrounding environment. Otherwise, researchers have focused on how to perceive the indoor environment to improve the efficiency and stability of 5G mmWave indoor networking.

The radar system, using frequency modulated continuous wave (FMCW) mmWave signals, is easy to identify the surrounding environment because it has the advantages of robustness, low computational complexity, strong penetration, etc. However, the current work is often done by reusing low-cost and ubiquitous 5G mmWave communication devices. Compared with radar devices, this device is difficult to identify the surrounding environment and thus is unable to handle complex environments. Therefore, how to reuse 5G signals to obtain the surrounding environment has become the key to the problem. According to the limitations of different devices and the way they use signals, we categorize the related works into threefold implemental technologies (categorized in Table 1).

2.1 Utilizing Both Phase and RSS

During the process of device localization for mmWave, we normally have the problem of knowing nothing about the initial surrounding environment. To solve this, PALACIOS et al. design JADE[2], which can estimate the location of a mobile user in indoor space without any prior knowledge. The JADE algorithm can be directly combined with a commercial device to extract information using the angle of arrival (AoA) of signals, so as to obtain the location of users. Then, they propose CLAM[7] to localize the mobile user, estimate the position of access points, and finally form a map of the environment. Using these two algorithms, they realize the localization and mapping of the mmWave network without knowledge of the initial environment.

▼Table 1. Comparison of the state-of-the-art works on sensing static indoor environment

Citation	Usage	Technology	Method
JADE	Position the client	An iterative algorithm	Both phase and RSS
CLAM	Map the environment	A distributed localization algorithm	Both phase and RSS
E-Mi	Boost the indoor network	A multi-path analysis framework	Both phase and RSS
Beam-Forcast	Improve the mobile links	Reverse engineering of E-Mi	Both phase and RSS
RSA	Position object & identify materials	Move Rx along trajectory while collecting	Only RSS
Ulysses	Image environment	Bind Tx & Rx together to collect reflection signals	Only RSS
RadarCat	Identify plenty of materials	Classification by learning on the radar signals	Only RSS
mmRanger	Sense environment	Automatically collect signals	RSS with a robot
miDroid	Improve mobile links	Set a network relay piggybacked on the robot	RSS with a robot

RSS: received signal strength Rx: receiver Tx: transmitter

CLAM combines experiments with simulation because JADE is entirely dependent on simulation results. Thus, the above two works are too dependent on the simulation experiments, which theoretically discuss the possibility of indoor environment perception. E-Mi[8] achieves the same function experimentally. Specifically, E-Mi proposes a multi-path analysis framework, using a customized 360 omnidirectional antenna to obtain the information of the <angle, length> of all reflection links through the RSS and phase information of the signal, so as to infer the surrounding environment according to the geometric relationship. Based on E-Mi's equipment and method, Beam-Forcast[9] makes a reverse engineering, i. e., how to accelerate the alignment of network link for a mobile client through the angle change of the client in a given environment.

2.2 Utilizing Only RSS

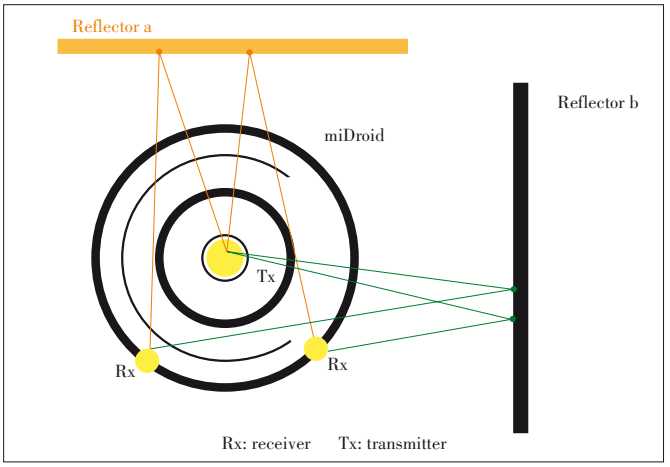
The above methods can acquire the AoA information directly, but commercial off-the-shelf (COTS) mmWave nodes in daily life cannot accurately extract phase information. This is because the daily mmWave nodes use ideal laser-like beams, which are generated by horn antennas. Due to its lack of ability to maintain the phase offset between the transmitter (Tx) and the receiver (Rx), a lot of small phase jitters may cause a large error in practical applications. To solve this question, researchers try to determine the AoAs of the link by adjusting the Rx's orientation (i.e., the receiving direction with the largest RSS).

ZHU et al. design RSA[1] which moves radio antennas while collecting reflection signals to create a synthetic aperture radar (SAR) to infer the object's boundaries, curvature, and surface material. Specifically, RSA fixes the Tx, a commodity HXI Gigalink 6 541 board, and moves the Rx with a deliberately designed path collecting reflection signals. In this process, RSA can extract the detected object's status including both its curve direction (by aligning its receiver orientation towards the strongest one) and its material according to the signal strength. Furthermore, Ulysses[10] binds Tx and Rx (the same as that used by RSA) together and moves them at the same time. According to this, the system can prevent collision in motion, which can be used in future scenarios, e.g., the unmanned driving and robot movement.

Uniquely, utilizing RSS of the reflected signals, RadarCat[11] can identify different materials of plenty of static objects, such as glasses, water, metals, and mobile phones. Specifically, RadarCat is based on a principle that the signals reflected from different materials are highly characteristic because the thickness and geometry of the object will scatter, refract and reflect the radar signals differently. Therefore, RadarCat feeds the 8-channel radar signal RSS and their statistics into a well-trained classifier to distinguish the different objects.

2.3 Utilizing RSS with a Programmable Robot

All of the above indoor environment mapping works (except RadarCat) need to manually adjust the position and orientation of the devices, so researchers try to automate the task of collecting signals with the help of a programmable sweeping robot as shown in Fig. 2, which is precise to position, steer and track. With the help of this robot, mmRanger[12] can use the commercial mmWave network cards to achieve the construction of and indoor environment by only utilizing the correlation between AoA and RSS. Specifically, mmRanger carries two mmWave cards on a commercial cleaning robot to make all of them as a whole, i.e., a smart mobile environment sen-



▲Figure 2. A programmable sweeping robot that works precisely to position, steer, and track

sor. The robot can sense the environment by exchanging mmWave signals between the two cards when it moves and rotates freely inside the target room. Upon this process, the geometry of the multiple reflection paths can be extracted by mmRanger from their RSS sequence, and then the environment layout can be reconstructed.

Actually, smart robots will eventually become a part of the home and enterprise environment to help automate our daily lives and improve productivity. Based on this vision, miDroid[13] binds the same network card node to the same robot, turning the robot into an indoor mmWave Wi-Fi relay and thereby achieving faster mobile client network optimization. More specifically, miDroid firstly analyzes the series of access point (AP) beacons to extract spatial factors and then finds out the AoA/angle of departure (AoD) of signal transmission paths, so as to map the environment. Then, miDroid proposes a real-time and adaptive path planning algorithm to instruct the navigation of the robot relay, thus improving the performance during the client's blockage when he changes orientation rapidly.

3 Sensing Human

Different from the last section, human movements simultaneously generate multiple reflection points each with different position, velocity, reflection intensity, etc. Therefore, two research methods are adopted according to the complex degree of the sensing movements (categorized in Table 2).

3.1 Human Movement Tracking

Several works are aimed only at small, uniform movements on a certain part of the human body, such as vertical finger tracking, regular breathing, and heartbeat. Specifically, researchers utilize horn antennas to focus on the detected target, thereby eliminating the interference of unrelated motions around, as shown in Fig. 3. In this way, any variation in the reflection signal represents the change of the target to sense changes in the target part of the human body.

To focus on a periodic movement, e.g., human chest variation caused by the breath and heartbeats, YANG et al.[14] propose mmVital, a system using 60 GHz mmWave signals for vital sign monitoring. In this system, they utilize two horn antennas fixing their orientation to continuously capture the subtle variation of the skin on one's chest by analyzing the signal modulation caused by the skin fluctuation. Specifically, they extract the periodic changes within the signal RSS to acquire the frequency of breath and heartbeat and filter the raw signals through the suited band-pass filter to achieve better accuracy. As a result, mmVital provides a mean estimation error of 0.43 breaths and 2.15 breaths per minute within 100 ms of dwell time on reflection.

On the other hand, for the non-periodic movement, e.g., tracking a finger, the horn antenna needs to determine the direction of the finger movement by analyzing the signal changes to realize finger tracking. WEI et al.[6] trace a rectangle area as the tracking region in their proposed tracking system mTrack. Then, they place a quasi-omni-directional (180° beamwidth) transmitter on one of the rectangle vertexes and two horn-antenna receivers on the adjacent sides facing the region respectively. In this region, finger movements in any direction will be transformed into two relative movements: approaching or moving away from the two receivers. Fortunately, these two movements can be detected by the phase changes of the signal thank to the small wavelength of 60 GHz mmWave. As a result, mTrack can track a vertical finger with a 90th percentile error below 8 mm which is sufficient for a virtual trackpad.

In addition to the above work on sensing a certain part of

▼Table 2. Comparison of the state-of-the-art works on sensing dynamic human movements

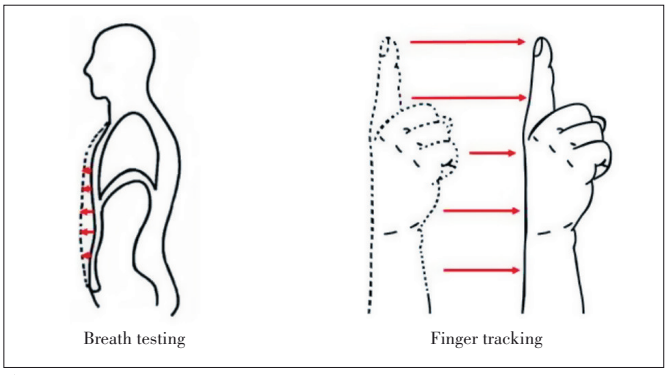
Citation	Usage	Equipment	Technology	Whether AI
Deep-soli	Gesture recognition	Soli	Range-Doppler images	Customized CNN
Ubi-soli	Gesture recognition	Soli	Multiple abstract representations	Random forest
HCI	Vehicular gesture recognition	Soli	Several physical features	Random forest
MengZhenGait	Identify different users	TI-IWR1443&6843	Three spatial features	MmGaitNet
MmVital	Monitor vital signs	Horn antennas	Extract periodic changes	No
MTrack	Track a vertical finger	Horn antennas	Combine target's angle and phase change	No
MmSense	Multi-human detection	Free	Features of 60 GHz signal	No
MID	Gesture recognition	Free	MmWave sensing	No
LowCostGes	Detect gesture	RIC60a	Extract power profile and AoA	No
MmASL	Home-assistant system	Free	Features of 60 GHz signal	No
MHomeGes	Smart home-usage	TI-IWR1443	MmWave sensing	No
MTranseSee	User recognition	Free	MmWave radar	No
Pantomime	Gesture recognition	Pantomime	MmWave sensing	No
MmMesh	Human mesh construction	Free	Deep learning framework	No

AI: artificial intelligence AoA: angle of arrival CNN: convolutional neural network HCI: human computer interface

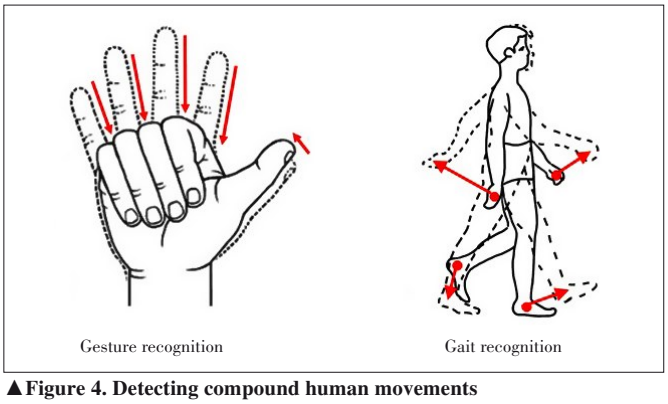
the human body, there are also some work on human detection. For example, GU et al.[15] propose mmSense, a device-free multi-person detection framework. During this work, they use the properties of 60 GHz signal for human bodies and objects to fingerprint the environments including and excluding humans. Then based on the monitored 60 GHz signals and generated fingerprints of environments, mmSense can simultaneously detect the presence and locations of multiple persons. Furthermore, by correlating the 60 GHz RSS series with the measurement of different people's outlines and vital signs, they propose a new method to identify multi-person. Finally, they demonstrate the effectiveness and low cost of this method through experiments. Besides, in the domain of user recognition, LIU et al.[16] propose mID, the first user identification approaches that utilize mmWave signals.

3.2 Complicated Behavior Sensing

As aforementioned, the movement of the human body creates a large number of reflection points with different spatial features, as shown in Fig. 4. So we have to treat the human body as a soft body instead of a rigid body (e.g., a wall). In order to obtain the position-scattering points, all of the work utilizes FMCW, which can separate the reflection points with their spatial features, i.e., distance, velocity, and angle. With the different features of the points on a human body, his/her movements can be recorded in detail. However, it is too complex to recognize human movements by analyzing different changing features with simple geometrical relationships. Therefore, the AI tool acting as a black box is utilized to help the system sense human movements.



▲Figure 3. Detecting simple human movements



▲Figure 4. Detecting compound human movements

To recognize gestures, WANG et al.[4] directly utilize a mature method, Range Doppler (RD), in the FMCW signal processing field for each fleeting period slice of a gesture. Then, to classify these RD sequences, they input the RD image of each slice into a customized convolutional neural network (CNN) and put the discrete recognition results from each slice into a recurrent neural network (RNN). As a result, this work achieved 87% accuracy on 11 gestures. Crucially, the hardware in this work is the first small-size mmWave radar called "soli"[5], which transmits omni-directional FMCW signals to sense the gesture of the nearby environment. To verify its gesture sensing ability, the researchers, in addition to the above RD metrics, also introduce two series of features, i.e., in-phase/quadrature (I/Q) statistics, and some tracking information. In particular, the former is useful for detecting micro motions and the latter is beneficial for recording the moving trend of the gesturing hand. Then, these features are fed into a random forest (RF) classifier due to its computational speed, low memory consumption, and generalization ability.

Totally, this work is sufficient to recognize four gestures with 92.10% accuracy. Furthermore, using the same chip, SMITH et al.[17] implement the gesture recognition into a human-car interface with also an RF classifier. Based on the above solutions, PATRA et al.[18] present a low-cost mmwave radar-based system to save the computation resource. They detect gestures only by the AoA and the power profile extracted from the measured signal. Specifically, they use two low-complexity classification algorithms: unsupervised self-organization mapping (SOM) and supervised learning vector quantization (LVQ). With these methods, gesture recognition can reach 87% accuracy for some gestures.

Gesture recognition can be applied to home assistants. For example, the user can wave one's hand at a distance to turn on a TV. To achieve gesture recognition in such home scenarios, there are several vital challenges: 1) With the increase of the detection range, the details of gestures will become difficult to be captured due to the severe attenuation of the mmWave transmission in the air. In this case, the reflection points will be sparser when the user is standing at a distance. 2) There are also plenty of non-target movements, i.e., they are not predefined gestures but are performed in our daily life. In this case, the points, which are generated by reflection on the non-target movements or even by the multipath reflection on the user's ambient appliances, will interfere with the original recognition.

To solve the problem, LIU et al.[19] propose mHomeGes, a real-time smart home gesture recognition system completely using mmWave. First, they obtain the position and dynamic variation of the gesture. Then they recognize fine-grained gestures

by using a lightweight CNN. Next, they propose a user-discovery approach to focus on target human gestures, eliminating the adverse effects of surrounding interference. Finally, they realize the continuous gesture recognition in real time. In the end, mHomeGes achieved more than 95.30% of high-precision recognition in real-time smart home scenarios, successfully solving these problems. On the other hand, PALIPANA et al.[20] use a Pointnet++ and LSTM combination to extract the spatio-temporal feature of point clouds. In this case, they build a 4D point cloud classification architecture that feeds on the point clouds directly to recognize the gestures. SANTHALINGAM et al.[21] also try to recognize the movements (e.g., American sign language recognition for the deaf and hard-of-hearing people) at a distance, but they highly rely on cumbersome devices. Based on the home-scenario gesture recognition, LIU et al. further propose mTranSee[22] to largely reduce the adaptation effort in a smart home scenario via transfer learning, which promotes the practicability of mmWave gesture sensing.

Based on the similar inspiration of the point cloud, MENG et al.[3] utilize FMCW mmWave signals to realize the gait identification by three spatial features, i.e., three-dimensional (3D) coordinates in space, speed relative to the radar, and the energy intensity of each reflection point on a human body, respectively. In order to adapt these three features, they also design a novel AI algorithm, i.e., GaitNet, which concatenates five separated attribute networks (the first three are for the 3D coordinates) and feeds them into a fusion network. As a result, mmGaitNet achieves 90% and 88% accuracy for single person and five co-existing person scenarios, respectively.

Since point clouds reflect the movement of the human body, and millimeter waves can detect tiny movements, can we construct the human body through millimeter wave signals? For this, XUE et al.[23] propose mmMesh, a real-time 3D human mesh estimation system. This system can accurately align the 3D points with the corresponding body segments. With this approach, the lost part due to the sparsity of mmWave point cloud is introduced from the information of the previous frame so that a dynamic human mesh is completed.

4 Discussion

According to different sensing tasks (environment/human body) and core methods (AI/non-AI), we cross-classify the existing methods across application scenarios (categorized in Table 3). Next, we will discuss the deficiency of mmWave sensing-related work and look forward to the other perception applications.

▼Table 3. Existing work and prospect for the future

	AI	Non-AI
Human dynamics	2,3,5,6	1,7,16,20,21,22,23
Indoor environment	8	9,11,12,13,14,15,17,18,19

AI: artificial intelligence

4.1 Deficiencies of State-of-the-Art Work

Deficiencies of the state-of-the-art work are introduced in the following scenes.

1) Sensing the indoor environment

For environment construction, the current work lacks the ability of complex environment construction, i.e., it is limited to smaller indoor environments. Moreover, the current work can only identify simple construction, i.e., it is difficult to distinguish the details of the surroundings (e.g., a safe in the corner), so the challenge is to study how to accurately sense the details. None of the methods have ever tried to combine efficient AI tools with indoor environment construction. There is only one work[11] that uses the AI method to perceive the materials. For material reflected, though mmWave-based material recognition is currently used in security inspections, there are still problems like the bulky and inefficient machines.

2) Sensing human dynamics

Gestures are currently sensed only in an ideal environment, i.e., the researchers only allow the user to perform the gestures in several fixed positions. For monitoring vital signs, current work only extracts the frequency of a person's breath rate and heartbeat in calm status. Therefore, when the solution is applied at home, it is difficult to track a moving human body. Moreover, the current work only determines the breath and heartbeat rates approximately through the periodicity of the signal. However, compared with more accurate medical devices upon bioelectric signals, it greatly lacks accuracy and credibility.

4.2 Future Outlook on Novel Applications

We also look into the future of novel applications as follows.

1) Sensing indoor environment

For environment sensing, since the current work has achieved the ability of simple indoor environment construction, in the future the same method will meet new challenges when applied to a wider space (e.g., larger office space). Furthermore, to improve the accuracy of recognition of the surrounding environment, smaller FMCW radars can be used to achieve the same ability. However, FMCW radars are sensitive to minor moving variations which may affect the detection ability of small metal objects. Moreover, since mmWave cannot finely image the sensing objects, it may provide important assistance for the vision of some certain unmanned systems like domestic robots which need to protect user's privacy security.

2) Sensing human dynamics

For human dynamics sensing, everyone making the same gesture will have slight differences in the motion habits, so we can also use the same gesture to authenticate different users. But it is important to extract the unique information differentiating host gestures from others. For monitoring vital signs, since the current work of measuring heartbeat or breath re-

quires the person to be still, one can explore how to measure these vital signs in a dynamic process. In the future, once implementing the mobile tracking monitoring method, we can also track the user's living habits and real-time health conditions, so as to better help people improve their living quality. Moreover, the data collected by medical devices can supervise the mmWave signal training on the AI algorithm, thereby judging the breath and heart rates more precisely, reliably and automatically.

As can be seen from Table 3, only Ref. [11] is based on the AI method and explores the indoor environment. Based on RadarCat's inspiration, we can also learn about changes in signal strength to detect changes in the body's blood sugar concentration "in air", therefore reducing the pain of patients. Similarly, this approach could be applied to screening passengers for sensitive metal objects at security checkpoints, reducing the work of staff. Furthermore, researchers can further explore this direction, e.g., AI algorithm can learn the changes in a signal state to judge the tendency of users to move in different positions in the room, so as to improve the performance of the mmWave network more quickly instead of the traditional iterative algorithms.

Besides, there are only two existing tasks based on non-AI and human dynamics, so researchers can design more targeted and simple algorithms, e.g., tracking the angle of the swiping hand to determine its direction and speed. Furthermore, the hand location in space can be obtained by using the range and angle information of its reflection point cloud to realize hand tracking. Based on this, researchers can also realize a low-power virtual reality (VR) game sensor based on mmWave.

5 Conclusions

The state-of-the-art mmWave sensing solutions have performed basic functions: constructing the general indoor environment and recognizing the dynamics of the human body. We classify the existing work according to their sensing tasks, i.e., static indoor environment and dynamic human movements, and then introduce the characteristics and advantages separately. Finally, we use a table to further classify the work and present a forward look at future work in this field.

References

[1] ZHU Y Z, ZHU Y B, ZHAO B Y, et al. Reusing 60 GHz radios for mobile radar imaging [C]//The 21st Annual International Conference on Mobile Computing and Networking. Paris, France: ACM, 2015: 103 – 116. DOI: 10.1145/2789168.2790112
[2] PALACIOS J, CASARI P, WIDMER J. JADE: Zero-knowledge device localization and environment mapping for millimeter wave systems [C]//IEEE Conference on Computer Communications. Atlanta, USA: IEEE, 2017: 1 – 9. DOI: 10.1109/INFOCOM.2017.8057183
[3] MENG Z, FU S, YAN J, et al. Gait recognition for co-existing multiple people using millimeter wave sensing [C]//The AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020, 34(01): 849 – 856. DOI: index.php/AAAI/article/view/5430
[4] WANG S W, SONG J, LIEN J, et al. Interacting with soli: exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum [C]//The 29th Annual Symposium on User Interface Software and Technology. Tokyo, Japan: ACM, 2016: 851 – 860. DOI: 10.1145/2984511.2984565
[5] LIEN J, GILLIAN N, KARAGOZLER M E, et al. Soli: ubiquitous gesture sensing with millimeter wave radar [J]. ACM transactions on graphics, 2016, 35(4): 1 – 19. DOI: 10.1145/2897824.2925953
[6] WEI T, ZHANG X Y. MTrack: high-precision passive tracking using millimeter wave radios [C]//The 21st Annual International Conference on Mobile Computing and Networking. Paris, France: ACM, 2015: 117 – 129. DOI: 10.1145/2789168.2790113
[7] PALACIOS J, BIELSA G, CASARI P, et al. Communication-driven localization and mapping for millimeter wave networks [C]//IEEE Conference on Computer Communications. Honolulu, USA: IEEE, 2018: 2402 – 2410. DOI: 10.1109/INFOCOM.2018.8485819
[8] WEI T, ZHOU A, ZHANG X. Facilitating robust 60 GHz network deployment by sensing ambient reflectors [C]//14th USENIX Symposium on Networked Systems Design and Implementation. Boston, USA: IEEE, 2017: 213 – 226
[9] ZHOU A F, ZHANG X Y, MA H D. Beam-forecast: facilitating mobile 60 GHz networks via model-driven beam steering [C]//IEEE Conference on Computer Communications. Atlanta, USA: IEEE, 2017: 1 – 9. DOI: 10.1109/INFOCOM.2017.8057188
[10] ZHU Y, YAO Y, ZHAO B Y, et al. Object recognition and navigation using a single networking device [C]//Annual International Conference on Mobile Systems, Applications, and Services. Niagara Falls, USA: ACM, 2017: 265 – 277. DOI: 10.1145/3081333.3081339
[11] YEO H S, FLAMICH G, SCHREMPF P, et al. RadarCat: radar categorization for input & interaction [C]//The 29th Annual Symposium on User Interface Software and Technology. Tokyo, Japan: ACM, 2016: 833 – 841.DOI: 10.1145/2984511.2984515
[12] ZHOU A F, YANG S Y, YANG Y, et al. Autonomous environment mapping using commodity millimeter-wave network device [C]//IEEE Conference on Computer Communications. Paris, France: IEEE, 2019: 1126 – 1134. DOI: 10.1109/INFOCOM.2019.8737624
[13] ZHOU A F, XU S Q, WANG S, et al. Robot navigation in radio beam space: Leveraging robotic intelligence for seamless mmWave network coverage [C]//The 20th ACM International Symposium on Mobile Ad Hoc Networking and Computing. Catania, Italy: ACM, 2019: 161 – 170. DOI: 10.1145/3323679.3326514
[14] YANG Z C, PATHAK P H, ZENG Y Z, et al. Monitoring vital signs using millimeter wave [C]//Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing. Paderborn, Germany: ACM, 2016: 211 – 220. DOI: 10.1145/2942358.2942381
[15] GU T B, FANG Z, YANG Z C, et al. MmSense: multi-person detection and identification via mmWave sensing [C]//The 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems. Los Cabos, Mexico: ACM, 2019: 45 – 50. DOI: 10.1145/3349624.3356765
[16] LIU H P, BAI X W, GAO H Y, et al. MID: accurate and robust user identification and authentication through hand-gesture sensing with mmwave radar [J]. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 2021.
[17] SMITH K A, CSECH C, MURDOCH D, et al. Gesture recognition using mm-wave sensor for human-car interface [J]. IEEE sensors letters, 2018, 2 (2): 1 – 4. DOI: 10.1109/LSENS.2018.2810093
[18] PATRA A, GEUER P, MUNARI A, et al. Mm-wave radar based gesture recognition: development and evaluation of a low-power, low-complexity system [C]//The 2nd ACM Workshop on Millimeter Wave Networks and Sensing Systems. New Delhi, India: ACM, 2018: 51 – 56. DOI: 10.1145/3264492.3264501
[19] LIU H P, WANG Y H, ZHOU A F, et al. Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing [J]. Proceedings of the ACM

on interactive, mobile, wearable and ubiquitous technologies, 2020, 4(4): 140. DOI: 10.1145/3432235
[20] PALIPANA S, SALAMI D, LEIVA L A, et al. Pantomime: mid-air gesture recognition with sparse millimeter-wave radar point clouds [J]. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 2021, 5(1): 1 – 27. DOI:10.1145/3448110
[21] SANTHALINGAM P S, HOSAIN A A, ZHANG D, et al. MmASL: environment-independent asl gesture recognition using 60 GHz millimeter-wave signals [J]. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 2020, 4(1): 1 – 30. DOI: 10.1145/3381010
[22] LIU H P, CUI K N, HU K Y, et al. Environment-independent mmWave sensing based gesture recognition via transfer learning [J]. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 2021
[23] XUE H, JU Y, MIAO C, et al. MmMesh: towards 3D real-time dynamic human mesh construction using millimeter-wave [C]//Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services. 2021: 269 – 282. DOI: 10.1145/3458864.3467679

Biographies

LIU Haipeng (fengxudi@qq.com) is a Ph.D. student of Beijing University of Posts and Telecommunications, China. He is directly recommended for a doctorate degree from undergraduate. His research interest lies in wireless networks and mobile sensing, with the emphasis on ubiquitous mobile sensing for enabling new Internet-of-Things applications.

ZHANG Xingyue is a second-grade student at the International School of Beijing University of Posts and Telecommunications, China. She is currently majoring in telecommunications engineering and management, and is about to enter the direction of multimedia learning.

ZHOU Anfu received his Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Science, China in 2012. Before that, he received B.S. degree from the Renmin University of China. He is now a professor at the School of Computer Science, Beijing University of Posts and Telecommunications, China. His research interest lies in mobile computing and wireless networking and IoT systems.

LIU Liang obtained his Ph.D. degree from Beijing University of Posts and Telecommunications, China in 2009, and conducted visiting research in TAMU in the US. His main research directions are multimedia and sensor networks. He has published nearly 40 papers in international authoritative journals and conferences such as *IEEE Transactions* and INFOCOM.

MA Huadong is an IEEE Fellow. He received his Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Science, China in 1995. He is now a Chang Jiang Scholar Professor and Director of Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia. He is also the Executive Dean of School of Computer Science, Beijing University of Posts and Telecommunications, China. His research interests include multimedia systems and networking, sensor networks, and the Internet of Things. He has authored over 200 papers and four books.

Using UAV to Detect Truth for Clean Data Collection in Sensor-Cloud Systems

LI Xiuxian[1], LI Zhetao[1], OUYANG Yan[2], DUAN Haohua[3], XIANG Liyao[3]

(1. Xiangtan University, Xiangtan 411100, China;
2. Central South University, Changsha 410000, China;
3. Shanghai Jiao Tong University, Shanghai 200000, China)

Abstract: Mobile edge users (MEUs) collect data from sensor devices and report to cloud systems, which can facilitate numerous applications in sensor-cloud systems (SCS). However, because there is no effective way to access the ground truth to verify the quality of sensing devices' data or MEUs' reports, malicious sensing devices or MEUs may report false data and cause damage to the platform. It is critical for selecting sensing devices and MEUs to report truthful data. To tackle this challenge, a novel scheme that uses unmanned aerial vehicles (UAV) to detect the truth of sensing devices and MEUs (UAV-DT) is proposed to construct a clean data collection platform for SCS. In the UAV-DT scheme, the UAV delivers check codes to sensor devices and requires them to provide routes to the specified destination node. Then, the UAV flies along the path that enables maximal truth detection and collects the information of the sensing devices forwarding data packets to the cloud during this period. The information collected by the UAV will be checked in two aspects to verify the credibility of the sensor devices. The first is to check whether there is an abnormality in the received and sent data packets of the sensing devices and an evaluation of the degree of trust is given; the second is to compare the data packets submitted by the sensing devices to MEUs with the data packets submitted by the MEUs to the platform to verify the credibility of MEUs. Then, based on the verified trust value, an incentive mechanism is proposed to select credible MEUs for data collection, so as to create a clean data collection sensor-cloud network. The simulation results show that the proposed UAV-DT scheme can identify the trust of sensing devices and MEUs well. As a result, the proportion of clean data collected is greatly improved.

Keywords: sensor-cloud system; truth detection; trust reasoning and evolution; mobile edge user; unmanned aerial vehicle

DOI: 10.12142/ZTECOM.202103005

https://kns.cnki.net/kcms/detail/34.1294.TN.20210818.1538.005.html, published online August 18, 2021

Manuscript received: 2021-06-08

Citation (IEEE Format): X. X. Li, Z. T. Li, Y. Ouyang, et al., "Using UAV to detect truth for clean data collection in sensor-cloud systems," *ZTE Communications*, vol. 19, no. 3, pp. 30 - 45, Sept. 2021. doi: 10.12142/ZTECOM.202103005.

1 Introduction

With the development of techniques on microprocessor industry, sensing-based devices are becoming smaller while their computation and capacities are strengthened gradually[1-3]. Therefore, sensing technologies are widely deployed in areas with on-demand monitoring processes. According to a survey, there were more than 20 billion devices connected to the Internet of Thing (IoT) in 2020 and the number is growing at a faster rate[4-6]. These IoT devices are equipped with numerous sensing devices to realize the perception of the surroundings[9-10]. Thus, the sensor-cloud systems (SCS), within which the IoT devices and cloud services are well combined, can be more productive and effective on its functionality and solve such problems as the

This work was supported in part by National Natural Science Foundation of China under Grant No. 62032020, Hunan Science and Technology Planning Project under Grant No.2019RS3019, and the National Key Research and Development Program of China under Grant 2018YFB1003702.

sharing of sensor nodes and large amounts of data analysis due to memory and energy limitations[9]. In an SCS, a huge number of sensing devices are deployed at the edge of the network to sense the surrounding environment[11–13], and then upload the sensed data to the cloud. Due to the excellent computing power, cloud services can perform sophisticated computation and analytics, as well as orchestrate various applications. For example, the supervisory control and data acquisition (SCADA) system is one of the SCS and composes of smart sensing devices spreading over a wide area in order to remotely monitor physical phenomena[14]. These smart sensing devices can be deployed on demand in the areas that require temporary testing, and then collect data into the cloud in various ways to initiate and build up various applications[15–16]. The method of data collection has also changed a lot from the traditional methods in the past. In traditional wireless sensor networks (WSNs), many nodes are deployed in specific areas and self-organize into a network. The sensed data is routed to a specific node called sink through multi-hop routing[17–18] and the sink is connected to the Internet by a wired network; in this way, the data are reported to the cloud. However, the time and economic cost of deploying the network to establish the connection with the sink will be relatively high, so this system is hardly used on some scenarios such as urgent events and scenarios without complete infrastructure. Thus, many researchers have proposed more flexible and convenient data collection schemes. For example, BONOLA et al. [19] proposed a method of data collection using opportunistic routing through mobile vehicles (MVs), and in this way, the roadside is deployed with sensing devices to monitor the status of street lights, smart trash cans, and roads and bridges on demand. With this solution, the sensing hardware will be simple, only a short-distance wireless communication capability be required, and installing expensive 5G communication hardware be not necessary[19]. The reason is that, in a smart city, there are a large number of MVs moving on the roads of the city, and when the MVs pass through the communication range of sensing devices, they can collect data and transmit the data to the cloud through 5G communications. In the research of HUANG et al. [20], numerous deployed sensor nodes can also self-organize into a network; the nodes on both sides of a road act as gateways, which are responsible for converging the entire network, and pass data to the cloud through MVs[21]. Therefore, this method may be widely used in smart cities. More related studies have been conducted[22–24]. In fact, except MVs[25–27], smartphones, tablets and smart watches can also act as data collectors[28]. They are called mobile edge users (MEUs) in the research of WANG et al. [28]. Because these MEUs have 5G communication capabilities, they can communicate directly with the cloud. The MVs are only on the road, but there are multiple types of MEUs in the market[28] with a wider moving range. When these MEUs pass through sensing devices with weak communication capabilities, they can collect data from sensing devices within their communication range and relay the data to the cloud. The use of MVs for data collection[19] is also a form of data collection approaches using MEUs. Therefore, in this paper, MEU is the general term for the devices that have 5G communication capabilities to perform data collection in a relay mode, and sensing devices or sensing nodes refer to a type of simple hardware that can only communicate over a short distance and needs to rely on MEUs to relay data to the cloud.

In order to incentivize MEUs to collect data, the incentive mechanism[29] is widely used, which enables cloud to initiate data collection tasks, grant a reward for collecting data, and incentivize MEUs to collect data[29]. This mechanism simplifies the deployment requirements of sensing devices and many sensor devices can be deployed on demand without 5G communication capabilities. Therefore, it facilitates a dramatic cost reduction of numerous sensor devices[30]. Moreover, the data collected by these sensing devices will be reported to the cloud through a huge number of MEUs, rather than specifically deploying a network for device connection. Such a system based on the incentive mechanism has strong adaptability and has been widely studied and used.

However, in such applications, the pivotal point is how to ensure the security of data collection. The factors affecting the security of data collection mainly come from MEUs and sensing devices[28, 30]. The impact of MEUs on the security of data collection is mainly manifested in some MVs reporting false data in order to obtain rewards, and there are even some malicious MVs that deliberately report offensive data, making data-based applications unusable[20–22]. For sensing devices, due to their simple hardware design, they are vulnerable to face attacks. Once these sensing devices are attacked, various problems will occur. For example, a black hole drops the data packets that are passing through it, so that the cloud platform cannot receive data[30]. According to statistics, there are more than 30 types of attacks on the sensing network, and these attacks will generate false data, tamper with data, or block the collection of data to damage the network[30]. Therefore, how to create a safe and clean environment of data collection as well as collecting authentic and credible data is a challenge deserved to concern with.

Although the use of MEUs is a cost-effective method[20], it is more challenging to ensure the security of data collection in such a data collection mode. In addition to the inherent unsafe factors in sensing devices, the use of MEUs for data collection may bring more threatening factors[15]. In particular, MEUs participate in data collection voluntarily with no identification, so it is difficult to ensure that MEUs are trustworthy in such an open-ended network environment[24]. What is more serious is that it is incredibly hard to verify whether the data reported by MEUs are true, which is known as an information elicitation without verification (IEWV) problem[31]. Due to the IEWV problem, even if the MV report reports false data in or-

der to obtain rewards, it is difficult to verify the data.

Using a credibility mechanism to choose trustable MVs for data collection is a feasible method. Because credible MVs will truthfully report the collected data, selecting credible MVs for data collection can improve the authenticity of the data[9, 15]. However, as mentioned earlier, it is difficult to verify the authenticity of data reported by MVs[32]; similarly, it is also difficult to identify the trustworthiness of MVs. In addition, for the sensing network, it is a major challenge to identify the credibility of these sensing devices[22, 24]. For a sensing network far from the edge of the network, it is very difficult to detect data attack[32]. Thus, we make the first attempt to deal with this challenge. In this paper, we propose a novel scheme that uses unmanned aerial vehicles (UAV) to detect the trust of sensing devices and MEUs (UAV-DT) to construct a clean data collection platform for SCS. The main contributions of this article are as follows:

1) We propose a framework using UAV to detect the trust of sensing devices and MEUs. In the proposed framework, the UAV is sent to the sensing network, deliver check codes to some selected sensing devices, and is required to route the code to the designated destination node. At the same time, the UAV collects information about data packets sent from sensing devices within a time span when passing through the sensing network. In this scheme, the checking code can act as a base truth indicator. If the UAV or cloud does not receive the verification code on the time that it should receive it, it can indicate that the verification code has been attacked during the data collection process. In this way, the IEWV problem that exists in this type of network can be effectively solved.

2) We propose an effective approach to sensing devices and MVs credibility computation. This method can construct a trusted data collection network environment. The information collected by the UAV will be checked by the cloud platform in two aspects to verify the credibility of sensor devices. On the one hand, the platform will check whether there is an anomaly in the data packet routing process for trust evaluation. It mainly checks whether the upstream and downstream nodes of the sensing devices receive and send data packets abnormally and therefore provide a performance evaluation about the trustworthiness. Besides, the data packets submitted by the sensing devices to the MEUs and those submitted by the MEUs to the platform will be checked and compared to verify the trustworthiness of the MEUs. On the other hand, according to the designed routing path of the verification code, it checks whether the verification code is successfully routed from the originated nodes to the MVs, and then submits the message to the cloud, which improves the trust of these sensing devices and MVs. The credibility computation method proposed in this paper enables accurate verification.

3) Based on the proposed framework, we propose a data collection strategy based on credibility magnitude and incentivation mechanism. The simulation results show that the proposed UAV-DT scheme can identify the credibility magnitude of sensing devices and MEUs, and the amount of clean data collected has been proportionally incremented. The classification rate for trusted sensing devices is as high as 98.9%. Meanwhile, a data collection rate of 89.9% on average can be achieved.

2 Related Work

To protect the security of networks in smart cities, various safety mechanisms, e.g., cryptographic schemes, authentication mechanisms and secure storage, were proposed in the past. However, using a trust-based model, the trust evaluation mechanism has the advantages of efficiency, lightweight and low overheads. The trust models that have been proposed provide a better choice in terms of network security and safety.

In general, the trust-based evaluation mechanisms for network security can be classified into two categories: the centralized and distributed. For the former, the trust value of nodes can be calculated by themselves. KIM and SEO[33] have proposed a trust computation method using fuzzy logic (TCFL) for WSN. They suggest a trust model using fuzzy logic in sensor network, in which trust is an aggregation of consensus given a set of past interaction among sensors. They calculate the trust value of the path through the trust of the nodes, then the path with the highest trust value is selected to transmit data packets[33]. However, in the great majority of applications, smart network system is distributed with a large number of nodes and a node in the system only focuses on the trustworthiness of its neighbor nodes. Besides, centralized approaches always make high energy consumption.

A distributed mechanism, the beta-based trust and reputation evaluation system for wireless sensor networks (BTRES), is proposed in Ref. [34]. BTRES is based on monitoring nodes' behavior and beta distribution is used to describe the distribution of nodes' credibility. Another distributed trust computation scheme, the parameterized and localized trust management scheme (PLUS), is proposed by YAO et al.[35]. In PLUS, each sensor node maintains highly abstracted parameters, and rates the trustworthiness of its interested neighbors to adopt appropriate cryptographic methods, identifing the malicious nodes and sharing the opinion locally. Distributed mechanisms have obvious disadvantages as well, which include the excess energy node and time costs due to the cooperation and communication with neighbors and increasing memory costs with the increase of network density caused by the lack of centralized management.

To overcome the defects above, WANG et al.[28] propose a crowdsourcing mechanism for trust evaluation based on mobile edge computing. In this mechanism, through close access to end nodes, mobile edge users can obtain various types of information of the end nodes and determine whether the node is trustworthy. HUANG et al.[20] propose a novel baseline data

based verifiable trust evaluation scheme, called BD-VTE, similar to the scheme in Ref. [28]. In BD-VTE, the trust of MVs is evaluated by sending UAVs to perceive IoT devices data as baseline data.

3 System Model and Problem Statement

3.1 System Model

Fig. 1 shows the SCS network model used in this paper. Our model includes sensing devices, MEUs and UAVs. The following is the description and symbol definition of each role.

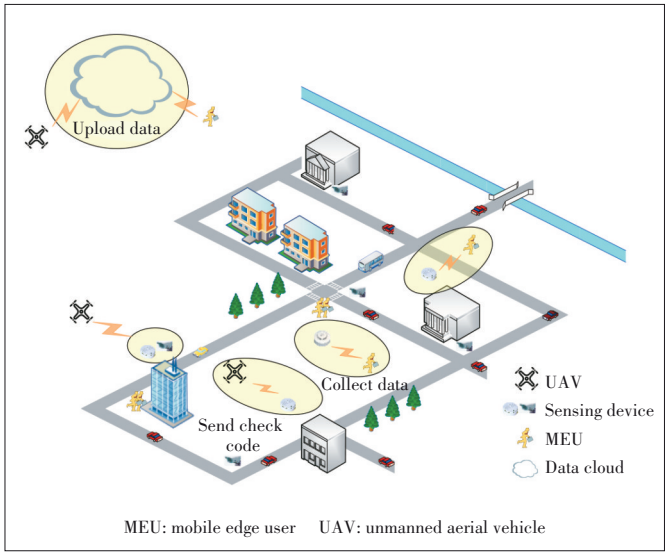
1) Sensing devices

As shown in Fig. 1, the SCS, the IoT devices are treated as sensor nodes and constitute the sensor network. There are N sensor nodes deployed in the network. The set of nodes is represented by $V = \{1,2,...,N\}$, and the sensor nodes perceive the environment in the city, output data and transmit them to the outside. A small number of nodes may be attacked and become malicious nodes, which is manifested as deliberate packet loss during data transmission. Suppose the number of malicious nodes is K and the set of malicious nodes is represented by $M = \{1,2,...,K\}$.

2) Unmanned aerial vehicles

The role of the UAVs is a bridge between the sensor network and the external network and they can communicate directly with the data center in the cloud. The UAVs can distribute the verification packets generated by the data center to the sensor nodes for transmission and directly check the transmission status of the nodes. In the scenario shown in Fig. 1, the UAVs pass the verification packet with a check code to a starting node, and then the node transmits it according to a certain routing rule.

3) Mobile edge users



▲Figure 1. Sensor-cloud system model

The number of MEUs with strong communication and storage capabilities distributed in the city far exceeds sensing devices. The MEU acts as a data collector in the system model and can directly communicate with the sensor node to obtain the data packet transmitted to the node. There is a total of L MEUs in the system, and their set is represented by $U = \{1,2,...,L\}$. Each MEU has its own active range, which is abstracted as a circle whose radius is r_i, and the abstract model of MEU is widely used by many researchers[28–29]. Within a certain period of time, the MEU can collect the data of all nodes covered in its active range, which means that the active range of the MEU indirectly refers to its ability to perform data collection tasks.

4) Transmission model

Considering communications between sender node n_1 and receiver node n_2, let p_{n_1} denote the transmitting power of n_1, and h_{n_1,n_2} denote the channel gain between n_1 and n_2. The channel gain follows the Rayleigh distribution. The distance between n_1 and n_2 is denoted by d_{n_1,n_2}, and the channel attenuation factor and Gaussian channel coefficient are donated by ϑ and h_0, respectively. Therefore, the channel gain holds as:

$$h_{n_1,n_2} = h_0 d^{-\vartheta}_{n_1,n_2}. \tag{1}$$

And the transmission rate between the sender node n_1 and receiver node n_2 can be denoted according to Shannon equation:

$$r_{n_1,n_2} = B\log_2\left(1 + \frac{p_{n_1} \times h_{n_1,n_2}}{p_0 + N_0}\right), \tag{2}$$

where B denotes the bandwidth, N_0 denotes the power spectral density of additive Gaussian white noise, and p_0 denotes the interference caused by reusing identical spectrum resources.

3.2 Problem Statement and Relevant Definition

The previous study has shown that using MEU can infer the trust value of the node according to its various states, e.g., the communication behavior, remaining battery, data content of target node, and so on[28]. In practice, it is difficult to obtain such information directly through the MEU, while obtaining data indirectly by monitoring neighbor nodes will add additional communication burden to each node, which will greatly reduce the life of the entire network. Due to the above limitations, it is unrealistic to directly or indirectly obtain the status of a node. Another problem that needs to be solved is the lack of an effective mechanism to ensure the authenticity of the data uploaded by MEUs. Therefore, we need to distinguish trusted nodes from malicious nodes in the network in an effective and realistic way. In general, when the nodes in the sensor network transmit data packets, trusted nodes can complete the data transmission task well. Occasionally, packet loss will occur when the network fluctuates greatly and the integrity of the data will not change significantly. However, malicious nodes will

frequently drop packets or tamper with data, which will compromise the validity of the data. Meanwhile, as the third-party data collector, the credibility of MEU also needs investigating. It is also necessary to distinguish between trusted and malicious MEUs and hire trusted users to complete data collection tasks, thereby ensuring the quality of the collected data by MEUs. Thus, in this paper, the MEU that plays the role of data collector is regarded as a mobile node and it is referred to as a node with sensing devices when there is no special distinction. We also need to minimize the cost of evaluating and classifying nodes. Hiring MEUs for data collection is the main cost of the system. Therefore, the data center should adopt an efficient MEU incentive mechanism to hire a set of trusted MEUs to complete data collection tasks with high quality. This paper reflects the performance of system through the following trust indicators and overall costs:

1) Difference of trust values between normal and malicious nodes D, which is defined as

$$D = \overline{\sigma_{\text{nor}}} - \overline{\sigma_{\text{mal}}}, \tag{3}$$

where $\overline{\sigma_{\text{nor}}}$ is the average trust value of normal nodes and $\overline{\sigma_{\text{mal}}}$ is the average trust value of malicious nodes. The difference D between the two averages can show the difference of benefits to the network. When D is large, it means that the distinction between the two is obvious. Therefore, one of the goals of our strategy is $\text{Max}(D) = \max(\overline{\sigma_{\text{nor}}} - \overline{\sigma_{\text{mal}}})$.

2) The discrimination rate of trusted nodes $\mathcal{R}_t$ and discrimination rate of malicious nodes $\mathcal{R}_m$ are defined as

$$\mathcal{R}_t = \frac{num_{\tilde{t}}}{num_t}, \tag{4}$$

$$\mathcal{R}_m = \frac{num_{\tilde{m}}}{num_m}. \tag{5}$$

These two indicators refer to the ratio of the correct number of nodes judged to be trusted $num_{\tilde{t}}$ and the total number of trusted nodes num_t, and the ratio of the correct number of nodes judged to be malicious $num_{\tilde{m}}$ and the total number of malicious nodes num_m. Both $\mathcal{R}_t$ and $\mathcal{R}_m$ reflect the system's ability to classify nodes. Then, the goals of our strategy include $\text{Max}(\mathcal{R}_t) = \max(\frac{num_{\tilde{t}}}{num_t})$ and $\text{Max}(\mathcal{R}_m) = \max(\frac{num_{\tilde{m}}}{num_m})$ as well.

3) Total cost of system P is defined as

$$P = \sum_{i=1}^{R}\sum_{j=1}^{L} f_{ij} \times p_{ij} + \sum_{i=1}^{R} \mathcal{L}_i \times v, \tag{6}$$

where f_{ij} indicates whether the user labeled j in the data collector set U in the ith round participates in the data collection task; $f_{ij} = 1$ indicates that the user participated in the data collection task, otherwise $f_{ij} = 0$ means not; p_{ij} represents the remuneration received by the user labeled as j in the data collector set U in the ith round; $\mathcal{L}_i$ represents the number of nodes that need UAVs to inspect in the ith round and v represents the cost of UAV verification of one node. Therefore, one of the purposes of our strategy is $\text{Min}(P) = \min(\sum_{i=1}^{R}\sum_{j=1}^{L} f_{ij} \times p_{ij} + \sum_{i=1}^{R} \mathcal{L}_i \times v)$.

In summary, all objectives in this paper are shown in Eqs. (7) – (10).

$$\text{Max}(D) = \max\left(\overline{\sigma_{\text{nor}}} - \overline{\sigma_{\text{mal}}}\right), \tag{7}$$

$$\text{Max}\left(\mathcal{R}_t\right) = \max\left(\frac{num_{\tilde{t}}}{num_t}\right), \tag{8}$$

$$\text{Max}\left(\mathcal{R}_m\right) = \max\left(\frac{num_{\tilde{m}}}{num_m}\right), \tag{9}$$

$$\text{Min}(P) = \min\left(\sum_{i=1}^{R}\sum_{j=1}^{L} f_{ij} * p_{ij} + \sum_{i=1}^{R} \mathcal{L}_i * v\right). \tag{10}$$

The notation of parameters in the model and for problem statement is shown in Table 1.

4 Proposed UAV-DT System

In this part, we present our UAV-DT scheme. The proposed scheme is divided into three parts: the UAV-assisted trust verification mechanism, trust reasoning mechanism based on communication behavior, and incentive mechanism based on cost performance and trust.

4.1 UAV-Assisted Trust Verification Mechanism

The most critical part of our scheme is to evaluate the trust value of nodes in the SCS, thereby distinguishing between trusted and malicious nodes. We propose a UAV-assisted trust verification mechanism to determine whether the communication behavior of the node is normal.

The mechanism is divided into four stages: generation and distribution of verification packets, result gathering of the tail node, review of data collection tasks, and verification of a suspect path. In each stage, UAVs play an important role.

At the beginning, the sensor network shown in Fig. 2(a) is untested. The following is an introduction to the four steps:

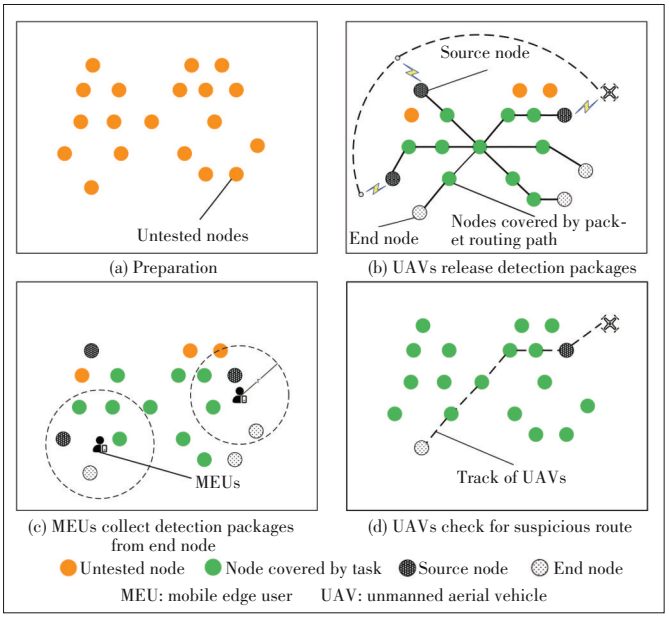
1) Generation and distribution of detection packets

At this stage, the UAV selects a certain number of source nodes in the network as the start of data packet delivery. Our trust evaluation is conducted in multiple rounds, so the trust value of a node will change after each round of calculation. In this process, when the trust value of the node is higher than $\sigma_{\max}$, the node is judged to be trusted. On the contrary, when

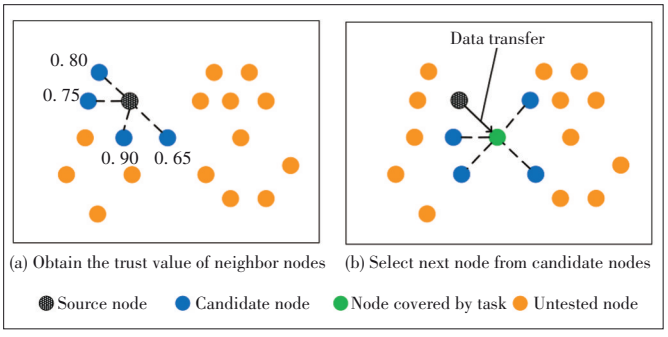
▼Table 1. Parameters in System Model and Problem Statement

Parameter	Meaning
V	Collection of sensor nodes
U	Collection of mobile edge users
M	Collection of malicious nodes
T	Number of verification tasks per round
R	Number of rounds of a verification task
D	Difference of trust values between normal and malicious nodes
P	Total cost of system
$\overline{\sigma_{nor}}$	Average trust of normal nodes
$\overline{\sigma_{mal}}$	Average trust of malicious nodes
$\mathcal{R}_t$	Discrimination rate of trusted nodes
$\mathcal{R}_m$	Discrimination rate of malicious nodes
$num_{\tilde{t}}$	Number of nodes judged to be trusted
num_t	Total number of trusted nodes
$num_{\tilde{m}}$	Number of nodes judged to be malicious
num_m	Total number of malicious nodes
$f_{i,j}$	Participation flag for user labelled as j in the ith round
$B_{i,j}$	Result flag for node labelled as j in the ith task
$p_{i,j}$	Payment for user labelled as j in the ith round
$\mathcal{L}_i$	Number of nodes that need UAVs to inspect in the ith round
υ	Cost of UAV verification of one node
N	Number of sensor nodes
K	Number of malicious nodes
L	Number of mobile edge users
σ_0	Initial trust value
$\sigma_{max}, \sigma_{min}$	Trust value threshold
σ_{com_i}	Communication trust value
$\sigma_{rec}(i,j)$	Cooperative recommendation coefficient between nodes labeled as i and j
σ_{rec_i}	Cooperative recommendation trust value
σ_{int_i}	Comprehensive trust value
bid_i	Bid of mobile edge user
b_i	Expected reward of mobile edge user
PoI_i	Number of task nodes in the active range
α_1, α_2	Weight coefficient of winning bids set selection algorithm
ω_1, ω_2	Weight coefficient of comprehensive trust

UAV: unmanned aerial vehicle



▲Figure 2. Using UAV to detect the trust of sensing devices and MEUs (UAV-DT)



▲Figure 3. Route generation strategy

the trust value of the node is lower than σ_{min}, the node is considered untrusted. Therefore, our criterion for selecting a source node is the node whose trust value is between σ_{min} and σ_{max} in the suspicious state.

As Fig. 2(b) shows, after a source node is selected, it is necessary to determine the route of the detection data packet transmission. We use the low-trust node diffusion strategy shown in Fig. 3 to generate the transmission path of the data packet. Starting from the source node, when determining the next hop node, the current trust value of each neighboring node is considered; the node that is closest to the initial trust value σ_0 is selected and the task is refused to repeat. It will be ensured that each node in the network receives a certain number of inspections.

The generation and distribution process of the verification data packet can be summarized in Algorithm 1.

Algorithm 1. Generation of detection packet algorithm

Input: R, T, σ_{max}, σ_{min}, σ_0, V
Output: S
Initialize $iter_{out} = 0$, $S = \varnothing$
While $iter_{out} < R$ **Do**
  $\rho = \varnothing$
  Randomly choose source node n in V and $\sigma_n < \sigma_{max}$ and $\sigma_n > \sigma_{min}$
  $\rho = \rho \cup n$
  $node_{cur} = n$
  $node_{source} = n$
  $iter_{in} = 0$
  While $iter_{in} < T$ **Do**
    Choose nxt in $neb(node_{cur})$ $\min |\sigma_{nxt} - \sigma_0|$ and $nxt \notin \rho$
    $\rho = \rho \cup nxt$

12: $node_{cur} = nxt$
13: $iter_{in} = iter_{in} + 1$
14: **End While**
15: $node_{end} = node_{cur}$
16: Generate check packet $p = (node_{source}, node_{end}, \rho)$
17: $S = S \cup p$
18: $iter_{out} = iter_{out} + 1$
19: **End While**
20: **Return** S

Algorithm 1 shows the process of multiple rounds of verification data packet generation and routing distribution. The input of the algorithm includes the number of verification task rounds R, the number of verification data packets in each round T, three constants related to trust value and node sets V. The outer loop (Lines 2 – 19) represents each round of verification tasks, and at the beginning of each mission, the system randomly chooses source node n in V. Then, the system generates a route for each verification task in the inner loop (Lines 9 – 14). Finally, the verification task set S is output.

2) Result gathering of the tail node

When the route of the verification packets is determined, the UAV distributes the data packets to the source node, and the packets are transmitted to the tail node in turn according to the routing path.

Then, we only need to collect the delivered data packets at the tail node to confirm whether there is any node loss behavior on the delivery path of the data packets. As shown in Fig. 2(c), MEUs are hired as data collectors to perform data collection tasks at the end nodes and hand data over to the data center in the cloud for further processing.

3) Review of data collection tasks

Since the data collector is not necessarily credible, we still need to further verify the validity and completeness of the data collector's collection results.

Obviously, not all collected results need to be verified. When issuing data collection tasks, the system clarifies which data packets at the sensing devices need to be collected, and the data collector does not know the content of the data packet and its check code in advance. Therefore, in the case that the verification packet is normally delivered to the tail node, we can consider that the result of this collection must be credible and no further confirmation is required if the submitted by the data collector is consistent with the original packet and is accompanied by a true verification code.

If a data package uploaded by the data collector is inconsistent with the original package distributed by the UAV, it is necessary to rely on the UAV to recollect the data at the tail node to make a judgment on the communication behavior between the data collector and the sensing devices on the routing path. The UAV compares the original verification packet with the data collected by itself to determine whether the data collector has performed the data collection task honestly, or the verification packet has been modified by one or more malicious nodes during the transmission process.

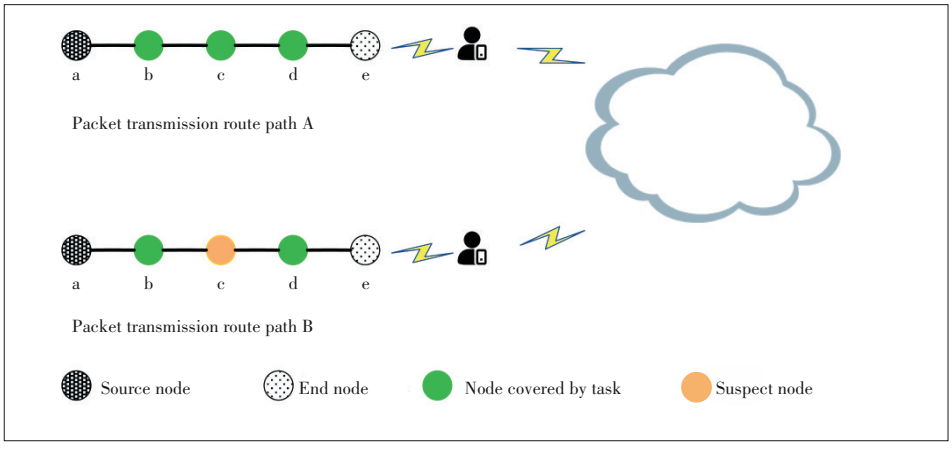
4) Verification of a suspect path

Through the previous stage, the system knows which verification packets have been modified during the transmission process, and defines such a routing path as a suspected path; that is, malicious nodes may appear on this path. In our system, the UAV checks the communication records of each node along the path, and judges the node(s) where the packet loss has occurred in the transmission process based on these records.

Then, we specifically describe the processing of nodes on the successful transmission path and the verification of suspected path (Fig. 4).

As shown in Fig. 4, in route path A, the verification packet is successfully transmitted to the tail node, and then collected by the MEUs faithfully, which means a successful transmission path. In this case, all the nodes on the transmission path honestly transmit the verification data packet to the next hop node. The system records the successful communication behavior once for Nodes a, b, c, and d. Since Node e is the tail node, it does not participate in the transmission process, so this verification task cannot perform trust evaluation on it.

On the contrary, after system verification in route path B, the verification data packet collected at the tail node has changed compared with the original data packet, which indicates that there is packet loss behavior by one or more nodes. In the figure, the node marked in orange is the node that has lost packets during transmission. Considering that there are network fluctuations, trusted nodes may also lose packets due to poor network communications, so the system cannot directly



▲**Figure 4. Successful transmission and suspect paths**

determine whether the node that has lost the packet during the data packet transmission is a malicious node, but can only define its communication behavior this time is malicious. According to the transmission result of the data packet, the system records the successful communication behavior for Nodes a, b, and d once, and correspondingly, it records the malicious communication behavior for Node c once. As mentioned above, Node e is the tail node and does not participate in the transmission of data packets.

When the MEU collects data packets at the tail node and uploads them to the data center in the cloud, it may also misrepresent the data. The system will also check the communication behavior of its uploaded data, and record the number of honest uploads and false uploads. As shown in Fig. 5, the real active range of an MEU is a light area with a radius of R_1, but it lies to the cloud data center that its active range is the dark area with a radius of R_2. Then the system assigns data collection tasks at four tail nodes, but the MEU only completes three collection tasks. The data at the node at the bottom right is not collected by the MEU because it exceeds its active range and at the same time it lies about a false result. Based on the above, the system will record the honest upload and false upload of the MEU.

After the above four steps finish, one round of a verification task is completed. With the obtained communication behaviors of the nodes and data collector, we can use various trust evaluation methods to calculate their trust values. The next section will focus on describing the trust reasoning mechanism used in our scheme.

4.2 Trust Reasoning Mechanism Based on Communication Behavior

The proposed trust reasoning mechanism in this paper is divided into three parts: the trust value initialization, trust evaluation, and trust state determination (Fig. 6).

4.2.1 Trust Value Initialization

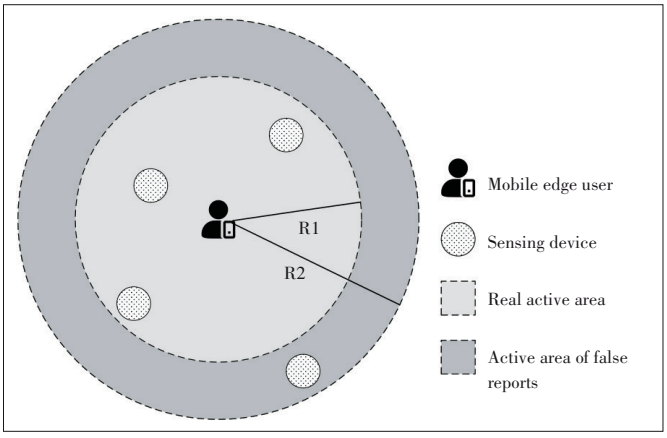
At the beginning, when all nodes and MEUs have not been fully checked, we cannot judge whether they are malicious or not, so we give each node the same initial trust value σ_0 and record it as a suspect state. After multiple rounds of trust inspection, the trust value of the nodes or MEUs participating in the transmission and collection of verification data packets will change through the trust reasoning mechanism and their status will increase or decrease accordingly.

4.2.2 Trust Evaluation

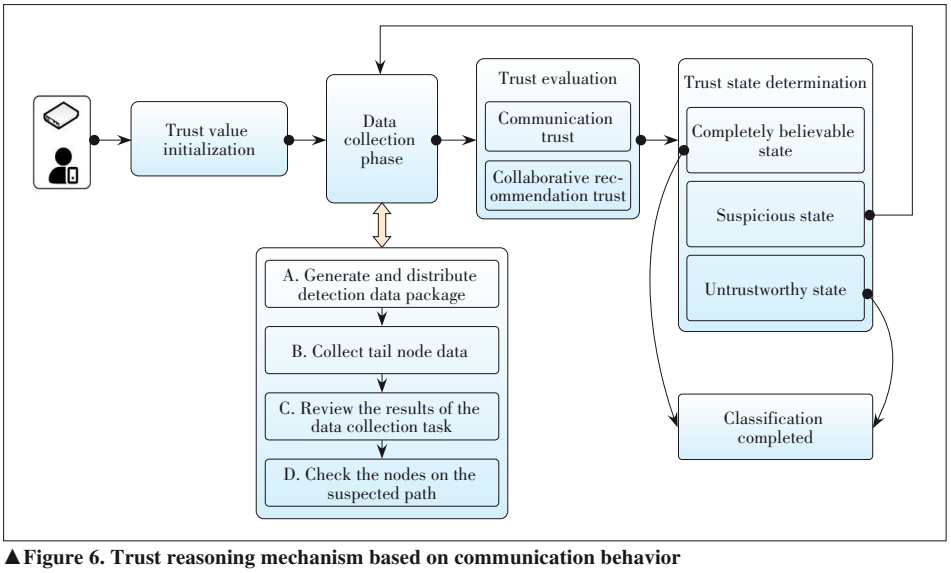
The trust value is the direct basis for judging whether the nodes and MEUs are malicious or not in our scheme. In our trust reasoning mechanism, two calculation methods are mainly used for trust value evaluation: communication behavior trust and collaborative recommendation trust.

1) Communication behavior trust

In each round of a trust verification task, the transmission of the verification data packet by the participating nodes is a communication behavior, and we can obtain the number of successful and malicious communication behaviors respectively. Similarly, we can also acquire two types of behaviors (honest upload behavior and false upload behavior) of MEUs participating in the data collection task. We can accordingly calculate the communication trust between the sensor nodes participating in the transmission task and the MEUs participating in the collection task, and mark them as $(\sigma_{com_i})^{wsn}$ and $(\sigma_{com_j})^{meu}$ respectively, in which i refers to the label of a sensor node in the node set V and and j refers to the label of an MEU in the user set U.



▲Figure 5. Mobile edge user (MEU) collects data package



▲Figure 6. Trust reasoning mechanism based on communication behavior

We then record successful communication behaviors of the sensor node and honest upload behaviors of the MEU as positive communication behaviors, and the malicious communication behaviors of the sensor node and the false reports uploaded by the MEU as negative communication behaviors. We use a subjective logic framework (SLF)[36] to describe communication behavior trust and the formula for calculating the trust value is as follows:

$$\sigma_{com_i} = \frac{2A + B}{2}, \tag{11}$$

where $A = p/(p + n + 1)$, $B = 1/(p + n + 1)$, p is the number of positive communication behaviors of node i, and n is the number of negative communication behaviors of node i.

2) Collaborative recommendation trust

As shown in Fig. 7, the transmission and collection of a verification data packet involves multiple sensor nodes and multiple MEUs (MEU is regarded as mobile nodes), and these nodes coordinate to complete this verification task. With the packet as an intermediary, a virtual connection is created between the nodes, which is called collaborative recommendation in our scheme.

In our example, the nodes labeled a, b, e and f and the MEUs labeled B and C perform their task honestly, and then there is a positive virtual connection between them. However, the nodes labeled c, d and the MEU labeled A do not complete the task faithfully, so there is a negative virtual connection between them. Similar to the communication behavior trust, the subjective logic framework is used in the collaborative recommendation value calculation, and the formula is as follows:

$$\sigma_{rec}(i,j) = \frac{2A + B}{2}, \tag{12}$$

where $A = p/(p + n + 1), B = 1/(p + n + 1)$, p is the number of positive virtual connections established by nodes i and j in multiple rounds of verification and connection tasks, and n is the number of negative virtual connections.

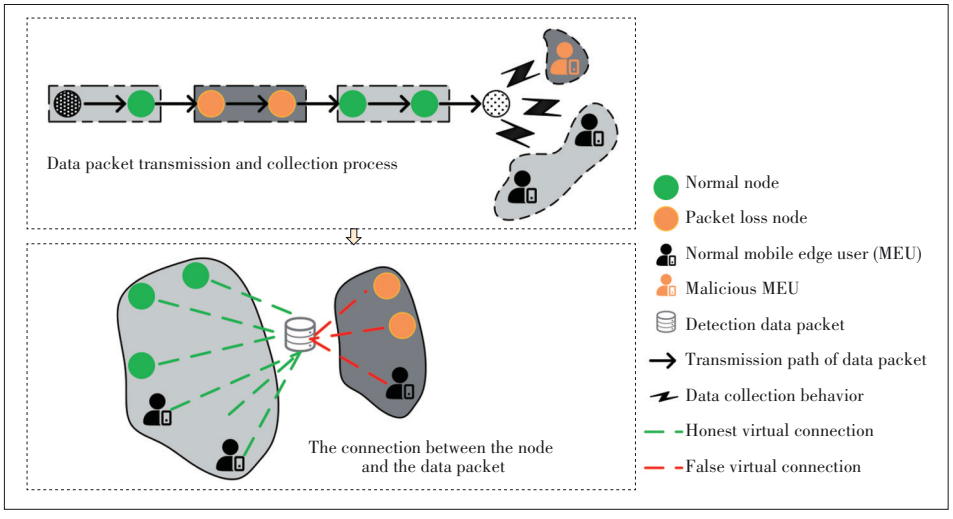
The collaborative recommendation trust coefficient between nodes i and j is $\sigma_{rec}(i,j)$, but if we want to calculate the recommendation trust value of node i, we need to synthesize the collaborative recommendation trust coefficients of all the nodes that have virtual connections with it. What's more, in order to ensure the reliability of recommendation, it is essential to consider the trust of the recommender's own communication behavior trust. In summary, the calculation formula of the node's collaborative recommendation trust is as follows:

$$\sigma_{rec_i} = \frac{\sum_j I_{i,j}(\sigma_{rec}(j,i))^2 \sigma_{com_j}}{\sum_j I_{i,j}\sigma_{rec}(j,i)}, \tag{13}$$

where $I_{i,j}$ is a status indicating whether there is a connection between nodes i and j. When $I_{i,j} = 1$, there is a connection between the two nodes, otherwise not; σ_{com_j} represents the communication behavior trust of node j and takes its own communication behavior trust as the weighting coefficient when node j recommends node i.

Algorithm 2. Algorithm of trust value evaluation (AoTVE) based on communication behavior

Input: ρ_I, V, U, T
Output: $\hat{V}$, $\hat{U}$
1: **Initialize** $\hat{V} = V, \hat{U} = U$
2: **For** each $\rho_{I,j} \in \rho_I$ **Do**
3: Detection packets are transmitted on the router $\rho_{I,j}$
4: **End For**
5: **For** each $node_i \in (V^{mod} \cup U^{mod})$ **Do**
6: Calculate h, m of $node_i$ by using the data collector and UAV
7: Calculate σ_{com_i} using h,m in Eq. (11)
8: Let x_i be a collection which node all in router path ρ_I
9: $Q = \varnothing$
10: **For** each j, *i.e.*, $node_j \in x_i$ **Do**
11: **If** $\sigma_{int_j} \neq \sigma_0$ and $\sum_{k=0}^{T} |B_{k,i} + B_{k,j}| \geqslant 2$ **Do**
12: Calculate p, n between $node_i$ and $node_j$
13: Calculate $\sigma_{rec}(i, j)$ using p, n in Eq. (12)
14: $Q = Q \cup \{\sigma_{rec}(i,j)\}$
15: **End If**
16: **End For**
17: Calculate σ_{rec_i} using Q in Eq. (13)
18: Calculate σ_{int_i} using σ_{rec_i}, σ_{com_i} in Eq. (14)
19: **If** $node_i$ in V **Do**



▲**Figure 7. Trust reasoning mechanism based on communication behavior**

20: $\hat{V}(i) = node_i$
21: **Else If** $node_i$ in U **Do**
22: $\hat{U}(i) = node_i$
23: **End If**
24: **End For**
25: **Return** $\hat{V}, \hat{U}$

Algorithm 2 is the algorithm of trust value evaluation (AoTV) based on communication behavior for two trust values in the trust reasoning mechanism. Line 11 (If $\sigma_{int_j} \neq \sigma_0$ and $\sum_{k=0}^{T} |B_{k,i} + B_{k,j}| \geqslant 2$ Do) restricts the conditions for node j to recommend node i. The restriction conditions require that node j has participated in the verification task before, that is, the comprehensive trust value is not the initial value ($\sigma_{int_j} \neq \sigma_0$), and node j has a virtual connection with node i. For example, suppose that node i and node j perform the task labeled 1 in this round and complete honestly, then $B_{1,i} = B_{2,j} = 1$. If they do not complete the task honestly, then $B_{1,i} = B_{2,j} = -1$. In both cases, $\sum_{k=0}^{T} |B_{k,i} + B_{k,j}| \geqslant 2$ is established, then node i and node j form a recommendation relationship with each other. After calculating the communication behavior trust and collaborative recommendation trust, the comprehensive trust of the node is obtained by the following formula:

$$\sigma_{int_i} = \omega_1 * \sigma_{com_i} + \omega_2 * \sigma_{rec_i}, \tag{14}$$

where ω_1 and ω_2 are aggregation constants and the best combination is found by subsequent experiments.

4.2.3 Trust State Determination

After each round of trust evaluation, the nodes participating in the task will update the trust value once. We take the comprehensive trust value of the node as the basis for its state judgment and use two constants σ_{max} and σ_{min} to divide the node into three states. When the comprehensive trust value of the node is greater than σ_{max}, the node is judged to be trusted. When the comprehensive trust value of the node is less than σ_{min}, the node is judged to be malicious state, and the node is classified as suspicious when it is in between σ_{max} and σ_{min}.

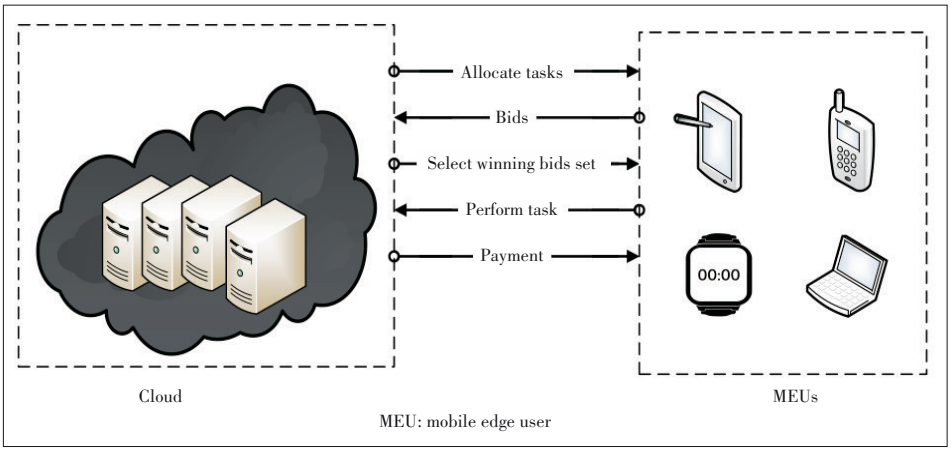
When a node is classified as a trusted state and a malicious state, for the sensor node, it no longer needs to be verified, but for the MEU, we should continuously verify its credibility because of its strong subjectivity. Besides, the higher the MEU's comprehensive trust, the higher the probability and rewards that it will be selected for the task.

4.3 Incentive Mechanism Based on Cost Performance and Trust

In our scheme, the final result of the data packet transmitted via the sensor node is to be collected by third-party users to reduce the flying distance of the UAV and improve the efficiency of the system. We use mobile crowdsourcing (MCS)-based data collection scheme and complete the task of data collection by hiring a large number of MEUs distributed in the city. The MEUs have strong data storage capacity, computing power and high mobility. They can complete multiple data collection tasks in a short time and make full use of the idle computing power of the equipment.

However, hiring MEUs needs to consider two aspects. On the one hand, it is impossible for MEUs to unconditionally participate in data collection tasks. Participants hope to get actual rewards from providing data, rather than volunteering to provide data for free. Because the perception of data needs to consume resources such as battery power, computing resources and data flow of participants' mobile devices, the participants in this process also need to pay time and labor. Without proper return, participants are not interested in staying active in the MCS-based network for a long time. On the other hand, we need to select MEUs to participate in the task reasonably and give them appropriate remuneration to make full use of the remuneration budget. In other words, we need to focus on the selection criteria of participants.

Therefore, we propose an incentive mechanism based on cost performance and trust. Our incentive mechanism uses a reverse auction framework to describe the relationship between data centers and MEUs. Our mechanism is divided into five steps (Fig. 8). At first, the system selects all tail nodes, and sends data collection tasks to idle MEUs according to each round of verification data packets. Then, the MEUs whose active scopes cover the target node give their own quotes and, after the system receives the quotation from the MEUs, it uses Algorithm 3 to select a set of suitable bids, which is recorded as the winning bids set, and determines the



▲**Figure 8. Reverse auction framework**

reward based on its performance. Subsequently, the selected MEUs perform data collection tasks within the scope of their activities and upload the collected results to the data center. When we design the winning bids set selection algorithm, we consider two selection criteria:

1) The ratio of an MEU's expected revenue to its data collection capacity. The quotes of an MEU can be expressed as a two-tuple $bid_i = (b_i, PoI_i)$, in which b_i is the expected reward of the MEU labeled as i, while PoI_i represents the number of task nodes covered in the active range of the MEU labeled as i. This ratio can directly reflect the cost-effectiveness of the data benefits we can obtain by providing remuneration to users.

2) The comprehensive trust value of an MEU. Not all MEUs are authentic and the data they submit may be biased. We can divide untrusted MEUs into two categories: "Greedy Users" who may report falsehood by exaggerating their scope of activities for their own benefit and "Real Malicious Users" who deliberately misrepresent data, thus affecting the true collection of data packets, and have a certain strategy.

Based on the above two criteria, we designed Algorithm 3. The input of the algorithm includes the MEU bid set BID and node set V. The output of the algorithm is the winner set S and their payment set P. Then the algorithm uses the greedy method to find the MEU with the maximum sensing performance-price ratio and their trust value in first loop (Lines 3 – 11). In second loop (Lines 13 – 26), for each MEU in the winner set S, the algorithm removes the MEU from S and continues to select other MEUs in $\overline{BID}$ to join $\bar{S}$ until all the nodes can be accessed. Finally, according to the element in $\bar{S}$, the algorithm gets the payment of each MEU.

Algorithm 3. Winning bids set selection and payment determination

Input: BID, V
Output: S, P
1: $S, P = \varnothing$
2: $\overline{BID} = BID$
3: **While** $PoI(S)$ is not contain all node of V **Do**
4: Select participant u from $\overline{BID}$ by using Eq. (15)
5: **If** $PoI(\{u\}) \subseteq PoI(S)$ **Then**
6: $\overline{BID} = \overline{BID} \setminus u$
7: **Else**
8: $S = S \cup \{u\}$
9: $BID = BID \setminus u$
10: **End If**
11: **End While**
12: $\bar{S} = \varnothing$
13: **For** each $u \in S$ **do**
14: $\bar{S} = S \setminus u$
15: **While** $PoI(\bar{S})$ is not contain all node of V **Do**
16: Select participant $\bar{u}$ from $\overline{BID}$ by using Eq. (15)
17: **If** $PoI(\{\bar{u}\}) \subseteq PoI(\bar{S})$ **Then**
18: $\overline{BID} = \overline{BID} \setminus \bar{u}$
19: **Else**
20: $S' = S' \cup \{\bar{u}\}$
21: $BID = BID \setminus \bar{u}$
22: **End If**
23: **End While**
24: Calculate p_b by using Eq. (16)
25: $P = P \cup \{p_b\}$
26: **End for**
27: **Return** S, P

Our incentive mechanism uses the following formula to select the current best participant:

$$u = \max_{i \in BID}\left(\alpha_1 * \frac{b_i}{PoI_i} + \alpha_2 * \sigma_{com_i}\right), \tag{15}$$

where b_i / PoI_i is the ratio of MEU's expected revenue to its data collection capacity and σ_{com_i} is comprehensive trust value of the participant labeled as i. We use proportional coefficients α_1 and α_2 to aggregate the participants' bid scores.

Algorithm 3 uses the ratio of the best data benefit in the alternative set S' to calculate the participant's payment (Lines 12 – 24). The calculation formula is:

$$p_b = \max_{j \in S'}\left(\frac{r_b}{r_j} * PoI_j\right), \tag{16}$$

where r_b is expected revenue of participant u_b, r_j is revenue of participant of u_j, and PoI_j is the number of task nodes in the active area of u_j.

5 Performance Analysis

5.1 Experiment Setup

We realized the UAV-DT scheme in Python 3.7 and ran a simulation experiment on IdeaPad Air 14 with 16 GB 2 133 MHz LPDDR4 RAM, whose CPU parameters is 2.10 GHz AMD Ryzen 5 4600U with Radeon Graphics.

The important parameters used in our experiments are listed in Table 2. Each experiment was carried out in a network area of 100×100 m^2, where 1 000 smart devices and 500 ME-

▼**Table 2. Experimental parameters**

Parameter	Value
Size of area/m^2	100 ×100
Number of sensor nodes	1 000
Number of MEUs	500
Active radius of MEU/m	[5, 10]
Payment of hiring MEU	[10, 25]

MEU: mobile edge user

Us were randomly deployed. We randomly created 20 different network scenarios in total and ran them once in each experiment. The results were averaged to ensure the robustness of our strategy in different network scenarios.

For a normal sensor device, there was a small probability of packet loss in the process of transmitting data packets due to network fluctuations. However, a malicious sensor device would deliberately discard a part of the data packet with a greater probability. We gave 5% and 20% probabilities for two different packet loss situations, which were reflected in the form of random functions in the simulation experiments. In the simulation, the data MEU reported had a 10% – 40% probability of being false.

In the experiments, 70 rounds of verification tasks were carried out in each scenario, and in each round, we used the drone to release 30 data packets starting with random sensor device. The length of the routing path of each packet was fixed to 10 nodes.

5.2 Discrepancy of Trust Values

In our scheme, the communication behavior trust and collaborative recommendation trust are aggregated into integration trust, then whether a behavior is malicious or not is determined by setting two thresholds (σ_{max} and σ_{min}), which involves two aggregation coefficients ω_1 and ω_2.

In the discrepancy experiments, we set five sets of coefficients to test the effect of different coefficients on the discrepancy of trust values. We set $\hat{\omega}_1 = (0.5, 0.5)$, $\hat{\omega}_2 = (0.6, 0.4)$, $\hat{\omega}_3 = (0.7, 0.3)$, $\hat{\omega}_4 = (0.4, 0.6)$, and $\hat{\omega}_5 = (0.3, 0.7)$, and guaranteed $\omega_1 + \omega_2 = 1$. Besides, since the optimal classification threshold has not been determined, so our experiments did not classify nodes.

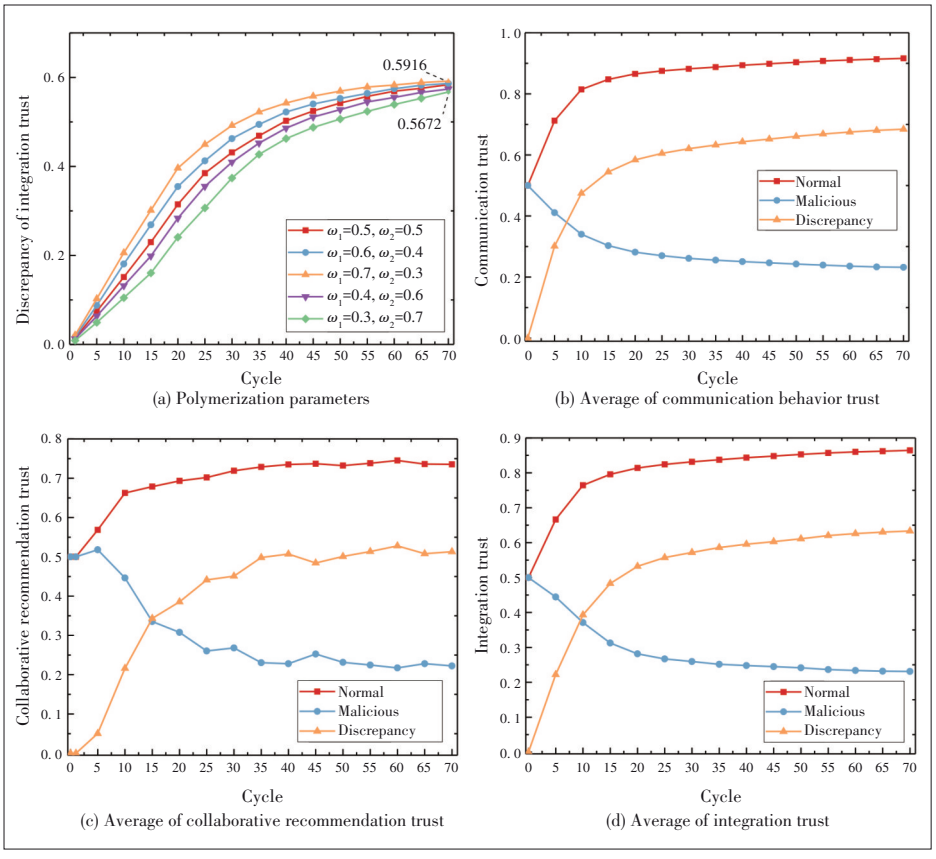
Fig. 9(a) shows that after 70 rounds of verification tasks are completed, the values of discrepancy between normal and malicious nodes in each set of experiments are in the interval between 0.5672 and 0.5916. Obviously, when we set $\hat{\omega} = \hat{\omega}_3 = (0.7, 0.3)$, the discrepancy reaches a peak, equal to 0.5916. The result shows that our scheme has a high degree of discrimination between normal and malicious nodes and the discrepancy between the two reaches a high value. Thus, we still use this set of aggregation coefficients in subsequent experiments.

We continue to advance the discrepancy experiment subsequently and Figs. 9(b), 9(c) and 9(d) show the discrepancy of communication behavior trust, collaborative recommendation trust, and integration trust between normal and malicious nodes. The three trust values of normal nodes are 0.916, 0.735 and 0.864 after 70 rounds of verification tasks, and the three trust values of malicious nodes are 0.684, 0.513 and 0.633. Compared with collaborative recommendation trust, the curve of communication behavior trust is smoother and the convergence speed is faster. After 20 rounds, the average of discrepancy in communication behavior trust has reached a high level. In terms of collaborative recommendation trust, the numerical curve has fluctuations in 70 rounds. As shown in Fig. 9(c), the average trust of normal nodes is not high enough so that the distinction between the two is not obvious.

The following is the conclusions of this group of experiments:

1) It is reasonable to trust a higher aggregation coefficient for communication behavior trust, and when the classification thresholds (σ_{max} and σ_{min}) are not set, we use the communication behavior trust to distinguish normal nodes from malicious nodes clearly.

2) The verification effect of communication behavior trust and collaborative recommendation trust can still be further optimized. We set classification thresholds σ_{max} and σ_{min}, and the system excludes the nodes that can be clearly identified as



▲Figure 9. Influence of different parameters on the average trust between normal and malicious nodes

the trustworthy or malicious from subsequent tasks, so that the left nodes with doubtful status (that is $\sigma_{\min} < \sigma_{int_i} < \sigma_{\max}$) can get a chance to be verified.

From Fig. 9(d), we can clearly see the average integration trust of normal nodes is 0.864 and has a slight upward trend. The average integration trust of malicious nodes is 0.231 and has a downward trend stably. Then we use 0.85 and 0.25 as the central values to find the best classification thresholds.

5.3 Discriminant Rate of Normal and Malicious Nodes

In this section, we set two groups of thresholds to conduct classification discrimination rate experiments. We use 0.85 and 0.25 as the central values of two groups ($\overline{\sigma_{\max}} = 0.85$, $\overline{\sigma_{\min}} = 0.25$), and find the best value in the interval between the upper and lower domains is 0.1. Then the interval of $\sigma_{\max}$ is $(\overline{\sigma_{\max}} - 0.05, \overline{\sigma_{\max}} + 0.05)$ and the interval of $\sigma_{\min}$ is $\left(\overline{\sigma_{\min}} - 0.05, \overline{\sigma_{\min}} + 0.05\right)$.

As shown in Fig. 10, in the first three sets of experiments, the discrimination rates of trusted nodes converge quickly and the final result is around 0.988, which means the correct discrimination rate of trustworthy results reaches 98.8%. Even if the result of the last set in more stringent conditions reaches 0.930, the correct discrimination rate can reach up to 93%. The results of the experiments on the discrimination rate of malicious nodes are shown in Fig. 11. The discrimination effect of groups 2 – 5 is better and the final discrimination rate is between 0.821 and 0.933. Under the most stringent threshold conditions, when $\sigma_{\min} = 0.20$ in the experiment, the discrimination rat is still higher than 65%.

From the experimental results, it can be seen that our scheme has excellent effect and robustness on the recognition ability of trusted and malicious nodes, and can obtain a high recognition rate even after 20 rounds. Then we choose $\sigma_{max} = 0.85$ and $\sigma_{\min} = 0.30$ as the best thresholds for our system to classify nodes. The node labeled as i will be treated as a trusted node when $\sigma_{int_i} > \left(\sigma_{\max} = 0.85\right)$ and a malicious node when $\sigma_{int_i} < (\sigma_{min} = 0.30)$.

Under the best classification threshold, the classification results are shown in Fig. 12. Trusted and malicious nodes are thoroughly classified after 30 rounds. Finally, the classification rate of trusted nodes is as high as 98.9%, while the classification rate of malicious nodes also reaches 94.2%. In other words, only 15 nodes are still in doubt status after 70 rounds in our simulation environment.

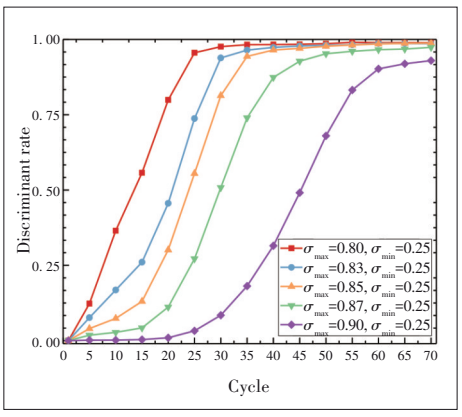
After the best classification thresholds ($\sigma_{max} = 0.85$ and $\sigma_{min} = 0.30$) are set, we repeat the experiment in Section 5.2 with $\omega_1 = 0.7$ and $\omega_2 = 0.3$. As shown in Figs. 13(a), 13(b) and 13(c), our scheme has a good improvement on the trust evaluation ability of nodes after setting the classification thresholds.

In terms of communication behavior trust, the average trust of normal nodes, the average trust of malicious nodes and the discrepancy between normal and malicious nodes are 0.929, 0.241, 0.688 respectively, which are slightly improved. In terms of collaborative recommendation trust, the average trust of normal nodes, the average trust of malicious nodes and the discrepancy between normal and malicious nodes have changed from 0.735, 0.222, and 0.513 to 0.890, 0.282 and 0.608, respectively. The discrepancy has increased by 18.5%, which means that the evaluation effect of collaborative recommendation trust has been improved. Combining the above two types of trust, the result of discrepancy in integration trust has increased by 4.8%.

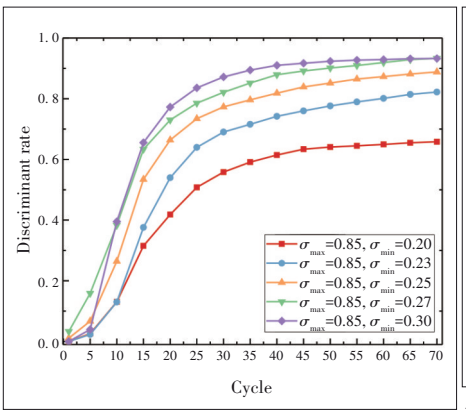
Based on the experiments, we determine the classification thresholds for node classification in our scheme, and the results prove that the node classification ability of the system is very significant.

5.4 Collection Rate

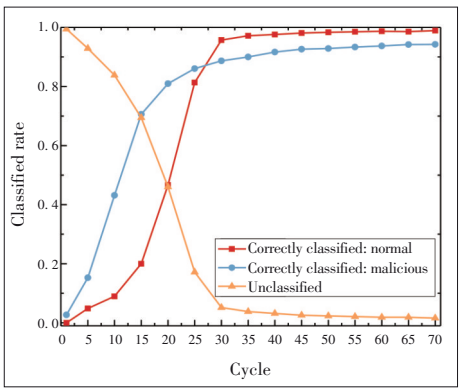
In our network scenario, there are some malicious nodes, which randomly discard some data packets passing through it with a certain probability. In this section, we generate the same number of regular data packets as the verification data packets, transmit them on the network, and hand them over to the MEU



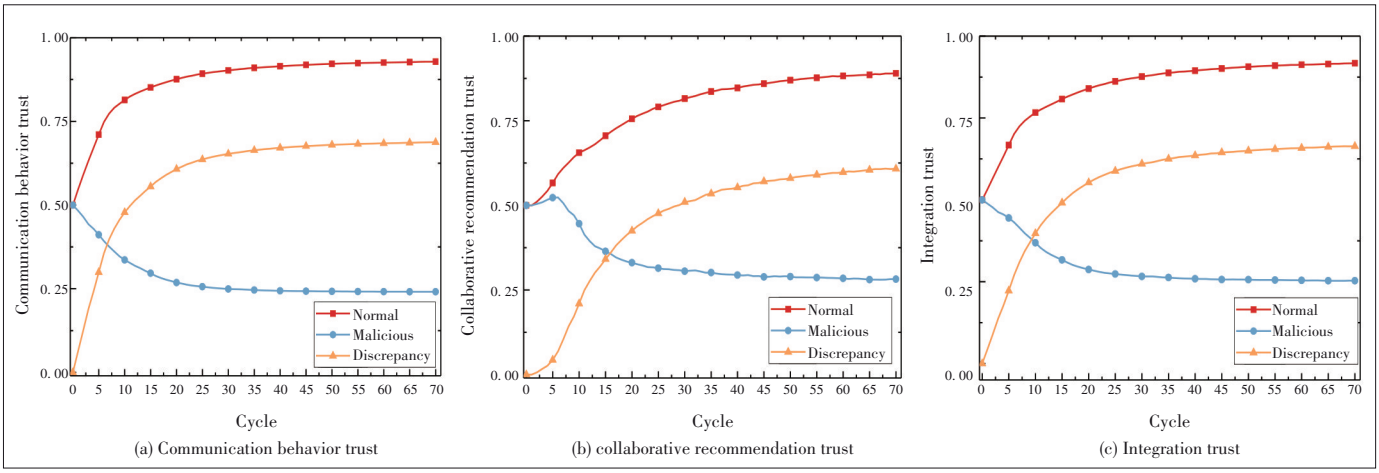
▲Figure 10. Discrimination rate of trusted nodes



▲Figure 11. Discrimination rate of malicious nodes



▲Figure 12. Discriminative ability under the optimal classification threshold



▲Figure 13. Trust evaluation capability under the optimal threshold

for collection. In contrast, we also simulated the collection of regular data packets in the original network scenario without our scheme, which is usually called unverified network.

The results are shown in Figs. 14 and 15. In our scheme, the collection rate curve first rises quickly and stabilizes in a very high value range, while the collection rate curve fluctuates at a relatively low position in the unverified network. After 14 rounds (when the most nodes in the network are classified), the collection rate of our scheme stays within the range 0.88 to 0.92, and the collection rate of unverified network maintains between 0.78 and 0.82. The average of the former is 0.899, while that of the latter is 0.808, that is, our scheme improves 11.2% compared with the unverified network.

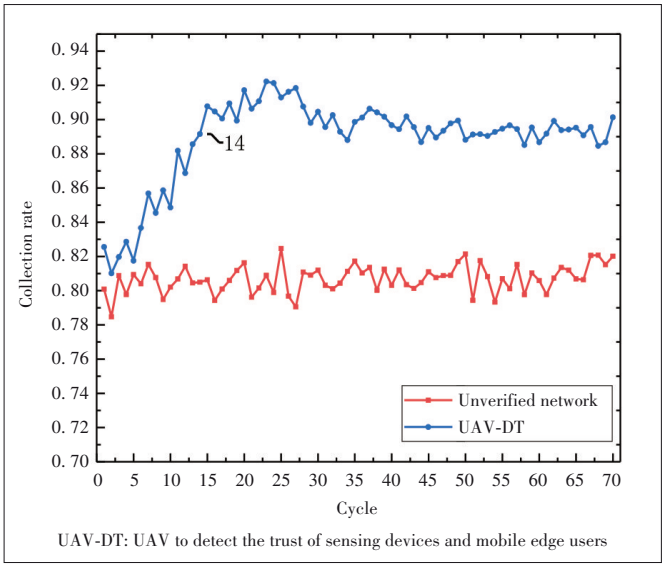
From another perspective, in our network scenario, there are two main reasons for packet loss: network fluctuations and the malicious node that deliberately loses packets. We also conducted two comparative experiments on the causes of packet loss.

As shown in Fig. 15, the proportions of the packet loss ratio caused by network fluctuations and malicious nodes are relatively stable in multiple rounds of experiments in the unverified network. The former is 26.6% in average and the latter is 73.4% correspondingly. In our scheme, the proportions of the two keep changing with the increase of rounds and the proportion of malicious nodes intentionally losing packets is slowly decreasing from 0.638 to 0.468.

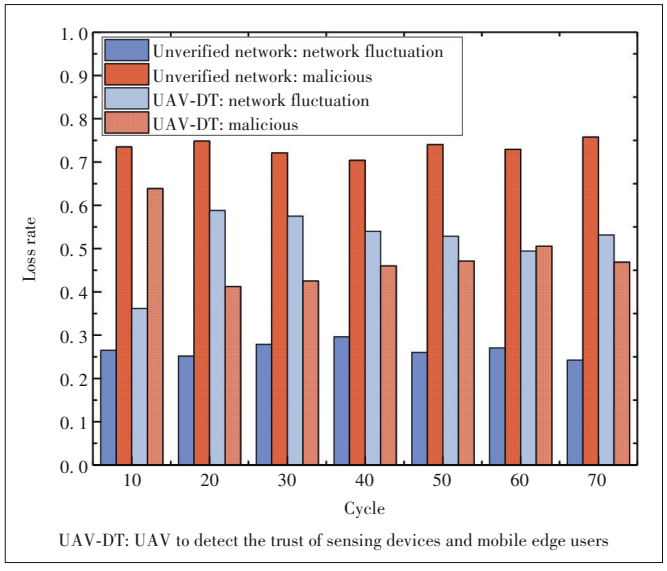
From the above experimental results, it can be seen that our scheme effectively detects a large number of trusted nodes and malicious nodes in the network, thereby avoiding malicious nodes during data packet transmission and improving the collection rate of data packets.

5.5 Cost

In this section, our experiments compare the winning bids



▲Figure 14. Collection rate



▲Figure 15. Rate of packet loss

set selection algorithm in our scheme with the conventional greedy algorithm (Fig. 16). The total cost of our system is divided into two parts: the cost of hiring MEUs and additional costs (the cost of sending drones for additional verification). We assume that the cost of each additional verification by the drone is as five times much as the cost of hiring an MEU.

According to the experimental results, the cost of hiring MEU in each round is much higher than the cost of additional verification of drones. The reason is: in our network scenario, there are more normal IoT devices than malicious ones and the frequency of transmission errors is relatively low compared to the total number of transmissions. Therefore, the cost of the system is mainly focused on hiring MEUs. In addition, it is obvious that the cost of our scheme for hiring MEUs is lower than that of the greedy strategy, with an average reduction of 23.4%. Although the additional costs are slightly higher, our scheme is still the best in terms of total cost, with an average reduction of 10.7%. Our UAV-DT scheme spends significantly less on employment than the greedy strategy. The greedy strategy does not consider the trust value in the selection range when selecting MEUs to participate in the data collection, which causes the suspected path shown in Fig. 4. This will inevitably lead to an increase in the cost of using UAV for review. On the contrary, UAV-DT uses the trust value of MEUs as the selection criterion shown in Algorithm 2.

6 Conclusions

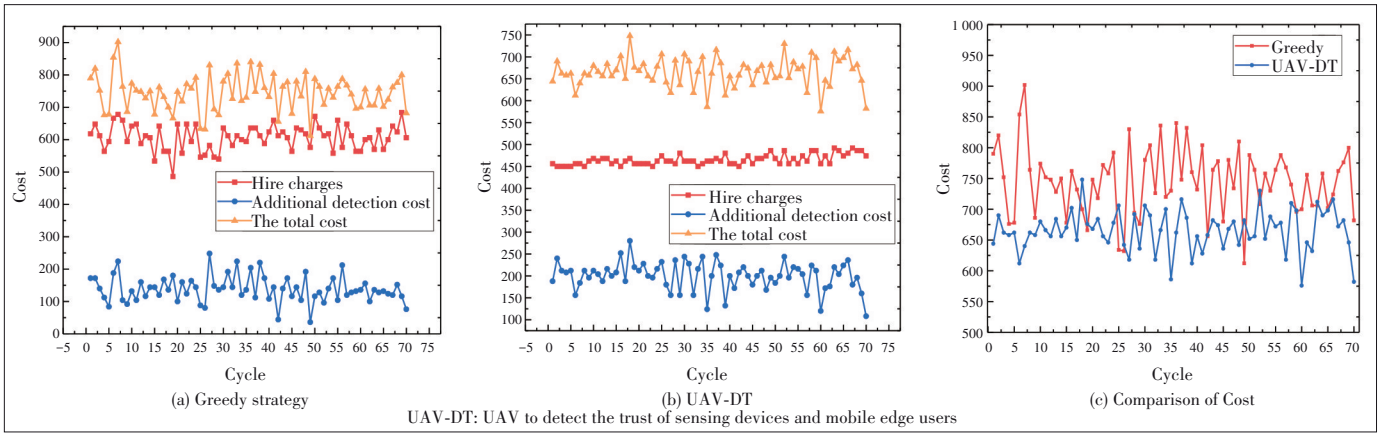
In this paper, we propose a low-cost and efficient UAV-DT security scheme, including the UAV-assisted trust verification mechanism, incentive mechanism based on cost performance and trust, and trust reasoning mechanism based on communication behavior. By continuously verifying the trust of the nodes in the network, the trusted and malicious nodes can be quickly distinguished by comparing their trust values.

Our experimental results show that our security scheme has high discrimination for malicious nodes, and provides an effective solution to efficient and safe data collection in the city.

However, we did not use a good path planning scheme in the UAV broadcast verification packet stage and the UAV secondary inspection stage. Further studies are needed in future and we will focus on how to develop an efficient UAV flight path in the future research.

References

[1] HILLS G, LAU C, WRIGHT A, et al. Modern microprocessor built from complementary carbon nanotube transistors [J]. Nature, 2019, 572(7771): 595 – 602. DOI: 10.1109/tcad.2015.2415492
[2] REN Y Y, WANG T, ZHANG S B, et al. An intelligent big data collection technology based on micro mobile data centers for crowdsensing vehicular sensor network [J]. Personal and ubiquitous computing, 2020: 1 – 17. DOI: 10.1007/s00779-020-01440-0
[3] YU M Y, LIU A F, XIONG N N, et al. An intelligent game based offloading scheme for maximizing benefits of IoT-edge-cloud ecosystems [J]. IEEE Internet of Things journal, 2020, early access. DOI: 10.1109/JIOT.2020.3039828
[4] Gartner. 20.4 billion connected things by 2020 [EB/OL]. (2017-2-09) [2021-05-01]. https://www.itp.net/611397-204-billion-connected-things-by-2020-gartner
[5] LI T, LIU W, ZENG Z W, et al. DRLR: a deep reinforcement learning based recruitment scheme for massive data collections in 6G-based IoT networks [J]. IEEE Internet of Things journal, 2021, early access. DOI: 10.1109/JIOT.2021.3067904
[6] HUANG M F, ZHANG K, ZENG Z W, et al. An AUV-assisted data gathering scheme based on clustering and matrix completion for smart ocean [J]. IEEE Internet of Things journal, 2020, 7(10): 9904 – 9918. DOI: 10.1109/JIOT.2020.2988035
[7] OUYANG Y, LIU A F, XIONG N X, et al. An effective early message ahead join adaptive data aggregation scheme for sustainable IoT [J]. IEEE transactions on network science and engineering, 2021, 8(1): 201 – 219. DOI: 10.1109/TNSE.2020.3033938
[8] LI A, LIU W, ZENG L J, et al. An efficient data aggregation scheme based on differentiated threshold configuring joint optimal relay selection in WSNs [J]. IEEE access, 2021, 9: 19254 – 19269. DOI: 10.1109/ACCESS.2021.3054630
[9] WANG T, ZHANG G X, BHUIYAN M Z A, et al. A novel trust mechanism based on fog computing in sensor-cloud system [J]. Future generation computer systems, 2020, 109: 573 – 582. DOI: 10.1016/j.future.2018.05.049
[10] LIU S, HUANG G S, GUI J S, et al. Energy-aware MAC protocol for data differentiated services in sensor-cloud computing [J]. Journal of cloud computing, 2020, 9(1): 1 – 33. DOI: 10.1186/s13677-020-00196-5
[11] LI F F, HUANG G S, YANG Q, et al. Adaptive contention window MAC protocol in a global view for emerging trends networks [J]. IEEE access, 2021, 9: 18402 – 18423. DOI: 10.1109/ACCESS.2021.3054015



▲Figure 16. System Costs

[12] HUANG C Q, HUANG G S, LIU W, et al. A parallel joint optimized relay selection protocol for wake-up radio enabled WSNs [J]. Physical communication, 2021, 47: 101320. DOI: 10.1016/j.phycom.2021.101320
[13] GUO J L, LI F F, WANG T, et al. Parameter analysis and optimization of polling-based medium access control protocol for multi-sensor communication [J]. International journal of distributed sensor networks, 2021, 17(4): 155014772110074. DOI: 10.1177/15501477211007412
[14] PALADINO, FISSORE, NEVIANI. A low-cost monitoring system and operating database for quality control in small food processing industry [J]. Journal of sensor and actuator networks, 2019, 8(4): 52. DOI: 10.3390/jsan8040052
[15] HUANG S B, ZENG Z W, OTA K, et al. An intelligent collaboration trust interconnections system for mobile information control in ubiquitous 5G networks [J]. IEEE transactions on network science and engineering, 2021, 8(1): 347 - 365. DOI: 10.1109/TNSE.2020.3038454
[16] ZHU X Y, LUO Y Y, LIU A F, et al. Multiagent deep reinforcement learning for vehicular computation offloading in IoT [J]. IEEE Internet of Things journal, 2021, 8(12): 9763 - 9773. DOI: 10.1109/JIOT.2020.3040768
[17] TENG H J, DONG M X, LIU Y X, et al. A low-cost physical location discovery scheme for large-scale Internet of Things in smart city through joint use of vehicles and UAVs [J]. Future generation computer systems, 2021, 118: 310 - 326. DOI: 10.1016/j.future.2021.01.032
[18] DENG Q Y, OUYANG Y, TIAN S J, et al. Early wake-up ahead node for fast code dissemination in wireless sensor networks [J]. IEEE transactions on vehicular technology, 2021, 70(4): 3877 - 3890. DOI: 10.1109/TVT.2021.3066216
[19] BONOLA M, BRACCIALE L, LORETI P, et al. Opportunistic communication in smart city: Experimental insight with small-scale taxi fleets as data carriers [J]. Ad hoc networks, 2016, 43: 43 - 55. DOI: 10.1016/j.adhoc.2016.02.002
[20] HUANG S B, LIU A F, ZHANG S B, et al. BD-VTE: A novel baseline data based verifiable trust evaluation scheme for smart network systems [J]. IEEE transactions on network science and engineering, 2021, 8(3): 2087 - 2105. DOI: 10.1109/TNSE.2020.3014455
[21] GUO J L, LIU A F, OTA K, et al. ITCN: an intelligent trust collaboration network system in IoT [J]. IEEE transactions on network science and engineering, 2021, early access. DOI: 10.1109/TNSE.2021.3057881
[22] LI T, LIU A F, XIONG N N, et al. A trustworthiness-based vehicular recruitment scheme for information collections in distributed networked systems [J]. Information sciences, 2021, 545: 65 - 81. DOI:10.1016/j.ins.2020.07.052
[23] HU L, LIU A F, XIE M D, et al. UAVs joint vehicles as data mules for fast codes dissemination for edge networking in smart city [J]. Peer-to-peer networking and applications, 2019, 12(6): 1550 - 1574. DOI: 10.1007/s12083-019-00752-0
[24] OUYANG Y, ZENG Z W, LI X, et al. A verifiable trust evaluation mechanism for ultra-reliable applications in 5G and beyond networks [J]. Computer standards & interfaces, 2021, 77: 103519. DOI: 10.1016/j.csi.2021.103519
[25] ZHU X Y, LUO Y Y, LIU A F, et al. A deep learning-based mobile crowdsensing scheme by predicting vehicle mobility [J]. IEEE transactions on intelligent transportation systems, 2021, 22(7): 4648 - 4659. DOI: 10.1109/TITS.2020.3023446
[26] HUANG W, OTA K, DONG M X, et al. Result return aware offloading scheme in vehicular edge networks for IoT [J]. Computer communications, 2020, 164: 201 - 214. DOI: 10.1016/j.comcom.2020.10.019
[27] SHEN M Q, LIU A F, HUANG G S, et al. ATTDC: an active and traceable trust data collection scheme for industrial security in smart cities [J]. IEEE Internet of Things journal, 2021, 8(8): 6437 - 6453. DOI: 10.1109/JIOT.2021.3049173
[28] WANG T, LUO H, ZHENG X, et al. Crowdsourcing mechanism for trust evaluation in CPCS based on intelligent mobile edge computing [J]. ACM transactions on intelligent systems and technology, 2019, 10(6): 1 - 19. DOI: 10.1145/3324926
[29] BAEK D, CHEN J, CHOI B J. Small profits and quick returns: An incentive mechanism design for crowdsourcing under continuous platform competition [J]. IEEE Internet of Things journal, 2020, 7(1): 349 - 362. DOI: 10.1109/JIOT.2019.2953278
[30] LIU Y X, DONG M X, OTA K, et al. ActiveTrust: secure and trustable routing in wireless sensor networks [J]. IEEE transactions on information forensics and security, 2016, 11(9): 2013 - 2027. DOI: 10.1109/TIFS.2016.2570740
[31] WAGGONER B and CHEN Y L. Output agreement mechanisms and common knowledge [C]//Second AAAI Conference on Human Computation & Crowdsourcing (HCOMP). Pittsburgh, USA: AAAI, 2014
[32] HUANG C, YU H R, BERRY R A, et al. Using truth detection to incentivize workers in mobile crowdsourcing [J]. IEEE transactions on mobile computing, 2020, early access. DOI: 10.1109/TMC.2020.3034590
[33] KIM T K, SEO H S. A trust model using fuzzy logic in wireless sensor network [J]. World academy of science, engineering and technology, 2018, 42: 63 - 66
[34] FUANG W D, ZHANG C L, SHI Z D, et al. BTRES: beta-based trust and reputation evaluation system for wireless sensor networks [J]. Journal of network and computer applications, 2016, 59: 88 - 94. DOI: 10.1016/j.jnca.2015.06.013
[35] YAO Z Y, KIM D Y, DOH Y M. PLUS: parameterized and localized trust management scheme for sensor networks security [C]//IEEE International Conference on Mobile Adhoc and Sensor Systems. Vancouver, Canada, 2006: 437 - 446. DOI: 10.1109/MOBHOC.2006.278584
[36] BALAKRISHNAN V, VARADHARAJAN V, TUPAKULA U. Subjective logic based trust model for mobile ad hoc networks [C]//4th International Conference on Security and Privacy in Communication Netowrks. Istanbul, Turkey: ACM, 2008: 1 - 11. DOI: 10.1145/1460877.1460916

Biographies

LI Xiuxian is currently pursuing his master's degree at School of Computer Science and School of Cyberspace Science from Xiangtan University, China. His research interests include mobile crowding sensing, IoT devices, and edge computing.

LI Zhetao (liztchina@hotmail.com) is a professor with the College of Computer, Xiangtan University, China. He received his B.Eng. degree in electrical information engineering from Xiangtan University in 2002, the M.Eng. degree in pattern recognition and intelligent system from Beihang University, China in 2005, and the Ph.D. degree in computer application technology from Hunan University, China in 2010. From December 2013 to December 2014, he was a postdoc in wireless network at Stony Brook University, USA. He is a member of IEEE and CCF.

OUYANG Yan is currently a postgraduate student with the School of Computer Science and Engineering, Central South University, China. Her research interests include crowd sensing networks and wireless sensor networks.

DUAN Haohua received his bachelor's degree in computer science and technology from Jilin University, China in 2020. He is currently pursuing his Ph.D. degree in computer science, Shanghai Jiao Tong University, China. His research interests include security and privacy in machine learning and blockchain.

XIANG Liyao received her B.Eng. degree in electrical and computer engineering from Shanghai Jiao Tong University, China in 2012, and Ph.D. degree in computer engineering from the University of Toronto, Canada in 2018. She is currently an assistant professor with Shanghai Jiao Tong University. Her research interests include security and privacy, privacy analysis in data mining, and mobile computing.

Artificial Intelligence Rehabilitation Evaluation and Training System for Degeneration of Joint Disease

LIU Weichen[1], SHEN Mengqi[2], ZHANG Anda[1], CHENG Yiting[2], ZHANG Wenqiang[1,2]

(1. Academy for Engineering and Technology, Fudan University, Shanghai 200433, China;
2. School of Computer Science, Fudan University, Shanghai 200433, China)

Abstract: Degeneration of joint disease is one of the problems that threaten global public health. Currently, the therapies of the disease are mainly conservative but not very effective. To solve the problem, we need to find effective, convenient and inexpensive therapies. With the rapid development of artificial intelligence, we innovatively propose to combine Traditional Chinese Medicine (TCM) with artificial intelligence to design a rehabilitation assessment system based on TCM Daoyin. Our system consists of four subsystems: the spine movement assessment system, the posture recognition and correction system, the background music recommendation system, and the physiological signal monitoring system. We incorporate several technologies such as keypoint detection, posture estimation, heart rate detection, and deriving respiration from electrocardiogram (ECG) signals. Finally, we integrate the four subsystems into a portable wireless device so that the rehabilitation equipment is well suited for home and community environment. The system can effectively alleviate the problem of an inadequate number of physicians and nurses. At the same time, it can promote our TCM culture as well.

Keywords: rehabilitation; Traditional Chinese Medicine; artificial intelligence; degeneration of joint disease

DOI: 10.12142/ZTECOM.202103006

https://kns.cnki.net/kcms/detail/34.1294.TN.20210818.1124.002.html, published online August 18, 2021

Manuscript received: 2021-06-15

Citation (IEEE Format): W. C. Liu, M. Q. Shen, A. D. Zhang, et al., "Artificial intelligence rehabilitation evaluation and training system for degeneration of joint disease," *ZTE Communications*, vol. 19, no. 3, pp. 46 - 55, Sept. 2021. doi: 10.12142/ZTECOM.202103006.

1 Introduction

Today, many policies support the inheritance and development of Traditional Chinese Medicine (TCM). It is popular that manufacturing a portable artificial intelligence (AI) rehabilitation evaluation and training system to improve the rehabilitation ability and promote the rehabilitation equipment industry of TCM. Degeneration of joint disease (DJD)[1] is a physiological and pathological degeneration process that occurs in the spine as the human body naturally ages. DJD can cause a variety of spinal-related disease syndromes and bring pain and stiffness, which will seriously affect patients' daily life. Severely, patients' nervous systems may be compressed and cause paralysis. Today, DJD has become one of the serious public health problems. The main clinical rehabilitation methods are traction therapy, infrared hyperthermia, percutaneous electrical stimulation, etc. ,

This work was supported by National Key R&D Program of China (No. 2019YFC1711800, 2020AAA0108300), National Natural Science Foundation of China (No. 62072112), Fudan University-CIOMP Joint Fund (No. FC2019-005).

but they are only suitable for some Grade A tertiary hospitals and specialist rehabilitation hospitals because of expensive and large equipment. In general, the doctor will advise the patient to take some conservative treatment unless the invasive surgical treatment must be taken. Therefore, we need to find an appropriate therapy as well as manufacture an efficient and inexpensive device for the therapy of DJD.

At present, there are various clinical treatments for DJD. HU[2] believes that moxibustion can relieve pain, replenish Qi and thus treat DJD. YANG et al.[3] believe that "Jin Gu Bing Ju, Chan Xuan Xiang Ji" and advocate the use of LI's Tuina method to treat DJD. The authors in Ref. [4 – 5] demonstrate that the combination of TCM and electromagnetic wave irradiation for the treatment of DJD has more prominent efficacy. The authors in Ref. [6 – 7] demonstrate that the combination of laser irradiation and physical traction therapy also performs better on the treatment of DJD, but it is only suitable for some Grade A tertiary hospitals and specialist rehabilitation hospitals because of expensive and large equipment.

Among all kinds of conservative therapies, TCM has a 5 000-year history of development for health care. TCM Daoyin[8] is guided by TCM theories such as Yin and Yang, five elements, meridians, and internal organs. It promotes functional recovery through breathing and exhalation, physical activities, and psychological regulation. It also has obvious therapeutic effects on the rehabilitation and prevention of soft tissue and bone and joint diseases. It is increasingly used in clinical treatment. Therefore, we choose to establish an artificial intelligence rehabilitation system based on TCM Daoyin for the therapy of DJD.

At the same time, we construct a background music database based on the five-tone theory of TCM[9]. It provides suitable background music for patients as an aid to therapy when practicing TCM Daoyin. TCM's five tones are "Gong Shang Jue Zhi Yu", the ancient way to recognize the sound, which originally belongs to the category of temperament. Huangdi Neijing (the Medical Classic of the Yellow Emperor) introduces five tones into TCM theory and forms a certain system. Five-tone therapy is based on the theory that the five tones correspond to the five organs in Huangdi Neijing. Studies[9 – 10] have shown that the five-tone therapy has a significant effect. Therefore, we embed the background music database into our system for the therapy of DJD.

In this paper, we propose a portable wireless system device that innovatively combines artificial intelligence and TCM for the rehabilitation of DJD. It replaces part of repetitive heavy physical labor in traditional rehabilitation. Furthermore, it realizes automatic, accurate, and intelligent rehabilitation and is very suitable for family and community clinics because of its small size, wireless, and portability. What's more, it could reduce the stress of physicians and rehabilitation trainers and increase the flexibility and effectiveness of patient rehabilitation training. Therefore, our system has strong innovation and practicability.

The remaining structure of this paper is as follows. Section 2 presents an overall overview of our system and describes the functions implemented in each subsystem separately. Section 3 elaborates the implementation principles, calculation formulas, and algorithm details for each of our subsystems. Section 4 describes the overall system integration application and the overall system workflow.

2 Functions

In this paper, we innovatively propose to develop a small-scale rehabilitation assessment and training system by combining the popular technologies in the field of artificial intelligence with TCM Daoyin. The system can provide comprehensive automatic assessment and assisted exercises for patients with DJD during their rehabilitation training. We will integrate all the modules of the system into a Mini-PC and transmitted the data via 5G signals to facilitate real-time rehabilitation training for patients and remote monitoring and guidance for doctors. As shown in Fig. 1, the system includes the following four subsystems:

1) SMA: spine movement assessment system. It automatically evaluates the range of motion of human cervical and lumbar vertebrae joints.

2) PRC: posture recognition and correction system. It can automatically compare and correct errors between patient actions and standard expert actions.

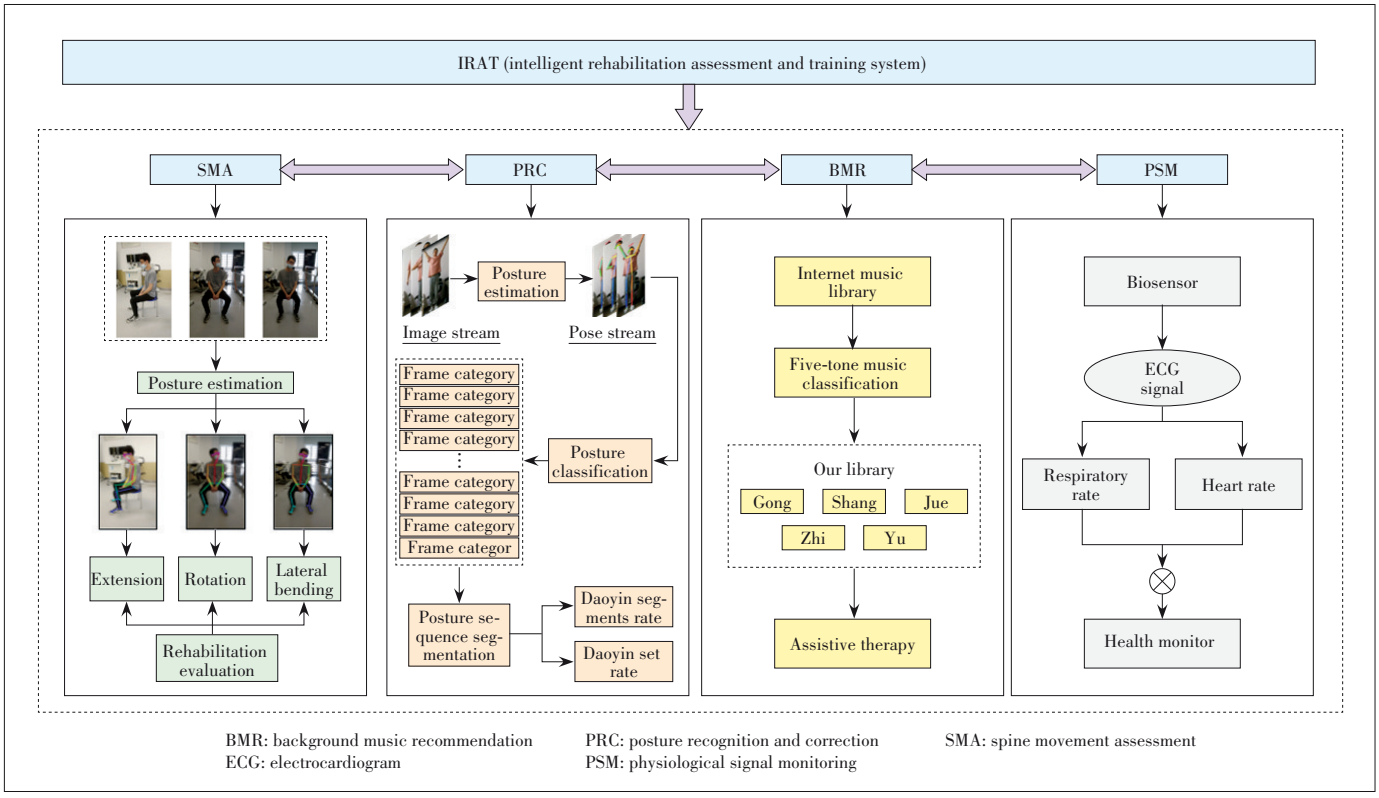
3) BMR: background music recommendation system. It filters and builds a library of TCM five-tone background music.

4) PSM: physiological signal monitoring system. It monitors and alerts the patient's respiratory rate and heart rate stability.

2.1 SMA: Spine Movement Assessment System

The spine movement assessment system is designed by combining popular techniques in computer vision. It can automatically measure the maximum joint mobility of the patient's cervical and lumbar vertebrae before the start and after the end of the patient's TCM Daoyin rehabilitation training. The joint mobility includes 6 types of subscales: cervical extension, lumbar extension, cervical lateral bending, lumbar lateral bending, cervical rotation, and lumbar rotation. Therefore, this subsystem can visually and quantitatively assess the effect of the patient's rehabilitation training.

First, the system takes an image of the patient in a specific computational scenario and uses maximum force to stretch the cervical or lumbar vertebrae. Then, the system performs image processing operations and inputs the results into a neural network model for posture estimation and inference. As a result, we can obtain information about the 2D skeletal coordinates corresponding to the patient image. We use different calculation or estimation formulas for three different types of joint mobility. Finally, we combine the three metric perspectives to give the patient's rehabilitation evaluation score.



▲**Figure 1. Intelligent rehabilitation assessment and training system**

2.2 PRC: Posture Recognition and Correction System

The function of the posture recognition and correction system is to identify and classify the movements of patients in real time, while they are practicing the TCM Daoyin. The system will compare and score each Daoyin movement with the standard movements of the expert group and provide real-time feedback. Finally, the patient's overall movements are analyzed when the patient has completed the entire Daoyin set. The system will give a score for each segment of different movements and an overall score for the entire set.

The system uses a neural network model to identify and classify the current rehabilitation Daoyin movements being practiced by the patient based on a feature learning approach. It will get the patient's historical posture sequence by tracking the key points of the patient's body at the same time. What's more, we collect and construct a small sample dataset to train the network models. Then, the system automatically segments the patient's historical pose sequence based on the patient's action category by the method of sequence similarity estimation. Finally, we obtain the similarity score of the patients' Daoyin segments by comparing each segment of the patient's posture sequence with the standard posture sequence of the expert.

2.3 BMR: Background Music Recommendation System

Many clinical studies show that TCM's five-tone therapy can effectively assist or even directly contribute to the rehabilitation process of patients. The overall rehabilitation efficiency of patients treated with the TCM five-tone therapy has been significantly improved.

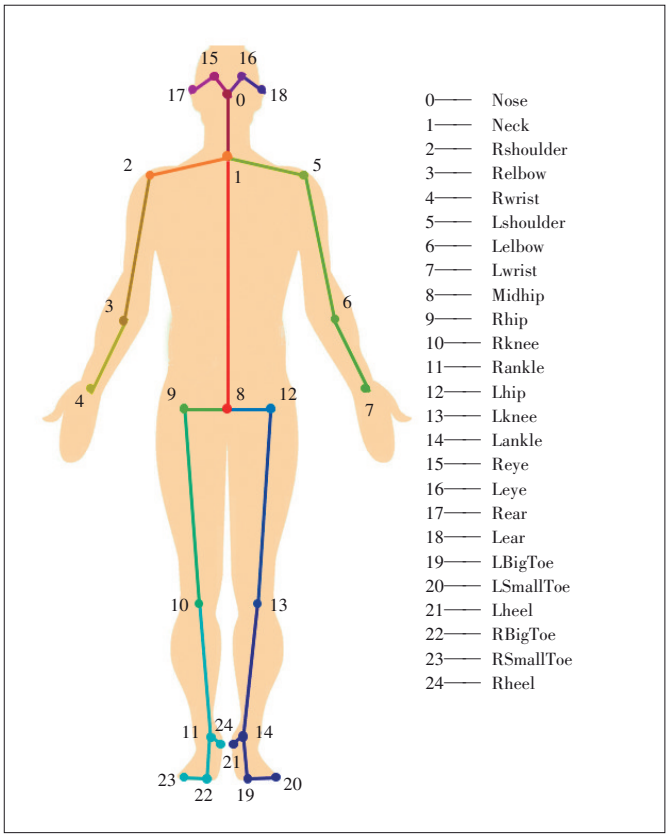
Most scholars think that "Gong Shang Jue Zhi Yu" corresponds to "Do, Re, Mi, So, La" in the modern numbered musical notation from a musical point of view.However, traditional music has indistinct tuning and frequent modulation. There are many different versions of the same piece and the tuning is inconsistent with each other. The current research application of five-tone therapy is too limited and lacks a holistic approach. Moreover, there is no systematic and standardized five-tone music dataset. The manual-based five-tone classification is laborious and inefficient because there is a huge amount of online music. Therefore, our background music recommendation system aims at addressing the shortcomings of the current clinical application of five-tone therapy. It uses computer audition technology to automatically filter and classify music from a large library of music online. Eventually, a five-tone database of background music for patients to perform TCM Daoyin exercises is constructed to assist patients' rehabilitation training.

2.4 PSM: Physiological Signal Monitoring System

Doctors often monitor patients' health based on some of their physiological signatures such as the heart rate and respiratory rate. We need to monitor the heart rate and respiratory rate in real time while the patient is practicing TCM Daoyin

training, ensure the patient's safety when practicing alone, and alerts the patient's family and doctor in case of abnormal conditions. The physiological signal monitoring system is based on a biosensor that uses a single-lead approach to acquire the patient's ECG signal. Then, the collected ECG signal data is further processed to achieve accurate and reliable heart rate detection.

For the respiratory rate, there is no non-invasive respiratory rate monitoring device available in the market. Traditional methods use large instruments with low accuracy, which are difficult to extend to home rehabilitation equipment. The respiratory rate is also an important physiological signal for the patient, so we have to find a method for detecting respiratory rate that would work with our mobile rehabilitation system. The physiological signal monitoring system captures the patient's respiratory rate using an algorithm, named ECG derived respiration (EDR), which extracts the derived respiratory signal from the ECG signal in real time. The system thus enables real-time monitoring of the heart rate and respiratory rate of patients during TCM Daoyin training.



▲**Figure 2. Common key points of the human body**

3 Algorithms

3.1 Algorithms of SMA System

3.1.1 Human Posture Estimation

Human posture estimation is to estimate the human pose by calculating the relative position of the human key points in 3D space and correctly linking the human key points that have been detected in the picture. The key points of the human body usually correspond to the joints with degrees of freedom of the human body, for example, neck, shoulder, elbow, wrist, knee, ankle, etc., as shown in Fig. 2.

CAO et al.[11] proposed the OpenPose multi-person pose estimation framework. It is a real-time approach to detect the 2D pose of multiple people in an image. We use this model to implement the estimation and scoring of the patient's spinal motion status. The model can automatically identify and match the key points of the patient's body in each frame of the received image. It consists of four main components: feature extraction, partial affinity fields (PAFs) prediction, key point location confidence map prediction, and pose graph matching. First, we input an image and extract features by convolutional neural networks (CNNs) to get a set of feature maps. Then we use CNNs to extract part confidence maps and part affinity fields respectively. Second, we use bipartite matching in graph theory to find the part association, which connects the joints of the same person. Finally, we merge them into the overall skeleton of a person. Details of the algorithm can be found in Ref. [11].

3.1.2 Calculation of Joint Mobility

Because the overall system is designed for simplicity and lightness, the hardware equipment conditions are somewhat limited. Under the condition of only one camera, we use different calculation or estimation formulas for different joint mobility.

As shown in Fig. 3, we use precise positioning to calculate the angles for the extension and lateral bending by the coordinates of the body's key points. Here we set the coordinate origin at the Midhip point. As shown in Table 1, we have run experiments and ensured that the error is within 3°, which allows us to score the patient's posture very well.

1) Cervical and lumbar extension angles

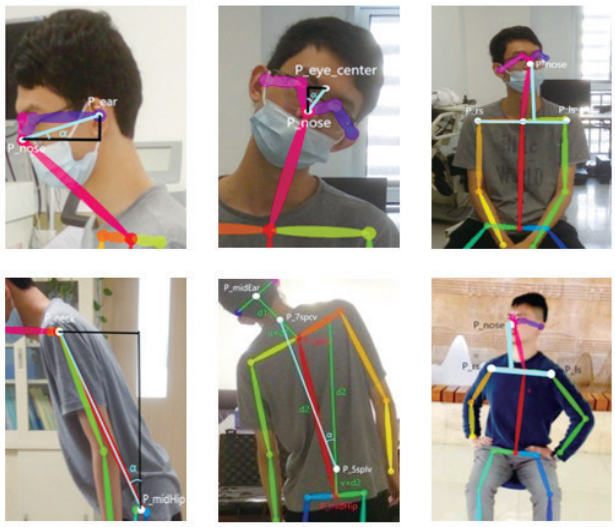
We use different formulas to calculate the extension angles of the patient's cervical and lumbar spine.

As for the patient's cervical extension angle, we calculate it from the coordinates of the nose and ear key points in the lateral view.

$$\propto_{ce} = \tan^{-1}\left(\frac{\left|y_{ear} - y_{nose}\right|}{\left|x_{ear} - x_{nose}\right|}\right), \quad (1)$$

where $\propto_{ce}$ is the cervical extension angle, x_i is the horizontal coordinate of key point i, and y_i is the vertical coordinate of key point i.

As for the patient's lumbar extension angle, we calculate it



▲Figure 3. (a) Extension angles; (b) lateral bending angles; (c) rotation angles

▼Table 1. Results of our spine movement assessment (SMA) system

Different Joints	True Value/°	Measured Value/°	Error/°
Cervical anterior extension	20.00	18.48	1.52
Cervical posterior extension	22.50	24.55	2.05
Cervical left bending	14.00	13.85	0.15
Cervical right bending	20.00	20.69	0.69
Cervical left rotation	30.00	28.82	1.18
Cervical Right Rotation	42.00	41.92	0.08
Lumbar Anterior Extension	32.00	31.06	1.64
Lumbar Posterior Extension	25.00	25.53	0.53
Lumbar Left Bending	27.00	22.52	4.48
Lumbar Right Bending	19.00	20.59	1.59
Lumbar Left Rotation	42.00	41.09	0.91
Lumbar Right Rotation	38.00	36.38	1.62

from the coordinates of the neck and midhip key points in the lateral view.

$$\propto_{le} = \tan^{-1}\left(\frac{\left|x_{neck} - x_{midHip}\right|}{\left|y_{neck} - y_{midHip}\right|}\right), \tag{2}$$

where, $\propto_{le}$ is the lumbar extension angle, x_i is the horizontal coordinate of key point i, and y_i is the vertical coordinate of key point i.

2) Cervical and lumbar lateral bending angles

We capture the patient's key points from the front and back of the patient. Accordingly, we calculate the patient's cervical and lumbar lateral bending angles.

As for the patient's cervical lateral bending angle, we calculate it from the coordinates of the nose key points and estimate eye_center key points in the front view.

The estimated coordinates of the eye_center key point are shown as follows.

$$x_{eye\_center} = \frac{\left(x_{l\_eye} + x_{r\_eye}\right)}{2}, \tag{3}$$

$$y_{eye\_center} = \frac{\left(y_{l\_eye} + y_{r\_eye}\right)}{2}, \tag{4}$$

where, x_i is the horizontal coordinate of key point i, and y_i is the vertical coordinate of key point i.

Calculation of the patient's cervical lateral bending angle is shown as follows.

$$\propto_{cl} = \tan^{-1}\left(\frac{\left|y_{eye\_enter} - y_{nose}\right|}{\left|x_{eye\_enter} - x_{nose}\right|}\right), \tag{5}$$

where $\propto_{cl}$ is the cervical lateral bending angle, x_i is the horizontal coordinate of key point i, and y_i is the vertical coordinate of key point i.

As for the patient's lumbar lateral bending angle, we calculate it from the estimated coordinates of the 7th cervical spinous process and the 5th lumbar spinous process key points in the back view.

We use the neck key point and the estimated mid_ear key point, where the mid_ear key point's calculation method is the same as the eye_center key point, to estimate coordinates of the 7th cervical spinous process key point.

$$x_{7spcv} = x_{neck} \pm \left|x_{mid\_ear} - x_{neck}\right| \times u, \tag{6}$$

$$y_{7spcv} = y_{neck} - \left|y_{mid\_ear} - y_{neck}\right| \times u, \tag{7}$$

where x_i is the horizontal coordinate of key point i and y_i is the vertical coordinate of key point i. Here $u = 0.5$ is used to calculate the 7th cervical spinous process.

We use the neck key point and the midhip key point to estimate coordinates of the 5th lumbar spinous process key point.

$$x_{5splv} = x_{midhip}, \tag{8}$$

$$y_{5splv} = y_{Midhip} - \sqrt{\left(x_{Neck} - x_{Midhip}\right)^2 + \left(y_{Neck} - y_{Midhip}\right)^2} \times v, \tag{9}$$

where x_i is the horizontal coordinate of key point i and y_i is the vertical coordinate of key point i. Here $v = 0.2$ is used to calculate the 5th lumbar spinous process.

Calculation of the patient's lumbar lateral bending angle is

shown as follows.

$$\propto_{ll} = \tan^{-1}\left(\frac{\left|x_{7spcv} - x_{5splv}\right|}{\left|y_{7spcv} - y_{5splv}\right|}\right), \tag{10}$$

where, $\propto_{ll}$ is the lumbar lateral bending angle, x_i is the horizontal coordinate of key point i, and y_i is the vertical coordinate of key point i.

3) Cervical and lumbar rotation angles

For the rotation angle, due to the limitation of the number of cameras condition, we take the estimation approach.

We assume that the patient's cervical spine is rotated by 90° and the angle between the nose key point in line with the midpoint of the Lshoulder and Rshoulder key points, and the Lshoulder key point in line with the Rshoulder key point denoted by ε, $\varepsilon_{90} = 45°$. Dis_{n_ms} is the distance between the horizontal coordinate of the nose key point and the horizontal coordinate of the Lshoulder and Rshoulder key points' midpoint. And Dis_{ls_rs} is the distance between the horizontal coordinate of the Lshoulder and Rshoulder key points. It can be calculated that Dis_{n_ms} is a quarter of Dis_{ls_rs}.

$$x_{ms} = \frac{\left(x_{rs} + x_{ls}\right)}{2}, \tag{11}$$

$$\varepsilon = \frac{\left|x_{nose} - x_{ms}\right|}{\left|x_{rs} - x_{ls}\right|}, \tag{12}$$

where, x_i is the horizontal coordinate of key point i.

Estimation of the rotation angle of the cervical spine is shown as follows.

$$\propto_{cr} = \sin\left(\frac{\varepsilon}{\varepsilon_{90}} \times \frac{\pi}{2}\right) \times 90, \tag{13}$$

where $\propto_{cr}$ is the cervical rotation angle.

As for the lumbar rotation angle, we simply replace the Lshoulder and Rshoulder key points with the Lhip and Rhip key points. We give no further explaination to keep the paper reasonably concise.

3.2 Algorithms of PRC System

The system first applies the same OpenPose multi-person pose estimation framework[11] as in Section 3.1 to estimate the patient's posture while training the TCM Daoyin. After that, we use the pose sequence tracking and classification algorithm and the pose sequence segmentation and similarity calculation algorithm to evaluate the patient's practice posture. Meanwhile, to improve the robustness of the overall network, we add a view adaptive (VA) module[12] to the bottom layer of the network. The VA module performs a two-dimensional transformation of the input pose utilizing the learned rotation matrix parameters and translation matrix parameters.

3.2.1 Pose Sequence Tracking and Classification Algorithm

Almost all existing video action classification algorithms take the whole video as input and get the classification result by a trained deep network. However, since the input must include the whole sequence, even if the inference speed of the deep network can reach real time, this does not meet the demand of real-time classification. In the real-time sign language detection task, Google proposes a lightweight sign language detection network based on pose recognition[13]. It can classify each frame of the video signal in real time. We borrow this idea to classify and track the patient training video in real time by using patient pose information and inter-frame pose optical flow information for each frame[14]. The human pose optical flow information obtained by computing inter-frame based on pose estimation is input to the long short-term memory (LSTM)[15]. It obtains real-time classification results at each frame based on the current inter-frame optical flow characteristics and historical optical flow characteristics. The structure of our network is shown in Fig. 4.

3.2.2 Pose Sequence Segmentation and Similarity Calculation

We obtain the category of each frame of the patient's training video by the algorithm of Section 3.2.1. Then, we segment the video subsequences belonging to the same category to obtain the pose sequence of the patient's current exercise. By the method of similarity calculation we can get the similarity between the patient's current pose sequence and the standard pose sequence. Finally, we score and correct the patient's movements in real time.

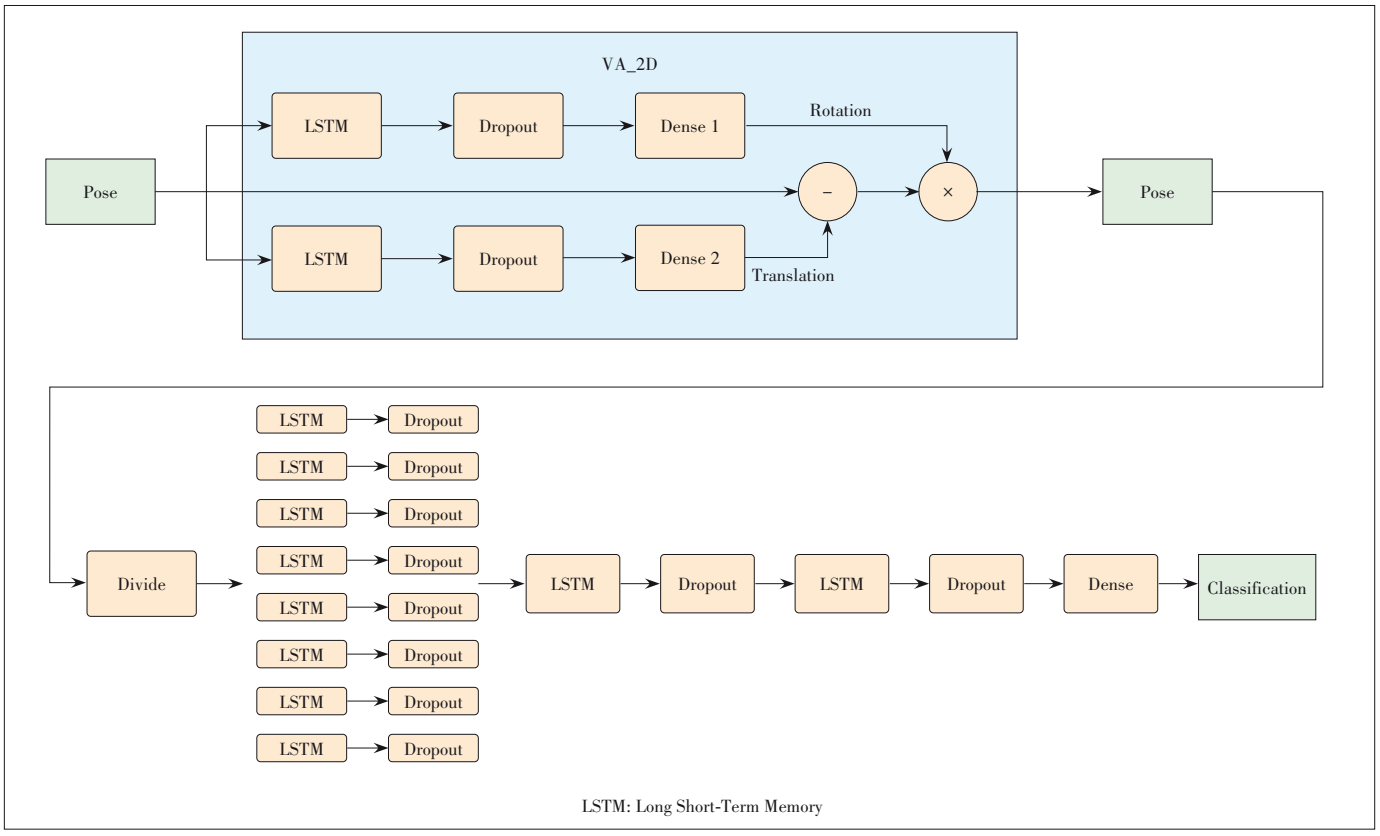
Because of the difficulty in aligning the patient's current pose sequence with the standard pose sequence, we use a dynamic time warping (DTW) algorithm[16]. It finds the best alignment path with the lowest total sequence cost by dynamic programming.

$$\begin{aligned}\mathrm{Cost}(i,j) = D(i,j) + \min\big[&\mathrm{Cost}(i-1,j), \mathrm{Cost}(i,j-1), \\ &\mathrm{Cost}(i-1,j-1)\big],\end{aligned} \tag{14}$$

$$D(i,j) = \sqrt{(x_i^Q - x_j^C)^2 + (y_i^Q - y_j^C)^2}, \tag{15}$$

where $\mathrm{Cost}(i,j)$ is the alignment cost of point i and point j, and $D(i,j)$ is the distance between point i and point j.

After obtaining the alignment path, we calculate the score of the segment based on the key points identified during the patient's practice sequence and the corresponding key points of the standard sequence. Since the movements of



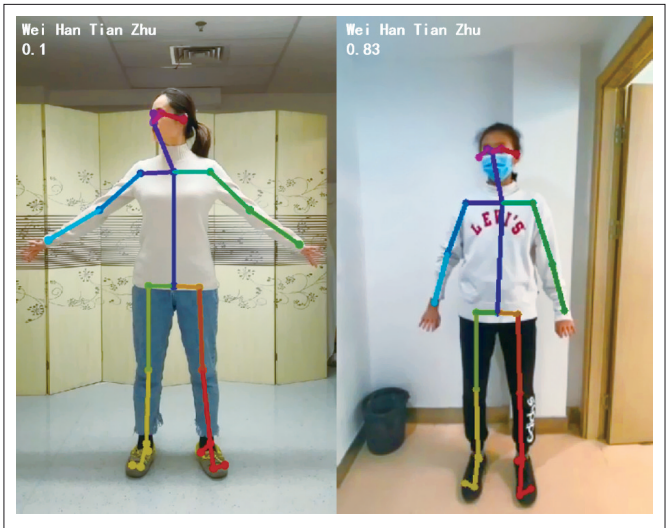
▲Figure 4. Structure of pose sequence classification

each video vary greatly, we choose different key points of attention and different calculation methods for different movements. For example, for the "Niao Shen Niao Fei" movement, the main focus is on the amplitude of the hands up and the amplitude of the single foot when lifted backward; for the "Wei Han Tian Zhu" movement, the main focus is on the horizontal coordinates of the nose and arms. The results are shown in Fig. 5.

3.3 Algorithms of BMR System

Ancient Chinese five-tone music is divided into five keys: Gong, Shang, Jue, Zhi, and Yu. However, because of the wide variety of music libraries currently available, it is not practical to manually classify songs. Therefore, we use LSTM[15] as a backbone and design an algorithm to implement an automatic judgment of the main key of the song. The algorithm is divided into three specific steps: the main melody transcription to determine the pitch of the ending, the tonality detection to determine the key number, and the main tone discrimination according to music theory.

Building a five-tone music database is highly complex due to the different meanings of traditional five-tone Chinese tuning and western music systems. Currently, there is a lack of a relatively complete dataset on the internet. We invite Dr. XIA from Sichuan Conservatory of Music to set up a working group



(a) The standard pose (b) The patient's pose

▲Figure 5. Results of posture recognition and correction (PRC) system

to help us annotate traditional five-tone music and build a small dataset. The dataset we build is used to train the network model to achieve the automatic classification of traditional Chinese music. Finally, a background music library of 500 songs is constructed to assist patients in the practice of TCM Daoyin.

3.4 Algorithms of PSM System

The detection of physiological signals is often used as a criterion for the physician's judgment during the patient's rehabilitation process. Therefore, our system needs to enable real-time measurement and monitoring during the patient's TCM Daoyin training. Our algorithm is developed to design a portable and miniaturized physiological signal monitoring device. ECG signals can be used to represent the electrical signals of heart activity[17] and are widely used in the detection of arrhythmia. In this paper, the BMD101 chip is selected to acquire and detect the heart rate signal of patients. At the same time, abnormal respiratory rates are sometimes one of the best independent predictors of cardiac arrest in clinical practice. The device also incorporates an algorithm to extract respiratory signals from ECG signals to monitor the patient's respiratory rates.

3.4.1 Heart Rate Detection and Judgment

The BMD101 chip[18] is a miniature device developed by NeuroSky specifically for bio-signal detection and processing. The device uses a single-lead detection method, which allows detection of ECG signals at the μV to mV level by simply placing the electrode pad on the chest. Moreover, it can realize signal pre-processing and calculate indexes such as mean heart rate and heart rate variability.

The chip's built-in algorithm parses the patient's current static heart rate from the data stream and calculates the average heart rate. When we get the patient's static heart rate, we can get the patient's current theoretical maximum heart rate and theoretical minimum heart rate based on the physician's clinical studies. And based on this, we can make a judgment about the patient's heart rate stability.

$$HR_{max} = \left(220 - Age - HR_{static}\right) \times 0.3 + HR_{static}, \tag{16}$$

$$HR_{min} = HR_{static} - 10, \tag{17}$$

where HR_{max} and HR_{min} indicate the theoretical maximum and minimum number of heart rates per minute, respectively. And HR_{static} indicates the number of resting heart rate per minute.

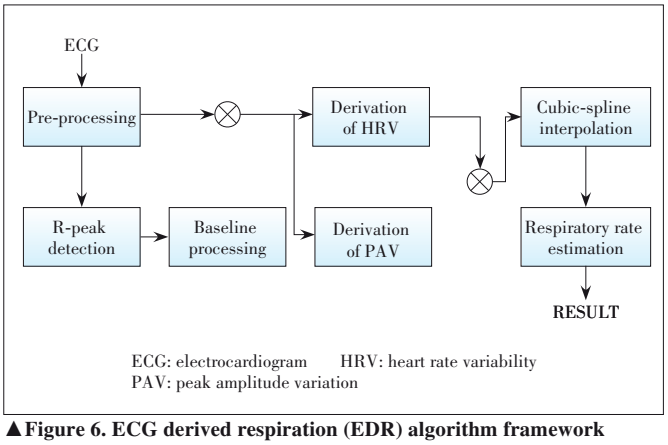
3.4.2 Deriving Respiratory Rates from ECG Signals

Previous studies have shown that respiratory rates can be derived from ECG signals by analyzing either the variability in R - R intervals during a respiratory cycle or the change in QRS (the Q, R, and S wave groups) amplitude during breathing. The EDR[19-20] is obtained using both the heart rate variability (HRV) method and the peak amplitude variation (PAV) method, as shown in Fig. 6.

Because the actual measurement to obtain the ECG signal is influenced by body temperature changes, industrial frequency interference, and visceral activity, it generates baseline drift and other noise. We obtain the original ECG signal by removing the baseline drift from the detected R-peaks using the cubic-spline interpolation method. And due to the low frequency of the respiratory signal (0.1 - 0.7 Hz), we need to downsample and smoothly filter the obtained signal to finally obtain the respiratory signal. Table 2 shows the results.

4 Device Integration

We integrate the subsystems into a mini-PC to make the system portable and operational and to ensure that patients can practice wireless self-rehabilitation at home and in the community. And the device is equipped with a Bluetooth module as well as a 5G module, which facilitates data transmission from sensors and ensures that doctors can monitor patients'



▲Figure 6. ECG derived respiration (EDR) algorithm framework

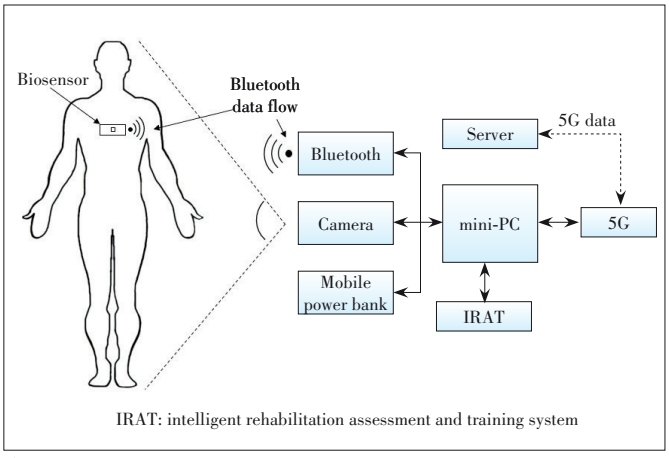
▼Table 2. Results of physiological signal monitoring (PSM) system

	Estimated Heart Rate/(beat/m)	Real Heart Rate/(beat/m)	Estimated Respiratory Rate/(breath/m)	Patient Status
1	89	86	16	Stabilization
2	85	87	16	Stabilization
3	87	85	17	Stabilization
4	84	83	15	Stabilization
5	82	83	16	Stabilization
6	83	85	16	Stabilization
7	87	85	17	Stabilization
8	90	85	18	Stabilization
9	87	86	16	Stabilization
10	82	86	15	Stabilization
11	83	86	16	Stabilization
12	88	85	17	Stabilization
13	94	85	18	Stabilization
14	85	88	16	Stabilization
15	85	89	16	Stabilization
16	90	90	18	Stabilization
17	93	90	18	Stabilization
18	95	90	18	Stabilization
19	87	88	16	Stabilization
20	86	86	17	Stabilization

rehabilitation training in real time on a remote server.

As shown in Fig. 7, we choose a Core i7-7820HQ processor and a GTX1650 graphics card with 4 GB of video memory to support the hardware requirements for our algorithm implementation. It achieves real-time feedback on the scoring of each index during patients' TCM Daoyin training. To improve the accuracy rate, we choose a high-definition camera to satisfy the needs of computer vision. Meanwhile, to collect the ECG signal from the patient, we embed a Bluetooth module in the device to enable remote wireless transmission of ECG signals. To maintain the overall portability of the device, we choose a removable power supply. The device also needs to ensure that the server receives and feeds back quickly and doctors could monitor and guide patients in time. Therefore, we embed a 5G communication module in our device.

When the equipment works, we set up a variety of options to provide a more comprehensive rehabilitation strategy for users with different needs. The specific process is shown in Fig. 8.

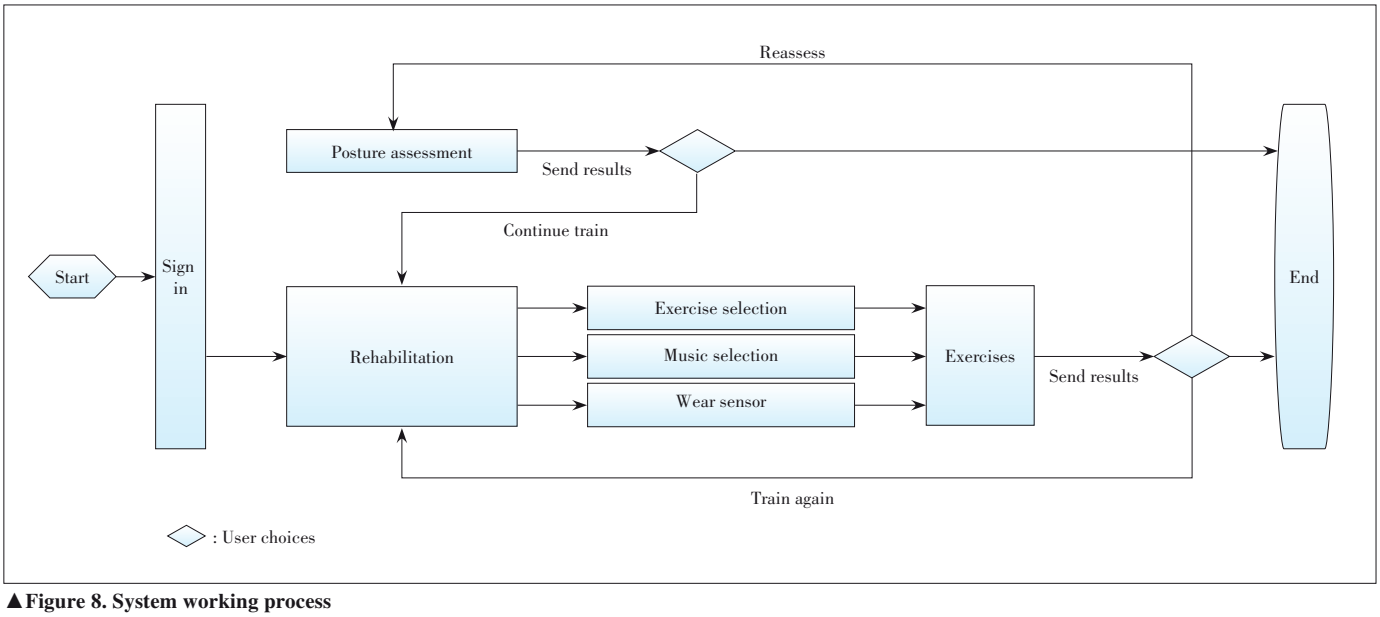


▲Figure 7. System hardware equipment

5 Conclusions

In this paper, we innovatively combine AI and TCM Daoyin for rehabilitation assessment and training of patients with DJD. We integrate the system into a mini PC to ensure the portability of our device. The system is equipped with a Bluetooth module and a 5G communication module, and therefore fast and wireless data transmission between the system and the server is achieved. Each of our subsystems has its innovative points. The SMA system detects the patient's posture by computer vision algorithm and gives a skeleton model to assess the patient's spine status by our innovatively proposed scoring formulas. The PRC system classifies the posture of patients during TCM Daoyin training and proposes different scoring methods based on different standard movements. What's more, we construct a small sample dataset. The BMR system is based on the five-tone theory of TCM to assist in patient therapy. We also build a small sample dataset to achieve the automatic classification of the five-tone music. The PSM system calculates the real-time heart rate of the patient by collecting the ECG signal of the patient. The real-time respiratory rate of the patient is separated from the ECG signal by the EDR algorithm, which achieves timely alerting of abnormal situations.

In all, our system is geared toward patients with DJD who are in remission. The system combines multiple algorithms of AI, and more importantly, it is wireless and portable so that it achieves the rehabilitation treatment of patients at home and in the community. The device has high practicality, a wide range of applications, and effective dissemination of TCM theory.



▲Figure 8. System working process

References

[1] ZHU D, ZHANG G, GUO X, et al. A new hope in spinal degenerative diseases: Piezo1 [J]. BioMed research international, 2021: 6645193. DOI:10.1155/2021/6645193
[2] HU J. The therapeutic and health effects of moxibustion on Degeneration of Joint Disease (in Chinese) [J]. Journal of Hunan University of Chinese Medicine, 2011, 31(2): 32 – 33
[3] YANG D G, LI P Z, SHAO C L, et al. Li Yefu's treatment of Degeneration of Joint Disease by applying the idea of "Jin Gu Bing Ju, Chan Xuan Xiang Ji" (in Chinese) [J]. Journal of Anhui university of Chinese medicine, 2018, 37 (02): 43 – 45
[4] LIU Y X, DAI L B. Therapeutic Effect of Chinese Medicine Packet Therapeutic Instrument Plus Low-frequency Pulsed Electromagnetic Field Therapeutic Instrument for Lumbar Disc Herniation (in Chinese) [J]. Journal of Guangzhou university of traditional Chinese medicine, 2018, 35(04): 655 – 658
[5] CUI Y Z. Clinical care study on the treatment of lumbar intervertebral disc herniation by combining specific electromagnetic wave irradiation with external application of Chinese medicine (in Chinese) [J]. Shanxi medical journal, 2016, 45 (19): 2336 – 2338
[6] YAN H Z, MAO G H. Effect of three-dimensional curvature traction instrument combined with intermediate frequency therapy instrument in patients with cervical spondylosis (in Chinese) [J]. Shandong medical journal, 2017, 57 (13): 90 – 92
[7] QIN Y, WANG Y X. Diode Laser irradiation on ganglion stellatum and traction in treatment of vertebral artery type cervical spondylosis [J]. Chinese journal of laser medicine & surgery, 2010, 19(02): 102 – 105
[8] CHEN M Y, YU Zhongshun. On the Application of Traditional Chinese Daoyin in the Elderly Health (in Chinese) [J]. Wushu studies, 2021,6 (01): 119 – 122
[9] ZHAO L Z, CHEN Y G. Research progress on TCM five-note therapy [J]. China journal of traditional Chinese medicine and pharmacy, 2016, 31(11): 4666 – 4668
[10] GONG Z Z, DU Y Y, XU M H, et al. Theoretical Analysis of TCM Wuyin Therapy for insomnia (in Chinese) [J]. Chinese journal of medicinal guide, 2021, 23 (02): 96 – 99
[11] CAO Z, SIMON T, WEI S H, et al. Realtime multi-person 2D pose estimation using part affinity fields [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017: 1302 – 1310. DOI: 10.1109/CVPR.2017.143
[12] ZHANG P F, LAN C L, XING J L, et al. View adaptive neural networks for high performance skeleton-based human action recognition [J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 41(8): 1963 – 1978. DOI: 10.1109/TPAMI.2019.2896631
[13] MORYOSSEF A, TSOCHANTARIDIS I, AHARONI R, et al. Real-time sign language detection using human pose estimation [EB/OL]. [2021-03-25]. https://www. researchgate. net/publication/343599175_Real-Time_Sign_Language_Detection_using_Human_Pose_Estimation. DOI: 10.1007/978-3-030-66096-3_17
[14] LUCAS B D, KANADE T. An iterative image registration technique with an application to stereo vision [C]//The 7th International Joint Conference on Artificial Intelligence. Vancouver, Canada: IJCAI. 1981
[15] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural computation, 1997, 9(8): 1735 – 1780. DOI: 10.1162/neco.1997.9.8.1735
[16] KEOGH E, RATANAMAHATANA C A. Exact indexing of dynamic time warping [J]. Knowledge and information systems, 2005, 7(3): 358 – 386. DOI: 10.1007/s10115-004-0154-9
[17] CHOI H S, LEE B, YOON S. Biometric authentication using noisy electrocardiograms acquired by mobile sensors [J]. IEEE access, 2016, 4: 1266 – 1273. DOI: 10.1109/ACCESS.2016.2548519
[18] YANG K, CONG L, HU W D, et al. Embedded wireless ECG monitoring system based on BMD101 [J]. Application of electronic technique, 2014, 40(1): 122 – 124. DOI: 10.16157/j.issn.0258-7998.2014.01.036
[19] SARKAR S, BHATTACHERJEE S, PAL S. Extraction of respiration signal from ECG for respiratory rate estimation [C]//Michael Faraday IET International Summit. Kolkata, India: IET, 2016
[20] GUO Y, SI Y J. Respiratory signal extraction algorithm based on ECG [J]. Journal of Jilin University (Information Science Edition), 2016, 34(03): 327 – 333

Biographies

LIU Weichen is pursuing his master's degree in the Academy for Engineering & Technology (AET), Fudan University, China. His research interests include machine learning and signal processing.

SHEN Mengqi is a graduate student of the School of Computer Science and Technology, Fudan University, China. His main research interests include action recognition, video understanding and human pose estimation.

ZHANG Anda is pursuing his master's degree of electronic information in AET, Fudan University, China. His research interests include artificial intelligence and deep learning related to Chinese Medicine.

CHENG Yiting received the M.S. degree in computer science from Fudan University, China in 2021. Her main research interests include unsupervised learning and computer vision.

ZHANG Wenqiang (wqzhang@fudan.edu.cn) is a professor with the School of Computer Science, Fudan University, China. He received his Ph.D. degree in mechanical engineering from Shanghai Jiao Tong University, China in 2004. His current research interests include computer vision and robot intelligence.

A Survey of Intelligent Sensing Technologies in Autonomous Driving

SHAO Hong[1], XIE Daxiong[1,2], HUANG Yihua[1]

(1. ZTE Corporation, Shenzhen 518057, China;
2. State Key Laboratory of Mobile Network and Mobile Multimedia, Shenzhen 518057, China)

Abstract: Intelligent perception technology of sensors in autonomous vehicles has been deeply integrated with the algorithm of autonomous driving. This paper provides a survey of the impact of sensing technologies on autonomous driving, including the intelligent perception reshaping the car architecture from distributed to centralized processing and the common perception algorithms being explored in autonomous driving vehicles, such as visual perception, 3D perception and sensor fusion. The pure visual sensing solutions have shown the powerful capabilities in 3D perception leveraging the latest self-supervised learning progress, compared with light detection and ranging (LiDAR)-based solutions. Moreover, we discuss the trends on end-to-end policy decision models of high-level autonomous driving technologies.

Keywords: autonomous vehicles; neuron network; automotive electronics; sensor fusion

DOI: 10.12142/ZTECOM.202103007

https://kns.cnki.net/kcms/detail/34.1294.TN.20210817.1538.004.html, published online August 18, 2021

Manuscript received: 2021-06-22

Citation (IEEE Format): H. Shao, D. X. Xie, and Y. H. Huang, "A survey of intelligent sensing technologies in autonomous driving," *ZTE Communications*, vol. 19, no. 3, pp. 56 - 63, Sept. 2021. doi: 10.12142/ZTECOM.2021030007.

1 Introduction

The next generation vehicles are transforming from mechanical-centric to software-defined. Since the Grand Challenge orchestrated by the Defense Advanced Research Projects Agency (DARPA)[1], autonomous driving (AD) technologies have been accelerating. Autonomous driving is considered to be a revolutionary technology that profoundly affects human society and transportation. To categorize these systems, the Society of Automobile Engineers (SAE) has defined six levels of automation ranging from 0 (no automation) to 5 (full automation). The deployment of full autonomy is still expected in years. Automotive manufactures are more inclined to gradually increase the level of autonomy from Advanced Driver Assistance Systems (ADAS) to full autonomous driving. The ADAS ranges on the spectrum of passive to active safety functions, such as forward collision warning (FCW), lane departure warning (LDW), blind spot monitoring (BSM), autonomous emergency braking (AEB), lane keeping assistance (LKA), adaptive cruise control (ACC), forward collision-avoidance (FCA), traffic jam assist (TJA), and traffic jam pilot (TJP).

Autonomous driving cars need to understand the surrounding environment and then take actions continuously. Autonomous vehicles rely on different sensors that work together to perceive the internal and external car environments. The most involved sensors in the car are radars, light detection and ranging (LiDAR) systems, cameras, ultrasonic and far-infrared sensors, etc.[2] These long-range and short-range sensors provide relevant data to interpret the surrounding scenes near the vehicle with a variety of solutions, such as from 8 Vision 1 Radar

(8V1R) to 15 Vision 5 Radar 3 LiDAR (15V5R3L). Sensor fusion processing is also deeply integrated into the algorithms of autonomous driving. Smart sensors combined with the extreme compute performance deployed in a car make the car more and more like a robot. This will be continuously increasing the complexity and bringing challenges for the automotive electronic architecture.

The goal of this paper is to provide a survey of sensing technologies on autonomous vehicles. Ref. [2] reviewed the most popular sensor technologies and their characteristics but did not analyze how the progress of the algorithms affected the configuration of vehicle sensors. Here we track some important improvements of the neural network and deep learning algorithms linked with perception in autonomous driving. The well-known mask region based convolutional neural network (Mask R-CNN) algorithm[3] achieves the best instance segmentation accuracy in 2D visual recognition. The popular vision algorithm You Only Look Once (YOLO)[4] is less accurate but much faster than Mask R-CNN and suitable for autonomous driving. YOLO is also extended to LiDAR 3D point clouds[5]. The fusion of multiple sensors like vision and LiDAR[6] has taken more advantages before Pseudo LiDAR technology[7] emerges and the latter is showing the power of pure vision in 3D perception. Unsupervised learning approaches of depth estimation[8] have further accelerated the utilization of pure vision in autonomous driving. Moreover, sensors should not only perceive the current environment, but also constantly predict the environmental context in the next few seconds. For example, Uber uses a convolutional neural network (CNN) model to predict possible trajectories of the surrounding actors[9].

The structure of this paper is arranged as follows. Section 2 explains the impact of intelligent sensing technology on automotive electronic architecture. Section 3 provides a detailed overview of the sensing algorithms. Section 4 discusses a decision-making model. Finally, Section 5 concludes this paper.

2 Impacts on Electrical/Electronic (E/E) Architecture

Autonomous driving requires processing of dozens of sensors with high performance computation. This brings new impacts on the traditional automotive electrical and electronic architecture. Firstly, centralized processing will replace the distributed processing to provide high computational power. Secondly, sensor data transmission will require higher communication bandwidth and time-sensitive networking (TSN) becomes the promising technology for it.

2.1 Centralized Computing

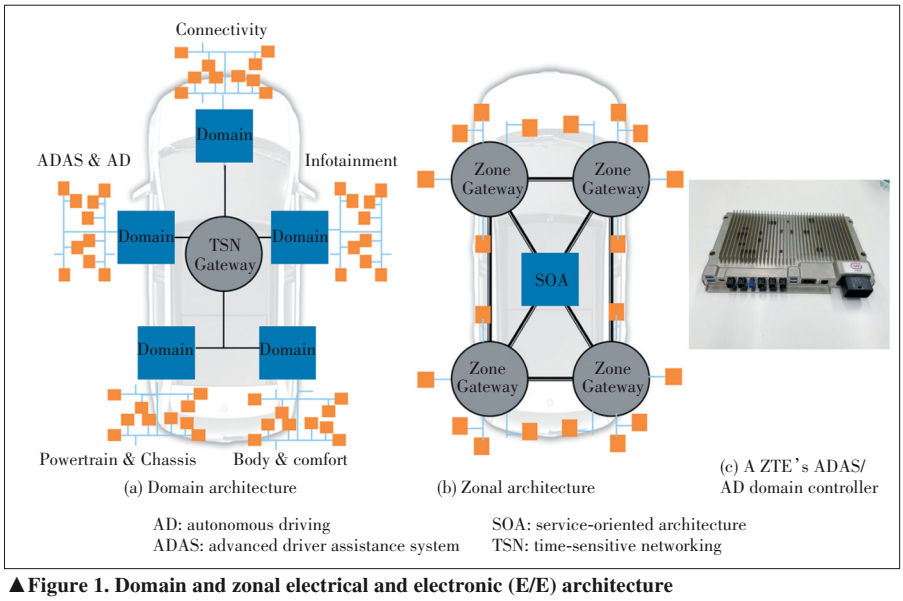
Traditional cars are composed of one-box one-function modular electronic control units (ECUs). However, due to the complexity of autonomous vehicles, the approach where ECUs are tightly coupled with firmware from hardware will encounter difficulties to meet the requirements of high computation power and software integration in intelligent perception. Regarding the increasing number of sensors and actuators in autonomous vehicles, there are several impacts on the legacy automotive E/E architecture such as complexity, harness, high bandwidth, and artificial intelligence (AI) computing.

The distributed modular ECU system needs to upgrade to an integrated centralized computing system for autonomous driving. The future trend is combining sensors and ECUs into the domain controller (Fig. 1). Then all domain controllers will be further merged into one centralized vehicle computing platform with functional redundancy to achieve functional centralization.

Fig. 1 shows the domain architecture and the zonal architecture with a centralized vehicle computing platform to further optimize the harness layout. The domain architecture consists of separated domain controllers according to the vehicle functions. The zonal architecture consists of gateways connected with the redundant computing platform that supports service oriented architecture (SoA) to process the vehicle functions. Fig. 1(c) shows a ZTE's ADAS/AD domain controller, which can be used in L2 and L3 autonomous driving scenarios.

2.2 Time-Sensitive Networking

Another impact is the high data rate sensors and actuators, such as raw data cameras, LiDARs and radars, which will need high bandwidth and deterministic real-time communication within the car. As early as 2006, the IEEE802.1 established the audio video bridging (AVB) working group and suc-



▲Figure 1. Domain and zonal electrical and electronic (E/E) architecture

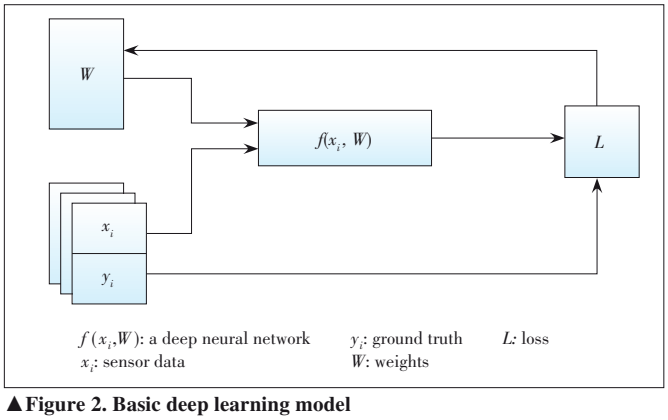
cessfully solved real-time synchronous data transmission in the following years. This immediately attracted the attention of the automotive industry. In 2012, The AVB working group was renamed by the TSN working group, focusing on enabling low-latency and high-quality transmission of streaming data. TSN aims to establish a "universal" time-sensitive mechanism for the Ethernet protocol to ensure the time determinism of network data transmission and the delay reaches the microsecond level. So it will be the ideal candidate communication backbone technology for the new automotive E/E architecture. As shown in Fig. 1, in the automotive backbone which requires high bandwidth and deterministic real-time communication, the TSN gateway is used to ensure that it can transmit between different domains with low latency and small jitters. The TSN node will also transmit a redundant frame for the high-performance sensors (e.g., high-resolution cameras). The Ethernet TSN can reach 1 Gbit/s or more, while the controller area network (CAN) and FlexRay are 1 Mbit/s and 20 Mbit/s respectively.

3 Sensing Algorithms

At present, most intelligent perception tasks are achieved by deep neural network models. Here we analyze several basic algorithms and their development.

3.1 Deep Learning Models

For intelligent sensing, a very basic deep learning model is well known as illustrated in Fig. 2. A deep neural network $f(x_i,W)$ consists of multiple layers. For example, it could be combinations of CNN, fully connected (FC) networks, residual networks (ResNet), long short term memory (LSTM) networks, even transformer networks, etc. An input data set $\{(x_i,y_i)\}_{i=1}^{n}$ is sampled from the collected data which could be historical driving data or synthesized data constructed from a simulator. Within the input data set where x_i is denoted as the sensor data that could be images, videos, or LiDAR point clouds and y_i denoted as the ground truth of the target or labels that are extracted from the car environment. The target y_i can be spatial information (like drivable area, lanes, roads, etc.), semantic information (traffic lights, traffic signs, turn indicators, on-road marking, etc.), or moving objects (pedestrians, cyclists, cars, etc.). We can manually label them or generate them by a simulator.



▲Figure 2. Basic deep learning model

The difference between the predicted result of the neural network f and the ground truth y_i is the loss function denoted as L. The training goal is to optimize L through iterations of the data set and adjust the weights W.

The weights W could be a very large tensor including all the weights of each deep network layer. The neural network f can do the inference once we get the optimal weight W^*:

$$W^* = \arg\min_W \frac{1}{n}\sum_{i=1}^{n} L(f(x_i;W),y_i), \tag{1}$$

where $W = \{W^{(0)},W^{(1)},\dots\}$.

Reasonably defining the loss function L is the most important work for designing deep learning model. For example, for the classification case, we use the binary cross-entropy to measure the loss:

$$L = \frac{1}{n}\sum_{i=1}^{n} y_i \log(f(x_i;W)) + (1-y_i)\log(1-f(x_i;W)). \tag{2}$$

And for the regression case, we use mean square error (MSE) to measure the loss:

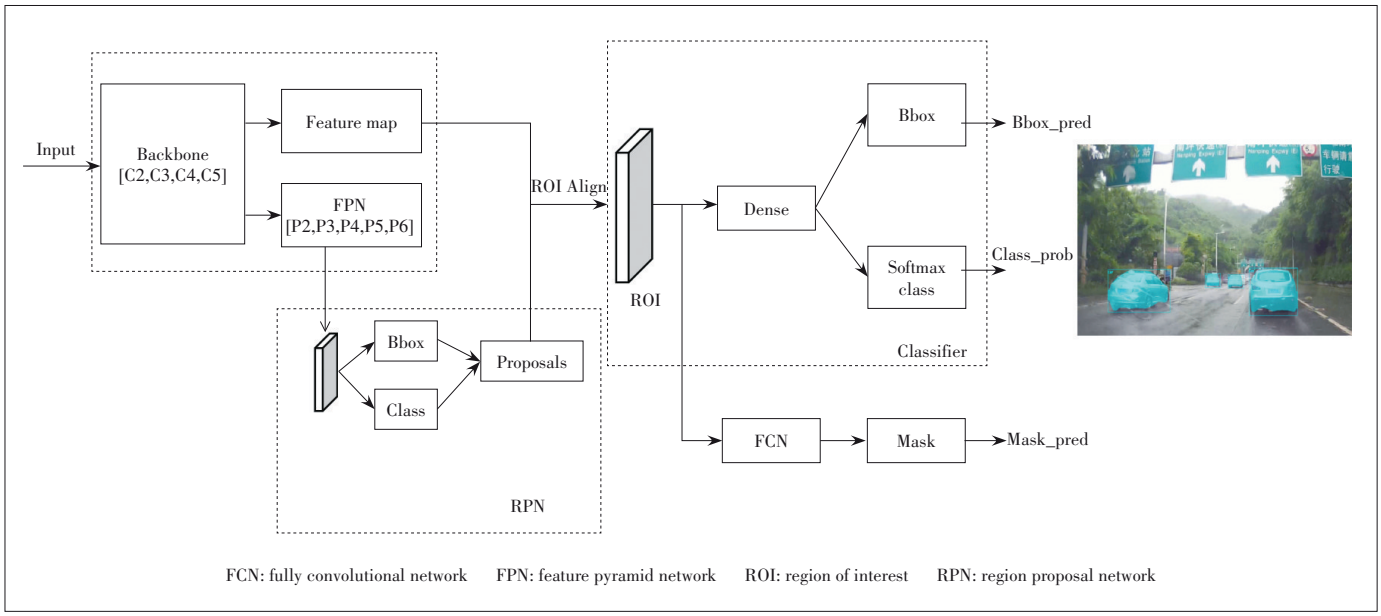
$$L = \frac{1}{n}\sum_{i=1}^{n} (y_i - f(x_i;W))^2. \tag{3}$$

We can also put extra terms in the loss function to reduce its generalization error but not its training error. These strategies are as known as regularization.

3.2 Visual Perception Understanding

Computer vision has almost become the foundation of intelligent perception for autonomous driving. Many visual recognition problems are related to autonomous driving, such as object detection, segmentation and instance detection. We has built a practice Mask R-CNN[3] for visual perception, as shown in Fig. 3.

We use a ResNet50[10] to construct the backbone of the Mask R-CNN. The original images are resized to a fixed size before entering the backbone network. The feature maps exacted from the backbone are C2, C3, C4, and C5, which construct the feature pyramid networks (FPN) in order to detect the objects from different scales. FPN has a bottom-up and top-down structure to connect the corresponding layer and generate a new feature map [P2, P3, P4, P5, P6], where P5 corresponds to C5, P4 corresponds to C4+ UpSampling2D of P5, P3 corresponds to C3+ UpSampling2D of P4, P2 corresponds to C2+ UpSampling2D of P3, and P6 corresponds to MaxPooling2D of P5. Then this feature map is put into a region proposal net-



▲Figure 3. A practiced mask region based convolutional neural network (Mask R-CNN)

work (RPN) to generate the region of interest (ROI) proposals and tune the coordinates.

This network completes three tasks simultaneously: 1) target localization, which directly predicts a target bounding box on the image. Here it is called Bbox_pred; 2) target classification, which is denoted as Class_prob; 3) pixel-level target segmentation, denoted as the Mask_pred for each ROI.

The total loss of the network is:

$$L = L_{cls} + L_{box} + L_{mask}, \quad (4)$$

where L_{cls} is the classification loss, L_{box} is the bounding-box loss, and L_{mask} is the average binary cross-entropy loss of pixel-wise mask for each instance.

Although the accuracy of Mask R-CNN is relatively high, its region proposal pipelines are still time-consuming. For autonomous driving, since the perception task requires real-time performance, we need to make the inference efficient and maintain good accuracy. So a single shot detector model like YOLO[4] will be a better choice than Mask-R-CNN. It can process videos at real time.

3.3 3D Perception

3D object detection in autonomous driving is a common task. A direct and reliable approach is employing the LiDAR sensor to provide the 3D point cloud reconstruction of the surrounding environment. The object detection and classification of LiDAR point clouds may be conducted in 2D bird-view by projecting the 3D point clouds into 2D or directly conducted in 3D space. Unlike visual systems, LiDAR point clouds lack of rich RGB information, the density of the point cloud is critical for small object detection.

For high resolution LiDAR, YOLO3D[5] introduced a 3D ob-

▼Table 1. YOLO3D network architecture[5]

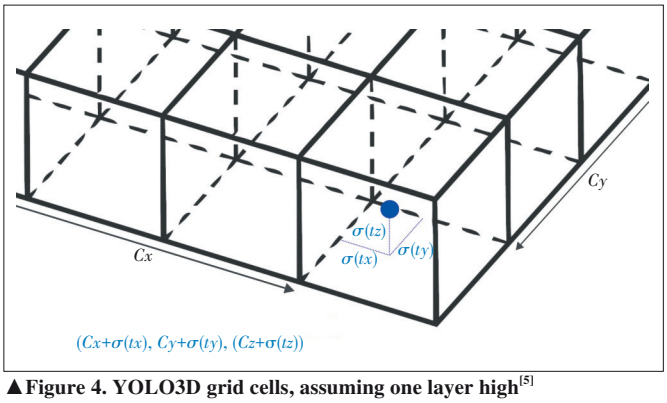
Layer	Filter	Size	Feature Maps
Conv2d	32	(3, 3)	608×608×2
Maxpooling		(size 2, stride 2)	
Conv2d	64	(3, 3)	
Maxpooling		(size 2, stride 2)	
Conv2d	128	(3, 3)	
Conv2d	64	(3, 3)	
Conv2d	128	(3, 3)	
Maxpooling		(size 2, stride 1)	
Conv2d	256	(3, 3)	
Conv2d	128	(3, 3)	
Conv2d	256	(3, 3)	
Maxpooling		(size 2, stride 2)	
Conv2d	512	(3, 3)	
Conv2d	256	(1,1)	
Conv2d	512	(3, 3)	
Conv2d	256	(1,1)	
Conv2d	512	(3, 3)	
Maxpooling		(size 2, stride 2)	
Conv2d	1 024	(3, 3)	
Conv2d	512	(1, 1)	
Conv2d	1 024	(3, 3)	
Conv2d	512	(1, 1)	
Conv2d	1 024	(3, 3)	
Conv2d	1 024	(3, 3)	
Conv2d	1 024	(3, 3)	
Conv2d	1 024	(3, 3)	
Conv2d	1 024	(1,1)	38×38×33
Reshape			38×38×3×11

ject detection algorithm that direct expands from the 2D algorithm YOLO[4]. This approach directly projects the LiDAR point cloud to the bird-view space for real-time classification and detecting 3D Object Bounding Box. The structure of the network is shown in Table 1.

The network output prediction is expanded by the YOLO regression to 3D dimensions regression output and target classification. It will return the object bounding box center (x, y, and z), the 3D dimensions (length, width, and height), the orientation in the bird-view space, the confidence, and the object class label. The YOLO3D grid expanding from YOLO is shown in Fig. 4.

The loss also extends from the YOLO 2D boxes (x, y, l, h) to 3D oriented boxes (x, y, z, w, l, h) and the orientation. The total loss includes the confidence score and the cross-entropy loss over the object classes.

The YOLO3D network is trained end to end because it is a single shot detector that ensures its real-time 3D performance in the inference path.



▲Figure 4. YOLO3D grid cells, assuming one layer high[5]

3.4 Sensor Fusion

If the density of the LiDAR point cloud is sparse, small objects like pedestrians and cyclists will be hard to recognize. The LiDAR needs to do a sensor fusion with the camera. The sensor fusion can take advantage of both LiDAR point clouds and camera images, which can preserve more semantic information to achieve higher object detection accuracy. Therefore, the autonomous driving cars are commonly equipped with multiple kinds of sensors like both LiDAR and camera.

The multi-view 3D (MV3D) network[6] gives an example of sensor fusion framework that takes 3D LiDAR point clouds and RGB images as input to predict 3D objects (Fig. 5).

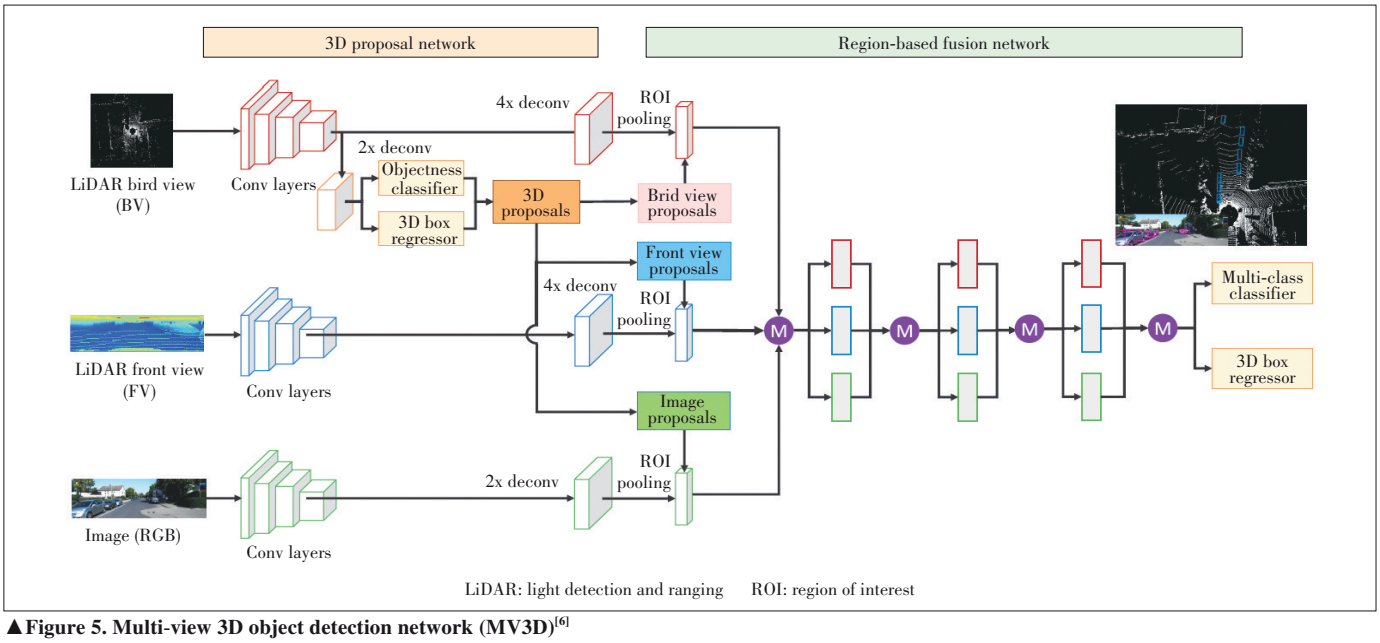
The MV3D network consists of two sub-nets: a 3D proposal network and a region-based fusion network. The 3D proposal network generates highly accurate 3D candidate boxes from the bird's eye view of a point cloud. The region-based fusion network deeply fuses multi-view features to predict the position, size, and orientations of the 3D target.

A multiple layer feature fusion is adopted to increase the selected ROI from the fusion between different view features. The fusion network can significantly improve the position accuracy and recognition accuracy of 3D perception.

3.5 Pure Vision vs. LiDAR

In autonomous driving cars, LiDAR or pure vision based solution has become a controversial topic. We mainly consider the perception algorithm to put aside the cost, weather environment and other factors.

There is a vast difference in perception between pure vision and the LiDAR solution. While in the LiDAR case, the vehicle detects the 3D object to avoid collisions of pedestrians, bicycles, vehicles, etc. and it also compares the features of real-time 3D point clouds with a pre-built high definition map, uti-



▲Figure 5. Multi-view 3D object detection network (MV3D)[6]

lizing the simultaneous localization and mapping (SLAM) algorithm to precisely localize the vehicle position and execute lane follow function. However, for pure vision cases, the camera creates 2D information and it is difficult to reliably and accurately reconstruct the 3D environment of each pixel, so the SLAM will not be conducted to directly predict the lane from the camera images. This limits the pure vision solution under the L3 autonomy.

With the progress of monocular or binocular camera 3D perception, the accuracy of depth estimation is continuously improved and even pseudo-LiDAR can be constructed, as shown in Fig. 6[7].

With the input of stereo or monocular images, the network can predict the depth map and back-project it into a 3D point cloud in the LiDAR coordinate system called a pseudo-LiDAR, so we can reuse the LiDAR-based algorithms in the pure visual solution and also implement high-level autonomy.

3.6 Self-Supervised Learning

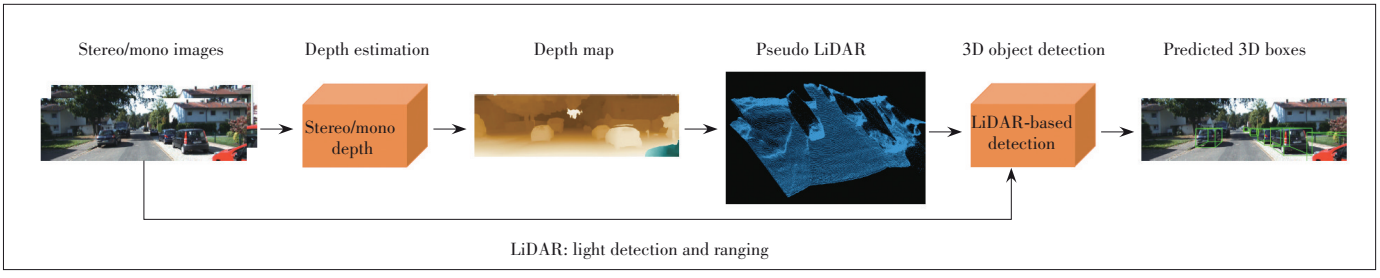
The supervised learning requires the provision of a data set with depth information as a ground truth. However, the ground-truth of depth information of visual data is more difficult to obtain, so pure visual depth perception technology mentioned in the previous section is limited to a restricted training data set, while the self-supervised method developed later does not require depth information annotation and directly uses video frames to complete the training, which is a great improvement.

The self-supervised learning method is to reconstruct the related pixels through geometric constraints between two frames of the multi-view as the supervision input so that there is no need to rely on the annotation of depth. In the backbone network, it is the same as the original network of supervised learning.

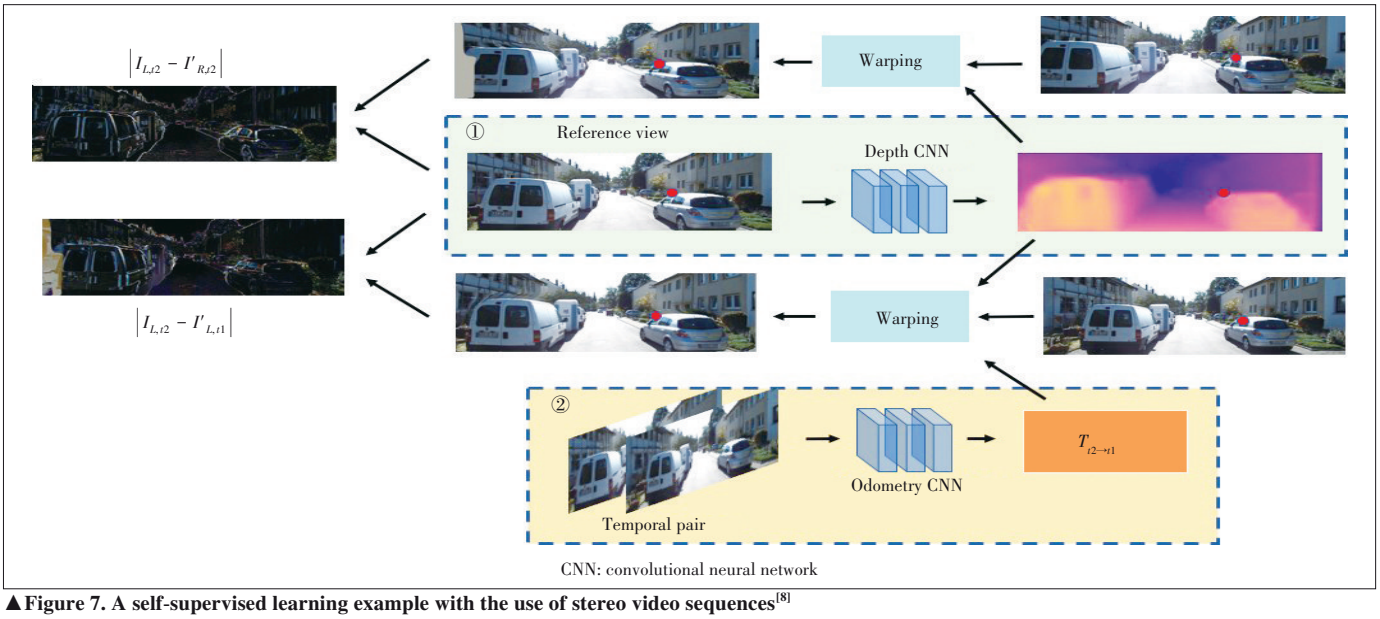
A self-supervised learning example[8] is shown in Fig. 7. It can estimate the depth and movement of the camera using stereo video sequences.

Ref. [8] enables the use of both spatial and temporal photometric warp errors, and constrains the scene depth and camera motion in a common real-world scale.

As shown in Fig. 8, a convolutional neural network for single view depth (CNN_D) and a convolutional neural network for visual odometry (CNN_{VO}) are used. For self-supervised learning, the fundamental supervision signal comes from the task of image reconstruction and the image reconstruction loss is used as a supervision signal to train CNN_D and CNN_{VO}.



▲Figure 6. Image-based 3D object detection[7]



▲Figure 7. A self-supervised learning example with the use of stereo video sequences[8]

4 End-to-End Decision Model

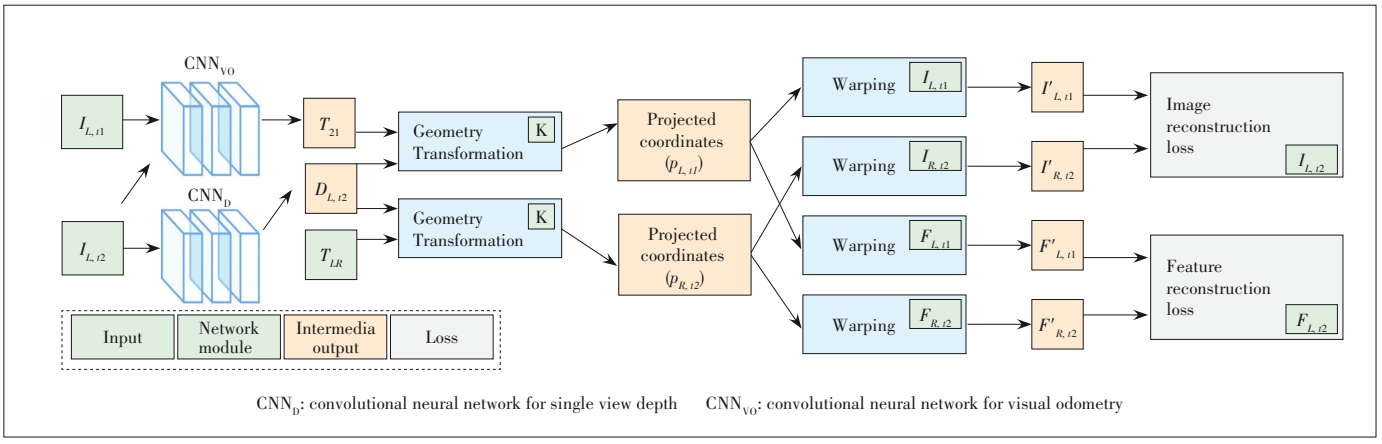
A typical autonomous intelligent vehicle system can be simply divided into three parts: the perception part, the planner and the controller. The perception part extracts the features from the environment and the planner outputs a driving trajectory for driving in the 3D space. The controller then executes this trajectory as the steering angle and acceleration within the physical constraint of the vehicle. We call the decision model end-to-end when it takes in sensing data and outputs how we should drive with fully end-to-end training approach to mimic human driving.

We usually implement autonomous driving by using a rule-based planner to make trajectory decisions. Engineers will manually write the planner for such an autonomous driving system. Due to the complexity of the driving problem, the manual rule-based planner may never enable the level of full autonomy because the edge cases, such as temporary road signals and traffic accidents, are constantly increasing in new scenarios. To build an end-to-end intelligent planner based on neural network is an idealist goal people want to achieve[9].

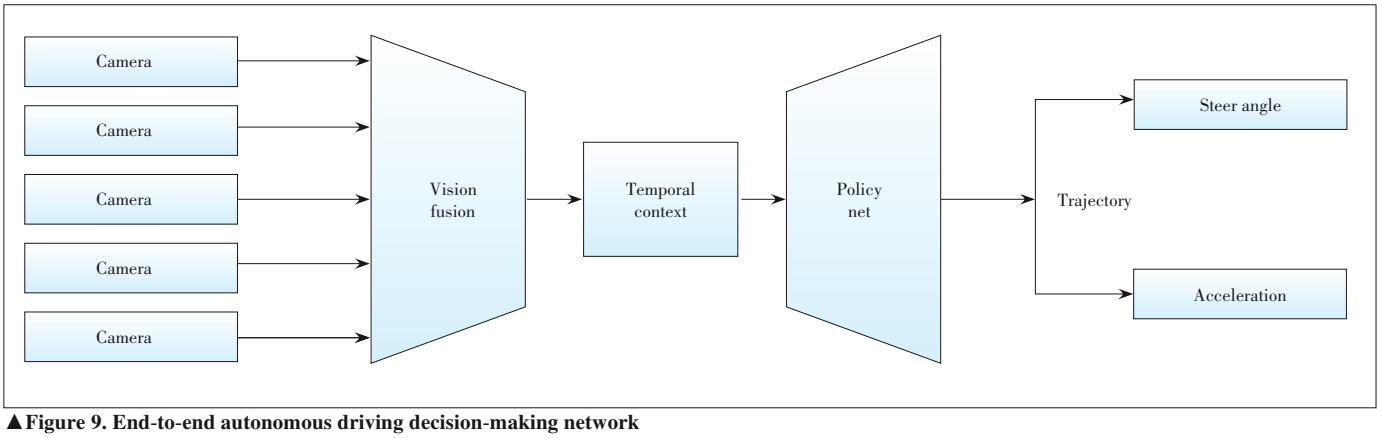
The reinforcement learning algorithms such as AlphaGo, AlphaZero and muZero[11] have shown powerful capabilities in policy searching for building an end-to-end decision model. However, reinforcement learning is still limited to being trained in a simulation environment or a game environment. Interacting with the true environment of autonomous driving is extremely expensive, slow and dangerous, which is completely unrealistic.

Recent development shows model-based offline reinforcement learning approaches are trying to learn a policy model from the environment dynamics. Just learning from observational data, the model may perform well in a real environment. Intuitively, we may extend this kind of model to build an end-to-end planner. Then human driving behaviors are collected and a model with observational data is trained to predict the trajectories for mimicking human driving. Fig. 9 shows the prototype of an end-to-end decision model we are studying. The videos from multiple cameras are input to a convolutional fusion network and a feature map is output. The features go into a temporal block, such as LSTM and GRU, to extract the sequence features, and then connect to the policy network and predict possible multiple trajectories.

However, the learning policies from purely observational data may not normally work because the data only cover a small region of the observed space[12]. Once a car deviates from the predicted best "human driving" trajectory, it is difficult for the car to recover from deviation and it will drift away from the



▲Figure 8. A self-supervised learning framework[8]



▲Figure 9. End-to-end autonomous driving decision-making network

ideal trajectory. The reason is that, unlike the learning in simulation environment where interaction and self-correcting are allowed, there is no actual interactive driving data for training on in this case.

In order to solve the problem, Ref. [12] proposes to train a policy by unrolling a learned model of the environment dynamics over multiple time steps while explicitly penalizing such costs as an uncertainty cost that represents its divergence from the states (trajectories) on which it is trained.

5 Conclusions

We discuss the application of intelligent sensors in autonomous vehicles and their impacts on automotive E/E architecture. The distributed ECU system will be replaced by centralized architecture to provide more computation power and integration. Moreover, for the sensors with high data rates, a TSN backbone plays a key role for E/E architecture. The algorithm of sensing perception based on neural networks is highly integrated with autonomous driving. The fusion of multiple sensors may enable better accuracy and robustness. Moreover, pure visual perception shows a powerful capability of 3D estimation versus LiDAR, and the visual-based pseud-LiDAR can reuse the existing LiDAR-based algorithms and improve the autonomy to a high level. Self-supervised learning is a more promising technology for cars in 3D perception.

It is also pointed out that the rule-based policy will never get over the edge cases, so the end-to-end policy seems to be a better approach to high-level autonomous driving and still needs further studying.

References

[1] BEBEL J C, HOWARD N, PATEL T. An autonomous system used in the darpa grand challenge [C]//7th International IEEE Conference on Intelligent Transportation Systems. Washington, USA: IEEE, 2004: 487 – 490. DOI: 10.1109/ITSC.2004.1398948

[2] MARTI E, DE MIGUEL M A, GARCIA F, et al. A review of sensor technologies for perception in automated driving [J]. IEEE intelligent transportation systems magazine, 2019, 11: 94 – 108. DOI: 10.1109/MITS.2019.2907630

[3] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [EB/OL]. (2018-01-24)[2021-04-28]. https://arxiv.org/abs/1703.06870v3

[4] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [EB/OL]. (2016-12-25)[2021-04-28]. https://arxiv.org/abs/1612.08242v1

[5] ALI W, ABDELKARIM S, ZAHRAN M, et al. YOLO3D: end-to-end real-time 3D oriented object bounding box detection from LiDAR point cloud [EB/OL]. (2018-08-07)[2021-04-28]. https://arxiv.org/abs/1808.02350v1

[6] CHEN X Z, MA H M, WAN J, et al. Multi-view 3D object detection network for autonomous driving [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. DOI: 10.1109/CVPR.2017.691

[7] WANG Y, CHAO W-L, GARG D, et al. Pseudo-LiDAR from visual depth estimation: bridging the gap in 3D object detection for autonomous driving [EB/OL]. (2020-02-22)[2021-04-28]. https://arxiv.org/abs/1812.07179v6

[8] ZHAN H Y, GARG R, WEERASEKERA C S, et al. Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction [EB/OL]. (2020-02-22)[2021-04-28]. https://arxiv.org/abs/1803.03893v3

[9] CUI H, RADOSAVLJEVIC V, F-CCHOU, et al. Multimodal trajectory predictions for autonomous driving using deep convolutional networks [EB/OL]. (2019-03-01)[2021-04-28]. https://arxiv.org/abs/1809.10732v2

[10] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [EB/OL]. (2015-12-10)[2021-04-28]. https://arxiv.org/abs/1512.03385

[11] SCHRITTWIESER J, ANTONOGLOU I, HUBERT T, et al. Mastering Atari, Go, chess and shogi by planning with a learned model [J]. Nature, 2020, 588: 604 – 609. DOI: 10.1038/s41586-020-03051-4

[12] HENAFF M, CANZIANI A, LECUN Y, Model-predictive policy learning with uncertainty regularization for driving in dense traffic [EB/OL]. (2019-01-08)[2021-04-28]. https://arxiv.org/abs/1901.02705v1

Biographies

SHAO Hong (shao.hong@zte.com.cn) received his M.S. degree from Harbin Institute of Technology, China and joined ZTE Corporation in 1994. He is currently the leader of the Innovation Team in the Corporation Development Department of ZTE Corporation. His research interests include deep learning, autonomous vehicles and robotics.

XIE Daxiong is currently the chairman of the board of supervisors of ZTE Corporation and the director of State Key Laboratory of Mobile Network and Mobile Multimedia Technology, China. He is a professorate senior engineer and a member of the Mobile and Satellite Expert Group of Communication Science and Technology Commission of Ministry of Industry and Information Technology of China. With ZTE Corporation, he led the successful research and development of CDMA, GoTa, WCDMA and other large-scale digital telecommunication systems.

HUANG Yihua received his M.S. degree from Huazhong University of Science and Technology, China. He is currently the manager of Corporation Development Department of ZTE Corporation. His research interests include new technology innovation, digital transformation and innovation strategy.

QoE Management for 5G New Radio

ZHANG Man, LI Dapeng,
LIU Zhuang, GAO Yin
(Algorithm Department, Wireless Product R&D Institute, ZTE Corporation, Shanghai 201203, China)

DOI: 10.12142/ZTECOM.202103008
http://kns.cnki.net/kcms/detail/34.1294.TN.20210730.1517.003.html, published online August 2, 2021
Manuscript received: 2021-05-19

Abstract: Quality of Experience (QoE) is used to monitor the user experience of telecommunication services, which has been studied for a long time. In universal terrestrial radio access network (UTRAN), evolved UTRAN (E-UTRA) and Long Term Evolution (LTE), QoE has also been specified for the improvement of user experience. The 5G New Radio (NR) technology is designed for providing various types of new services, and therefore operators have strong demand to continuously upgrade the 5G network to provide sufficient and good QoE for corresponding services. With new emerging 5G services, 5G QoE management collection aims at specifying the mechanism to collect the experience parameters for the multimedia telephony service for IP multimedia subsystem (IMS), multimedia broadcast and multicast service (MBMS), virtual reality (VR), etc. Taking LTE QoE as a baseline, generic NR QoE management mechanisms for activation, deactivation, configuration, and reporting of QoE measurement are introduced in this paper. Additionally, some enhanced QoE features in NR are discussed, such as radio access network (RAN) overload handling, RAN-visible QoE, per-slice QoE measurement, radio-related measurement, and QoE continuity for mobility. This paper also introduces solutions to NR QoE, which concludes the progress of NR QoE in the 3rd Generation Partnership Project (3GPP).

Keywords: NR; QoE; measurement collection; activation; deactivation; mobility; RAN visible QoE; radio-related information

Citation (IEEE Format): M. Zhang, D. P. Li, Z. Liu, et al., "QoE management for 5G new radio," *ZTE Communications*, vol. 19, no. 3, pp. 64–72, Sept. 2021. doi: 10.12142/ZTECOM.202103008.

1 Introduction

The multimedia services have experienced swift development over the past decades with more services, such as streaming services, virtual reality (VR), and extended reality (XR), coming to people's daily life. The emergence of various services, along with the convergence of multimedia communication, has brought more challenges to the radio access network (RAN), pushed the competition between network service providers, and aroused the user expectations of network services[1]. Quality of Experience (QoE) is used to reflect the quality experience of a user when he or she uses the telecommunication service. The definition of QoE in the 3rd Generation Partnership Project (3GPP) is "the collective effect of service performances which determines the degree of satisfaction of a user of a service"[2]. Since the 5G New Radio (NR) has been put into use, NR QoE is designed for different service types and complex scenarios. In December 2019, 3GPP started a study project for NR QoE with the description named "Study on NR QoE management and optimizations for diverse services"[3]. In February 2021, with many discussions, the study item of NR QoE has been completed and a new work item (WI) was started in May 2021 to specify the technical details of NR QoE solutions.

QoE has gained much attention in the research and industry area, not only in 3GPP. In 2019, a survey[4] of mobile edge caching was provided, with the demand for greater QoE and performance. In the same year, a paper[5] discussed some QoE improvement issues based on artificial intelligence technology. This paper mainly focuses on the progress in 3GPP.

The rest of the article is organized as follows. In Section 2, an overview of QoE is given. Section 3 introduces the basic solutions to NR QoE, such as the activation/deactivation procedures, the triggering and stopping of QoE, etc. In Section 4, some further solutions to NR QoE are discussed, such as handling in RAN overload, RAN visible QoE, radio-related information, per-slice QoE, etc. Section 5 concludes the paper.

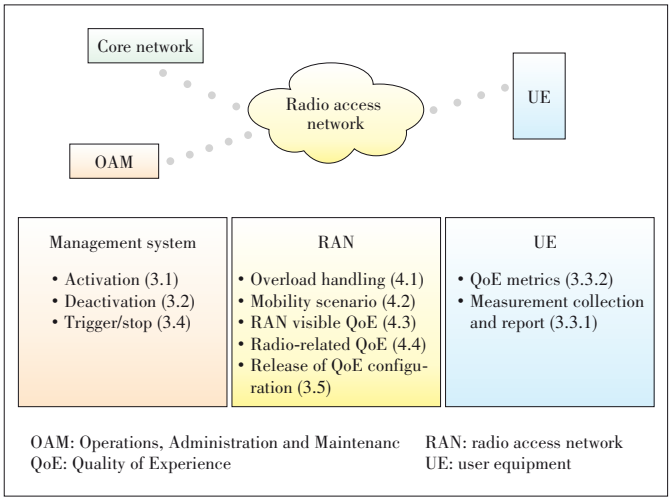
2 Overview

The QoE measurement collection for LTE has been specified in 3GPP. It is vital to guarantee the end users' experience, especially when the mobile network operators provide some real-time services which require high data rate and low latency, e.g. streaming services[6]. To support various new ser-

vice types and scenarios in 5G NR, the enhancement for QoE is discussed in Rel-17. NR QoE takes LTE QoE as a baseline. The main features of LTE QoE can be categorized into three parts as follows. The first is that both signaling based and management based cases are allowed. The second is that the LTE QoE feature is activated by trace function. The last is about the transfer of application layer measurement configuration and application layer measurements. To be specific, the application layer measurement configuration received from Operations, Administration and Maintenance (OAM) or core network (CN) can be encapsulated in a transparent container, which is forwarded to user equipment (UE) in a downlink radio resource control (RRC) message; the application layer measurements received from UE's higher layer can be encapsulated in a transparent container and sent to the network in an uplink RRC message[7].

NR QoE supports the functionality of application layer measurement collection, which could collect the application layer measurement results of diverse services. Since there have been more newly introduced service types provided by operators, the supported service types of QoE have also been expanded. The currently supported service types in NR QoE include streaming services[8], multimedia telephony service for IP Multimedia Subsystem (IMS) services[9], AR[10], multimedia broadcast and multicast service (MBMS)[11], and XR. Meanwhile, the support for additional service types is not precluded.

The potential solutions to NR QoE include some basic procedures, such as the activation and deactivation procedures, QoE measurement triggering and stopping, and the release of QoE measurement configuration. There are some additional solutions aside from the basic procedures, including QoE measurement handling at RAN overload, supports for mobility, RAN visible QoE measurement, radio-related measurement and information, and per-slice QoE measurement. The remaining of the paper would describe the QoE solutions and procedures in detail. The main QoE mechanisms are depicted in Fig. 1, based on the primary part which performs the mechanism.



▲Figure 1. Diagram for NR QoE mechanisms

3 Management Mechanisms for NR QoE Measurement Collection

3.1 Activation

The activation of NR QoE could be divided into two types: one is signaling-based activation, the other is management-based activation. The main difference between the two types is that, the signaling based activation is configured by OAM and triggered by the core network (CN), while the management-based activation is configured and triggered by OAM.

In signaling-based activation of QoE, specifically, OAM sends the QoE measurement configuration to CN, and it is CN to initiate the activation of QoE and to send the QoE measurement configuration to the new generation radio access network (NG-RAN) node. The NG-RAN node sends the QoE measurement configuration to the UE access stratum (AS) layer via RRC message. After receiving the QoE measurement configuration, the UE AS layer sends it to the UE APP layer.

In management-based activation, the OAM sends the QoE measurement configuration directly to the NG-RAN node, without the involvement of CN. The NG-RAN node would find multiple qualified UE or UE that meets the criteria for QoE measurement (e.g. area scope, application layer capability, service type, etc.). Then the NG-RAN node sends the QoE measurement configuration to the specific UE or multiple qualified UE via RRC message. The steps after that are similar to the signaling-based activation. The procedures for QoE activation are shown in Fig. 2.

Multiple simultaneous QoE measurements for UE could be supported, the details of the technique would be discussed in the normative phase, and whether each QoE measurement could be configured per service type will also be discussed then.

3.2 Deactivation

In accordance with the activation procedures, the deactivation of QoE could also be divided into signaling-based deactivation and management-based deactivation. Similarly, the signaling-based deactivation is configured by OAM and triggered by CN, while the management-based deactivation is configured and triggered by OAM.

In signaling-based deactivation, the OAM sends a deactivation indication to CN to configure the deactivation for QoE. After receiving the deactivation indication, the CN initiates the deactivation of QoE measurement and sends the deactivation indication to the NG-RAN node. The deactivation indication is used to indicate which QoE measurement is to be deactivated. After receiving the deactivation indication, the NG-RAN node sends the deactivation indication to the UE AS layer via RRC message, and the AS layer sends the deactivation indica-

tion to the UE APP layer.

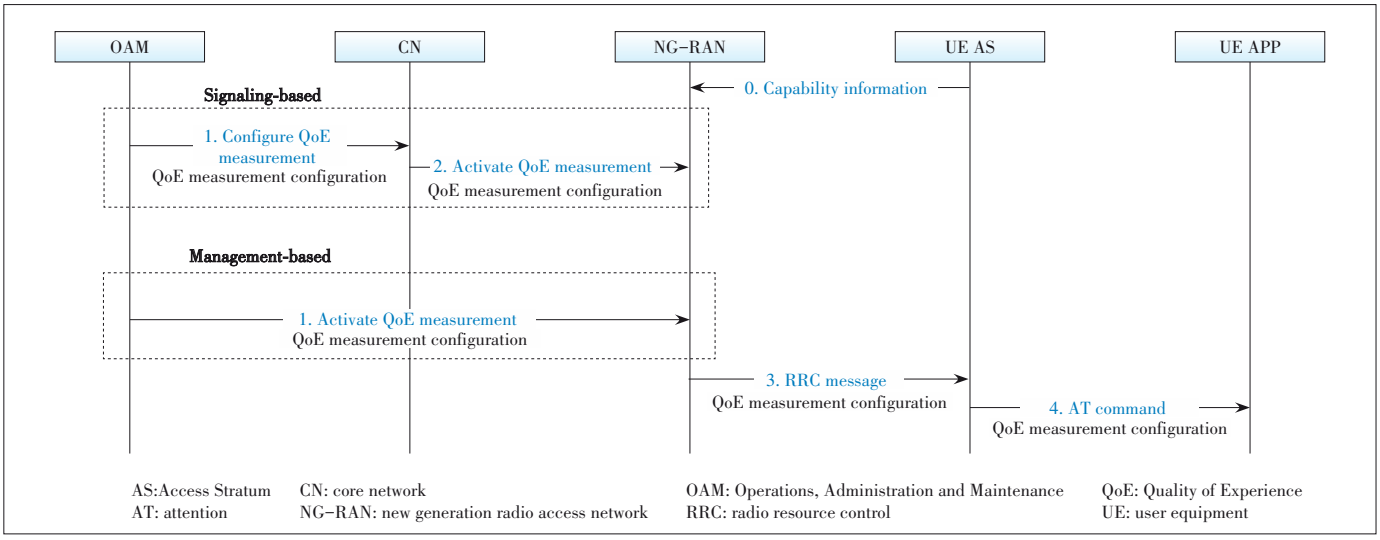
In management-based deactivation, the OAM configures the deactivation of QoE and sends the deactivation indication to the NG-RAN node directly without the involvement of CN. After receiving the deactivation indication from the OAM, the NG-RAN node sends the deactivation indication to UE AS layer via an RRC message. The following steps are similar to the signaling-based deactivation. The procedures for QoE deactivation are shown in Fig. 3.

3.3 QoE Measurement Reporting

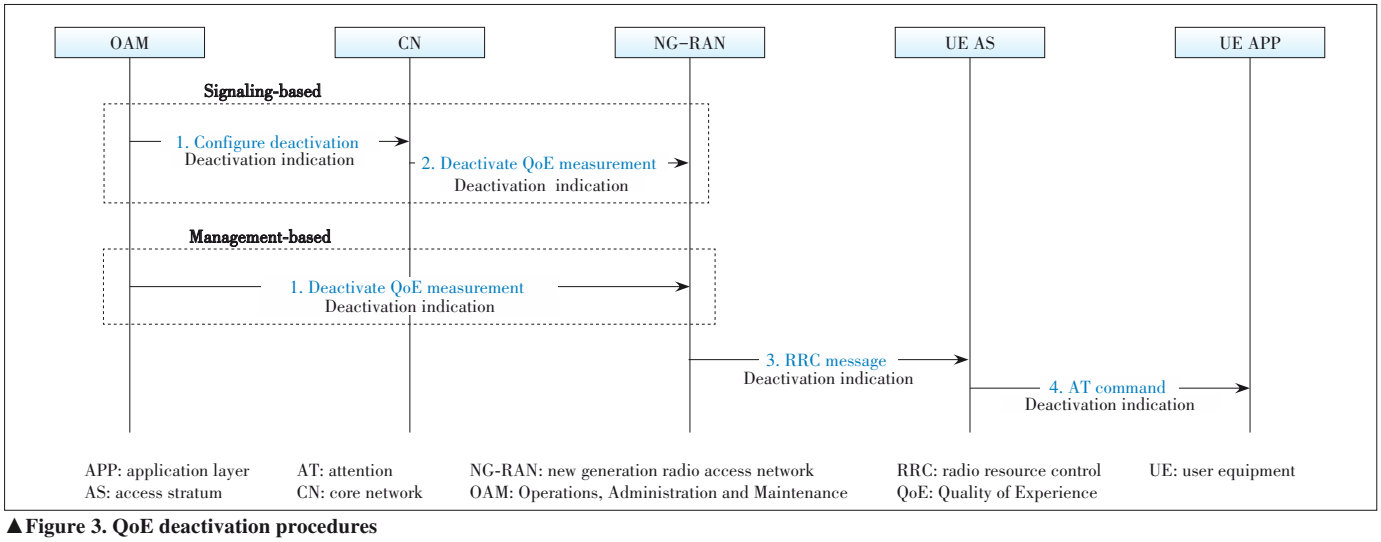
3.3.1 QoE Measurement Reporting Procedure

After receiving the QoE measurement configuration, the UE APP layer starts QoE measurement collection based on the configuration. UE APP layer generates the QoE reports according to the QoE measurement results, and once it is the time configured to report QoE measurement, the UE APP layer sends the QoE report to the UE AS layer. UE AS sends the QoE report to NG-RAN node via RRC message, and in particular, via a separate signalling radio bearer (SRB) aside from current SRBs. The separate SRB is used to differentiate the QoE report from other SRB transmissions, because the priority of the QoE report is lower than any other SRB transmission.

As stated in the overview, the QoE report sent via RRC message from UE's higher layer can be transparent containers with the measurement results encapsulated together. Hence, from the NG-RAN side, the contents in the QoE report container are totally transparent, i.e., the NG-RAN node is not able to get the QoE measurement results for the optimization of itself. Another discussion about RAN visible QoE, focusing on QoE measurement visible at the NG-RAN side, will be introduced in the next part of this paper. The NG-RAN node would send the QoE report container to the final destination configured by OAM, e.g., the trace collection entity (TCE) or measurement collection entity (MCE). The procedure is shown



▲Figure 2. QoE activation procedures



▲Figure 3. QoE deactivation procedures

in Fig. 4.

3.3.2 QoE Metrics

The QoE report is the result of a series of QoE metrics data collected by the UE. QoE metrics are valid for the quality of the supported services provided by the operators. Examples of some QoE metrics for MTSI[9] are provided in Table 1.

3.4 QoE Measurement Triggering and Stopping

It has been stated above that the activation and deactivation configurations are generated by OAM. Accordingly, the criteria for triggering and stopping QoE measurement are also configured by OAM. The currently approved criteria to realize the triggering and stopping of QoE are time-based and threshold-based criteria.

• Time-based criteria: When it is the configured time to trigger or stop the QoE measurement, the QoE measurement would be triggered or stopped. This is implemented by reusing the mechanisms for the start and stop of QoE specified in LTE. Fig. 5 depicts the time-based criteria.

• Threshold-based criteria: When the given thresholds to trigger or stop the QoE are passed, the QoE measurement would be triggered or stopped consequently. Fig. 6 depicts the threshold-based criteria.

3.5 Release of QoE Measurement Configuration

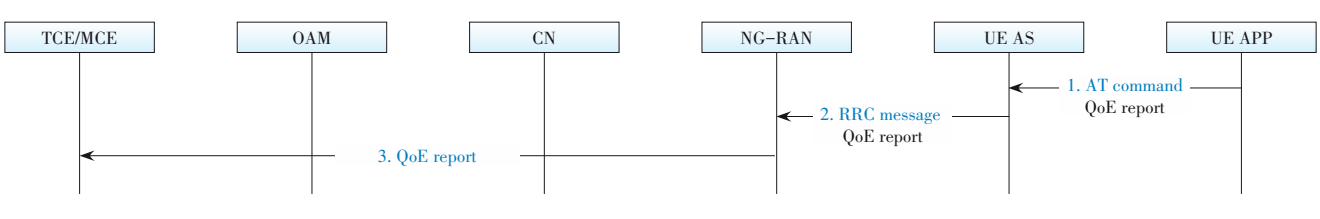
The QoE measurement collection and reporting would not last forever, i.e., the QoE measurement configuration for QoE measurement reporting should be released at some particular timing. The NG-RAN node has such ability to issue a release of QoE measurement configuration for UE, as long as the session for QoE measurement reporting is completed. In another case where UE hands over to a network that does not support QoE, RAN may need to release an ongoing QoE configuration.

4 Enhanced Features and Solution to NR QoE

4.1 QoE Measurement Handling at RAN Overload

Due to the extended bandwidth and various service types in 5G era, the RAN is confronted with more challenges to support the high burden of radio access. RAN overload, as depicted in Fig. 7, is an alarming problem that could probably happen, especially when QoE is supported, with extended pressure to transfer the QoE configuration and report.

Since RAN overload is an important situation, the reaction to RAN overload should not simply deactivate the QoE measurements. Instead, the QoE data information at RAN overload is necessary to be collected, and could be reported later to prevent aggravating the overload. In other words, it is vital to capture QoE data during periods of RAN overload. But on the other hand, persistent reporting might contribute to RAN overload[12]. Temporary stop and restart of QoE reporting could



▲Figure 4. QoE measurement reporting procedure

▼Table 1. Multimedia telephony service for IMS (MTSI) QoE metrics

QoE Metric	Description
Corruption duration metric	The time period from the NTP time of the last good frame before the corruption to the NPT time of the first subsequent good frame
Successive loss of RTP packets	The number of RTP packets lost in the successive per media channel
Frame rate	The playback frame rate is equal to the number of frames displayed during the measurement resolution period divided by the time duration (in seconds) of the measurement resolution period
Jitter duration	Jitter happens when the absolute difference between the actual playback time and the expected playback time is larger than JitterThreshold milliseconds
Sync loss duration	Sync loss happens when the absolute difference between value A and value B is larger than SyncThreshold milliseconds
Round-trip time	The RTT consists of the RTP-level round-trip time, plus the additional two-way delay due to buffering and other processing in each client
Average codec bitrate	The bitrate is used for coding "active" media information during the measurement resolution period
Codec information	The codec information metrics contain details of the media codec settings used in the receiving direction during the measurement resolution period

NTP: network time protocol QoE: Quality of Experience RTP: real-time transport protocol RTT: round-trip time

be useful functionality to handle RAN overload.

At the current stage, as shown in Fig. 8, RAN could take actions in the following three aspects to handle RAN overload:

- Stop new QoE measurement configurations
- Release existing QoE measurement configurations
- Pause QoE measurement reporting.

A specific diagram for procedures of RAN pausing the QoE reporting is shown in Fig. 9. Firstly, when RAN detects overload, it should send a request to UE AS to pause the QoE reporting. After that, RAN should send an indication to the management system about the temporary stop of the QoE report. When UE AS receives the request, it should inform the UE APP to temporarily stop the QoE reporting. After the RAN overload issue has eased off, RAN could resume the QoE measurement reporting that has been paused, indicating the UE to send the stored QoE report during the RAN overload.

Future works will focus on the detailed technical specifications of the pause/resume mechanisms, including whether to pause/resume for all QoE reports or per-QoE configuration, the time limit for UE to store the QoE reports, the limit for the stored report size, etc.

4.2 Support for Mobility

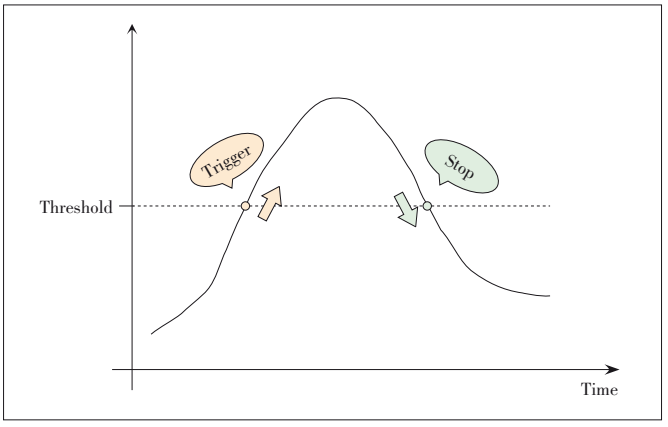
4.2.1 Intra-System Intra-RAT Mobility

With the development of modern traffic such as shared bikes, subway and high-speed railways, it is easier for people to move from one spot to another. Based on that consideration, the mobility scenario is a feature that should be enhanced in NR QoE. In the normative phase of Rel-17, the QoE continuity of intra-RAT mobility would be specified, to enable measurement of the impact of the mobility on the application and users' QoE. The intra-system intra-RAT mobility scenarios are categorized into intra-node and inter-node scenarios depicted in Figs. 10(a) and 10(b).

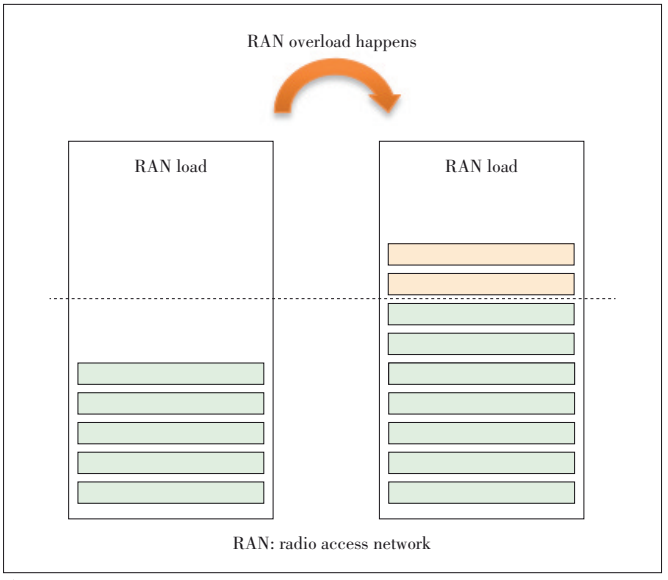
For the intra-node mobility scenario, the QoE measurement reporting for signaling-based and management-based QoE should both be supported. No matter before or after the UE handover, the served NG-RAN node is the same. Hence, there is no need to consider how to transfer the QoE measurement configuration from the source to the target. Both signaling-based and management-based QoE could be supported without much difficulty.

For the inter-node mobility scenario, UE

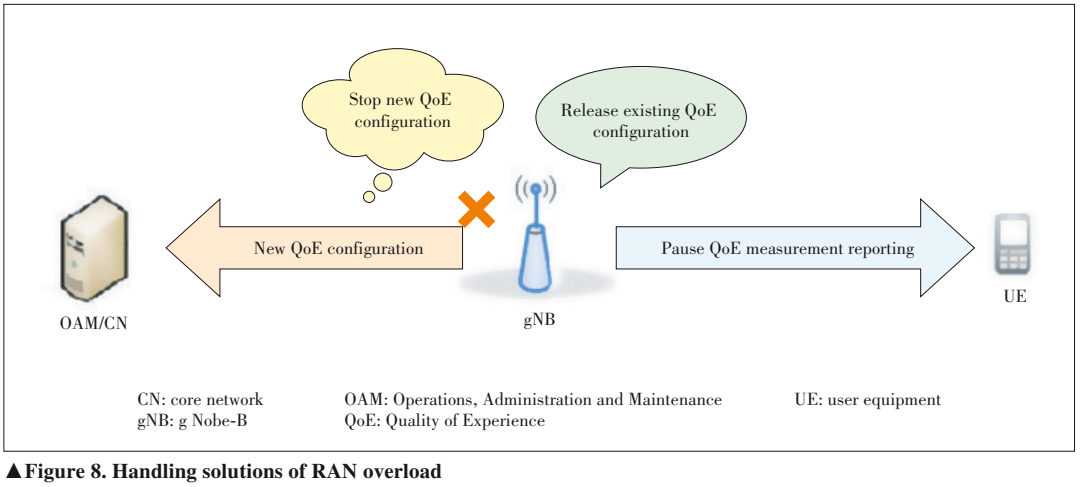
▲Figure 5. Time-based criteria



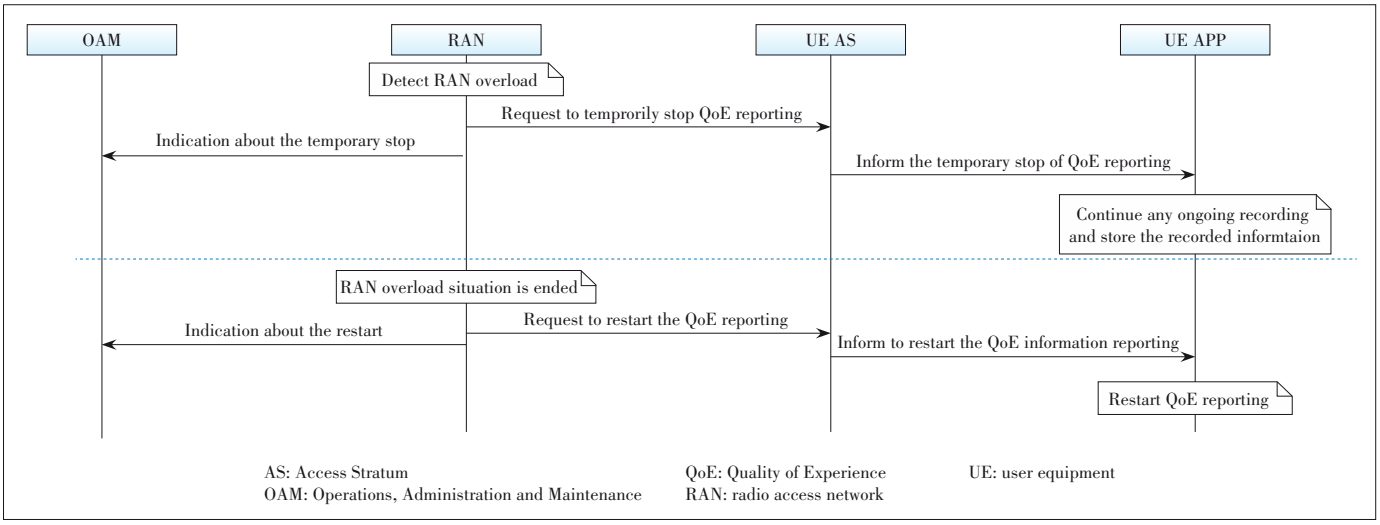
▲Figure 6. Threshold-based criteria



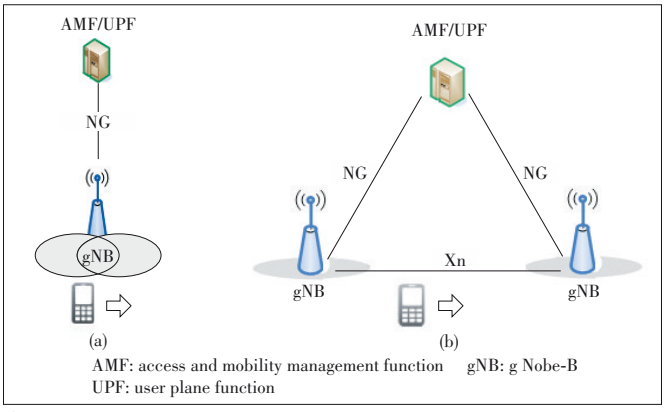
▲Figure 7. RAN overload



▲Figure 8. Handling solutions of RAN overload



▲Figure 9. Procedure of handling solutions to RAN overload



▲Figure 10. (a) Intra-node mobility and (b) inter-node mobility

hands over from the source NG-RAN node to the target NG-RAN node. In the case of signaling based QoE, CN can get the QoE measurement configuration, so that the QoE measurement could be transferred by the CN from the source NG-RAN node to the target NG-RAN node via NG interface. In the case of management-based QoE, the CN is not able to get the QoE measurement configuration, which makes it more complicated to transfer the configuration from OAM. Therefore, for the inter-node mobility scenarios, at least signaling-based QoE should be supported, while management-based QoE is to be studied in the normative phase.

The inter-system and inter-RAT mobility would be deprioritized in Rel-17.

4.2.2 Supported RRC State

In NR, the mobility for QoE measurements in RRC_CONNECTED state should be supported. As mentioned in the overview that NR QoE takes LTE QoE as a baseline, which uses the trace function to activate the QoE feature. To support mobility for UE RRC_CONNECTED, the trace function is also utilized in NR QoE. To be specific, the QoE measurement is transferred via the Xn and NG interfaces, with the part of UE application layer measurement configuration IE inside the Trace Activation IE.

QoE measurements in RRC_INACTIVE and RRC_IDLE state could be supported for multicast and broadcast service (MBS). To keep the QoE measurement configuration in RRC_INACTIVTAE state mobility, the UE could fetch the QoE measurement configuration from the node that hosts the UE context. Whether the QoE measurement configuration will be saved after UE goes to RRC_INACTIVATE state is still not decided. Once UE goes back to RRC_CONNECTED state, the QoE measurement configuration which has been stored when UE moves to RRC_INATCIVATE state might be useful. However, the necessity of doing this has not been confirmed.

4.2.3 Area Handling

There are some requirements that push the design of solutions for area handling. These requirements indicate that both the network side and the UE side could have the ability to check the UEs' location for a specific area requested for QoE measurement collection, which is called geographical filtering[6, 9]. But one thing to be noted is that if the network side is configured with geographical filtering, the related configuration should not be included in the QoE measurement configuration.

Based on the requirements, there are three possible solutions for area handling:

1) The network is responsible for keeping track of whether the UE is inside or outside the area and configures/releases configuration accordingly.

2) The network is responsible for keeping track of whether the UE is inside or outside the area, and the UE is responsible for the managing and start/stop of QoE accordingly.

3) The UE is responsible for area checking (UE has the area configuration) and the managing and start/stop of QoE accordingly.

4.3 RAN Visible QoE

As QoE measurements are transferred as a transparent container to the RAN side, the NG-RAN node cannot get the contents in the QoE measurement report. However, there are cases where RAN is in need of the QoE measurement results (e.g., buffer level) for the optimizations of itself. The QoE measurement information required by RAN could be designed to be visible to the RAN node, which is the so-called RAN visible QoE information.

The QoE metrics which could be visible to RAN are called RAN visible QoE metrics. The RAN visible QoE metrics for the streaming service could be round-trip time, jitter duration, corruption duration, average throughput, initial playout delay, device information, rendered viewports, codec information, etc.[13].

The message flow of RAN visible QoE information reporting is shown in Fig. 11.

1) The NG-RAN node sends the RAN visible QoE configuration to the UE, which may be sent along with the QoE measurement configuration.

2) The UE receives the RAN visible QoE configuration and/or the QoE measurement configuration. The UE collects QoE measurement results according to the received configuration and provides the RAN visible QoE report, along with the QoE report container to the RAN.

3) The NG-RAN node reads the RAN visible QoE report and/or forwards the QoE report container to the QoE server accordingly.

4.4 Radio-Related Measurement and Information

Since the radio-related measurement technology like MDT has been specified in 5G NR, it is possible that NR QoE could bring the existing mechanisms to do some optimization. Radio-related measurements and information are accordingly taken into consideration in NR QoE.

Radio-related measurements are measurements on the radio layer, with the purpose of helping networks to further evaluate and improve the QoE. The RAN can trigger radio-related measurement towards certain UE, based on the measurement configuration from the OAM.

Radio-related measurements are strongly associated with MDT mechanisms in several aspects. Firstly, radio-related measurements could be triggered by using existing mechanisms such as the MDT procedure. Secondly, the collection of radio-related measurements could also be done with the support of MDT[14]. Furthermore, the reporting of radio-related measurement is for all types of services that are supported, which include MDT-like measurements and additional measurements related to radio interface. If new radio-related measurements corresponding to currently specified MDT measurements are required for NR QoE management, these additional radio-related measurements will be specified as a part of MDT measurements.

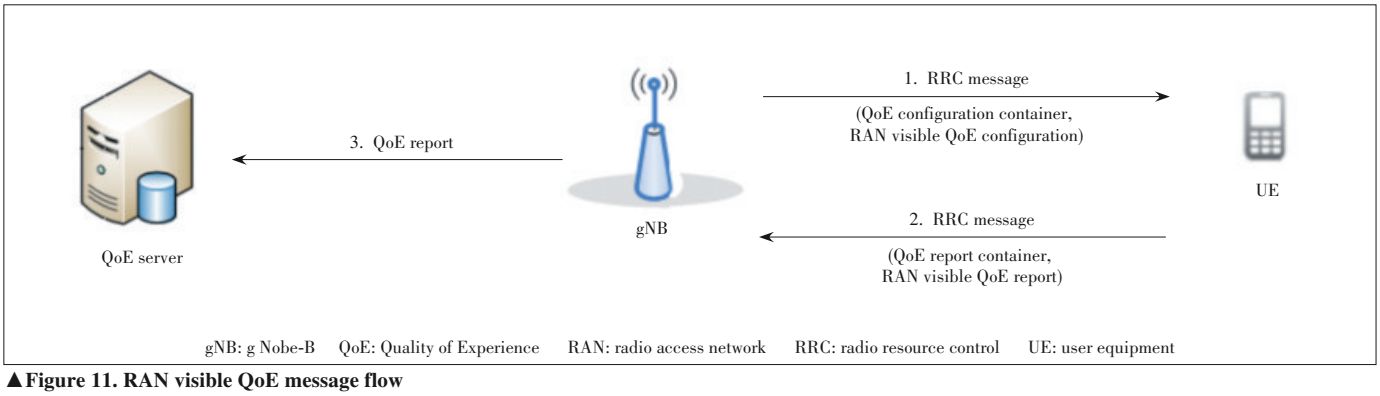
There are some further requirements for radio-related measurements, with respect to application-related measurements. Application-related measurements are only collected when the application is ongoing. Based on this fact, if the radio-related measurements are used for assisting application-related QoE measurements, it is beneficial and efficient if the measurement collection and reporting could start at the same time. If they are configured together, e.g. using the same trace function, or based on timestamps, correlation of the results could be done by post processing[15].

Radio-related information is information other than radio-related measurements, e.g., feature information, mobility history information, or dual connectivity status. Radio-related information may also be collected and reported, aside from radio-related measurements. Radio-related information may even be reported when radio-related measurements are not triggered over the radio.

One important requirement for radio-related measurement and radio-related information is that, both of them should be aligned and correlated with the QoE report, if they are reported. For example, trace ID could be used to align them with the QoE report.

4.5 Per-Slice QoE

With the application of slicing, the same service type could use different slices. Therefore, QoE measurements for each service type are not enough for the QoE management, while per-slice QoE is needed.



▲Figure 11. RAN visible QoE message flow

As shown in Fig. 12, there are three scenarios concerned by 3GPP at the current stage[16]:

· Scenario 1: different service types using different slices (e. g., UE1 & UE2)

· Scenario 2: different service types using the same slice (e. g., UE 1 & UE3)

· Scenario 3: the same service type using different slices (e. g., UE2 & UE3).

Three solutions[16] are approved for per-slice QoE measurement, with details left to be discussed in the future.

Solution 1: RAN is responsible for mapping slice scope to protocol data unit (PDU) session list. After UE sends the report with Slice ID, RAN could remap the PDU session ID back to Slice ID. The whole procedure is shown in Fig. 13.

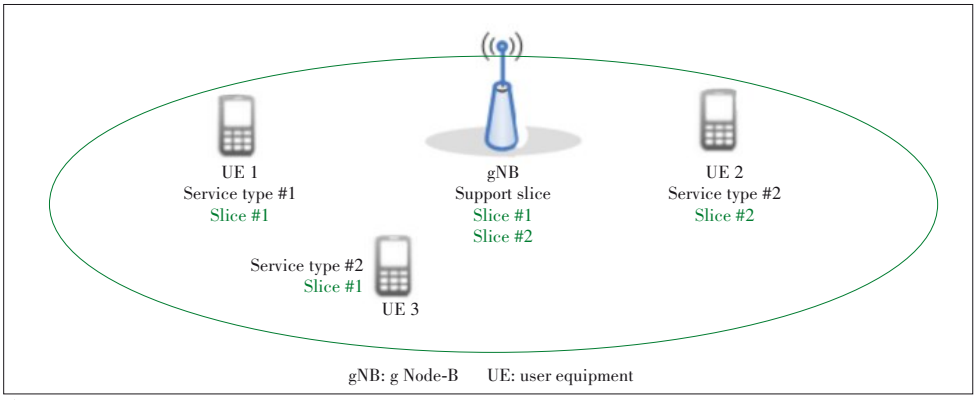
Solution 2: UE is responsible for mapping Slice Scope to QoE report. Whether the application layer or AS layer performs the mapping can be discussed at the normative stage. The whole procedure is shown in Fig. 14.

Solution 3: RAN is responsible for mapping the Slice ID to QoE measurements, with the Slice ID included in the QoE measurement configuration. The whole procedure is shown in Fig. 15.

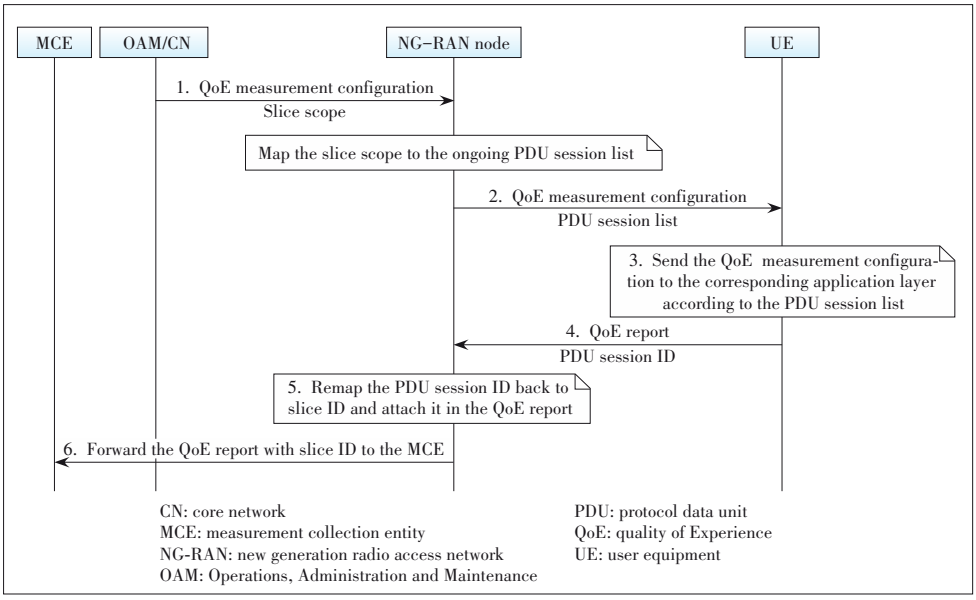
For all solutions mentioned above, Slice Scope is outside the QoE configuration container. When it comes to the mobility scenario, for signaling-based QoE, Slice Scope (e.g. list of single-network slice selection assistance information) should be transmitted to the target gNB during mobility. For management-based QoE, Slice Scope can be clarified at the normative stage.

5 Conclusions

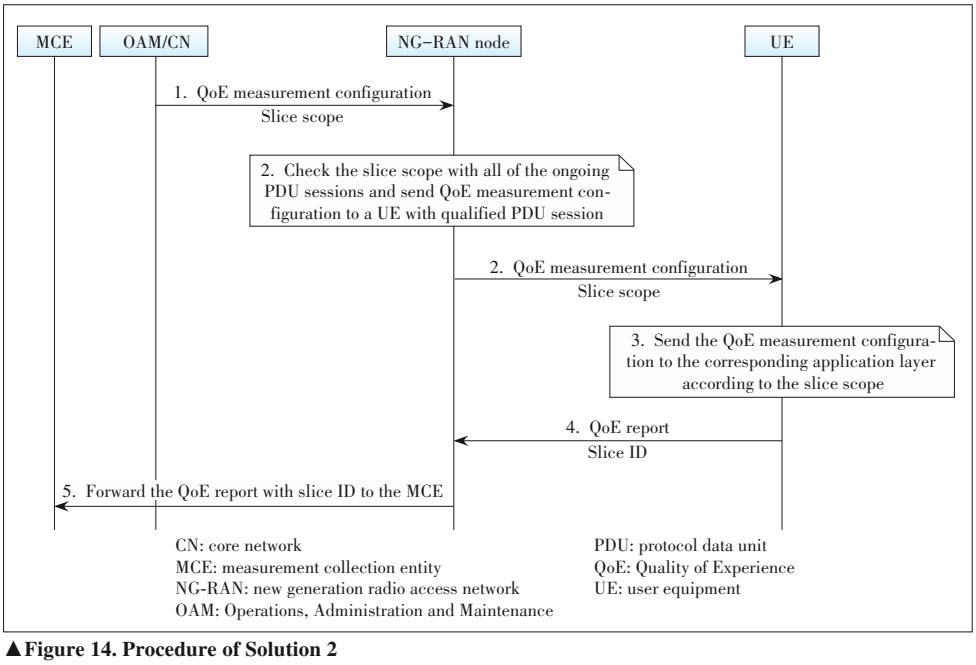
This paper provides the current



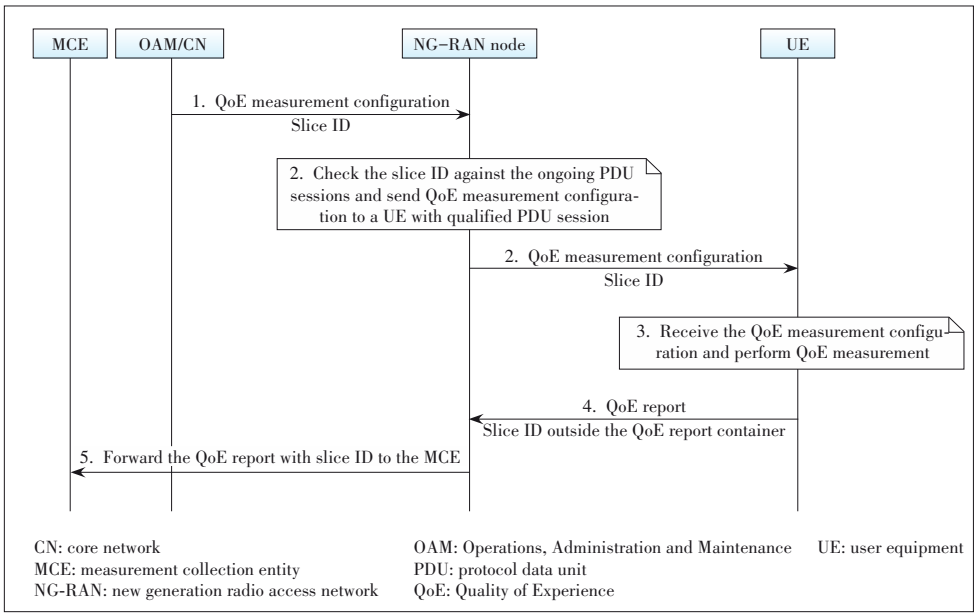
▲Figure 12. An example of different scenarios



▲Figure 13. Procedure of Solution 1



▲Figure 14. Procedure of Solution 2



▲**Figure 15. Procedure of Solution 3**

process of NR QoE from the perspective of various solutions agreed at 3GPP, which are divided into basic and additional solutions. The basic solutions and procedures include activation and deactivation of QoE, QoE measurement reporting, triggering and stopping, release of QoE measurement configuration, and QoE measurement handling at RAN overload. The additional solutions mainly contain the NR QoE mobility support, RAN visible QoE, radio-related measurement and information, and per-slice QoE. Further details of these solutions will be discussed in the future.

References

[1] JULURI P, TAMARAPALLI V, MEDHI D. Measurement of quality of experience of video-on-demand services: a survey [J]. IEEE communications surveys & tutorials, 2016, 18(1): 401 – 418. DOI:10.1109/comst.2015.2401424
[2] 3GPP. Vocabulary for 3GPP specifications: 3GPP TS21.905 [S]. 2020
[3] 3GPP. Study on NR QoE management and optimizations for diverse services: 3GPP RP-193256 [R]. 2019
[4] ANOKYE S, SEID M, SUN G L. A survey on machine learning based proactive caching [J]. ZTE communications, 2019, 17(4): 46 – 55. DOI: 10.12142/ZTECOM.201904007
[5] GAO Y, WEI X, ZHOU L. QoE improvement issues based on artificial intelligence [J]. ZTE technology journal, 2019, 25(6): 59 – 64. DOI: 10.12142/ZTETJ.201906010
[6] 3GPP. Telecommunication management; Quality of Experience (QoE) measurement collection; concepts, use cases and requirements: 3GPP TS28.404 [S]. 2020
[7] 3GPP. Technical Specification group radio access network; study on NR QoE management and optimizations for diverse services: 3GPP TR38.890 [S]. 2021
[8] 3GPP. Transparent end-to-end packet-switched streaming service (PSS); progressive download and dynamic adaptive streaming over HTTP (3GP-DASH): 3GPP TS26.247 [S]. 2020
[9] 3GPP. IP Multimedia subsystem (IMS); multimedia telephony; media handling and interaction: 3GPP TS26.114 [S]. 2021
[10] 3GPP. Virtual reality (VR) profiles for streaming applications: 3GPP TS26.118 [S]. 2021
[11] 3GPP. Multimedia broadcast/multicast service (MBMS); protocols and codecs: 3GPP TS26.346 [S]. 2021
[12] SA4. LS reply on QoE measurement collection: 3GPP S4-201600 [S]. 2020
[13] Ericsson. PCR for TR 38.890: QoE visibility at the RAN: 3GPP R3-211342 [S]. 2020
[14] UnicomChina. TP for TR update (RAN2): 3GPP R2-2102483 [S]. 2020
[15] Huawei. TP to 38.890 on open issues of QoE configuration and reporting: 3GPP R3-211358 [S]. 2020
[16] ZTE. TP for per slice QoE measurement: 3GPP R3-211341 [S]. 2016

Biographies

ZHANG Man (zhang.man4@zte.com.cn) received her master's degree in communication engineering from Shanghai Jiao Tong University, China in 2020. She is currently a technology pre-research engineer at the Algorithm Department, ZTE Corporation. Her research focuses on next generation radio access network.

LI Dapeng received the M.S. degree in computer science from University of Electronic Science and Technology of China in 2003. He is currently a senior researcher in ZTE Corporation, China and mainly focus on research and implementation of wireless access network system.

LIU Zhuang received his master's degree in computer science from Xidian University, China in 2003. He is currently a senior 5G research engineer at the R&D center of ZTE Corporation and the State Key Laboratory of Mobile Network and Mobile Multimedia, China. His research interests include 5G wireless communications and signal processing. He has filed more than 100 patents.

GAO Yin received her master's degree in circuit and system from Xidian University, China in 2005. Since 2005 she has been with the research center of ZTE and engaged in the study of 4G/5G technology. She has authored or co-authored about hundreds of proposals for 3GPP meeting and journal papers in wireless communications and filed more than 100 patents. In May, 2021 she was elected as 3GPP RAN3 Chairman.

Super Resolution Sensing Technique for Distributed Resource Monitoring on Edge Clouds

Abstract: With the vigorous development of mobile networks, the number of devices at the network edge is growing rapidly and the massive amount of data generated by the devices brings a huge challenge of response latency and communication burden. Existing resource monitoring systems are widely deployed in cloud data centers, but it is difficult for traditional resource monitoring solutions to handle the massive data generated by thousands of edge devices. To address these challenges, we propose a super resolution sensing (SRS) method for distributed resource monitoring, which can be used to recover reliable and accurate high-frequency data from low-frequency sampled resource monitoring data. Experiments based on the proposed SRS model are also conducted and the experimental results show that it can effectively reduce the errors generated when recovering low-frequency monitoring data to high-frequency data, and verify the effectiveness and practical value of applying SRS method for resource monitoring on edge clouds.

Keywords: edge clouds; super resolution sensing; distributed resource monitoring

YANG Han, CHEN Xu, ZHOU Zhi
(Sun Yat-Sen University, Guangzhou 510275, China)

DOI: 10.12142/ZTECOM.202103009
http://kns.cnki.net/kcms/detail/34.1294.TN.20210722.1305.002.html, published online July 23, 2021
Manuscript received: 2021-04-17

Citation (IEEE Format): H. Yang, X. Chen, and Z. Zhou, "Super resolution sensing technique for distributed resource monitoring on edge clouds," *ZTE Communications*, vol. 19, no. 3, pp. 73 - 80, Sept. 2021. doi: 10.12142/ZTECOM.202103009.

1 Introduction

In recent decades, with the advent of the era of Internet of Things and the rapid development of mobile networks, the number of edge devices and the number of data generated at the edge have been growing exponentially, leading to higher and higher requirements for network bandwidth. At the same time, driving modern intelligent mobile applications, deep learning has attracted much attention from scientists and IT enterprises. New applications based on deep neural networks have achieved great success in computer vision, speech recognition, natural language processing and intelligent robots. Although complex computing tasks are completed and reliable and accurate results are obtained, the massive data generated by these new applications will put forward higher requirements on network bandwidth and time delay. According to IDC's prediction, the global data volume would be more than 40 ZB before 2020, and the data generated at the edge will account for 45% of the total[1]. In new edge cloud scenarios, the traditional cloud computing technologies are difficult to process the billion-scale data generated by edge devices[2], and only using the strong computing power and storage resources of the cloud data center to solve the computation and storage problems can no longer adapt to the needs of the new era, because it has two main shortcomings: high latency and bandwidth limitation. The latency requirements of deep learning-based applications at the edge are very high. Such an application needs to transfer the data to the cloud and the data will be processed and then returned to the device side, which may significantly increase the processing delay of the application. For example, a car running at a high speed has millisecond-level delay requirements, but if the processing delay of the application increases due to the change of network conditions, the consequences will be unimaginable. Besides, affected by the edge device resource limitations, the data generated by the device will be transmitted to the cloud in real time; for example, the amount of data generated by the aircraft per second will exceed 5 GB[3]. However, the bandwidth limitations prevent this real-time transmission in edge cloud scenarios.

In recent years, with the development of cloud computing, virtualization and containerization technologies, companies such as eBay, Facebook, Google and Microsoft have made a lot of investment in large-scale data centers supporting cloud services[4]. Servers are the core part of data centers and monitoring server resources aims to guarantee the smooth operation of data centers. At present, resource monitoring is one of the key technologies to support cloud computing platforms, which

mainly includes cloud resource management, fault analysis, resource scheduling, statistical analysis and anomaly warning. The existing resource monitoring system is widely deployed in cloud data centers, but with the increasing of edge devices, resources need to be monitored on edge clouds. A large number of resource monitoring data will make the network bandwidth a bottleneck affecting system performance, but these high-frequency resource data are the basis for the monitoring system to conduct reliable and accurate early warning.

The traditional solutions to recovering low-frequency resource monitoring data into high-frequency data include linear interpolation, cubic interpolation and compressed sensing-based methods[5]. However, with the rapid development of deep learning technology, the traditional methods have two obvious shortcomings. One is that they need to either collect the original high-frequency data at the edge device end or consume additional computing power to calculate the low-frequency monitoring data, and then upload the data to the cloud, which will obviously increase the computing cost of the edge device. The other one is that the accuracy of high-frequency monitoring data recovered by the reconstruction algorithm based on the traditional methods cannot meet the needs of some online services; however, with the increasing number of online services in cloud data centers, the accuracy of high-frequency monitoring data is greatly important.

To tackle the challenges discussed above, novel methods are needed to support high-frequency sensing of monitoring systems on edge clouds. In this paper, we achieve this goal by applying the super resolution sensing (SRS) technology. The SRS technology based on deep learning is used to avoid collecting original high-frequency data at the edge device and reduce unnecessary calculation cost. It only involves low-frequency data at the transmission and storage stage of monitoring data and information reconstruction can be carried out by using this technology only when high-frequency data with high precision is needed, which can significantly reduce the cost of calculation and communication. The proposed SRS process is mainly divided into three stages: feature extraction, relationship mapping and information recovery. The feature extraction module is used to obtain the intrinsic features of the low-frequency data, followed by a gated recurrent unit network based on the attention mechanism to find the potential relationship between the low-frequency data and the high-frequency data in the relationship mapping stage, and finally, the high-frequency surveillance data are recovered based on the learned feature information in the information recovery stage.

The rest of the paper is organized as follows. We briefly review the related work in Section 2. We propose an SRS system for resource monitoring and present its network structure in Section 3. Then we demonstrate the effectiveness of the proposed approach by simulations in Section 4. Finally, we provide conclusions and some future work directions in Section 5.

2 Related Work

At present, super-resolution sensing techniques can be broadly classified into three categories: interpolation, reconstruction, and deep learning-based reconstruction methods.

The interpolation-based methods are mainly based on the relationship between the values of neighboring pixel points in the image and the positions of other pixel points around them, and the missing values of the pixel points on the high-resolution image are complemented by interpolation, and finally the high-resolution image is recovered by noise reduction and deblurring. The common interpolation-based methods include nearest neighbor interpolation, bilinear interpolation based on wavelet domain[6] and cubic interpolation[7]. On this basis, some researchers have further proposed interpolation based on gradient features, interpolation based on image features, etc. The interpolation based on bilateral filter proposed by TOMASI et al.[8] uses bilateral filtering as a constraint term to reduce the edge noise generated by the reconstructed image. To further reduce the effects of blurring and ringing in the recovered images, LI et al.[9] proposed an interpolation algorithm based on edge orientation, which uses a geometric pairwise method to interpolate the specified edge regions and highly textured regions in the image orientation, thus significantly improving the quality of image reconstruction. To reduce the artifacts in the recovered image, BELAHMIDI et al.[10] introduced partial differential equations and data fidelity for directional interpolation of edges, but the effectiveness of these algorithms is affected by the edge regions. Therefore, adaptive interpolation algorithms based on texture partitioning have also been proposed by some researchers[11].

Although the interpolation-based method is simple, it fails to introduce any priori knowledge and its information recovery capability is insufficient. Therefore, a reconstruction-based reconstruction method is used, which is more concerned with the image degradation itself than the interpolation method. STARK and OSKOUI proposed the convex set projection method[12], which, based on the set theory, first defines a set of convex constraint sets for the solution space of the image and seeks the points that satisfy all the conditions of the constrained convex sets by stepwise iteration to complete the reconstruction of the high-resolution image. Another typical algorithm is the maximum a posteriori (MAP) probability estimation method[13–15], a method proposed based on probability statistics, which treats the known low-resolution image and the high-resolution image to be recovered as two independent stochastic processes, and requires the design of a reasonable statistical prior model to maximize the posterior probability of image recovery after reconstruction. It has the advantages of direct incorporation of a priori constraints, high convergence stability and strong noise reduction capability, but the disadvantages are large computational effort and slow convergence speed. Another common reconstruction-based method is the iterative back-projection method[16], which back-projects the er-

ror between the degraded image and the low-resolution image of the reconstructed image and uses this error to correct the current reconstructed image. A super-resolution sensing method based on maximum likelihood estimation and convex set projection was proposed later[17], which makes full use of the prior knowledge and possesses good convergence stability.

With the continuous improvement of computer computing power, deep learning has become a popular topic for many researchers and scholars, and the application of neural networks in the field of image and signal processing has become a new development trend. Convolutional neural networks (CNN) have become a representative structure in the field of computer vision after KRIZHEVSKY et al.[18] applied CNN to image classification and achieved amazing results in 2012. In 2014, DONG et al.[19] first used CNN for super-resolution sensing of images and proposed a CNN-based super-resolution reconstruction network (SRCNN), which learns the features from low-resolution images to high-resolution images by pre-processing the input low-resolution images with linear interpolation and then recovering the high-resolution images through feature extraction, nonlinear mapping and image reconstruction. This model learns the feature mapping relationship from low-resolution images to high-resolution images, which is the pioneer of super-resolution image reconstruction based on deep learning, and has a great improvement in the quality of recovered images compared with traditional methods. In 2016, DONG et al.[20] proposed the faster SRCNN (FSRCNN), which eliminates the interpolation preprocessing step, takes the low-resolution image as the input of the model directly, and uses the deconvolution operation to enlarge the feature map at the end, which greatly reduces the computation of the model and accelerates the operation speed. Subsequently, SHI et al. [21] proposed an efficient sub-pixel convolutional neural network (ESPCN), which uses sub-pixel convolution layers instead of deconvolution to scale up the learned feature maps, further reducing the computational effort of the model and providing better image recovery quality compared with FSRCNN. Later, with the emergence of the residual network and the recurrent neural network (RNN), KIM et al. successively proposed the deeply-recursive convolution network (DRCN) [22] and very deep super resolution (VDSR)[23] models. The DRCN utilizes a recursive approach to share the parameters of the network layers and reduce the number of parameters of the model, while the VDSR model utilizes the global jump connection property of the residual structure to connect the input layer for better image recovery quality. LEDIG et al.[24] proposed super resolution residual network (SRResNet), a super-resolution sensing algorithm based on deep residual networks, which adds multiple modules for local residual learning and increases the number of layers of the network to learn more low-resolution to high-resolution feature mapping information, and achieves further improvement in the quality of image recovery. With the popularity of generative adversarial networks (GAN) [25], LEDIG also proposed the super resolution generative adversarial network (SRGAN) model with the main improvement of changing the loss function to adversarial loss and content loss, which can better recover the texture details of images. LIM et al.[26] proposed enhanced deep super resolution model (EDSR) based on SRResNet in 2017, the batch normalization layer in the model was removed, the network parameters were reduced, and it was applied to scenes with multi-scale recovery of low-resolution images.

3 Super Resolution Sensing for Resource Monitoring on Edge Clouds

3.1 System Model

At present, super-resolution sensing technology is mostly applied in the field of image and video processing, which can reconstruct low-resolution images into high-resolution images by deep learning methods and has great improvement in accuracy compared with traditional methods, but it is rarely applied to the reconstruction of low-frequency time series. Inspired by this, this paper proposes a super-resolution sensing algorithm for resource monitoring.

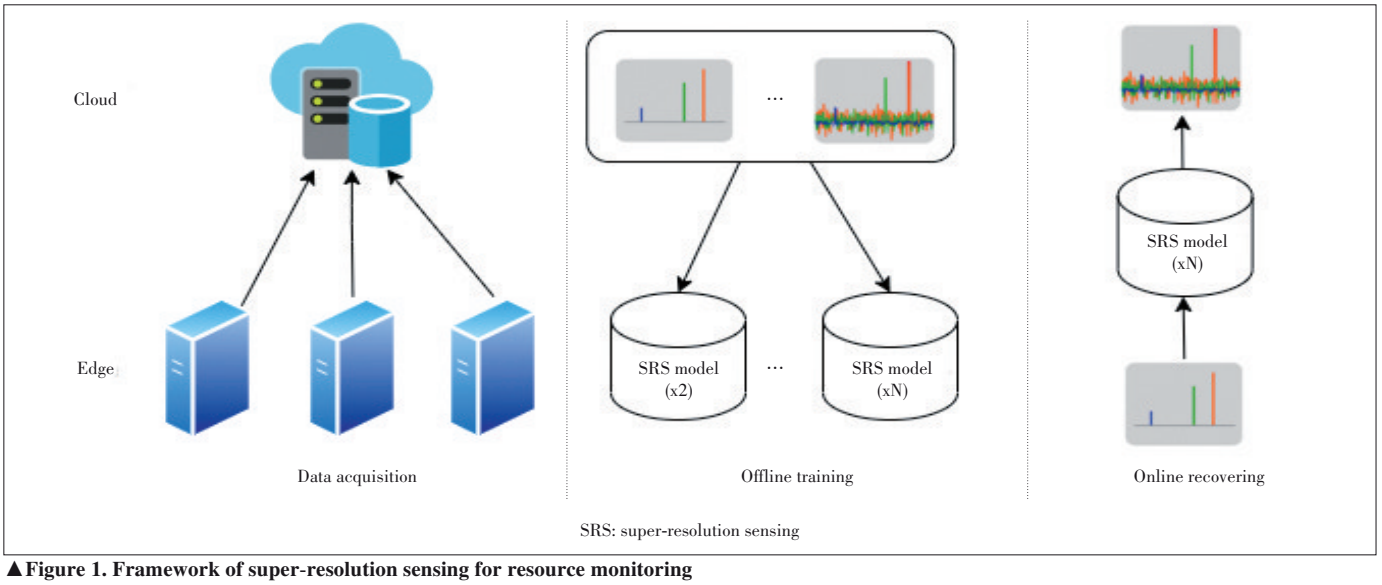
Fig. 1 shows the framework of the algorithm on edge clouds. It can be seen that there are three main phases included in the edge cloud application scenario, which are the data acquisition phase, offline training phase and online recovery phase. First, in the data collection phase, low-frequency and high-frequency monitoring data, which can be CPU utilization, memory usage or bandwidth status, need to be collected from each edge device side in a distributed manner and then received and stored uniformly by the cloud proxy server after the collection is completed. Then, in the offline training phase, the collected low-frequency and high-frequency resource data are used to train the corresponding SRS models, which are responsible for information reconstruction tasks with different sensing factors, such as super resolution sensing models with sensing factors of 2, 5, and 10. Once the training is completed, the trained models can be packaged into containers to be deployed as online services in the resource monitoring system in the cloud. Finally, the online recovery phase is to recover the reliable and accurate high-frequency data from the low-frequency data collected at the edge device side through the trained super-resolution sensing model, and the resource monitoring system will select the model with the appropriate sensing factors for recovery when and only when high-frequency data is needed, which can significantly reduce the communication and storage cost of resource data.

3.2 SRS Network Structure

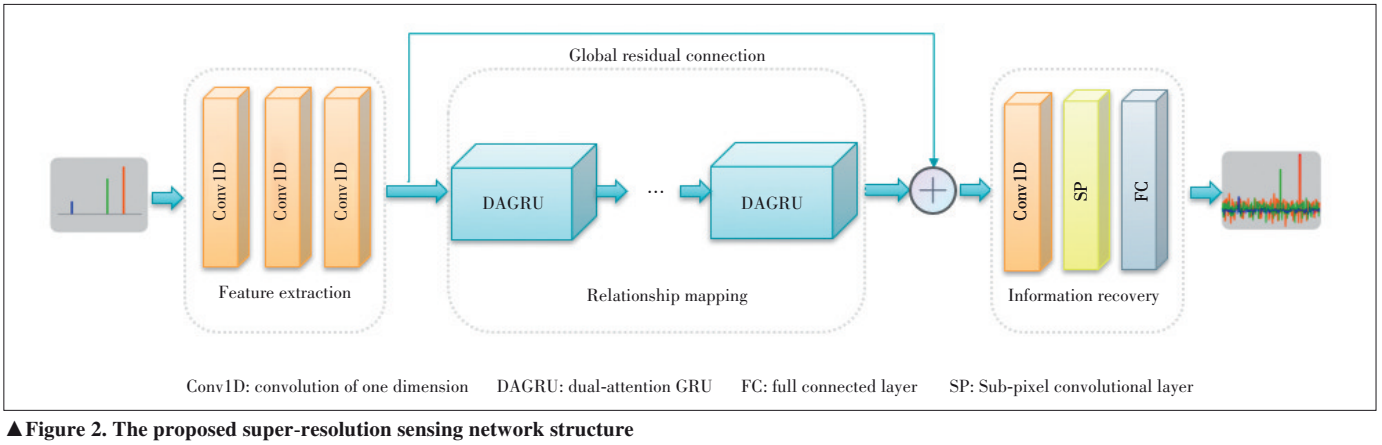
Based on the attention mechanism and gate recurrent unit (GRU) network, this paper proposes a super-resolution sensing model for resource monitoring, which can extract feature

information after inputting low-frequency data, reconstruct high-frequency information through relational mapping, and finally recover the corresponding high-frequency signal. The specific model structure is shown in Fig. 2. This super-resolution sensing model is mainly divided into three stages: the feature extraction, relational mapping and information recovery. In the feature extraction stage, multiple one-dimensional convolutional layers act as global feature extractors of low-frequency information to extract abstract features of the input signal and represent them as feature vectors. And the relational mapping layer consists of a series of GRU network blocks dual-attention GRU (DAGRU) based on the temporal attention mechanism and feature attention mechanism, in which global residual blocks and local residual blocks are also added. The sub-network composed of many DAGRU blocks can effectively learn the intrinsic connection between low-frequency information and high-frequency information, and the information lost by low-frequency information can be made up by relational mapping. Finally, in the information recovery stage, the information inferred by the relationship mapping layer is passed through the convolutional layer for feature extraction of potential relationships, while the feature map dimension is reconstructed, and then the low-frequency information is reconstructed into the high-frequency information corresponding to the sensing factors by multiple sub-pixel convolutional layers, followed by the recovery of the complete high-frequency resource monitoring information using the fully connected layer.

The most important part of the super-resolution sensing model proposed in this paper is the relational mapping layer and the feature information F_L obtained from the low-frequency resource signal through the feature extraction layer is used as input, which can complete the missing information of the feature vector in the low-frequency data after the relational inference by multiple DAGRU networks. As a variant of recurrent neural network, GRU can solve the problem of gradient disappearance and explosion than the standard RNN structure and can extract long-term dependencies in temporal sequences. It requires less computational resources, has fewer model parameters and has good convergence than the widely popular LSTM unit. Therefore, in this paper, GRU is used as a basic



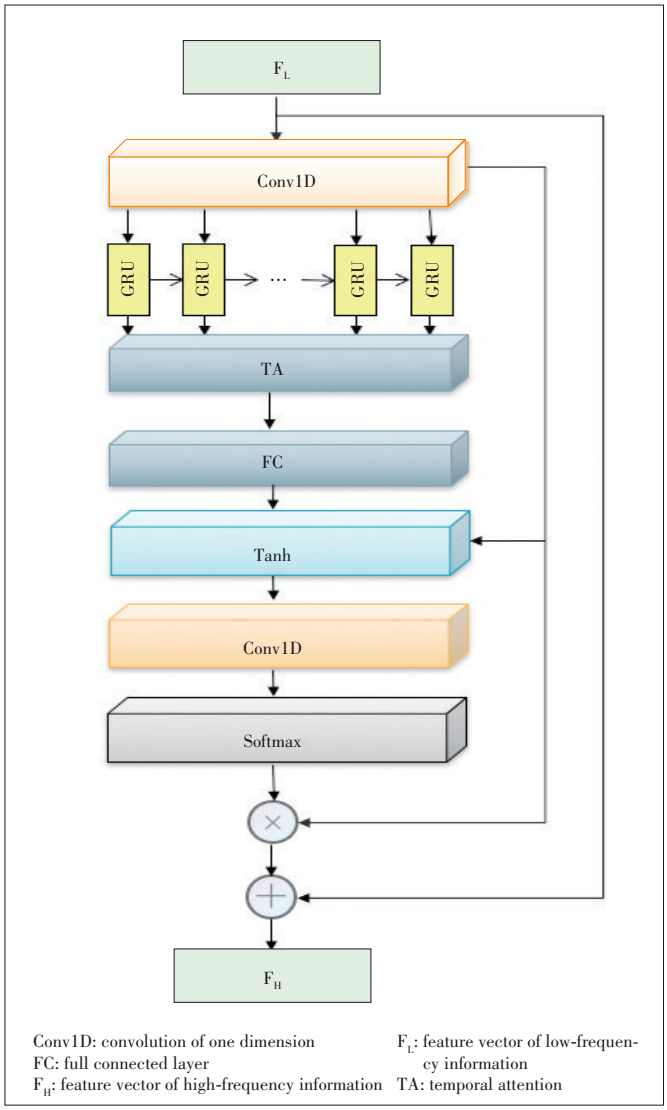
▲Figure 1. Framework of super-resolution sensing for resource monitoring



▲Figure 2. The proposed super-resolution sensing network structure

unit in the relational mapping layer and the DAGRU network is designed by combining the attention mechanism (Fig. 3).

The attention mechanism can be basically divided into hard attention mechanism and soft attention mechanism according to the degree of attention to important regions. The hard attention mechanism refers to the targeted selection of some features in the input information for learning while directly ignoring other unselected features. It can be implemented in two ways. One is to select input information with the highest frequency and the other can be obtained by performing random sampling on the attention distribution. Hard attention can greatly reduce the size of the parameters in a neural network and lower the requirement for computational resources by learning only some of the key regions and discarding other irrelevant information. However, it is usually based on random sampling to determine the input information, so it leads to a non-derivable relationship between the attention distribution and the loss function. Therefore, it is difficult to optimize the loss function by back-propagation methods. On the other hand, the soft attention mechanism can be used to learn the overall features according to the importance of different regions in the information by using weighted averaging without directly discarding some irrelevant regions, and the degree of each region being attended to can be expressed by a value between 0 and 1. Therefore, it is a microscopic process, which can be used in the training of neural networks by forward propagation to perform relational mapping and back propagation to perform parameter optimization. Based on the soft attention mechanism, in this paper, two attention mechanisms are designed in the DAGRU sub-network, which are the temporal attention mechanism (TA) and the feature attention mechanism.

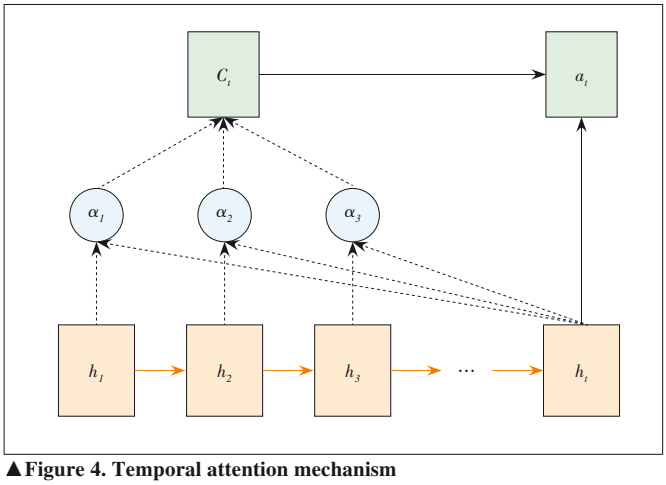


▲ **Figure 3. Structure of dual-attention gate recurrent unit (DAGRU network based on attention mechanism)**

Since the fluctuation of resource monitoring information is affected by the historical state, GRU network can be used to extract long-term dependencies, but this is not enough to meet the requirements for the accuracy of the trained model. After the analysis of some monitoring resource sequences, it is found that, for example, CPU utilization may rise suddenly in a period of time due to the sudden need to deal with computationally intensive tasks. This shows that the degree of influence of each historical moment on the current moment state is different, so a temporal attention mechanism is added on the basis of GRU network, which can adaptively decide the degree of influence of historical moments on the current moment information. The specific structure is shown in Fig. 4.

The hidden layer state values h_i containing information about the historical moments are taken as inputs in Fig. 4 and $i \in [1, t-1]$. Their degrees of influence on the current moment h_t are analyzed, which can be achieved by the scoring mechanism, and the scoring function f is in the form of a dot product as follows:

$$f(h_t, h_i) = h_t^T W h_i, \tag{1}$$



▲ **Figure 4. Temporal attention mechanism**

where W is a weight matrix about the hidden values of the historical states, followed by a softmax function to find the degree of influence of each historical moment on the current moment defined as α_i with the following formula:

$$\alpha_i = \frac{\exp\left(f\left(h_t, h_i\right)\right)}{\sum \exp\left(f\left(h_t, h_i\right)\right)}. \tag{2}$$

Then, according to the derived attention distribution α_i, a weighted average is done for each input h_i, and the input information is encoded to obtain the context vector $\boldsymbol{c}_t$ described by the formula:

$$\boldsymbol{c}_t = \sum_{i=1}^{t-1} \alpha_i h_i. \tag{3}$$

After obtaining the context vector $\boldsymbol{c}_t$, the final state value a_t is obtained with the hidden layer state value h_t by an additive model and by the tanh activation function with the following formula:

$$a_t = \tanh\left(W_c\left[\boldsymbol{c}_t ; h_t\right]\right). \tag{4}$$

It can be seen from the above equations that the temporal attention mechanism of the hidden layer states in the GRU unit can determine the degrees of correlation of historical state values at different moments. Combined with the current moment state values, this mechanism enables a kind of adaptive extraction of long-term dependency features, which can effectively suppress the less influential historical state values and help improve the quality of the recovered signal.

4 Evaluation

To verify the effectiveness of the super-resolution model for resource monitoring proposed in this paper, we select the 2018 cluster resource dataset named cluster-trace-v2018 publicly available from Alibaba Group, which contains resource monitoring information of about 4 000 servers over eight days.

4.1 Quantitative Comparison

In order to study the characteristics and effectiveness of the proposed SRS method, we conduct experiments using three metrics: the mean absolute percentage error (MAPE), peak signal to noise ratio (PSNR) and dynamic time warping (DTW)[27]. MAPE is commonly used to measure the temporal similarity of different time series and a higher MAPE value means a larger difference between the true value and the recovery value. PSNR represents the ratio of the maximum power of the signal to the average power of noise and a higher PSNR value indicates that the recovery value contains a smaller noise. DTW distance is a popular metric to obtain an optimal alignment that can be used to measure the similarity of shape features, which can be relatively robust to interference factors. We compare our SRS method with traditional upsampling methods including the linear interpolation, cubic interpolation, and compressive sampling matching pursuit (CoSaMP)[28] based on compressed sensing (CS). The quantitative comparison results between SRS and the other methods are shown in Table 1. It is obvious that the proposed SRS method is significantly better than other methods in different frequencies. From the perspective of DTW distances, the results of the SRS method are smaller than the results obtained by the other methods, which demonstrates that the proposed SRS method can effectively recover the shape characteristics of high-frequency resource data.

4.2 Qualitative Comparison

Figs. 5 and 6 show the qualitative comparison results. For each visualization result, it is obvious that the interpolation and compressed sensing methods cannot recover the missing details in the high-frequency data, resulting in the loss of information during the degradation process. Compared with the proposed SRS method, the other methods can only recover the rough shape of waveform, which is difficult to restore the peak values of the resource information. The proposed SRS method can effectively recover the shape of waveforms and peak value of high-frequency resource information.

▼**Table 1. Comparison results attained by SRS and other methods**

f_l	f_h	SRF	Linear Interpolation (MAPE/PSNR/DTW)	Cubic Interpolation (MAPE/PSNR/DTW)	CS (MAPE/PSNR/DTW)	SRS (MAPE/PSNR/DTW)
1/20	1/10	2	4.78%/33.58/334.62	4.66%/33.73/317.26	11.30%/27.04/678.56	**4.15%/34.78/296.52**
1/50	1/10	5	7.23%/29.89/493.52	7.16%/29.86/472.10	23.03%/20.93/1 499.84	**6.32%/30.95/442.34**
1/100	1/10	10	8.99%/27.99/623.26	9.10%/27.89/600.04	30.85%/19.06/2 157.87	**8.09%/28.72/548.06**
1/200	1/10	20	11.93%/26.02/781.07	11.02%/26.52/757.24	36.45%/18.42/2 649.60	**9.85%/27.03/696.51**
1/400	1/10	40	13.16%/24.30/936.30	12.62%/24.35/918.10	106.59%/15.34/8 549.43	**11.09%/26.01/817.40**
1/100	1/50	2	5.33%/32.66/78.78	5.33%/32.69/74.23	14.90%/24.85/191.03	**4.84%/33.91/70.99**
1/100	1/20	5	7.72%/29.34/264.58	7.81%/29.92/252.15	24.42%/20.77/819.10	**6.98%/30.29/241.26**
1/200	1/20	10	9.78%/27.41/347.59	9.99%/27.19/334.94	31.70%/16.64/1 159.87	**8.94%/28.03/318.82**
1/400	1/20	20	11.23%/26.38/428.85	11.13%/26.03/419.25	82.49%/16.16/3 458.83	**10.26%/26.79/379.41**

f_l: low sampling frequency of CPU usage f_h: high sampling frequency of CPU usage CS: compressed sensing DTW: dynamic time warping
MAPE: mean absolute percentage error PSNR: peak signal to noise ratio SRF: super resolution factor SRS: super resolution sensing

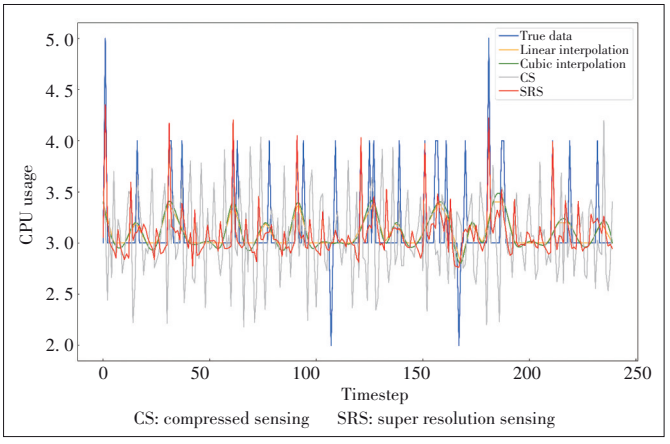
4.3 Analysis of Latency Under Dynamic Bandwidth

The delay variations of different algorithms under dynamic bandwidth environment with certain recovery accuracy are shown in Fig. 7. It can be seen that the super-resolution sensing model proposed in this paper can reduce the delay requirement of sending monitoring data from the edge to the cloud by allowing the edge devices to collect resources at a lower frequency and guaranteeing certain recovery accuracy at a lower bandwidth. In addition, along with the increase of bandwidth, the delay variation is basically in a smooth state while the other methods will generate higher delay on low-bandwidth network.

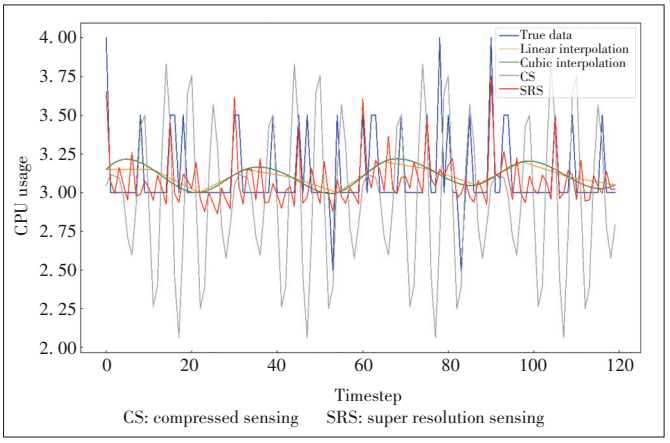
5 Conclusions and Future Work

With the advent of the IoT era, computing power has started to gradually move from the cloud down to the edge and the traditional cloud computing models are quietly changing. The emergence of thousands of edge devices will greatly share the pressure of cloud computing, so it is also necessary to provide a perfect resource monitoring scheme for the new edge cloud scenario. However, the existing resource monitoring schemes basically serve the cloud data center and there is no need to consider the computational overhead and bandwidth cost when collecting the resource monitoring data. Therefore, this paper proposes a super-resolution sensing model for resource monitoring by studying the existing signal reconstruction techniques including compressed sensing and super-resolution sensing for the edge cloud application scenario, which is mainly divided into three structures: the feature extraction layer, the relationship mapping layer and the information recovery layer. In the feature extraction layer, the low-frequency resource monitoring information is extracted by three-layer one-dimensional convolution. Then in the relationship mapping phase, the mapping relationship between low-frequency data and high-frequency data is mined by a GRU network based on the temporal attention mechanism and feature attention mechanism, which is used as the input of the information recovery layer. Finally, the low-frequency feature information is reconstructed into reliable and accurate high-frequency information in the multi-scale sub-pixel convolution layer.

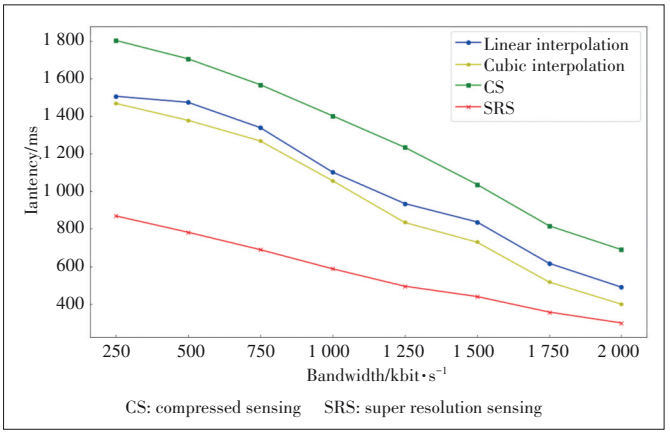
Aiming at meeting the requirement of real-time monitoring, the super-resolution sensing technique for resource monitoring proposed in this paper reduces the amount of transmitted resource data by reducing the sampling frequency at the edge devices. Due to the limitation of the experimental environment, a finite number of edge devices are used in our experiments, and subsequent studies can be migrated to larger edge clusters for future experiments. In addition, in the dynamic bandwidth environment, the data acquisition module can be allowed to adjust the acquisition frequency adaptively to collect monitoring resources, and the super-resolution sensing model with corresponding sensing factors can be trained in the cloud to increase the robustness of real-time resource monitoring by adjusting the sensing factors. In addition, only a single type of resource information is used to train the model in this paper, and subsequent research can also analyze the similarity of different types of resource information to collaborate the recovery process from low-frequency data to high-frequency data, which can also effectively improve the quality of resource information recovery and the generalization ability of the SRS model.



▲Figure 5. SRS results of the experiment with f_l = 1/50 Hz and f_h = 1/10 Hz



▲Figure 6. SRS results of the experiment with f_l = 1/200 Hz and f_h = 1/20 Hz



▲Figure 7. Results of latency under dynamic bandwidth obtained by SRS and other methods

Reference

[1] ZWOLENSKI M, WEATHERILL L. The digital universe rich data and the increasing value of the Internet of Things [J]. Australian journal of telecommunications and the digital economy, 2014, 2(3): 47. DOI: 10.7790/ajtde.v2n3.47
[2] SHI W S, ZHANG X Z, WANG Y F, et al. Edge computing: state-of-the-art and future directions [J]. Journal of computer research and development, 2019, 56 (1): 69 – 89
[3] FINNEGAN M. Boeing 787s to create half a terabyte of data per flight, says virgin atlantic [J]. Computerworld UK, 2013, 6: 1 – 2
[4] GREENBERG A, HAMILTON J, MALTZ D A, et al. The cost of a cloud: research problems in data center networks [J]. ACM SIGCOMM computer communication review, 2008, 39(1): 68 – 73. DOI: 10.1145/1496091.1496103
[5] DONOHO D L. Compressed sensing [J]. IEEE transactions on information theory, 2006, 52(4): 1289 – 1306. DOI: 10.1109/TIT.2006.871582
[6] CHANDRASEKARAN V, SANGHAVI S, PARRILO P A, et al. Rank-sparsity incoherence for matrix decomposition [J]. SIAM journal on optimization, 2011, 21(2): 572 – 596. DOI: 10.1137/090761793
[7] LERTRATTANAPANICH S, BOSE N K. High resolution image formation from low resolution frames using Delaunay triangulation [J]. IEEE transactions on image processing, 2002, 11(12): 1427 – 1441. DOI: 10.1109/TIP.2002.806234
[8] TOMASI C, MANDUCHI R. Bilateral filtering for gray and color images [C]// Sixth International Conference on Computer Vision. Mumbai, India: IEEE, 1998: 839 – 846. DOI: 10.1109/ICCV.1998.710815
[9] LI X, ORCHARD M T. New edge-directed interpolation [J]. IEEE transactions on image processing, 2001, 10(10): 1521 – 1527. DOI: 10.1109/83.951537
[10] BELAHMIDI A, GUICHARD F. A partial differential equation approach to image zoom [C]//International Conference on Image Processing, ICIP '04. Singapore, Singapore: IEEE, 2004: 649 – 652. DOI: 10.1109/ICIP.2004.1418838
[11] ZWART C M, FRAKES D H. Segment adaptive gradient angle interpolation [J]. IEEE transactions on image processing, 2013, 22(8): 2960 – 2969. DOI: 10.1109/TIP.2012.2228493
[12] STARK H, OSKOUI P. High-resolution image recovery from image-plane arrays, using convex projections [J]. Josa A, 1989, 6(11): 1715 – 1726
[13] SCHULTZ R R, STEVENSON R L. Improved definition video frame enhancement [C]//International Conference on Acoustics, Speech, and Signal Processing. Detroit, USA: IEEE, 1995: 2169 – 2172. DOI: 10.1109/ICASSP.1995.479905
[14] SCHULTZ R R, STEVENSON R L. Video resolution enhancement [C]//Proc SPIE 2421, Image and Video Processing III. San Jose, United States: SPIE, 1995, 2421: 23 – 34. DOI: 10.1117/12.205488
[15] LIU C, SUN D Q. On Bayesian adaptive video super resolution [J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 36(2): 346 – 360. DOI: 10.1109/TPAMI.2013.127
[16] IRANI M, PELEG S. Improving resolution by image registration [J]. CVGIP: graphical models and image processing, 1991, 53(3): 231 – 239. DOI: 10.1016/1049-9652(91)90045-L
[17] ADAMCZYK K, WALCZAK A. Digital images interpolation with wavelet edge extractors [C]//3rd International Conference on Human System Interaction. Rzeszow, Poland: IEEE, 2010: 399 – 405. DOI: 10.1109/HSI.2010.5514539
[18] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84 – 90. DOI: 10.1145/3065386
[19] DONG C, LOY C C, HE K M, et al. Learning a deep convolutional network for image super-resolution [C]//13th European Conference on Computer Vision (ECCV). Zurich, Switzerland, 2014: 184 – 199. DOI: 10.1007/978-3-319-10593-2_13
[20] DONG C, LOY C C, TANG X O. Accelerating the super-resolution convolutional neural network [C]//14th European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 391 – 407. DOI: 10.1007/978-3-319-46475-6_25
[21] SHI W Z, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 1874 – 1883. DOI: 10.1109/CVPR.2016.207
[22] KIM J, LEE J K, LEE K M. Deeply-recursive convolutional network for image super-resolution [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 1637 – 1645. DOI: 10.1109/CVPR.2016.181
[23] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 1646 – 1654. DOI: 10.1109/CVPR.2016.182
[24] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017: 105 – 114. DOI: 10.1109/CVPR.2017.19
[25] MIRZA M, OSINDERO S. Conditional generative adversarial nets [EB/OL]. (2014-11-06)[2020-12-21]. https://arxiv.org/abs/1411.1784v1
[26] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution [C]//IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, USA: IEEE, 2017: 1132 – 1140. DOI: 10.1109/CVPRW.2017.151
[27] SALVADOR S, CHAN P. Toward accurate dynamic time warping in linear time and space [J]. Intelligent data analysis, 2007, 11(5): 561 – 580. DOI: 10.3233/ida-2007-11508
[28] NEEDELL D, TROPP J A. CoSaMP: iterative signal recovery from incomplete and inaccurate samples [J]. Applied and computational harmonic analysis, 2009, 26(3): 301 – 321. DOI: 10.1016/j.acha.2008.07.002

Biographies

YANG Han received the B.Eng. degree from Sun Yat-Sen University, China in 2019. He is currently pursuing his master's degree at Sun Yat-Sen University. His research interests include edge computing and edge intelligence.

CHEN Xu (chenxu35@mail.sysu.edu.cn) is a full professor at Sun Yat-Sen University, China, and the vice director of National and Local Joint Engineering Laboratory of Digital Home Interactive Applications. He received the Ph.D. degree in information engineering from The Chinese University of Hong Kong, China in 2012, and worked as a post-doctoral research associate at Arizona State University, USA from 2012 to 2014 and a Humboldt Scholar Fellow at Institute of Computer Science of University of Goettingen, Germany from 2014 to 2016. He is currently an area editor of *IEEE Open Journal of the Communications Society*, an associate editor of the *IEEE Transactions Wireless Communications*, *IEEE Internet of Things Journal* and *IEEE Journal on Selected Areas in Communications (JSAC)* series on network softwarization and enablers.

ZHOU Zhi received the B.S., M.E. and Ph.D. degrees in 2012, 2014 and 2017, respectively, all from the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST), China. He is currently an associate professor in the School of Computer Science and Engineering, Sun Yat-Sen University, China. In 2016, he was a visiting scholar at University of Goettingen, Germany. He was nominated for the 2019 CCF Outstanding Doctoral Dissertation Award, the sole recipient of the 2018 ACM Wuhan & Hubei Computer Society Doctoral Dissertation Award, and a recipient of the Best Paper Award of IEEE UIC 2018. His research interests include edge computing, cloud computing, and distributed systems.

Semiconductor Optical Amplifier and Gain Chip Used in Wavelength Tunable Lasers

Abstract: The design concept of semiconductor optical amplifier (SOA) and gain chip used in wavelength tunable lasers (TL) is discussed in this paper. The design concept is similar to that of a conventional SOA or a laser; however, there are a few different points. An SOA in front of the tunable laser should be polarization dependent and has low optical confinement factor. To obtain wide gain bandwidth at the threshold current, the gain chip used in the tunable laser cavity should be something between SOA and fixed-wavelength laser design, while the fixed-wavelength laser has high optical confinement factor. Detailed discussion is given with basic equations and some simulation results on saturation power of the SOA and gain bandwidth of gain chip are shown.

Keywords: external cavity; gain chip; saturation power; semiconductor optical amplifier; tunable laser

SATO Kenji[1,2], ZHANG Xiaobo[1]

(1. ZTE Photonics, Nanjing 210000, China;
2. Southeast University, Nanjing 211189, China)

DOI: 10.12142/ZTECOM.202103010

http://kns.cnki.net/kcms/detail/34.1294.TN.20210806.1000.002.html, published online August 9, 2021

Manuscript received: 2021-01-01

Citation (IEEE Format): K. Sato and X. B. Zhang, "Semiconductor optical amplifier and gain chip used in wavelength tunable lasers," *ZTE Communications*, vol. 19, no. 3, pp. 81 - 87, Sept. 2021. doi: 10.12142/ZTECOM.202103010.

1 Introduction

A wavelength tunable laser (TL) is one of the most popular light sources in digital coherent optical fiber telecommunication systems, because the use of TL can reduce the required number of inventory lasers, thereby reducing inventory cost[1]. A wavelength channel can be selected on ITU grid channels, but it must be very accurate. In early days, many kinds of tuning mechanisms were proposed, but recently there exist only a few kinds of integrable tunable laser assembly (ITLA) products in the market, because it is very difficult to achieve such performance as high output power and wavelength accuracy and low cost. The modulation format used in those coherent systems is an advanced one, such as Dual Polarization Quadrature Phase Shift Keying (DP-QPSK), 16 Quadrature Amplitude Modulation (16QAM), and other multi-level formats[2-3]. Because the insertion loss of these advanced modulators may be large, higher optical output power of ITLA is strongly required. In addition, for the use in small form-factor pluggable optical modules like C Form-factor Pluggable 2 (CFP2), Octal Small Form-factor Pluggable (OSFP), or Quad Small Form-factor Pluggable (QSFP), a part of light power is split to a coherent receiver for a light as a local oscillator due to the small package. Therefore, the ITLA with an optical amplifier in front of a laser unit has been recently becoming more and more popular. This kind of semiconductor optical amplifier (SOA) has some interesting features, such as polarization dependent gain and high saturation power. Its design can be similar to the design of a conventional laser and the SOA is sometimes monolithically integrated with a tunable laser[4].

Moreover, a gain chip is the other important topic in this paper. A gain chip is used in external cavity (EC) tunable laser configuration, and it works as a laser. Equivalently, this can be treated as a Fabry-Perot (FP) laser with a wavelength bandpass filter. Because this is a tunable laser, the gain bandwidth of the gain chip should be very wide and cover the whole C-band[5-7]. Therefore, the design of the gain chip is not exactly the same as a laser, but it is not the same as that of SOA. The design concept of the gain chip is also discussed in this paper. Moreover, the gain bandwidth of Erbium-doped fiber amplifier (EDFA) is being extended over the conventional band[8-9]. In the near future, a wider tuning range of tunable lasers will also be required than the conventional one.

The purpose here is to write a tutorial paper on design concept with basic equations and detailed discussion for the SOA and gain chip that are used for wavelength tunable lasers. To the best of our knowledge, the paper on such a topic has not yet existed. In Section 2, the design concepts of these two devices are discussed. The difference from conventional design is shown. In Section 3, details of tunable lasers with SOA or gain chip are explained. In Section 4, some simulation results

and comparison of design concepts are discussed. Finally, conclusions are given in Section 5.

2 Design Concepts of SOA and Gain Chip

The structure of SOA is actually very important. The design concepts are basically based on a conventional SOA or a laser, which are slightly different.

2.1 SOA Design Concept

The design concept of SOA for tunable laser is similar to an inline SOA, except polarization dependence[10-11]. Inline SOAs are widely used in network systems. Here it is briefly explained. SOA is used for a one-pass gain medium, so it is reasonable to define the gain G_s in single-pass of SOA as

$$G_s = \exp\left((\Gamma g - \alpha_i)L\right), \tag{1}$$

where Γ is the optical confinement factor, g is the gain in unit length, α_i is the internal optical loss, and L is the length of SOA.

2.1.1 Residual Reflection

The gain g is assumed as a function of carrier density (see Section 2.1.2). As the injection current increases, the carrier density N is also increasing. The facet reflectivity is normally very low on antireflection (AR) coatings, but in reality, residual reflection exists. The reflectivity is $R = \sqrt{R_1 R_2}$, where R_1 and R_2 are front and rear facet reflectivities, respectively. Considering the round-trip gain, the laser oscillation condition is $(G_s R)^2 = 1$. Because the gain cannot be increased due to lasing, R should be as low as possible to obtain higher gain. This is the reason why both facets of SOA are AR-coated with angled waveguides. Recently, the facet reflectivity has reached lower than 10^{-4}, which is required to operate SOA properly. The influence of residual reflection may cause the gain ripple, which is a small dependence of gain on wavelength. This is also known as a cause of noise figure. If the residual reflectivity is lower than 10^{-4}, G_s would be 40 dB. In a practical use for amplification of tunable laser output power, the small signal non-saturated gain will be around 20 dB, for instance. In this case, the gain ripple and noise are negligible for the use in tunable laser source. According to Ref. [12], the gain ripple depth m can be described as

$$m = \frac{1 + \sqrt{R_1 R_2}\, G_s}{1 - \sqrt{R_1 R_2}\, G_s}. \tag{2}$$

According to Eq. (2), the depth m is a function of facet reflectivity and single-pass gain G_s. For example, m is 0.09 dB with R_1 and R_2 of 10^{-4} and G_s of 20 dB. To achieve such a low reflectivity, angled waveguides on both coated facets should be applied[13]. Larger than 7-degree angle is sufficient to reduce facet reflectivity.

2.1.2 Saturation Power

To design SOA, it is useful to understand the following carrier rate equation[14]:

$$\frac{dN}{dt} = \frac{J}{qd} - \frac{N}{\tau_S} - A(N - N_0)\frac{I}{h\nu}, \tag{3}$$

where J is the injection current density, q is the elementary electric charge, d is the thickness of active layer, N is the carrier density, N_0 is the transparent carrier density of gain medium, τ_S is the carrier relaxation time, A is the differential gain, h is the Planck's constant, and ν is the frequency of light.

The intensity of light P (W/cm^2) can be described by the following propagation equation:

$$\frac{dP}{dz} = \Gamma A\left(N - N_0\right)P - \alpha_i P. \tag{4}$$

Here to make the phenomena simplified, the gain is proportional to the carrier density at the bulk active layer. The steady solution for $dN/dt = 0$ from Eq. (3) can be substituted in Eq. (4), and then dP/dz can be derived as

$$\frac{dP}{dz} = \Gamma A\frac{\frac{\tau_S J}{qd} - N_0}{1 + P/P_0}P - \alpha_i P = \left(\frac{\Gamma g_0}{1 + P/P_S} - \alpha_i\right)P, \tag{5}$$

where P_S is the saturation intensity, described as $P_S = h\nu/\tau_S A$, and g_0 is a non-saturated gain coefficient, which is described as $g_0 = A(\tau_S J/qd - N_0)$. When the intensity of light P is much smaller than P_S, the gain coefficient will be nearly equal to g_0. As the intensity is becoming larger, the gain coefficient will be getting smaller, obeying the term $1/(1 + P/P_S)$. To avoid such saturation of gain, P_S must be enlarged in the design. As the light is amplified in SOA, the carrier density will be consumed for stimulated emission, especially near the end. Therefore, the number of carriers will not be enough to amplify it any more. To avoid the saturation, τ_S or A should be small as seen in the description of P_S. However, it is difficult to control τ_S and A in reality.

In an ideal model of SOA, the output power of SOA can be described with only non-saturated small signal gain, by integrating dP/dz in Eq. (5):

$$P_{out} = P_{in}G_S = P_{in}\exp\left[\left(\Gamma g_0 - \alpha_i\right)L\right], \tag{6}$$

where P_{out} is the output signal power, P_{in} is the input signal power, and L is the SOA cavity length. As seen in Eq. (6), the single-pass small signal gain is dependent on $\Gamma g_0 L$. As the carrier density is increased, g_0 becomes larger and will reach saturation. However, saturation will be suppressed for the

small optical confinement factor Γ or short cavity length L while keeping the gain G_S. This means the carrier density can be increased and then the saturation power will become larger. The saturation power[15] is described as

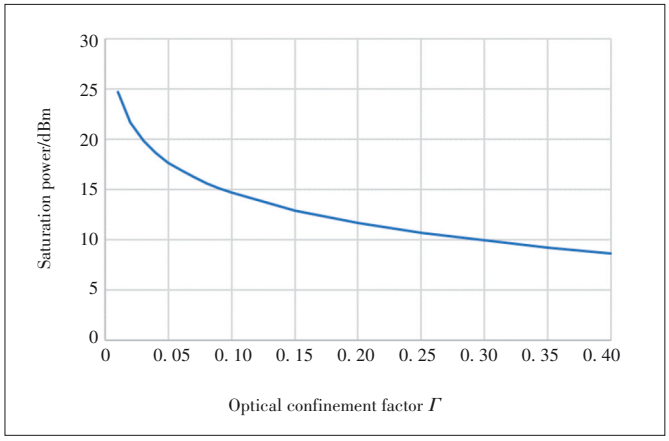
$$P_s = \frac{WdE}{\Gamma \tau_S A} \ln 2 , \tag{7}$$

where W is an active region width, d is the thickness, and E is photon energy as a function of wavelength λ. As shown in Eq. (7), the saturation power is inversely proportional to the optical confinement factor. One simulation result on gain saturation as a function of optical confinement factor is shown in Fig. 1, where W is 2 μm, d is 0.1 μm, λ is 1.55 μm and the product of $\tau_s A$ is 6.04×10^{-26} s·cm^2. As the confinement factor decreases, the saturation power will become larger.

2.1.3 Polarization Issue

Low residual reflection on facet and high saturation power have been discussed above. Because this kind of SOA is used just in front of the tunable laser, some other features are also required. We focus on polarization dependence here, rather than wide gain bandwidth that is very natural to be obtained by injecting high current to the SOA.

A very common structure of an active layer contains a multi-quantum well (MQW). This is a quantum well with multiple very thin layers, so the electrons' energy level to vertical direction can be quantized and then the gain has direction dependence to the quantum well structure. There exist heavy and light holes in the quantum well. The heavy holes can generate only Transverse Electric (TE) mode polarization, while the light holes can generate both TE and Transverse Magnetic (TM) mode polarizations. The inline SOA is normally required to obtain gain both for TE and TM modes as polarization independence[16-17]; however, the SOA used just in front of tunable laser should have gain only for TE mode polarization. To enhance transition from the electron to the heavy hole band, a compressively strained quantum well is used here. This can be very similar to the active layer of gain chip, but it is easier to monolithically integrate the SOA on the tunable laser. Because the injection current density for SOA is very different from that for the gain chip and the optical confinement factor of SOA should be low for high saturation power, the active layer should be different rigorously. Therefore, the active layers should be different from each other, so the chips should be integrated hybridly.



▲**Figure 1. Calculated saturation power as a function of optical confinement factor**

2.2 Gain Chip Design Concept

The MQW as an active layer is the most common gain medium for the SOA or laser. It is good to know how to design quantum wells with a quantum mechanics theory. Only a rough image of gain design with basic parameters is introduced in this paper. Gain per length is a function of the current density of the active layer, so it can be described as:

$$\text{Bulk active layer: Linear gain: } g = \Gamma A (N - N_0) , \tag{8}$$

$$\text{Quantum well: Logarithmic gain: } g = G_0 \ln\left(\frac{J}{J_0}\right) . \tag{9}$$

A linear gain function can be applied for the bulk active layer. On the other hand, due to a quantum well characteristic, the equation should be physically a logarithmic function of the current density[18]. However, it is a function of the current density and derived from an experience law. It is widely used because it can fit the characteristic very well. Laser designers are interested in the parameters G_0 and J_0. G_0 is an indicator kind of differential gain and the optical confinement factor Γ is already included. J_0 physically means a transparent current density.

For the waveguide quality, popular parameters used widely are α_i and η_i. α_i is an internal loss per length and η_i is a quantum internal efficiency. For α_i , it comes from an absorption of light and light scattering in waveguide and its design issue is to reduce internal loss in the waveguide. Besides, η_i means a rate of photon generation to one carrier seed. This value is very close to 1.0 because of a good quality of quantum well media, but it is decreased due to current leakage and a lost carrier for non-radiative recombination.

To reach the laser threshold, we must have

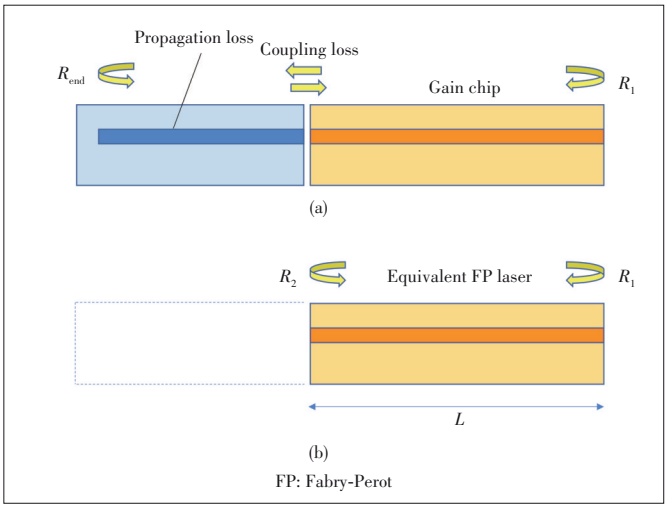
$$\Gamma g_{th} = \alpha_i + \alpha_m , \tag{10}$$

$$\alpha_m = -\frac{\ln(R_1 R_1)}{2L} : mirror\ loss . \tag{11}$$

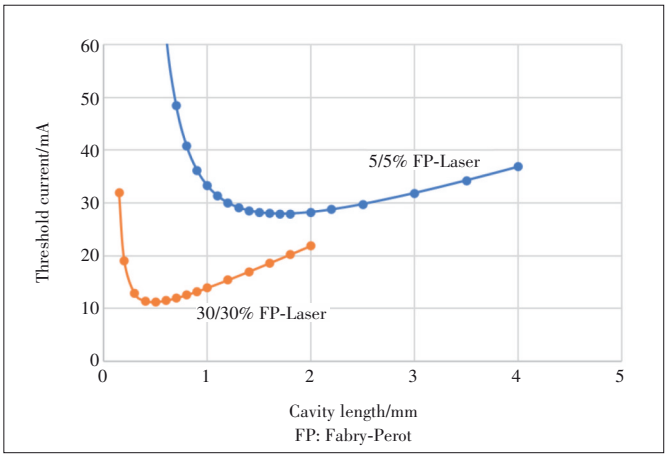
That is, the total of internal and mirror loss should be compensated by gain. Here the mirror loss is an out-going light per length in the gain region. In Eq. (10), all parameters are

defined for a unit length. The cavity length is also an important parameter. The external cavity laser can be handled as an equivalent FP laser, as shown in Fig. 2. The equivalent rear reflectivity R_2 includes the coupling loss between the gain chip and external cavity, propagation loss in external waveguide, and reflection at the end of external waveguide.

The threshold current I_{th} can be varied as a function of cavity length L, as shown in Fig. 3, where Γ is 5%, G_0 is 350 cm^{-1}, α_i is 8 cm^{-1}, and J_0 is 9.5×10^4 $A \cdot cm^{-2}$. It is well known that there is an optimum cavity length for the minimum threshold current. In Fig. 3, the ordinal FP-laser with 30/30% facet reflectivity shows the minimum threshold current at length around 0.5 mm. On the other hand, the 5/5% FP-laser which is equivalent to external cavity tunable laser shows the minimum at 1.5 mm. The effective reflectivity from the external cavity is assumed to be around 5%, and the confinement factor of the 5/5% FP-laser has two thirds of that of the 30/30% FP-laser to have higher threshold carrier density. This also implies that the designed effective reflectivity of the external cavity should be carefully considered. An example of the external cavity tunable laser can be found in Ref. [19].



▲**Figure 2. Schematic block diagram of (a) external cavity laser and (b) equivalent FP laser to external cavity with R_1 and R_2 with cavity length of L**



▲**Figure 3. Calculated threshold current as a function of cavity length with R1/R2 of 30/30% and 5/5% reflectivities**

Then the output power can be described as

$$P = \frac{\hbar\omega}{q}\left(I - I_{th}\right)\eta_i \frac{\alpha_m}{\alpha_i + \alpha_m}. \tag{12}$$

Therefore, the power is proportional to the current minus threshold current. If the gain is higher than the threshold gain, the light output will continue to increase and carrier density must clamp at its threshold value. That is, the energy goes to light output from the excess energy over threshold current. The carrier density cannot go higher in laser operation. On the other hand, the carrier density in SOA can go higher, due to a very high threshold current with small R_1 and R_2. This is the biggest difference between the SOA and the laser design. It is generally said that the design principle of SOA is opposite to the laser design. An SOA needs a lot of carriers to amplify the light in one direction and should avoid the carrier depletion in the cavity.

The gain bandwidth is also an important specification in gain chip. To realize very wide gain bandwidth, the carrier density at the threshold should be larger than that of an ordinary laser, but it should not be too high. The reflectivity of the cavity is rather low, like 5%, so naturally the threshold density increases; however, it is not enough. To tune the carrier density, it is good to take a smaller number of quantum wells or low optical confinement factor at the active layer, compared with an ordinary laser.

2.3 Summary

The gain chip design concept is to obtain wide gain bandwidth with low optical confinement factor and low residual reflectivity at the facet on the tunable filter side. The SOA design concept is to achieve high threshold current with low residual reflectivity and low optical confinement factor.

3 Examples of Tunable Lasers with SOA

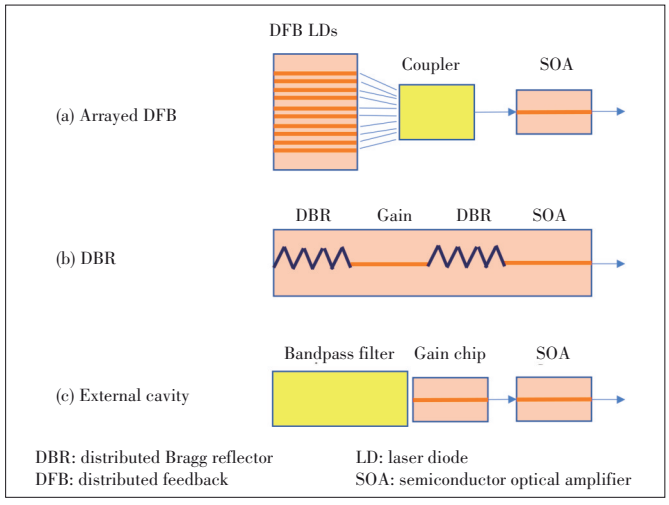
Some examples of tunable lasers with the gain chip and SOA are shown in this section. Generally, there are three major categories of tunable laser configuration with SOA, as shown in Fig. 4.

The first type is the arrayed distributed feedback (DFB) laser with integrated SOA[20 - 23]. This type has a long history of development. In early days of this research, the number of DFB lasers were limited, and the optical coupler was an multimode interference (MMI) coupler with SOA or Micro Electro Mechanical Systems (MEMS) coupler without SOA[24]. In the case of MMI couplers, the optical loss is large and then the SOA is needed. This type with SOA is suitable for monolithic integration and its wavelength tuning mechanism is simple to

tune only temperature of the chip. For monolithic integration, waveguides of active and passive regions are separately grown by Butt-joint technology. The same active layer is utilized both for laser and SOA regions. This means the freedom of SOA design is slightly limited.

The second type is the distributed Bragg reflector (DBR) tunable laser with a Vernier effect of two grating regions, implemented by multiple methods, such as sampled grating[25], digitally sampled (DS) DBR[26], and chirped sampled grating[27]. In this paper, the sampled grating is not the subject, so it is not discussed. Because the reflectivity of DBR must be rather high, the output power directly from the DBR tunable laser will be not high enough. Therefore, an SOA after the DBR laser is usually needed. In the same manner as the DFB arrayed type, the SOA in the other two types can be monolithically integrated by Butt-joint or Ion-implantation technology[28] on the same chip as a tunable laser part. The same active layer is also utilized.

The third one is an external cavity tunable laser with a gain chip[5, 18, 29]. The tunable laser is configured with an external tunable bandpass filter with a reflector and a gain chip. This gain chip can be handled as a FP-laser with a front facet mirror and a rear effective mirror of external filter, as mentioned above. In the case of silicon tunable filters, the gain chip and the filter chip are hybridly integrated. Because it is also difficult to obtain high output power directly from the gain chip, the SOA is also hybridly integrated to amplify the optical power. The gain chip and SOA cannot be monolithically integrated due to the facet reflectivity of the gain chip. Therefore, the freedom of SOA design is not limited to have high saturation power with lower optical confinement factor. One drawback is difficult optical coupling between chips, such as the interface of the filter and gain chip, and the gain chip and SOA. An advanced mounting technology is necessary for realizing low coupling loss.

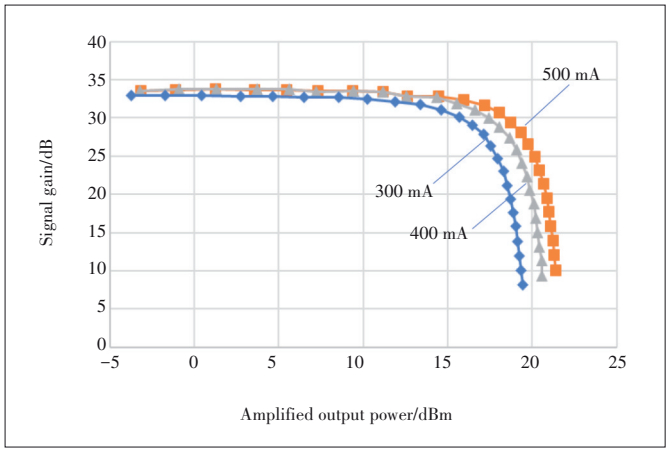


▲Figure 4. Tunable lasers with semiconductor optical amplifier (SOA)

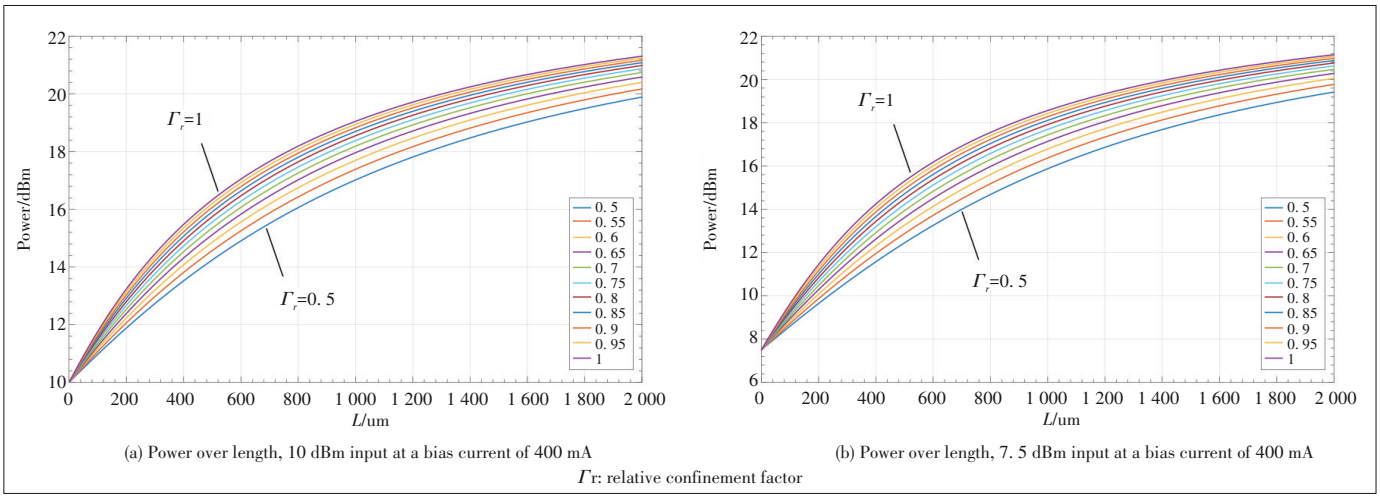
4 Simulation Examples and Discussion

In this section, the issue on optical confinement factor is discussed. Fig. 5 shows an example of measured optical signal gain as a function of amplified output power. As output power increases, the signal gain is becoming saturated. The point where the gain is decreased by 3 dB is called the saturation gain. As the injection current increases, the maximum output power will also increase, but it is also saturated even if the injection current keeps increasing.

In order to investigate saturation, internal power distribution in the SOA along the cavity axis with various optical confinement factors is calculated (Fig. 6). In this simulation, the SOA cavity is divided into small sections where the gain is saturated and slightly different from each other. The saturated gain in the i-th section can be described as $g_i = g/(1 + \varepsilon P)$, which depends on power P and saturation coefficient ε, where ε is $1/P_s$ and saturation power P_s is 22 dBm. Those parameters of confinement factors in Fig. 6 mean relative confinement factor Γ_r, where the actual confinement factor is a product of variable Γ_r and fixed Γ_0 of about 4% or so in reality. As shown in Fig. 6, a low relative confinement factor shows more linear amplification along the cavity than high confinement. In the case of amplification for the tunable laser, the input power to SOA is high, so it cannot be treated as a small signal. When an output power higher than 20 dBm is required, the light is amplified in almost saturation region with an amplification factor of around 10 dB for the input power of 10 dBm, as shown in Fig. 6 (a). When the input power is 7.5 dBm in Fig. 6 (b), the output power is about 20 dBm. Even if the input power changes, the output power is almost constant. This implies that confinement factor should be properly designed to suppress gain saturation, otherwise the consumption power of the amplification will increase. In Fig. 6, additional 1.5 dB gain is obtained to the half of the normalized optical confinement factor.



▲Figure 5. Measured optical signal gain for 1.55 μm light as a function of output power from a 2 mm long semiconductor optical amplifier (SOA)



(a) Power over length, 10 dBm input at a bias current of 400 mA

(b) Power over length, 7. 5 dBm input at a bias current of 400 mA

Γr: relative confinement factor

▲Figure 6. Calculated internal power distribution in SOA

Finally, the design concepts are summarized in Table 1, where the gain chip design concept can be placed between the SOA and laser design. The key parameter is the optical confinement factor. It is easy to convert the optical confinement factor into the number of quantum wells in the practical design. The active layer of the gain chip design is very similar to that of SOA except facet reflectivity, however the optical confinement factor should be slightly different from that of SOA, as mentioned above. Monolithic integration of SOA with the laser part almost works well. On the other hand, better performance should be obtained from hybrid integration with slightly different design, according to the principle.

5 Conclusions

The design concepts of the SOA and gain chip for wavelength tunable lasers are presented in this paper. The optical confinement factor of SOA should be low to realize high saturation power and be optimized only for TE mode polarization. The optical confinement factor of the gain chip should be rather lower than that of an ordinary laser and its length should be rather longer. It is also important to consider proper effective reflectivity of the external cavity to optimize performance of external cavity lasers.

▼Table 1. Comparison of design concepts

Items	SOA	Gain Chip for TL	Laser
Threshold	High	Moderate	Low
Facet reflectivity	Very low (for example, AR/AR)	AR/LR	Certain reflectivity (for example, HR/LR)
Optical confinement factor	Very low (3~5%)	~5%	Rather high (~10%)
Length	Long	Middle – long	Short – middle
Gain bandwidth	Wide	Wide	Narrow
Carrier density	High	Moderate	Low
Polarization	Independent for in-line SOA or dependent for use in TL	Dependent	Dependent

AR: anti-reflection LR: low-reflection TL: tunable laser
HR: high-reflection SOA: semiconductor optical amplifier

References

[1] BUUS J, MURPHY E J. Tunable lasers in optical networks [J]. Journal of lightwave technology, 2006, 24(1): 5 – 11. DOI: 10.1109/JLT.2005.859839
[2] TSUKAMOTO S, LY-GAGNON D S, KATOH K, et al. Coherent demodulation of 40-Gbit/s polarization-multiplexed QPSK signals with 16-GHz spacing after 200-km transmission [C]//Optical Fiber Communication Conference (OFC). Anaheim, USA: IEEE, 2005. DOI: 10.1109/OFC.2005.193207
[3] ZHOU X, YU J J. Multi-level, multi-dimensional coding for high-speed and high-spectral-efficiency optical transmission [J]. Journal of lightwave technology, 2009, 27(16): 3641 – 3653
[4] COLDREN L A. Monolithic tunable diode lasers [J]. IEEE journal of selected topics in quantum electronics, 2000, 6(6): 988 – 999. DOI: 10.1109/2944.902147
[5] CHAPMAN W B, DAIBER A, MCDONALD M, et al. Temperature tuned external cavity diode laser with micromachined silicon etalons [C]//Conference on Lasers and Electro-Optics (CLEO). San Francisco, USA: OSA, 2004: paper CWC2
[6] SATO K, MIZUTANI K, SUDO S, et al. Wideband external cavity wavelength-tunable laser utilizing a liquid-crystal-based mirror and an intracavity etalon [J]. Journal of lightwave technology, 2007, 25(8): 2226 – 2232
[7] TAKAHASHI M, DEKI Y, TAKAESU S, et al. A stable widely tunable laser using a silica-waveguide triple-ring resonator [C]//Optical Fiber Communication Conference, (OFC). Anaheim, USA: IEEE, 2005. DOI: 10.1109/OFC.2005.193197
[8] LEI C M, FENG H L, MESSADDEQ Y, et al. Investigation of C-band pumping for extended L-band EDFAs [J]. Journal of the optical society of America B, 2020, 37(8): 2345 – 2352. DOI: 10.1364/josab.392291
[9] DE BARROS M, ROSOLEM J, ROCHA M, et al. Transmission in the L+ band for metropolitan applications [C]//Optical Fiber Communications Conference (OFC), 2003. Atlanta, USA: IEEE, 2003: 93 – 94. DOI: 10.1109/OFC.2003.1247513
[10] BUUS J, PLASTOW R. A theoretical and experimental investigation of

Optical confinement factor Γ

Fabry-Perot semiconductor laser amplifiers [J]. IEEE journal of quantum electronics, 1985, 21(6): 614 – 618. DOI: 10.1109/JQE.1985.1072710
[11] SIMON J. GaInAsP semiconductor laser amplifiers for single-mode fiber communications [J]. Journal of lightwave technology, 1987, 5(9): 1286 – 1295. DOI: 10.1109/JLT.1987.1075637
[12] EISENSTEIN G, JOPSON R M, LINKE R A, et al. Gain measurements of InGaAsP 1.5 μm optical amplifiers [J]. Electronics letters, 1985, 21(23): 1076 – 1077. DOI: 10.1049/el: 19850764
[13] COLLAR A J, HENSHALL G D, FARRE J, et al. Low residual reflectivity of angled-facet semiconductor laser amplifiers [J]. IEEE photonics technology letters, 1990, 2(8): 553 – 555. DOI: 10.1109/68.58046
[14] MARCUSE D. Computer model of an injection laser amplifier [J]. IEEE journal of quantum electronics, 1983, 19(1): 63 – 73. DOI: 10.1109/JQE.1983.1071725
[15] O'MAHONY M J. Semiconductor laser optical amplifiers for use in future fiber systems [J]. Journal of lightwave technology, 1988, 6(4): 531 – 544. DOI: 10.1109/50.4035
[16] YOKOUCHI N, YAMANAKA N, IWAI N, et al. Tensile-strained GaInAsP-InP quantum-well lasers emitting at 1.3 um [J]. IEEE journal of quantum electronics, 1996, 32(12): 2148 – 2155. DOI: 10.1109/3.544762
[17] MAGARI K, OKAMOTO M, YASAKA H, et al. Polarization insensitive traveling wave type amplifier using strained multiple quantum well structure [J]. IEEE photonics technology letters, 1990, 2(8): 556 – 558. DOI: 10.1109/68.58047
[18] WHITEAWAY J E A, THOMPSON G H B, GREENE P D, et al. Logarithmic gain/current-density characteristic of InGaAs/InGaAlAs/InP multi-quantum well separate confinement heterostructure lasers [J]. Electronics letters, 1991, 27(4): 340 – 342. DOI: 10.1049/el: 19910215
[19] KOBAYASHI N, SATO K, NAMIWAKA M, et al. Silicon photonic hybrid ring-filter external cavity wavelength tunable lasers [J]. Journal of lightwave technology, 2015, 33(6): 1241 – 1246. DOI: 10.1109/JLT.2014.2385106
[20] OOHASHI H, SHIBATA Y, ISHII H, et al. 46.9-nm wavelength-selectable arrayed DFB lasers with integrated MMI coupler and SOA [C]//13th International Conference on Indium Phosphide and Related Materials (IPRM). Nara, Japan: IEEE, 2001: 575-578. DOI: 10.1109/ICIPRM.2001.929216
[21] KIMOTO T, KUROBE T, MURANUSHI K, et al. Reduction of spectral-linewidth in high power SOA integrated wavelength selectable laser [C]//19th International Semiconductor Laser Conference. Matsue, Japan: IEEE, 2004: 149 – 150. DOI: 10.1109/ISLC.2004.1382801
[22] BOUDA M, MATSUDA M, MORITO K, et al. Compact high-power wavelength selectable lasers for WDM applications [C]//Optical Fiber Communication Conference (OFC). Baltimore, USA: IEEE, 2000: 178 – 180. DOI: 10.1109/OFC.2000.868407
[23] YASHIKI K, SATO K, MORIMOTO T, et al. Wavelength-selectable light sources fabricated using advanced microarray-selective epitaxy [J]. IEEE photonics technology letters, 2004, 16(7): 1619 – 1621. DOI: 10.1109/LPT.2004.828544
[24] ZOU S, YOFFE G W, LU B, et al. 100 mW phase-shifted 1 550 nm BH DFB arrays with a 10-micron pitch [C]//Optical Fiber Communication Conference (OFC). Anaheim, USA: IEEE, 2005. DOI: 10.1109/OFC.2005.192825
[25] JAYARAMAN V, CHUANG Z M, COLDREN L A. Theory, design, and performance of extended tuning range semiconductor lasers with sampled gratings [J]. IEEE journal of quantum electronics, 1993, 29(6): 1824 – 1834. DOI: 10.1109/3.234440
[26] ROBBINS D J, BUSICO G, PONNAMPALAM L, et al. A high power, broadband tunable laser module based on a DS-DBR laser with integrated SOA [C]// Optical Fiber Communication Conference (OFC). Los Angeles, USA, 2004.
[27] YOSHINAGA H, YANAGISAWA M, KANEKO T, et al. Single-stripe tunable laser with chirped sampled gratings fabricated by nanoimprint lithography [J]. Japanese journal of applied physics, 2014, 53(8S2): 08MB05. DOI: 10.7567/jjap.53.08mb05
[28] COLDREN L A, CORZINE S W, MAŠANOVIĆ M L. Diode lasers and photonic integrated circuits [M]. Hoboken, USA: John Wiley & Sons, Inc., 2012. DOI: 10.1002/9781118148167
[29] GAO Y K, LO J C, LEE S, et al. High-power, narrow-linewidth, miniaturized silicon photonic tunable laser with accurate frequency control [J]. Journal of lightwave technology, 2020, 38(2): 265 – 271. DOI: 10.1109/JLT.2019.2940589

Biographies

SATO Kenji (zuoteng.jianer@zte.com.cn) received his B.E. degree in electrical engineering, M.E. degree in electronic engineering, and Ph.D. degree in electronic engineering from the University of Tokyo, Japan in 1991, 1993 and 1996, respectively. From 1993 to 1994, he was with the Department of Information Technology, Faculty of Applied Science, University of Ghent, Belgium, under a Flemish Government Scholarship. In 1996, he joined the Central Research Laboratories, NEC Corporation, Japan, where he had been engaged in the research and development of semiconductor laser diodes, photodetectors, modulators and silicon-photonic external cavity lasers for optical fiber communications. Since 2020, he has been a guest professor at Southeast University, China and concurrently a chip-design expert at ZTE Photonics. He is the author or coauthor of more than 60 technical and conference papers and has more than 20 registered patents.

ZHANG Xiaobo received his B.E. degree in optical engineering from Huazhong University of Science and Technology, China in 2014 and M.E. degree in optical engineering from Zhejiang University, China in 2017. In 2017, he joined ZTE Photonics, where he has been engaged in the research and development of silicon-photonic external cavity lasers for optical fiber communications.

Feedback-Aware Anomaly Detection Through Logs for Large-Scale Software Systems

Abstract: One particular challenge for large-scale software systems is anomaly detection. System logs are a straightforward and common source of information for anomaly detection. Existing log-based anomaly detectors are unusable in real-world industrial systems due to high false-positive rates. In this paper, we incorporate human feedback to adjust the detection model structure to reduce false positives. We apply our approach to two industrial large-scale systems. Results have shown that our approach performs much better than state-of-the-art works with 50% higher accuracy. Besides, human feedback can reduce more than 70% of false positives and greatly improve detection precision.

Keywords: human feedback; log-based anomaly detection; system log

HAN Jing[1], JIA Tong[2], WU Yifan[2], HOU Chuanjia[2], LI Ying[2]

(1. ZTE Corporation, Shenzhen 518057, China;
2. Peking University, Beijing 100091, China)

DOI: 10.12142/ZTECOM.202103011

http://kns.cnki.net/kcms/detail/34.1294.TN.20210730.1102.001.html, published online July 30, 2021

Manuscript received: 2021-02-04

Citation (IEEE Format): J. Han, T. Jia, Y. F. Wu, et al., "Feedback-aware anomaly detection through logs for large-scale software systems," *ZTE Communications*, vol. 19, no. 3, pp. 88 – 94, Sept. 2021. doi: 10.12142/ZTECOM.202103011.

1 Introduction

Large-scale software systems face one particular challenge which is anomaly detection. System logs provide a straightforward and common information source for anomaly detection. Typically, administrators manually check log files and search for problem-related log entries, which is error-prone and time-tedious. To reduce human efforts, researchers have proposed many automatic log-based anomaly detectors[1-19]. However, these detectors are ineffective in real-world industrial systems. First, most detectors typically operate by identifying statistical outliers. The utility of a particular detector for a system depends on how well its statistical outliers align with system anomaly symptoms. In general, the gap between statistical outliers and real system anomalies can result in high false-positive rates and easily render an anomaly detector unusable. Second, new types of anomalies may arise during system updates and conflict with existing anomaly detectors to produce false positives. Third, heterogeneous and complex log data contains massive noise. This noise may mislead detectors and further increase false positives.

One way to reduce the false-positive rate is to build domain knowledge into a detector. For example, a designer might apply domain expertise to label training logs that are more likely to produce correct anomalies and/or filter anomalies based on semantically defined white lists. Unfortunately, this requires significant expertise in both the system and anomaly detection. Besides, a large number of logs from industrial large-scale systems are almsot impossble to label; e.g., a Microsoft online service system even generates over one petabyte (PB) of logs every day[20].

In this paper, we consider an approach to reduce false positives based on incorporating human feedback. In our settings of feedback-aware anomaly detection, humans only provide feedback about whether the detected anomaly is false positive or not. This feedback is used by the detector to adjust the anomaly detection model structure. This approach has the advantage of an easy and concise feedback process with little overhead on time. The main contributions of this paper includes:

1) To the best of our knowledge, we are the first to incorporate human feedback to reduce false positives for the log-based anomaly detection task.

2) We propose a feedback-aware online anomaly detection approach that builds a graph model from an online log stream and adjusts the graph structure through human feedback.

3) We apply our approach to two industrial large-scale systems. Results have shown that human feedback can reduce

This work was supported by ZTE Industry-University-Institute Cooperation Funds under Grant No. 20200492.

most false positives and greatly improve detection precision.

The rest paper is organized as follows. Section 2 discusses the related work, Section 3 proposes the approach, Section 4 shows the experiment results, and Section 5 concludes this paper.

2 Related Work

2.1 Human-in-the-Loop Anomaly Detection

In existing works, incorporating human feedback into anomaly detection has been introduced. These works leverage the idea of active learning and focus on tuning the weights and scores in machine learning models. For instance, the online mirror descent (OMD) algorithm[21] associates a convex loss function to each feedback response which rewards the anomaly score. Active anomaly discovery (AAD) algorithm[22-23] defines and solves an optimization problem based on all prior feedback, which results in new weights for the model.

2.2 Log-Based Anomaly Detection

Log-based anomaly detection first parses logs into log templates based on static code analysis or clustering mechanism, and then builds anomaly detection models. These models include template frequency-based model, graph-based model, and deep learning-based model. The template frequency-based model[1-4] usually counts the number of different templates in a time window and sets up a vector for each time window. Then it utilizes methods such as machine learning algorithms to distinguish outliers.

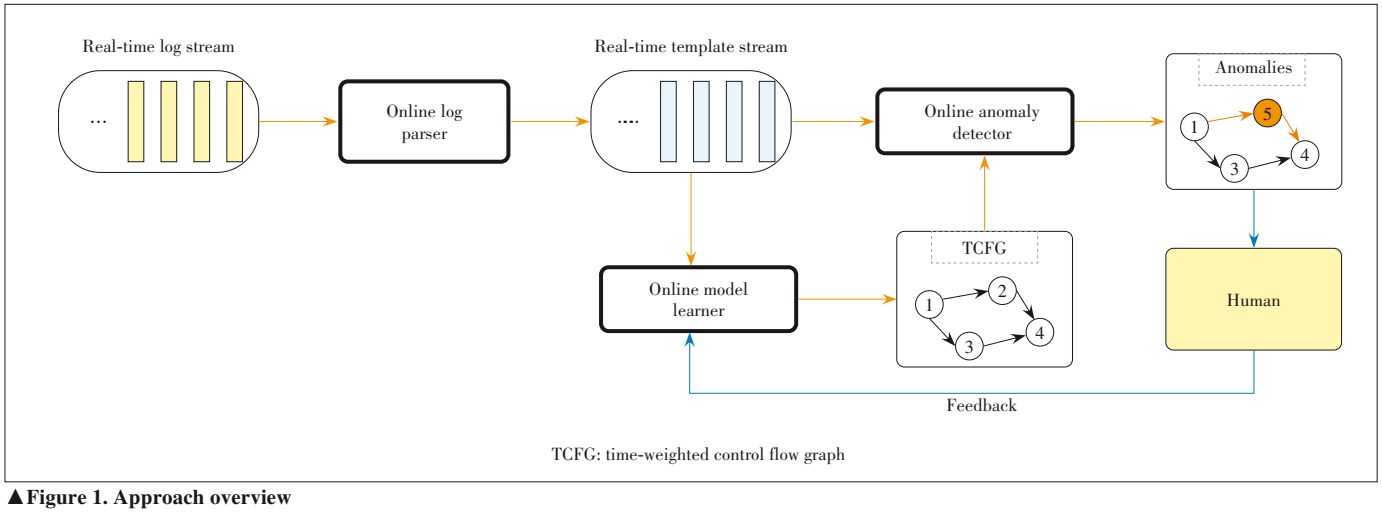
This model sacrifices the abundant information and the diagnosis ability of logs and it is not accurate and efficient. Thus it cannot provide help for problem identification and diagnosis. The graph-based model[5-17] is the current research hotspot. It extracts template sequence at first and then generates a graph-based model to compare with log sequences in the production environment to detect conflicts. This model has three advantages. First, it can diagnose problems deeply buried in log sequences, for example, performance degradation. Second, it can provide engineers with the context log messages of problems. Third, it can provide engineers with the correct log sequence and tell engineers what should have happened. The deep learning-based model[18-19] leverages long short-term memory (LSTM) to model the sequence of templates. This model takes a long time for training and inference, and thus cannot support online anomaly detection and diagnosis.

3 Approach

3.1 Overview

To solve the problems mentioned above, we design a human feedback-aware anomaly detection approach, called LogFlash, as shown in Fig. 1. The input is an online log stream $\boldsymbol{l} := (l_1, l_2, l_3, \ldots)$, which is a log entry. Our approach consists of three main components, namely the online log parser, the online model learner, and the online anomaly detector. In the online log parser, multiple log templates are mined from the log stream and each log entry is replaced by its corresponding template. In this way, the log stream is transformed into a template stream $\boldsymbol{p} := (p_1, p_2, p_3, \ldots)$. This template stream then goes through online model learner and online anomaly detector concurrently. The online model learner infers and updates a graph model called time-weighted control flow graph (TCFG) through mining the template stream. The online anomaly detector utilizes the latest TCFG model to detect anomalies in the template stream. Humans provide false positives in anomalies as feedback to the online model learner. The learner then adjusts the TCFG structure based on the feedback.

We leverage the existing online template mining algorithm[24] in the online log parser. Due to space limitations, we will only describe the TCFG model, online model learner, online anomaly detector, and human feedback loop.



▲Figure 1. Approach overview

3.2 Time-Weighted Control Flow Graph

A TCFG is a directed graph consisting of edges and nodes and each edge has a time weight recording the transition time. The TCFG model stitches together various log templates and represents the healthy state of the baseline system. It is used to flag deviations from expected behaviors at runtime. A template is an abstraction of a print statement in a source code, which manifests itself in logs with different embedded parameter values in different executions. Represented as a set of invariant keywords and parameters (denoted by parameter placeholder *), a template can be used for summarization of multiple log lines. The TCFG is such a graph where the nodes are templates and the edges represent the transition from one template to another. Besides, every log has a timestamp indicating its print time, and thus the difference between two log timestamps represents the program execution time between the two logs. The time weight on each edge in the TCFG records the longest normal transition time between two templates. If the execution time between two logs exceeds the time weight, it means the system is suffering from performance problems.

Fig. 2 shows an example of log templates and TCFG model. Each log has some invariant keywords and some variable parameters (shown in green), and log templates only reserve invariant keywords. Nodes in the TCFG are different log templates. Edges represent how each request flow passes between nodes, and the weight of edges indicates the transition time between two nodes.

3.3 Online Model Learner

We aim to construct a TCFG model in a black-box manner with only the template stream p. Our key idea is to define a dynamic pairwise transition rate $\alpha_{j,i}$ which models how frequently a request flows from template j to template i and trains/updates the transition rate $\alpha_{j,i}$ overtime with template stream p.

We further define $f\left(t_i|t_j,\alpha_{j,i}\right)$ to be the conditional likelihood of transition between template j and template i, where t_j and t_i are the timestamps of two occurrences of template j and template i in p. We assume the conditional likelihood depends on the transition time $\left(t_j,t_i\right)$ and the transition rate $\alpha_{j,i}$. To model this parametric likelihood, we first conduct a statistical analysis of the distribution of template transitions.

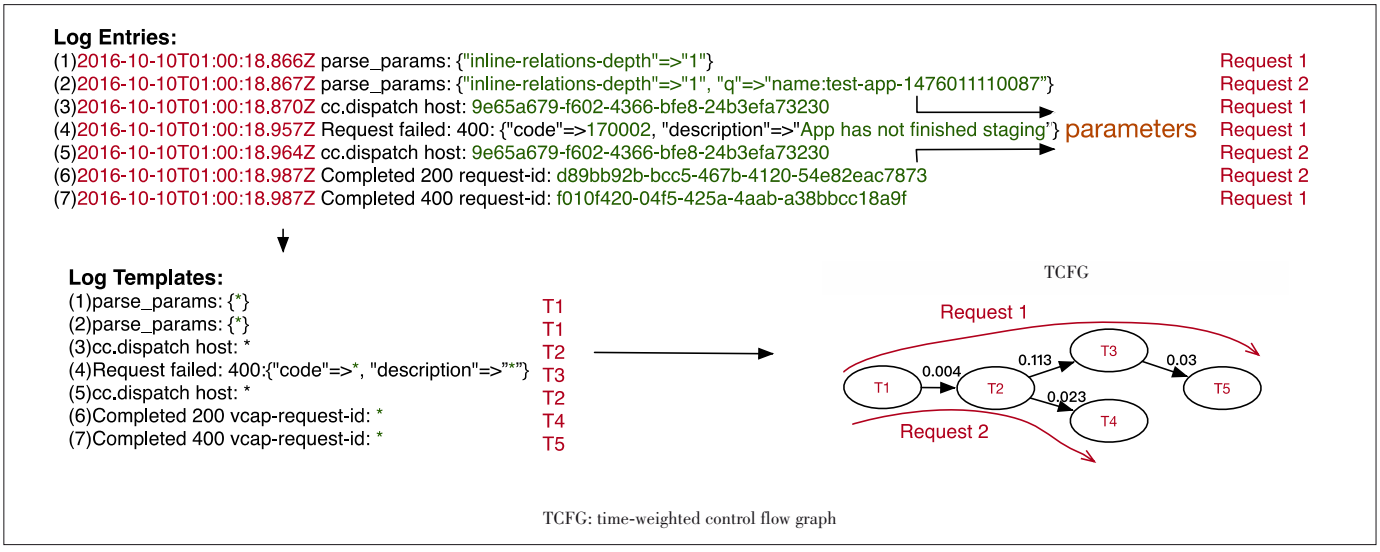
We collect system logs of 5 minutes from an industrial cloud system Ada. Then we record the transition time between every occurrence of two neighboring templates in the same request by calculating the difference of their timestamps. Next, we count the number of occurrences with the same transition time and plot the distribution of each template transition. Results are shown in Fig. 3. The distributions of these transitions show obvious long-tail distribution characteristics and the most transitions cost less than 0.2 norm-value of time.

Based on the above observations, the power-law likelihood is appropriate to model $f\left(t_i|t_j,\alpha_{j,i}\right)$, that is:

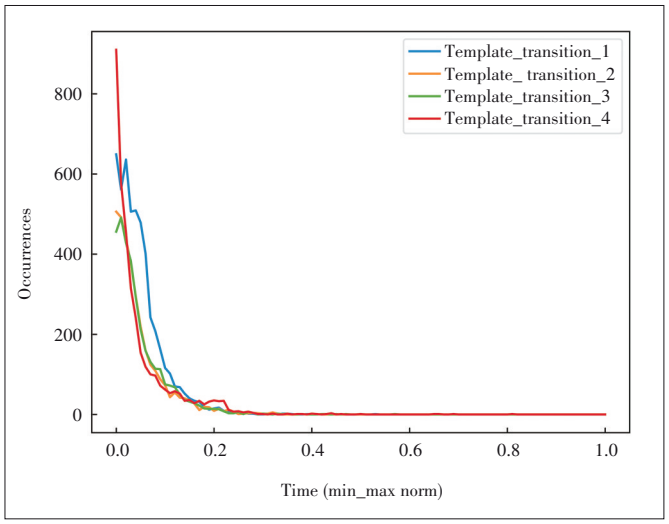
$$f\left(t_i|t_j,\alpha_{j,i}\right)=\begin{cases}\dfrac{\alpha_{j,i}}{\delta}\left(\dfrac{t_i-t_j}{\delta}\right)^{-1-\alpha_{j,i}} & \text{if } t_j+\delta<t_i\\ 0 & \text{otherwise}\end{cases}, \tag{1}$$

where δ states the minimum transition time from template j to template i. In Section 4, the power-law distribution proves to be generic enough to adapt anomaly detection methods to testing logs from diverse industrial systems. Then we apply network inference algorithm to train the structure of TCFG.

1) Template stream likelihood. In the template stream p, transitions from different templates to a certain template are independent, that is, each occurrence of template i can only be transmitted from the occurrence of one parent template. Then the likelihood of occurrence of template i at time t_i, giv-



▲Figure 2. Log templates and TCFG model



▲ **Figure 3. Template transition distributions of an industrial software system Ada**

en a collection of previous occurred templates $(t_1,...,t_N|t_k \leqslant t_i)$, results from summing over the likelihood of the mutually disjoint transition from each previously occurred template to template i:

$$f\left(t_i|t_1,...,t_{N_{t_i}},A\right)=\sum_{j:t_j<t_i} f\left(t_i|t_j,\alpha_{j,i}\right)\times\prod_{k:k\neq j,t_k<t_i} S\left(t_i|t_k,\alpha_{k,i}\right), \quad (2)$$

where $A=\{\alpha_{j,i}|i,j=1,...,N,i\neq j\}$, and $S\left(t_i|t_k,\alpha_{k,i}\right)$ is a defined survival function of transition $j\rightarrow i$ as

$$S\left(t_i|t_k,\alpha_{k,i}\right)=1-F\left(t_i|t_k,\alpha_{k,i}\right), \quad (3)$$

where $F\left(t_i|t_k,\alpha_{k,i}\right)=\int_{t_j}^{t_i} f\left(t|t_j,\alpha_{j,i}\right)dt$ is the cumulative transition density function computed from the transition likelihood.

To simplify the modeling process, we assume that transitions are conditionally independent, given a set of parent templates. The likelihood of all transitions in the template stream is

$$f\left(\boldsymbol{t}^{\leqslant T},A\right)=\prod_{t_i\leqslant T} f\left(t_i|t_1,...,t_{N_{t_i}},A\right), \quad (4)$$

where $\boldsymbol{t}^{\leqslant T}$ denotes that the time of template stream is up to T. After plugging Eq. (2) into Eq. (4) and removing the condition $k\neq j$, the product result is independent of j:

$$f\left(\boldsymbol{t}^{\leqslant T},A\right)=\prod_{i:t_i\leqslant T}\prod_{k:t_k<t_i} S\left(t_i|t_k,\alpha_{k,i}\right)\times\sum_{j:t_j<t_i}\frac{f\left(t_i|t_j,\alpha_{j,i}\right)}{S\left(t_i|t_j,\alpha_{j,i}\right)}. \quad (5)$$

The fact that some templates are not shown in the observation window is also informative. We therefore add multiplicative survival terms to Eq. (5) and rearrange it with hazard function[25] or instantaneous transition rate of transition $j\rightarrow i$ as $H\left(t_i|t_j,\alpha_{j,i}\right)=f\left(t_i|t_j,\alpha_{j,i}\right)/S\left(t_i|t_j,\alpha_{j,i}\right)$. Then the likelihood of the template stream is reformulated as

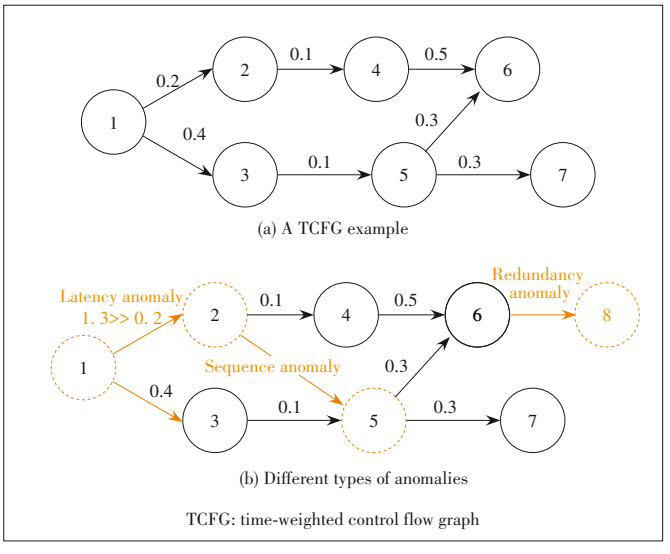
$$f\left(\boldsymbol{t},A\right)=\prod_{i:t_i\leqslant T}\prod_{m:t_m>T} S\left(T|t_i,\alpha_{i,m}\right)\times\prod_{k:t_k<t_i} S\left(t_i|t_k,\alpha_{k,i}\right)\left(\sum_{j:t_j<t_i} H\left(t_i|t_j,\alpha_{j,i}\right)\right). \quad (6)$$

2) TCFG structure inference problem. Our purpose is to infer a TCFG structure that is most possible to generate the template stream p. Given a TCFG with constant edge transition rate A, the TCFG structure inference problem problem reduces to solving a maximum likelihood problem:

$$\begin{aligned}&\text{maximize}_A \quad \log f\left(t,A\right)\\&\text{subject to} \quad \alpha_{j,i}\geqslant 0, i,j=1,...,N,i\neq j\end{aligned}, \quad (7)$$

where $A=\{\alpha_{j,i}|i,j=1,...,N,i\neq j\}$ are the edge transitions we aim to train. The edges in TCFG are those pairs of templates with transition rates $\alpha_{j,i}\geqslant 0$.

To support online model update, we generalize the inference problem to dynamic TCFG structure with edge transition rates $A(t)$ that may change over time. To this aim, we first split the template stream p to a set of sub-streams $\boldsymbol{c}=\left(c_1,c_2,c_3,...\right)$ based on the arrival of new templates. Given a time window size w, each time a template i arrives, we split out a sub-stream in which i is the latest template. An example is shown in Fig. 4. At time t_1, log stream in the red block is the current sub-stream. At time t_2, a new template T_2 is observed and the current sub-stream becomes $\{T_3,T_4,T_3,T_2\}$. When it comes to time t_3 when T_5 is observed, the current sub-stream becomes $\{T_4,T_3,T_2,T_5\}$. In this way, at any given



▲ **Figure 4. A TCFG example and different types of anomalies**

time t, we solve the maximum likelihood problem over the set of sub-streams:

$$\begin{aligned} &\text{maximize}_{A(t)} \quad \sum_{c \in \boldsymbol{c}} f(t^c, A(t)) \\ &\text{subject to} \quad \alpha_{j,i}(t) \geqslant 0, i,j = 1,...,N, i \neq j, \end{aligned} \tag{8}$$

where $c \in \boldsymbol{c}$. Next, we show how to efficiently solve the above optimization problem for all time points t.

3) Training method. The problem defined by Eq. (8) is serious for the power-law transition model. Therefore, we aim to find optimal training solution at any given time t. Since in the condition of power-law model, the edge transition rates usually vary smoothly, classical stochastic gradient descent[26] can be a perfect method for our training as we can use the inferred TCFG structure from the previous time step as initialization for the inference procedure in the current time step. The training phase uses iterations of the form:

$$\alpha_{j,i}^k(t) = \left(\alpha_{j,i}^{k-1}(t) - \gamma \nabla_{\alpha_{j,i}} L_c\left(A^{k-1}(t)\right)\right)^+, \tag{9}$$

where k is the iteration number, $\nabla_{\alpha_{j,i}} L_c(\cdot)$ is the gradient of the log-likelihood $L_c(\cdot)$ of sub-stream c with respect to the edge transition rate $\alpha_{j,i}$, γ is the update step size, and $(z)^+ = \max(0,z)$. The computations of log survival function, hazard function and gradient of sub-stream c for power-law model in Eq. (1) are given in Table 1.

Importantly, in each iteration of the training phase, we only need to compute the gradients $\nabla_{\alpha_{j,i}} L_c\left(A^k\right)$ for edges between template j and template i, as node j has been observed in sub-stream $\boldsymbol{c}$, and the iteration cost and convergence rate are independent of $|c|$.

3.4 Online Anomaly Detector

The basic idea for anomaly detection is to compare the log stream with TCFG to find the deviation. We first define three types of deviations/anomalies, namely sequence anomaly, redundancy anomaly, and latency anomaly. A sequence anomaly is raised when the log that follows the occurrence of a parent node cannot be mapped to any of its children. A redundancy anomaly is raised when unexpected logs that cannot be mapped to any node in the TCFG occur. A latency anomaly is raised when the child of a parent node is seen but the transition time exceeds the time weight recorded on the edge. Fig. 4 shows an example of different types of anomalies. Fig. 4(a) is an example of TCFG with 7 nodes. As shown in Fig. 4(b), suppose the transition time between Node 1 and Node 2 exceeds the time weight 0.2, they suffer from a latency anomaly. Node 5 appears after Node 2 unexpectedly and suffers from a sequence anomaly. Node 8 appears after Node 6 while Node 8 is a new template that has not been recorded in the TCFG, and thus a redundancy anomaly occurs.

▼**Table 1. Computations of transition likelihood for power-law model**

Computation Entity	Computation Method
Log survival function: $\log S\left(t_i \mid t_k, \alpha_{k,i}\right)$	$-\alpha_{j,i} \log\left(\frac{t_i - t_j}{\delta}\right)$
Hazard function: $H\left(t_i \mid t_j, \alpha_{j,i}\right)$	$\alpha_{j,i} \cdot \frac{1}{t_i - t_j}$
Gradient for unobserved ones in $\boldsymbol{c}$: $\nabla_{\alpha_{j,i}} L_c(A)$	$\log\left(\frac{T - t_j^c}{\delta}\right)$
Gradient for observed ones in $\boldsymbol{c}$: $\nabla_{\alpha_{j,i}} L_c(A)$	$\log\left(\frac{t_i^c - t_j^c}{\delta}\right) - \frac{(t_i^c - t_j^c)^{-1}}{\sum_{k: t_k^c < t_i^c} \alpha_{k,i} (t_i^c - t_k^c)^{-1}}$

3.5 Human Feedback Handling

As mentioned before, users report false positives in anomalies through detection results webpage as human feedback. The online model learner receives the feedback and adjust the TCFG based on the feedback. For different types of anomalies, the online model learner takes different operations. These operations are shown in Algorithm 1.

Algorithm 1. Human Feedback Handling Algorithm

Input: Human feedback **Anomaly**.
Definition: TCFG denotes the current TCFG model
1. **if Anomaly**. type = "Sequence"
2. **then TCFG**. addEdge (**Anomaly**. parentNode, **Anomaly**. childNode)
3. **if Anomaly**. type = "Redundancy"
4. **then TCFG**. addNode (**Anomaly**. redundantNode)
5. **if Anomaly**. type = "Latency"
6. **then TCFG**. setTimeWeight (**Anomaly**. parentNode, **Anomaly**. childNode, **Anomaly**. transitionTime)

• Sequence anomaly. A sequence anomaly is raised when the log that follows the occurrence of a parent node cannot be mapped to any of its children. If a sequence anomaly is a false positive, it means the log that follows the occurrence of the parent node should be its child. In other words, the transition between the parent and the child has not been correctly learned. For instance, in Fig. 4(b) Node 5 appears after Node 2 unexpectedly and suffers from a sequence anomaly. If it is a false positive, it means there should be a transition edge from Node 2 to Node 5. Therefore, the online model learner takes the operation to add a transition edge from the parent and the child in TCFG.

• Redundancy anomaly. A redundancy anomaly is raised when unexpected logs occur that cannot be mapped to any node in TCFG. If a redundancy anomaly is a false positive, it means the template of the unexpected log should be in TCFG. For instance, in Fig. 4(b) Node 8 appears unexpectedly and raises a redundancy anomaly. If it is a false positive, it means Node 8 should be in the path. Therefore, online model learners take the operation to add the template of the unexpected log to TCFG.

• Latency anomaly. The time weight on each edge in the

TCFG records the longest normal transition time between two nodes. An intuitive way to determine time weight is to update the time weight once it meets a longer transition time in the log stream. However, it is hard to determine whether the longer transition time is a real latency anomaly or normal latency fluctuation. Therefore, we rely on human feedback to update the time weight. If a latency anomaly is a false positive, it means the time weight is too small and should be updated. Therefore the online model learner updates the time weight once it receives feedback of latency anomaly. For instance, in Fig. 4(b) the transition time between Node 1 and Node 2 exceeds the time weight 0.2, and then they suffer a latency anomaly. If it is a false positive, it means the transition time 1.3 is normal. Therefore, the online model learner takes the operation to update the time weight from 0.2 to 1.3.

4 Experiment and Evaluation

4.1 Experiment Setup

We test our approach on three industrial large-scale systems called Ada, Bob, and Dockerd. Ada is an online image identification and analytics system that serves thousands of users. Bob is a software distribution system for 5G stations and chipboards. It distributes upgrade or bug fixing patches to thousands of 5G chipboards. We collect system logs of two days from Ada and Bob and use logs for training. Dockerd is a component of a Platform as a Service (PaaS) platform which contains 10 components and 957 nodes producing more than 8.1 million system logs per day. We collect logs of 20 days with a size of 52.94 G to verify the effectiveness of our approach.

We choose the state-of-the-art log-based anomaly detection DeepLog[18] and LogSed[6] as baselines. DeepLog leverages LSTM to model template sequences and detect anomalies through computing the distance between observed templates and predicted templates. LogSed first proposes TCFG model and infers the TCFG model based on frequent sequence mining.

We use typical Recall and Precision as our evaluation metrics, which are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{10}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{11}$$

where TP, FP, TN, FN are referred to as true positive, false positive, true negative, and false negative.

4.2 Evaluation Results

Logs in real-world industrial systems are much more complex and heterogeneous than lab systems, and it is very hard for today's anomaly detectors to produce satisfying results. Table 2 shows the evaluation results of Ada. Both LogSed and DeepLog show poor precision which means they produce lots of false positives. Our approach outputs 74 anomalies in which 31 anomalies are true positives leading to a precision of 0.42 without human feedback (statistics in parentheses). After we incorporate human feedback, our approach produces 36 anomalies. We also record the times of human feedback during training. Experts labeled 28 false positives as feedback to guide the system and the labeling task only costs about 5 minutes.

Bob is even more complex than Ada. Each 5G station and chipboard are in different environments with different network status, load status, etc. Logs of many processes such as network test, heartbeat, software download, reconnect, software security and consistency check are interleaved together. It is almost impossible for existing detectors to learn a usable model from such noisy system logs. As shown in Table 3, LogSed and DeepLog show a precision of 0.04 and 0.09 separately. Our approach detects 137 anomalies without human feedback in which 10 anomalies are true positives. After incorporating human feedback, the number of detected anomalies reduces to 13 leading to 0.77 precision. During training, experts label 52 false positives as feedback in total that costs about 15 minutes.

As evaluating the performance of the framework on Dockerd logs, our approach detects 1 515 sequence anomalies without human feedback, of which less than 160 are true positives. After dropping duplicates and incorporating human feedback, the accuracy rate increases to 0.82. In summary, our approach achieves much better precision than baseline works. Incorporating human feedback effectively reduces false positives and significantly improves model performance. Besides, the feedback process is very easy for experts and saves a lot of time.

5 Conclusions and Future Work

In this paper, we propose a feedback-aware anomaly detection approach. It builds a TCFG model to describe normal system status and incorporates human feedback to adjust the graph structure to reduce false positives. Experiment results on two industrial large-scale systems show that our approach enjoys much better precision than baselines. Besides, human feedback can significantly reduce false positives and improve model performance.

▼Table 2. Evaluation results of Ada

Approaches	Precision/%	Recall/%	#Human Feedback
LogSed[18]	0.34	1.00	/
DeepLog[6]	0.45	1.00	/
Our approach	0.86(0.42)	1.00	28

▼Table 3. Evaluation results of Bob

Approaches	Precision/%	Recall/%	#Human Feedback
LogSed[18]	0.04	0.89	/
DeepLog[6]	0.09	0.99	/
Our approach	0.77(0.07)	0.96	52

In the future, we will improve the human feedback handling process and perform more sophisticated tuning on the model with human feedback.

References

[1] LOU J G, FU Q, YANG S Q, et al. Mining invariants from console logs for system problem detection [C]//USENIX Annual Technical Conference. Berkeley, USA: USENIX, 2010
[2] OLINER A J, AIKEN A. Online detection of multi-component interactions in production systems [C]//2011 IEEE/IFIP 41st International Conference on Dependable Systems & Networks (DSN). Hong Kong, China: IEEE, 2011: 49 – 60. DOI: 10.1109/DSN.2011.5958206
[3] CHEN C, SINGH N, YAJNIK S. Log analytics for dependable enterprise telephony [C]//2012 Ninth European Dependable Computing Conference. Sibiu, Romania: IEEE, 2012: 94 – 101. DOI: 10.1109/EDCC.2012.14
[4] DU S Z, JIAN C. Behavioral anomaly detection approach based on log monitoring [C]//2015 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC). Nanjing, China: IEEE, 2015: 188 – 194. DOI:10.1109/BESC.2015.7365981
[5] NANDI A, MANDAL A, ATREJA S, et al. Anomaly detection using program control flow graph mining from execution logs [C]//The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA: ACM, 2016: 215 – 224. DOI:10.1145/2939672.2939712
[6] JIA T, YANG L, CHEN P F, et al. LogSed: anomaly diagnosis through mining time-weighted control flow graph in logs [C]//2017 IEEE 10th International Conference on Cloud Computing (CLOUD). Honolulu, USA: IEEE, 2017: 447 – 455. DOI:10.1109/CLOUD.2017.64
[7] JIA T, CHEN P F, YANG L, et al. An approach for anomaly diagnosis based on hybrid graph model with logs for distributed services [C]//2017 IEEE International Conference on Web Services (ICWS). Honolulu, USA: IEEE, 2017: 25 – 32. DOI:10.1109/ICWS.2017.12
[8] FU Q, LOU J G, WANG Y, et al. Execution anomaly detection in distributed systems through unstructured log analysis [C]// The 9th IEEE International Conference on Data Mining. Miami Beach, USA: IEEE, 2009: 149 – 158. DOI: 10.1109/ICDM.2009.60
[9] BABENKO A, MARIANI L, PASTORE F. AVA: automated interpretation of dynamically detected anomalies [C]//The 18th International Symposium on Software Testing and Analysis. Chicago, USA: ISSTA, 2009: 237 – 248. DOI: 10.1145/1572272.1572300
[10] YEN T F, OPREA A, ONARLIOGLU K, et al. Beehive: large-scale log analysis for detecting suspicious activity in enterprise networks [C]//The Annual Computer Security Applications Conference. New Orleans, USA: ACM, 2013: 199 – 208. DOI:10.1145/2523649.2523670
[11] ZHAOX, Y.ZHANG, LIOND, et al. lprof: a non-intrusive request flow profiler for distributed systems [C]//Usenix Symposium on Operating System Implementation & Design. Broomfield, USA: OSDI, 2014, 629 – 644
[12] YU X, JOSHI P, XU J W, et al. CloudSeer [J]. ACM SIGPLAN notices, 2016, 51(4): 489 – 502. DOI:10.1145/2954679.2872407
[13] TAK B C, TAO S, YANG L, et al. LOGAN: problem diagnosis in the cloud using log-based reference models [C]//2016 IEEE International Conference on Cloud Engineering (IC2E). Berlin, Germany: IEEE, 2016: 62 – 67. DOI: 10.1109/IC2E.2016.12
[14] AALST WVAN DER, WEIJTERS T, MARUSTER L. Workflow mining: discovering process models from event logs [J]. IEEE transactions on knowledge and data engineering, 2004, 16(9): 1128 – 1142. DOI:10.1109/TKDE.2004.47
[15] LOU J G, FU Q, YANG S Q, et al. Mining program workflow from interleaved traces [C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge discovery and data mining. Washington, USA: ACM, 2010: 613 – 622. DOI:10.1145/1835804.1835883
[16] YUAN D, MAI H H, XIONG W W, et al. SherLog [J]. ACM SIGARCH computer architecture news, 2010, 38(1): 143 – 154. DOI:10.1145/1735970.1736038
[17] FU Q, LOU J G, LIN Q W, et al. Contextual analysis of program logs for understanding system behaviors [C]//The 2013 10th Working Conference on Mining Software Repositories (MSR). San Francisco, USA: IEEE, 2013: 397 – 400. DOI:10.1109/MSR.2013.6624054
[18] DU M, LI F F, ZHENG G N, et al. DeepLog: anomaly detection and diagnosis from system logs through deep learning [C]//The 2017 ACM SIGSAC Conference on Computer and Communications Security. Dallas, USA: ACM, 2017: 1285 – 1298. DOI:10.1145/3133956.3134015
[19] MENG W B, LIU Y, ZHU Y C, et al. LogAnomaly: unsupervised detection of sequential and quantitative anomalies in unstructured logs [C]//The 28th International Joint Conference on Artificial Intelligence. Macao, China: IJCAI, 2019: 4739 – 4745. DOI:10.24963/ijcai.2019/658
[20] LIN Q W, ZHANG H Y, LOU J G, et al. Log clustering based problem identification for online service systems [C]//Proceedings of the 38th International Conference on Software Engineering Companion. Austin, USA: ACM, 2016: 102 – 111. DOI:10.1145/2889160.2889232
[21] SIDDIQUI M A, FERN A, DIETTERICH T G, et al. Feedback-guided anomaly discovery via online optimization [C]//The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, United Kingdom: ACM, 2018: 2200 – 2209. DOI:10.1145/3219819.3220083
[22] DAS S, WONG W K, DIETTERICH T , et al. Incorporating expert feedback into active anomaly discovery [C]//16th International Conference on Data Mining (ICDM): IEEE, 2017, 853 – 858
[23] DAS S, WONG W K, FERN A, et al. Incorporating feedback into tree-based anomaly detection [EB/OL]. [2021-01-20]. https://arxiv.org/abs/1708.09441
[24] HE P J, ZHU J M, ZHENG Z B, et al. Drain: an online log parsing approach with fixed depth tree [C]//2017 IEEE International Conference on Web Services (ICWS). Honolulu, USA: IEEE, 2017: 33 – 40. DOI: 10.1109/ICWS.2017.13
[25] GOMEZ RODRIGUEZ M, LESKOVEC J, SCHÖLKOPF B. Structure and dynamics of information pathways in online media [C]//The sixth ACM international conference on Web search and data mining - WSDM'13. Rome, Italy: ACM, 2013: 23 – 32. DOI: 10.1145/2433396.2433402

Biographies

HAN Jing (han.jing28@zte.com.cn) received her master's degree from Nanjing University of Aeronautics and Astronautics, China. She has been with ZTE Corporation since 2000, where she had been engaged in 3G/4G key technologies from 2000 to 2016. She has become a technical director responsible for intelligent operation of cloud platforms and wireless networks since 2016. Her research interests include machine learning, data mining, and signal processing.

JIA Tong is a doctoral researcher at Department of Computer Science and Technology, Peking University, China. His research interests include distributed computing and algorithmic IT operations.

WU Yifan is pursuing his doctorate at School of Software and Microelectronics in Peking University, China. His research mainly focuses on distributed computing and algorithmic IT operations.

HOU Chuanjia is pursuing his master's degree at School of Software and Microelectronics in Peking University, China. His research focuses on algorithmic IT operations.

LI Ying is a researcher at National Engineering Research Center for Software Engineering, Peking University, China. She is also a professor of School of Software and Microelectronics, Peking University. Her research interests include distributed computing and trusted computing.

ZTE Communications Guidelines for Authors

Remit of Journal

ZTE Communications publishes original theoretical papers, research findings, and surveys on a broad range of communications topics, including communications and information system design, optical fiber and electro-optical engineering, microwave technology, radio wave propagation, antenna engineering, electromagnetics, signal and image processing, and power engineering. The journal is designed to be an integrated forum for university academics and industry researchers from around the world.

Manuscript Preparation

Manuscripts must be typed in English and submitted electronically in MS Word (or compatible) format. The word length is approximately 3 000 to 8 000, and no more than 8 figures or tables should be included. Authors are requested to submit mathematical material and graphics in an editable format.

Abstract and Keywords

Each manuscript must include an abstract of approximately 150 words written as a single paragraph. The abstract should not include mathematics or references and should not be repeated verbatim in the introduction. The abstract should be a self-contained overview of the aims, methods, experimental results, and significance of research outlined in the paper. Five carefully chosen keywords must be provided with the abstract.

References

Manuscripts must be referenced at a level that conforms to international academic standards. All references must be numbered sequentially in-text and listed in corresponding order at the end of the paper. References that are not cited in-text should not be included in the reference list. References must be complete and formatted according to *ZTE Communications Editorial Style*. A minimum of 10 references should be provided. Footnotes should be avoided or kept to a minimum.

Copyright and Declaration

Authors are responsible for obtaining permission to reproduce any material for which they do not hold copyright. Permission to reproduce any part of this publication for commercial use must be obtained in advance from the editorial office of *ZTE Communications*. Authors agree that a) the manuscript is a product of research conducted by themselves and the stated co-authors; b) the manuscript has not been published elsewhere in its submitted form; c) the manuscript is not currently being considered for publication elsewhere. If the paper is an adaptation of a speech or presentation, acknowledgement of this is required within the paper. The number of co-authors should not exceed five.

Content and Structure

ZTE Communications seeks to publish original content that may build on existing literature in any field of communications. Authors should not dedicate a disproportionate amount of a paper to fundamental background, historical overviews, or chronologies that may be sufficiently dealt with by references. Authors are also requested to avoid the overuse of bullet points when structuring papers. The conclusion should include a commentary on the significance/future implications of the research as well as an overview of the material presented.

Peer Review and Editing

All manuscripts will be subject to a two-stage anonymous peer review as well as copyediting, and formatting. Authors may be asked to revise parts of a manuscript prior to publication.

Biographical Information

All authors are requested to provide a brief biography (approx. 100 words) that includes email address, educational background, career experience, research interests, awards, and publications.

Acknowledgements and Funding

A manuscript based on funded research must clearly state the program name, funding body, and grant number. Individuals who contributed to the manuscript should be acknowledged in a brief statement.

Address for Submission

http://mc03.manuscriptcentral.com/ztecom

ZTE COMMUNICATIONS

中兴通讯技术(英文版)

ZTE Communications has been indexed in the following databases:

- Abstract Journal
- Cambridge Scientific Abstracts (CSA)
- China Science and Technology Journal Database
- Chinese Journal Fulltext Databases
- Index of Copurnicus
- Inspec
- Ulrich's Periodicals Directory
- Wanfang Data

ZTE COMMUNICATIONS
Vol. 19 No. 3 (Issue 75)
Quarterly
First English Issue Published in 2003

Supervised by:
Anhui Publishing Group

Sponsored by:
Time Publishing and Media Co., Ltd.
Shenzhen Guangyu Aerospace Industry Co., Ltd.

Published by:
Anhui Science & Technology Publishing House

Edited and Circulated (Home and Abroad) by:
Magazine House of ZTE Communications

Editorial Correspondence:
Add: 12F Kaixuan Building, 329 Jinzhai Road,
Hefei 230061, P. R. China
Tel: +86-551-65533356
Email: magazine@zte.com.cn
Website: http://zte.magtechjournal.com

Annual Subscription: RMB 80
Printed by:
Hefei Tiancai Color Printing Company
Publication Date: September 25, 2021

China Standard Serial Number: ISSN 1673-5188
CN 34-1294/TN