

# 深度学习进展及其在图像处理领域的应用

## Advances in Deep Learning and Its Applications in Image Processing

刘涵/LIU Han<sup>1</sup>  
贺霖/HE Lin<sup>2</sup>  
李军/LI Jun<sup>1</sup>

(1. 中山大学, 广东 广州 510275;  
2. 华南理工大学, 广东 广州 510640)  
(1. Sun Yat-Sen University, Guangzhou  
510275, China;  
2. South China University of Technology,  
Guangzhou 510640, China)

在过去 10 年左右的时间里, 深度学习对信息技术的许多方面都产生了重要影响。诸多关于深度学习的描述普遍存在两个重要的共同点: 包含多层或多阶非线性信息处理的模型; 使用了连续的更高、更抽象层中的监督或无监督学习特征表示的方法。深度学习是以神经网络为基础, 包含人工智能、图模型、最优化等技术在内的交叉领域。它之所以如此受关注, 主要源于 3 个方面: 芯片硬件处理性能的巨大提升, 为深度神经网络的复杂计算提供了基础; 用于训练的数据呈爆炸性增长, 为复杂网络的学习提供了可能; 机器学习和图像处理等方面研究取得了很大进展。

### 1 深度学习的发展

以前, 绝大多数机器学习和信号处理技术都是基于浅层结构, 如高斯

收稿日期: 2017-05-23

网络出版日期: 2017-07-12

基金项目: 国家自然科学基金 (61571195); 广东省自然科学基金 (2016A030313516)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0036-005

**摘要:** 深度学习一般通过 3 种方式进行: 有监督学习、无监督学习和混合深度学习。以“无监督或生成式特征学习”以及“有监督特征学习和分类”为例, 讨论了深度学习及其在图像处理等领域的进展及未来可能的研究方向。认为深度学习打破了传统机器学习和信号处理技术普遍基于浅层结构的局限。得益于相关非凸优化等问题的逐步解决, 深度学习已经在图像处理等领域取得了一些突破性的进展。

**关键词:** 深度学习; 图像处理; 分层结构

**Abstract:** Deep learning methods are usually divided into three different categories, including supervised learning, unsupervised learning and mixed-structured learning. In this paper, "unsupervised or generative feature learning" and "supervised feature learning and classification" are taken for examples to illustrate the research advances in deep learning, and several promising research lines in image processing. Deep learning technique has overcome some limitations of shallow structures commonly used in traditional machine learning and signal processing. Due to the progress in some aspects, such as relevant nonconvex optimization, deep learning has achieved remarkable development.

**Keywords:** deep learning; image processing; hierarchical structure

混合模型 (GMM)、线性或非线性动力系统、条件随机场 (CRF)、最大熵模型 (MaxEnt)、支持向量机 (SVM)、逻辑回归 (LR)、核回归以及多层感知器 (MLP) 等。这些结构一般包含最多一到两层的非线性特征变换。已有研究表明: 浅层结构在解决简单的或者约束较多的问题上效果明显, 但是由于其建模和表示能力有限, 在对实际应用中一些较为复杂自然信号 (比如人类语音、自然声音和语言、自然图像和视觉景色) 进行处理时会遇到一些困难。人类的听觉和视觉信息等的处理机制一般可以用深度结构描述, 通过该结构可以从感官输入信息中提取复杂结构并构建内部

表示。如果能实现有效和高效的深度学习算法, 那么对于各种自然信号的处理技术而言, 其性能会得到很大提升。

深度学习的概念一般被认为来源于对神经网络的研究。前馈神经网络或具有多隐层的多层感知器是深度神经网络 (DNN) 的典型模型。反向传播 (BP) 算法是解决其学习问题的广泛运用的典型算法。遗憾的是, 仅仅使用 BP 算法在实际学习隐层数目较多的网络时往往效果不是很好<sup>[1]</sup>。在优化目标为非凸函数的深度神经网络学习中, 通常存在局部最优解等问题。BP 算法基于局部梯度信息, 往往从一些随机的初始点

开始寻优,当使用批量梯度下降或随机梯度下降的BP算法时,目标函数经常会陷入局部最优。随着网络层数的加深,局部最优的情况也就会变得越来越严重。虽然相关研究者对小规模的神经网络的探究从未间断过,但是在很多机器学习和信号处理方法中,研究者们将研究重点从对神经网络本身的研究转移到对具有凸损失函数的浅层模型的研究,这些模型以降低建模准确度为代价,达到快速高效地收敛到全局最优化的目的。所以,深层网络本身还存在着易于陷入局部最优等缺陷,有待于进行更深入的研究。

Hinton等在2006年左右提出了一种高效的基于深度置信网络(DBN)的无监督学习算法<sup>[2-3]</sup>,他们利用经验性的方法处理了与深度模型相关的最优化难题。DBN是一种深度生成式模型,由一组受限玻尔兹曼机(RBMs)堆叠而成,它的核心部分是贪婪式的逐层学习,这种算法可以最优化DBN中的权重,且其时间复杂度与网络的大小和深度呈线性关系。最近,相关研究者对于DNN与DBN进行了更加细致的研究,如可使用DBN来初始化DNN的权值等。在DNN中,多隐层的使用不仅显著提高了网络的表示能力,而且可得到一些较优解。然而,在训练过程中使用深而宽的神经网络需要依赖于强大的计算性能。随机梯度下降(SGD)算法就是一种在训练集较大且冗余的情况下较为有效的学习算法<sup>[4]</sup>。已有的研究表明:SGD可以有效地实现并行方式的运算。该并行运算主要通过两种方式实现:一种方式是通过异步模式使用多台计算机<sup>[5]</sup>;另一种方式是使用多图形处理器(GPU)的流水线型的BP算法<sup>[6]</sup>。另外,从单个或小批量样本中估计得到的随机性梯度使得SGD通常能跳出局部最优解。其他的一些学习算法,如Hessian free<sup>[7]</sup>、Krylov subspace<sup>[12]</sup>方法等,都表现出了类似的学习能力。对

于DNN学习中涉及的非凸优化问题,更好的参数初始化和学习技术都会学习出更好的模型。

DBN预训练并不是唯一可对DNN进行有效初始化的方法。基于降噪自动编码器的方法对DNN进行逐层地预训练,将每两层视为一个降噪自编码器,该编码器再通过将输入节点的随机子集设置为零进行正则化<sup>[8]</sup>。另一种方法则是使用压缩自编码器<sup>[4]</sup>,该编码器通过使输入变量具有更好的鲁棒性来达到同样的目的。此外,Ranzato等开发了稀疏编码对称机(SES<sub>M</sub>)<sup>[9]</sup>,其在构建DBN模块中具有和RBM非常类似的架构,它也可以用来有效地初始化DNN的训练过程。除了使用贪婪方法逐层地进行无监督预训练,有监督的预训练(有时称为判别式预训练)也被证明是比较有效的<sup>[10-11]</sup>。有监督的预训练的基本思路是从一个经过BP算法训练的单个隐层MLP开始,每一次需要添加一个新的隐层时,用一个随机初始化的新的隐层和输出层替换输出层,并用BP算法训练全新的MLP(或DNN)。在RBM发展的同时,出现了另外两种较有代表性的非概率的、非生成式的深度模型:一种是基于自编码器(AE)的改进模型,其使用与DBN训练相似的贪婪分层方法进行训练;另一种是基于能量的模型,其利用稀疏表示来进行非监督学习。与DBN相似,其也可对深度神经网络进行高效的预训练。

## 2 3种深度学习网络

深度学习是一类应用广泛的机器学习技术和架构,其特点是采用多层的非线性结构进行信息处理,这种方法在本质上是分层实现的。根据不同应用领域的任务目标及对应的不同深度学习网络结构,我们可以大致把已有深度学习结构分为3类:

(1)无监督的或生成式学习的深度网络。该结构针对模式分析和合成任务,用于在没有目标类标签信息

的情况下捕捉可见数据的高阶相关性。已有的无监督特征学习或表达学习指的就是这一类深度网络。

(2)有监督学习的深度网络。这种网络直接提供用于模式分类目的的判别能力,它的特点是描述了在给定可见数据的条件下不同类别的后验概率分布。对于这种有监督的学习,目标数据的类别标签总是以直接或间接形式给出,所以它们也被称作判别式深度网络。

(3)混合式深度网络。其目标是实现判别式模型的效果,往往以生成式或无监督深度网络的结果作为重要辅助,通过更好地优化和正则化以上类别(2)中的深度网络来实现,也可以通过使用判别式准则对以上类别(1)中所述的深度生成式或无监督深度网络进行参数估计来实现。

从传统机器学习的角度,深度学习模型可分为深度判别式模型和生成式/无监督模型。然而这种模型分类方法忽略了深度学习研究中的一个重要观点,即生成式和无监督学习模型可通过更好地正则化与优化来提高深度判别网络的训练效果。因此,深度学习网络有时会以混合式的结构形式出现。

## 3 深度学习在图像目标识别和计算机视觉中的应用

多年来,计算机视觉和图像目标识别等任务长期依赖人工设计的特征,如尺度不变特征变换(SIFT)和方向梯度直方图(HOG)等。此类特征仅仅是对图像中低级别的边缘信息进行描述与表征,若要描述图像中高级信息例如边缘交叉和局部外观等,其往往显得力不从心。深度学习可以通过无监督和有监督的学习方法直接从数据中获得层级化的视觉特征,从而提供一套更为有效的解决方案。深度学习方法经常可从无监督和有监督两个角度进行讨论:无监督特征学习,该类方法通常将深度学习用于特征提取,然后这些特征会被直

接送入后续分类算法;有监督的特征学习,当存在大量有标签样本时,此类方法通过端到端的学习策略实现特征提取与分类器的联合优化。

### 3.1 无监督或生成特征学习

当有标签样本相对缺乏时,无监督学习算法可用于学习视觉特征层级结构。如 Hinton 和 Salakhutdinov 等最早提出将无监督深度自编码方法应用于 DBN 模型的预训练<sup>[4]</sup>。他们利用该方法在仅有 60 000 个训练样本的美国国家标准技术研究所修正 (MNIST) 数据库上成功实现了图像的识别和降维(编码)任务。此外, Nair 和 Hinton 提出了一种改进的 DBN,该 DBN 的顶层使用了一个三阶的 RBM<sup>[12]</sup>。当这种 DBN 被应用于 NORB 数据库(一个三维目标识别任务数据库)上时,其错误率几乎下降到了目前所公布的最低水平,这再次表明了 DBN 在很大程度上是优于类 SVM 这样的浅层模型。随后,还出现一些对 DBN 的改进文献。其他的一些无监督深度特征学习方法还包括稀疏自编码器及基于深度稀疏编码的模型等<sup>[13]</sup>。

### 3.2 有监督特征学习和分类

CNN 是一种受到广泛关注的有监督深度学习结构。有监督 CNN 结构获得广泛关注始于 2012 年 10 月 ImageNet 竞赛,这主要是由于大量的有标签样本及高性能 GPU 计算平台的出现使得大规模 CNN 的高效训练成为可能。图 1 给出了文献<sup>[4]</sup>中所描

述的 CNN 的基本结构。为了实现典型图像像素的空间位置相对不变性这一特点,CNN 使用了一个带有局部连接和共享权值的卷积层,该层的输出通过一个非线性激活函数来获得激活响应,接着通过一个非线性池化层来减小数据量,最后再将池化层的输出连接到若干个全连接层。这种结构也常被称作深度卷积神经网络。

CNN 在 2012 年的 ImageNet 竞赛中取得了瞩目的成绩。在该场比赛中,使用深度 CNN 进行建模的方法获得了前所未有的低错误率。该深度 CNN 模型包含 6 千万个权值,65 万个神经元节点以及结合 5 个卷积层的最大池化层。此外,两个全连接层也被用于这个 CNN 模型的最顶层。另外,还有两个额外的因素也起到了很重要的作用:首先,是一个称为“dropout”<sup>[15]</sup>的强大的正则化技术;第二个重要因素是通过激活函数  $f(x) = \max(x, 0)$  所实现的整流线性单元 (ReLU) 的应用,使得整个训练过程的效率被极大地提高,尤其是通过 GPU 并行运算实现后取得的效果更加明显。其后,基于更大规模的模型以及更多的训练数据,CNN 得到了进一步的改进。不少深度 CNN 模型和方法的强大学习能力在各年的 ImageNet 竞赛上得到了验证。

深度 CNN 已被证明在图像目标识别任务中具有卓越的分类性能,关于其机理也逐渐有了一些解释,如 Zeiler 等利用基于反卷积网络的可视化技术对 CNN 的机理进行了一些讨论<sup>[16]</sup>。图 2 示意了反卷积机理,该反

卷积网络通过 CNN 中相应前馈计算的相反连续操作,其中包括反池化、校正和滤波,使得特征图谱之上的活动得以重建。在实现反池化的过程中,最大池化操作的非可逆性通过近似逆向逼近的方法得以解决。

除了深度 CNN 结构外,DNN 结构也在大量的计算机视觉任务上获得了成功<sup>[17-19]</sup>。目前,基于深度 CNN 结构的有监督学习模式及其相关的分类技术已在相关研究领域产生很大影响,这尤其体现在 2012—2013 年的 ImageNet 比赛中。这些方法不仅可以用于图像目标识别任务,同样还可以应用于其他一些计算机视觉的任务中。当然,关于 CNN 深度学习方法的机理及其局限性等,仍有很多问题需要探讨。

## 4 结论及展望

深度学习作为当前机器学习领域最热门的技术之一,已经在不少领域获得了应用,并且展现出巨大的前景。根据网络结构的不同,它大致可以分为无监督、有监督和混合神经网络 3 种类别。作为一种从本质上来说是分层非线性结构的深度模型,它所构建和学习的深层特征表示无疑极大地提升了传统浅层模型的泛化能力;但是随之而来的明显非凸的优化目标形式却长期困扰着该领域的研究者,如其导致的局部最优化等问题,阻碍着基于梯度的 BP 算法的有效实施。得益于无监督预训练方法对优化初始点的改善,该问题有了一些经验性且行之有效的解决方案。

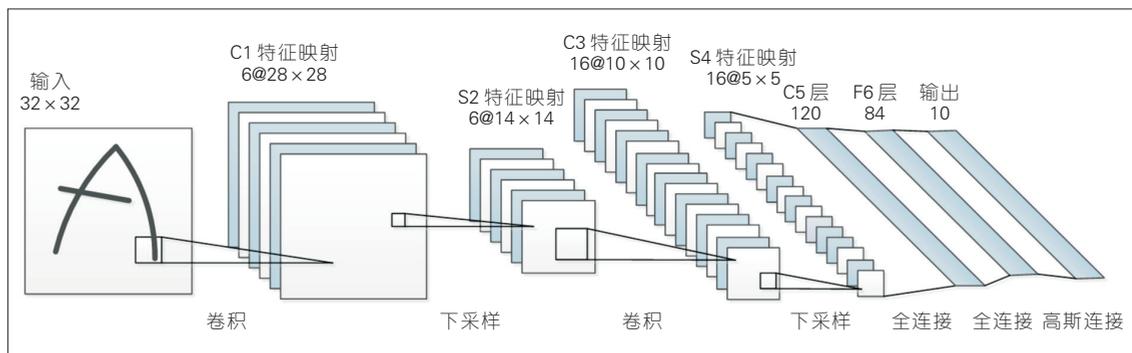
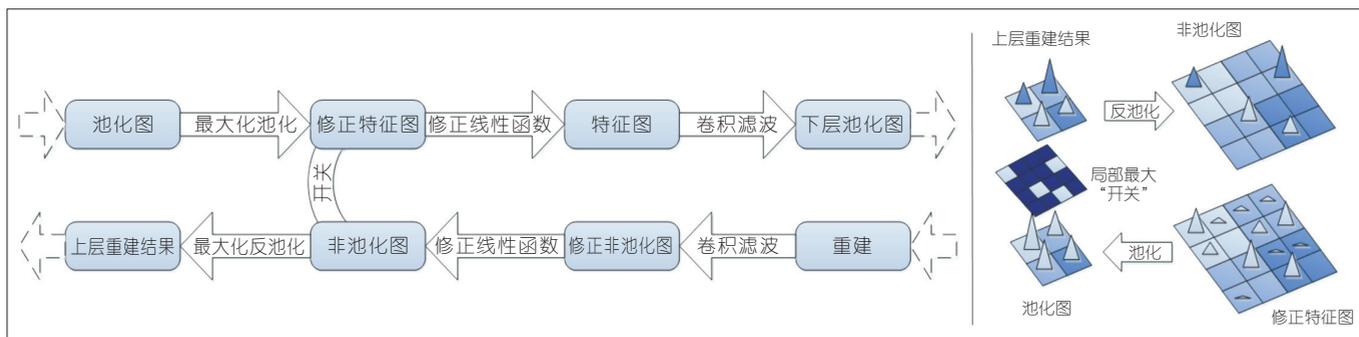


图 1 一种典型的 CNN 结构



▲图2 反卷积网络机理

虽然该类方法缺乏稳固的理论基础,却成功促成了深度学习方法在学术界和工业界的大规模成功应用。通过最近的研究发现,现有深度架构在优化技术等方面存在着巨大的提升空间<sup>[7,10,20-23]</sup>。

另一方面,如果训练数据集足够大,理论上讲,模型的泛化能力将会得到较大的提升,那么通过深度置信网络等预训练方法所带来的良好优化初始点的重要性必然会显著降低。然而,要实施针对大规模数据集的应用,强大的计算能力是必不可少的。当前来看,有效且可扩展的并行算法是训练庞大数据集的关键所在。然而,常用的基于 mini-batch 的梯度下降技术很难并行实现。最近出现的异步梯度下降等技术为这一领域相关问题的解决带来了一些新思路,并在 CPU 集群<sup>[7,13]</sup>和 GPU 集群<sup>[24]</sup>中得到了初步的实现。未来,对于并行学习和新型的大规模优化算法仍需要进行有针对性的理论研究。

目前,阻碍深度模型发展的另一主要问题在于超参数的合理选择。众所周知,基于神经网络的深度学习技术有着数量众多且自由度极大的超参数,如网络架构的层数以及每层的单元数、正则化强度、学习速率以及学习速率衰减率等。基于传统的网格搜索等技术的解决方案无论从效率还是成本的角度上来讲对于超参数的设定都是不可行的。此外,不同的超参数之间通常存在着相互依赖性,且微调代价巨大。这些问题决

定了我们需要开展进一步的研究来探索更有效的解决方案。令人欣慰的是,近来以随机采样<sup>[25]</sup>和贝叶斯优化过程<sup>[26]</sup>为代表的方法给我们带来了一些新思路。虽然一些经验性的超参数设置即可在一定程度上反映出深层结构所具有的强大泛化能力,但是超参数的最优化配置在对一些特定应用方面进一步提高深度模型性能上仍具有很大的研究价值。

领域知识的应用是深度学习方法成功的另一大关键因素。根据不同任务的特点设计不同的不变性特征提取方法以及正则化方法是当前所流行的一种应用方式。研究者们也在基于域适应的迁移学习对于深度学习的辅助作用方面进行了一些研究,但是能够普遍适用于各类分类任务的深度学习技术仍然是不存在的,例如:当前较为通用的生成式预训练伴随判别式微调的学习策略在一些特定的任务(如语音识别)中表现并不理想。因此,更加有效的域适应技术和新型的通用学习架构对于图像处理等领域中的一些较为复杂的问题而言是极其重要的。

深度学习理论还有其他一些方面的基础性理论问题值得关注,例如:如何通过关注数据变化中潜在因素的分分布式表示问题来设计更合理的深度学习架构,进而提取更高效的特征表示;在深度结构的输入输出表示中同时引入结构信息<sup>[27-29]</sup>,使得大多数传统的深度学习技术只能用于“扁平结构”表示的缺陷得到改善。

最后,为了实现理想中的“强人工智能”,从而实现具备类似人类大脑水平的智慧,传统的以信号处理和机器学习为主要技术基础的人工智能研究应更多地寻求同神经计算等领域的合作,通过借助于前沿生物领域对人类大脑分层结构的最新研究成果来改善当前的系统计算模型。

#### 参考文献

- [1] BENGIO Y. Learning Deep Architectures for AI [J]. Foundations and Trends® in Machine Learning, 2009, 2(1):1-127. DOI: 10.1561/2200000006
- [2] HINTON G E, OSINDRO S, TEH Y. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7):1527-1554. DOI: 10.1162/neco.2006.18.7.1527
- [3] HINTON G E, SALAKHUTDINOV R R. Reducing the Dimensionality of Data with Neural Networks[J]. science, 2006, 313(5786): 504-507. DOI: 10.1126/science.1127647
- [4] BOTTOU L, CUN Y L. Large Scale online Learning[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2004
- [5] DEAN J, CORRADO G, MONGA R, et al. Large Scale Distributed Deep Networks[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2004
- [6] CHEN X, EVERSOLE A, LI G, et al. Pipelined Back-Propagation for Context-Dependent Deep Neural Networks[C]//Interspeech 2012. USA: IEEE, 2012. DOI: 10.1.1.649.218
- [7] MARTENS J. Deep Learning via Hessian-Free Optimization[C]//Proceedings of the 27th International Conference on Machine Learning (ICML-10). USA: IEEE, 2010. DOI: 10.1.1.170.2846
- [8] VINCENT. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. Journal of Machine Learning Research, 11(Dec): 3371-3408. DOI: 10.1561/2200000006

- [9] BOUREAU Y, CUN Y L. Sparse Feature Learning for Deep Belief Networks[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2008
- [10] BENGIO Y. Greedy Layer-Wise Training of Deep Networks[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2007:153
- [11] YU K, LIN Y. Learning Image Representations from the Pixel Level via Hierarchical Sparse Coding[C]//2011 IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2011. DOI: 10.1109/CVPR.2011.5995732
- [12] MNIH V, KAVUKCUOGLU, SILVER D, et al. Playing Atari with Deep Reinforcement Learning[EB/OL].(2017-05-22). <https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf>
- [13] LE Q V. Building High-Level Features Using Large Scale Unsupervised Learning[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing. USA: IEEE, 2013. DOI: 10.1109/ICASSP.2013.6639343
- [14] LE C Y. Gradient-Based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998. 86(11): 2278-2324. DOI: 10.1109/5.726791
- [15] HINTON G E, SRIVASTAVA N, KRIZHEVSKY, et al. Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors[EB/OL].(2017-05-22). <https://arxiv.org/pdf/1207.0580.pdf>
- [16] ZEILER M D, FERGUS R. Visualizing and Understanding Convolutional Networks[C]//European Conference on Computer Vision. EU: Springer, 2014
- [17] CIRESAN D C. Deep, Big, Simple Neural Nets for Handwritten Digit Recognition[J]. Neural Computation, 2010, 22(12): 3207-3220. DOI: 10.1162/NECO\_a\_00052
- [18] CIRESAN D. Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2012
- [19] CIREGAN D and MEIER U. Multi-Column Deep Neural Networks for Image Classification[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2012. DOI:10.1.1.367.484
- [20] MARTENS J, SUTSKEVER I. Learning Recurrent Neural Networks with Hessian-Free Optimization[C]//The 28th International Conference on Machine Learning. USA: IEEE, 2011
- [21] LE V Q, NGUAN J, COATES A, et al. On Optimization Methods for Deep Learning[C]//The 28th International Conference on Machine Learning. USA: IEEE, 2011
- [22] SAINATH T N. Optimization Techniques to Improve Training Speed of Deep Neural Networks for Large Speech Tasks[J]. IEEE Transactions on Audio, Speech, and Language Processing, 21(11): 2267-2276. DOI: 10.1109/TASL.2013.2284378
- [23] WRIGHT S J. Optimization Algorithms and Applications for Speech and Language Processing[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(11): 2231-2243. DOI: 10.1109/TASL.2013.2283777
- [24] COATES A, HUVAL B, WANG T, et al. Deep Learning with COTS HPC Systems[C]//International Conference on Machine Learning. USA: IEEE, 2013
- [25] BERGSTRA J, BENGIO Y. Random Search for Hyper-Parameter Optimization [J]. Journal of Machine Learning Research, 2012, 13: 281-305. DOI: 10.2307/1268522
- [26] SNOEK J, LAROCHELLE H, ADAMS P R. Practical Bayesian Optimization of Machine Learning Algorithms[C]//Advances in Neural Information Processing Systems. USA: IEEE, 2012
- [27] SOCHER R. New Directions in Deep Learning: Structured Models, Tasks, and Datasets[C]//Neural Information Processing Systems. USA: NIPS, 2012
- [28] DENG L. Design and Learning of Output Representations for Speech Recognition[C]//Neural Information Processing Systems. USA: NIPS, 2013
- [29] SRIVASTAVA N, SALAKHUTDINOV R R. Discriminative Transfer Learning with Tree-Based Priors[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2013

## 作者简介



刘涵, 中山大学地理科学与规划学院学生; 主要研究领域为遥感图像分类及混合像元分解等。



贺森, 华南理工大学副教授; 主要研究领域为统计模式识别、统计信号处理和多维图像处理等; 主持国家级及省部级科研项目多项。



李军, 中山大学教授; 主要研究领域为多维图像处理和机器学习等; 主持国家级及省部级科研项目多项; 国家“青年千人计划”入选者。