

车辆属性识别及跨场景标注方案

A Solution for Vehicle Attributes Recognition and Cross-Scene Annotation

董振江/DONG Zhenjiang¹
高燕/GAO Yan¹
吴文熙/WU Wenxi²

(1. 中兴通讯股份有限公司, 广东 深圳 518057)

(2. 中山大学, 广东 广州 510006)

(1. ZTE Corporation, Shenzhen 518057, China)

(2. Sun Yat-Sen University, Guangzhou 510006, China)

智能交通^[1-2]领域中牵涉到很多计算机视觉的任务,需要对海量的视频、图像数据进行处理、分析。在基于车辆视频图像的分析任务中,传统技术不仅效率较低,耗费巨大的人力、物力、财力,而且欠缺对数据的感知表达能力,没有充分利用数据、挖掘数据特性,以致分析精度不够高。引入车辆属性识别的方法,可以增强算法对目标车辆特征的理解和表达能力,大大提高分析任务的可靠性和效率。

随着计算机视觉和机器学习技术的发展,以它们为技术基础的新一代智能分析技术,在目标检测、识别、检索等任务中表现优越,能够极大地减少人力成本,提高监测控制的效率,有巨大的应用潜力。

自2012年ImageNet^[3]国际大规模视觉识别比赛以来,业界掀起了一波基于计算机视觉和深度学习的识别技术研究和应用热潮。深度学习是一门大数据驱动的技术,我们通常认

收稿日期: 2017-05-25
网络出版日期: 2017-07-04

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0020-005

摘要: 提出采用一种基于深度学习的识别方法,来辅助获取训练模型中所需要的车辆属性标签。该方法首先利用海报图像构建大规模车辆属性数据集(SYSU-ZTE-CARS),训练基于卷积神经网络(CNN)的识别模型,再将模型迁移到监测控制场景中进行标注测试,间接获取属性标签。采用CNN+softmax分类器的结构作为基本框架,引入细粒度识别技术以进一步优化识别性能。利用5种常用车辆属性进行测试,实验结果表明:所提出的方案不仅在SYSU-ZTE-CARS数据集上的识别精度高,而且在监测控制场景下的标注结果也很可靠。

关键词: 精细化属性识别; CNN; 深度学习; 计算机视觉

Abstract: The deep learning-based solution is proposed to obtain the vehicle attributes data in recognizer training. Firstly, with poster images collected, a large scale vehicle attributes dataset—SYSU-ZTE-CARS, is built to train the convolutional neural network(CNN)-based recognition model. Then the model is moved to the monitor and control scene to mark test, and access attribute indirectly. CNN + softmax classifier is used as the basic framework, and the fine-grained identification is also used to further optimize the recognition performance. The experimental results on recognition of five different attributes show that our solution is not only with high accuracy in SYSU-ZTE-CAR dataset, but also reliable in auto-annotation under real-world surveillance.

Keywords: fine-grained attributes recognition; CNN; deep learning; computer vision

为:提供训练的数据量越大,所训练得到的模型性能越好;而在视频监控控制的应用场景下,海量的视频图像数据会源源不断地产生,因此将深度学习技术应用在视频监控控制领域,恰能发挥其先天优势以及视频数据的作用。

基于深度学习的属性识别技术,通常以卷积神经网络(CNN)为基本工具。CNN模型参数的学习,目前使用最广泛也是最稳定的方法是有监督的训练策略。这种策略需要使用大量车辆图片及其属性标签作为训练样本。目前面临最大的挑战是:在现实监测控制场景下的车辆属性标

签很难获取。虽然有些属性标签可以通过人工标注的方式获取,如颜色、视角、车类型、品牌等,但是这种方式同时也带来了新的问题,如:颜色易受光照影响,在极端光照环境下,人眼无法正确判别车辆颜色;品牌往往需要通过车标来确定,一旦车标模糊、被遮挡、不可见,则需要通过人眼来判别车辆品牌。

另一方面,我们注意到:在互联网上有大量带有属性标签的车辆海报图片。为了获取监测控制场景下的车辆属性标签,我们考虑了一种跨场景自动标注的技术方案。该方案中,我们先利用带有属性标签的车辆

海报图片训练属性识别模型,再将此模型迁移到现实的监测控制场景中,自动标注出车辆属性。

1 车辆属性识别的相关工作

1.1 车辆属性识别的研究

目前,对于车辆属性的研究工作和应用都已经相当广泛,主要集中于视角识别、颜色识别、车型识别、品牌识别等几个方面。

在视角识别方面,学术界主要在KITTI基准测试平台^[1]上进行。KITTI上的研究绝大部分是基于CNN的,其识别效果精度都非常高,有些优秀的工作如多尺度CNN(MSCNN)^[2]、循环滚动卷积(RRC)^[3]等,以及一些利用三维信息的工作,能够在直接预测车辆偏转角度的具体数值的任务上有可靠的性能表现。

在颜色识别方面,研究难点在于:光照影响对颜色判断的干扰特别严重。目前主要的研究多是基于强判别性区域辅助的方式,具体实现手段主要包括利用关键点、利用局部分割等。在颜色特征表达上,主要有基于阈值分割、颜色直方图和CNN等。

在车型识别方面,可以分成两大类:基于非视觉的方法和基于视觉的方法。基于非视觉的方法,业界通常采用如雷达、超声波、红外线等传感器信号开展研究应用;基于视觉的方法则通过通用摄像头采集视频图像数据,再通过计算机视觉技术进行识别。目前的研究主要有基于直方图特征、低维投影特征和稀疏表示特征等传统方法。

在品牌与子品牌识别方面,研究成果同样也可以分为传统方法和深度学习方法两大类。传统方法主要是通过传统特征工程的手段人工设计特征,如Curvelet变换、方向梯度直方图(HOG)特征、分层梯度直方图(PHOG)特征、边缘方向直方图(EOH)特征、Gabor小波等,也有基于融合特征进行稀疏编码的解决方

案。深度学习方法,主要是基于CNN特征对属性进行判别的,也有一些针对细粒度识别的工作,如Part-based R-CNN模型、基于分割图辅助的CNN模型、双线性CNN模型等。

总体来说,基于传统方法的车辆识别技术都有特征设计过程反复、特征表达能力鲁棒性差、模型泛化能力差的缺点;而基于深度学习的车辆属性识别技术,则具有强大的数据描述能力,且在泛化性、识别精度、稳定性上都比传统方法更胜一筹。

1.2 CNN网络的研究

属性识别往往通过分类技术实现,而分类问题是作为基于CNN的视觉识别中一项最基本任务,其性能最主要体现在CNN网络结构之上,CNN网络的发展直接就能够反映分类精度变化。

在大规模视觉识别挑战(ILSVRC)2012比赛中,AlexNet的成功再次掀起了学者们对CNN的研究热潮。CNN结构上的发展体现在两个主要的维度上——深度和广度。在深度上,从8层的AlexNet,到16层的VGG^[7]、24层的GoogLeNet,再到152层的ResNet。在Cifar数据集上,ResNet还成功实现了极深网络,层数达到1 001层。在2016年的ImageNet比赛中,商汤科技-香港中文大学团队更是使用了1 200层的极深网络。在广度上,GoogLeNet^[8]首先设计出Inception模块,在同一层网络里使用了更多子网络。

研究表明:CNN的层数越深,网络表达能力越强。文章中,我们基于ResNet系列的网络结构开展实验,同时考虑到任务难度,也采用了轻量级网络,如AlexNet;兼顾精度和测试效率,我们也采用了中等深度网络,如VGG16。

1.3 相关数据集的情况

关于车辆属性图片数据集,目前比较有代表性的主要有CompCars^[9]、

BoxCars^[10]、城镇交通运输监测控制系统数据集(UTS)^[11]。

CompCars数据集由香港中文大学所提出,是目前学术界规模最大的车辆精细化属性数据集。数据集分为两部分:一部分来自于各大汽车门户网站中的图片;另一部分来自于道路监测控制视频截取出的图片。第1部分收集161个汽车品牌、1 687个子品牌汽车,子品牌下还细分了汽车的生产年份,其中整车图片有136 727张,汽车部分细节图片有27 618张,这些图片涵盖车辆的5种视角、12种类型;第2部分有50 000张汽车正面的监测控制图像,同样标注了车型、品牌和子品牌,除此之外,第2部分数据还标注了10种颜色。BoxCars数据集也是汽车海报图片,标注内容和方式与CompCars相似,但其规模远不如CompCars。UTS数据集收集的是监测控制场景下的原始图片,是目前在该场景下标注内容最丰富、最贴近现实应用场景的图片。UTS数据集共3 600张图片,涵盖了6种场景,在每种场景内包含多辆车,它的标注内容包括车辆位置、5种视角、9种颜色和6种常见车型。

虽然UTS数据集非常贴近现实应用场景,但是由于标注量较小、属性丰富性欠缺等原因,使得无法训练得到泛化能力较强的属性识别模型。然而,CompCars的属性标签也未与现实场景完全吻合,如CompCars并没有覆盖到如“货车”“公交车”等常见车型,其第2部分数据虽然来自于卡口监测控制场景,但视角单一。为了弥补CompCars和UTS数据集的不足,文中我们自行构建了大规模的车辆属性数据集SYSU-ZTE-CARS。

2 基于深度学习的车辆属性识别方案

2.1 基于CNN的基本技术框架

基于深度学习的车辆属性识别方案,主要以CNN作为基本工具。

CNN是一种参数可学习的模型,通常是利用反向传播(BP)算法对网络参数进行微调更新,有着数据拟合能力强,训练简便,支持端到端的训练和测试的特点。在属性识别技术方案(如图1所示)中,CNN的作用是作为特征提取器,对输入的图片/视频帧进行特征表达,得到其具有较强判别性的特征向量;再利用分类器对特征向量进行分类,以此实现对图片/视频帧中车辆属性的识别。

文中所提出的方案选用了 AlexNet^[12]、VGG16^[7]、ResNet^[13]等经典CNN模型,利用有监督学习的方法,训练了多种属性识别网络模型,对不同属性的识别性能进行了测试评估和分析。

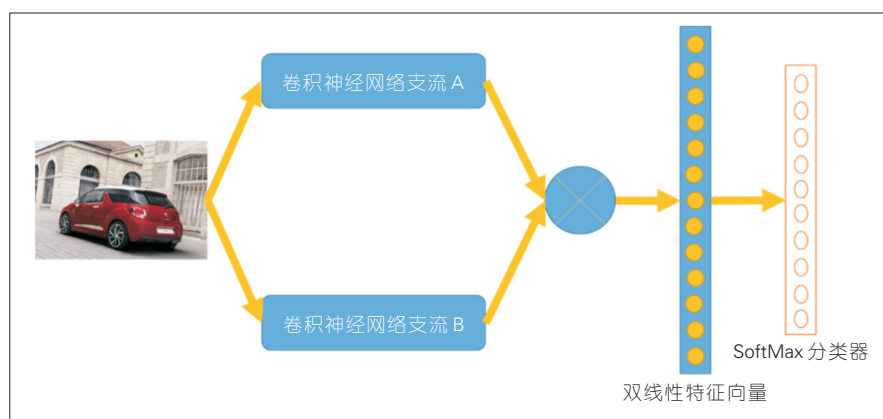
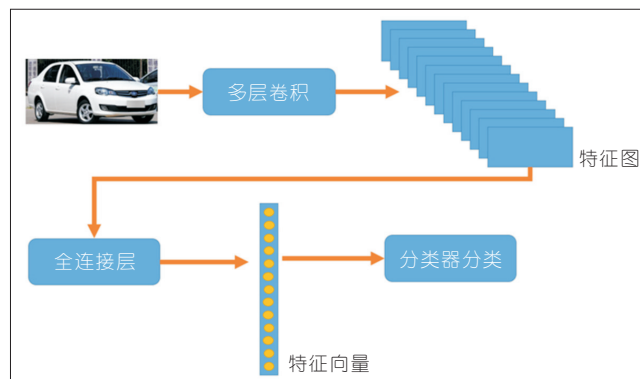
2.2 基于细粒度的识别优化方案

属性识别,本质上是类内区分。对于类内判别性较强的属性,如视角、颜色、车类型,直接采用“CNN+分类器”的方法已能够获得比较好的识别效果;但对于类内区分度不明显的属性,如品牌和子品牌,我们通常需要使用基于细粒度的识别技术来提高识别的精确度。文中,我们采用双线性CNN(B-CNN)^[14]的方法,使用B-CNN网络框架(如图2所示),增强特征向量对于细粒度分类识别的表达能力。

B-CNN延续了一般物体分类的做法,即通过CNN提取特征,然后用向量表示特征,最后用分类器分类向量。一般物体分类的做法是用全连接层将特征图转化为一条特征向量,作者认为这种做法导致特征向量的辨别程度不高,难以辨识相似度高的物体。在工作中,使用的是双流CNN,它分别提取了两个特征图,并且将在这两个特征图每个位置算外积,然后用sum-pooling池化,最后再拉直成一条特征向量^[14]。用这个特征向量进行细粒度的分类识别,效果有显著提升。

假设图1某个区域*i*经过两个

图1
基于CNN的属性识别
基本技术框架



▲图2 B-CNN原理

CNN对应的特征图分别是矩阵 A ($C \times M$ 维)和 B ($C \times N$ 维),其中 M 和 N 是两个特征图的深度, C 可以是两位的数对,代表宽和高。

那么此时两矩阵相乘 $A^T \times B$ 的结果是矩阵 P ($M \times N$ 维),则

$$\Phi(I) = \sum_{i,j} P_{ij} \quad (1)$$

将 $\Phi(I)$ 拉直成一条向量 L ($MN \times 1$ 维),即图2中的双线性特征向量。对这个特征向量做标准化处理后便用作分类。在对细粒度属性表达能力上,该方法所得到的特征向量比全连接层输出的特征向量的更强。

2.3 车辆属性识别演示系统

文中,我们研究并实现了一个车辆属性识别演示系统,此原型系统的设计主要参考MVC框架,将界面、业务逻辑、数据三者分层,在修改业务逻辑时只需要修改控制层的部件,在修改用户界面(UI)时,只需要对定义

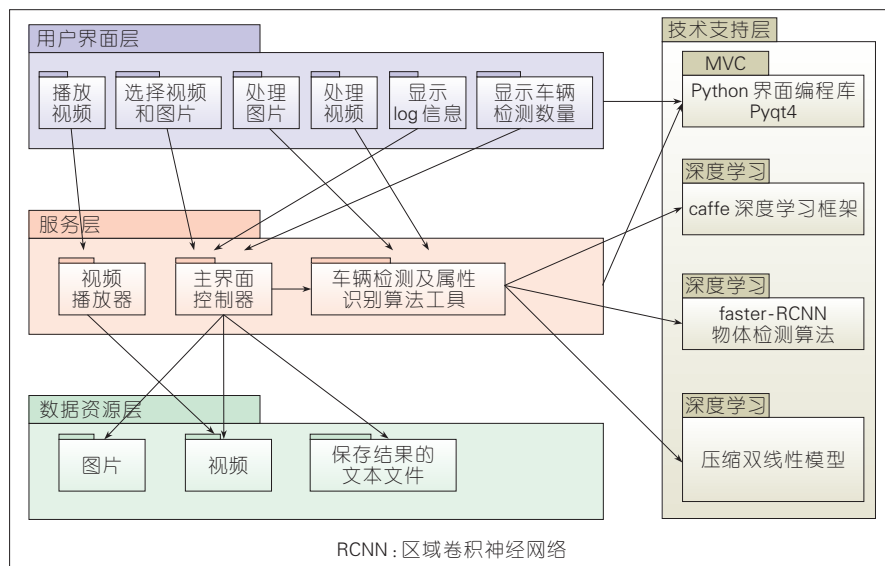
UI的xml文件做出修改,不需要改变控制层的代码,降低了代码的耦合性,增加重用性。系统的总体架构如图3所示:最上层是UI层,中间是业务逻辑层,下面是数据资源层,而右边则是该系统的技术支持层。

演示系统使用python+pyqt4开发,界面简洁。系统主要功能是对输入的图像和视频做车辆检测和属性识别处理,并输出结果。系统的深度学习框架采用Caffe^[15],Caffe是采用C++语言开发的高清晰、高可读性、高效的深度学习框架,支持多语言接口。

3 车辆属性识别实验结果与分析

3.1 SYSU-ZTE-CARS数据集

我们使用有监督的训练方法训练属性识别模型,该方法需要训练样本,即图片及其属性类别标签。因此,我们参考CompCars数据集的构建



▲图3 演示系统架构

方式和经验,利用网络爬虫技术,从互联网获取了大量车辆图片,并对所有图片进行了属性标注,构建了用于训练属性识别模型的大规模数据集 SYSU-ZTE-CARS,图4是该数据集的样例图片。

SYSU-ZTE-CARS 数据集共有 169 610 张的车辆图片,共标注了 5 种常用车辆属性,其属性标签标注情况如表1所示。其中,属性标签包括正前、正后、正侧、前侧和后侧共 5 种车辆视角;包括黑、白、红、蓝、黄、绿、紫、棕、灰、银共 10 种颜色;包括“MPV”“SUV”“掀背车”“轿车”等 15 种类型,以及 215 种品牌和 1 609 种子品牌。

3.2 车辆属性识别实验结果

在算法的性能测试实验中,我们初步采用单任务学习框架,在 SYSU-ZTE-CARS 数据集上对属性识别算法性能进行评估,如表2所示。所谓的单任务学习框架,即对每一种属性,分别训练一个 CNN 模型,各属性识别模型之间的参数不共享。

评测指标我们采用 Top-1 精度评价策略,即把模型预测结果按置信度从高到低排序,只考虑置信度最高的预测结果是否正确。

实验结果表明:对于简单的属



▲图4 SYSU-ZTE-CARS 数据集样例

▼表1 SYSU-ZTE-CARS 数据集属性标签标注情况

属性	视角	颜色	车类型	品牌	子品牌
种类	5	10	15	215	1 609

▼表2 各种属性在 SYSU-ZTE-CARS 数据集上的初步测评效果

方法	top-1 精度 / %				
	视角	颜色	车类型	品牌	子品牌
AlexNet+Softmax	87.10	89.01	88.30	76.66	60.44
VGG16+Softmax	93.61	91.38	93.23	88.70	66.69
ResNet50+Softmax	92.88	91.45	93.64	90.89	67.15
ResNet101+Softmax	93.95	92.07	94.05	93.51	74.32
BCNN-ResNet50+Softmax	92.73	91.47	93.74	94.27	83.91
BCNN-ResNet101+Softmax	93.90	92.03	93.98	95.18	88.84

BCNN: 双线性卷积神经网络

性,如视角、颜色、车类型,采用轻量级网络(AlexNet)或中等深度网络(VGG16),已经有比较可靠的识别精度。采用极深网络(ResNet系列)虽然对识别精度有小幅提升,但同时也增加了模型的参数规模,增大了计算量。细粒度识别技术对于简单属性的识别精度提升不但帮助不大,反而增加了模型运算开销。对于较困难的属性,如品牌,轻量级网络的特征判别性表达能力不足,增加网络深度,可以明显提高识别精度。采用细粒度识别优化后,其识别精度有小幅上升。对于困难的属性,如子品牌,仅仅增加网络深度并不能获得较好的识别性能,而须要通过细粒度的识别技术来增强特征的可判别性,才能达到可靠的识别精度。

SYSU-ZTE-CARS						
视角	正侧	前侧	正后	后侧	正前	前侧
颜色	黑色	黑色	白色	银色	金色	红色
车类型	SUV	SUV	SUV	Hatchback	Sedan	Sedan
品牌	Jeep	MIN_JCW	Toyota	Volkswagen	Volkswagen	Baojun
子品牌	Compass	PACEMAN	Toyota_RAV4	GOL	Passat	Baojun_630
现实监测控制场景						
视角	正前	正前	正前	正前	正前	正前
颜色	银色	灰色	紫色	蓝色	红色	黄色
车类型	Sedan	Sedan	Minibus	Hatchback	Hatchback	Sedan
品牌	Lexus	Nissan	Wuling	Suzuki	Nissan	Hyundai
子品牌	Lexus_NX	Nissan_Teana	Wulingrongguang	Suzuki_Alto	Nissan_Tiida	Hyundai_Rena

▲图5 算法识别效果样例

图5是算法识别效果样例,上面两行分别是SYSU-ZTE-CARS数据集上的样例图片和算法预测的属性标签,下面两行分别是在现实应用场景中的测试样例图片和算法预测的属性标签。预测标签中,黑色字体表示正确预测,红色字体表示预测错误。SYSU-ZTE-CARS上正确属性通过构建数据集的时候标注获得,由于现实场景中并没有预标注的属性标签,我们将预测结果与网上搜索结果进行对比来判断预测的正确性。我们抽取了1000张真实场景的图片作为测试样例,人工评估的结论是:这种跨场景自动标注的方法在非极端条件下的识别精确度较高,比较可靠;在极端条件下,则仍须要人工检验。

4 结束语

基于跨场景标注的思路,通过构建大规模车辆属性数据集,在车辆海报图片场景中训练属性识别模型,再将模型迁移到现实监测控制场景中进行自动标注,以此方式间接获得检测控制场景下的车辆属性标签。通过初步的实验测试评估,该思路可行性较高。

致谢

本研究得到中山大学林惊教授和李冠彬博士的大力支持和帮助,谨

致谢!

实验工作是由中山大学人机物智能融合实验室完成的,周启贤、陈日全同学承担了大量的数据集建立工作和实验,对他们谨致谢意!

参考文献

- [1] 潘秋菱, 许晓杰, 肖乾友. CNGI上的一种电信业务:智能交通[J]. 中兴通讯技术, 2005, 11(5):16-19
- [2] 王添宇, 付超贤. 交通2.0——智慧交通的关键[J]. 中兴通讯技术, 2014, 20(4):11-15. DOI: 10.3969/j.issn.1009-6868.2014.04.003
- [3] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252
- [4] GEIGER A, LENZ P, URSTASUN R. Are We Ready for Autonomous Driving? The Kitti Vision Benchmark Suite[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA: IEEE, 2012: 3354-3361
- [5] CAI Z, FAN Q, FERIS R S, et al. A Unified Multi-Scale Deep Convolutional Neural Network for Fast Object Detection[C]// European Conference on Computer Vision. Germany: Springer International Publishing, 2016: 354-370
- [6] REN J, CHEN X, LIU J, et al. Accurate Single Stage Detector Using Recurrent Rolling Convolution[EB/OL].[2017-04-27]. https://arxiv.org/abs/1704.05776
- [7] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[EB/OL].[2017-04-27]. https://arxiv.org/abs/1409.1556
- [8] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper with Convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2015: 1-9. DOI: 10.1109/CVPR.2015.7298594
- [9] YANG L, LUO P, CHANGELOY C, et al. A Large-Scale Car Dataset for Fine-Grained Categorization and Verification[C]// Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition. USA:IEEE, 2015: 3973-3981. DOI: 10.1109/CVPR.2015.7299023

- [10] SOCHOR J, HEROUT A, HAVEL J. BoxCars: 3D Boxes as CNN Input for Improved Fine-Grained Vehicle Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA:IEEE, 2016: 3006-3015. DOI: 10.1109/CVPR.2016.328
- [11] ZHOU Y, LIU L, SHAO L, et al. DAVE: A Unified Framework for Fast Vehicle Detection and Annotation[C]//European Conference on Computer Vision. Germany: Springer International Publishing, 2016: 278-293
- [12] KRIZHEVSKY A, SUSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks[C]// Advances in Neural Information Processing Systems. USA: IEEE, 2012: 1097-1105
- [13] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2016: 770-778. DOI: 10.1109/CVPR.2016.90
- [14] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN Models for Fine-Grained Visual Recognition[C]//Proceedings of the IEEE International Conference on Computer Vision. USA:IEEE, 2015: 1449-1457
- [15] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[C]//Proceedings of the 22nd ACM international conference on Multimedia. USA:ACM, 2014: 675-678

作者简介



董振江, 中兴通讯股份有限公司战略与技术专家委员会业务专家组组长、云计算及IT研究院副院长, 中国人工智能学会常务理事; 主要研究方向为云计算与大数据、新媒体、移动互联网等; 曾获国家科技进步二等奖2次、省部级奖项多次, 主持基金项目10余项; 已发表论文20余篇, 出版专著1部。



高燕, 中兴通讯有限公司高级工程师; 主要研究领域包括人工智能、图像和视频分析; 主持5项中兴通讯产学研合作项目, 参与3项国家重大科技项目; 发表论文多篇。



吴文熙, 中山大学数据科学与计算机学院研究生; 主要研究领域为深度学习与计算机视觉, 主要研究课题为大规模物体识别; 在Kaggle、ImageNet等国际计算机视觉比赛中均有优秀表现。