

DOI: 10.11992/tis.201706078

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180328.1448.010.html>

基于卷积神经网络的遥感图像分类研究

李亚飞, 董红斌

(哈尔滨工程大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001)

摘 要: 遥感图像分类是模式识别技术在遥感领域的具体应用, 针对遥感图像处理中的分类问题, 提出了一种基于卷积神经网络 (convolutional neural networks, CNN) 的遥感图像分类方法, 并针对单源特征无法提供有效信息的问题, 设计了一种多源多特征融合的方法, 将遥感图像的光谱特征、纹理特征、空间结构特征等按空间维度以向量或矩阵的形式进行有效融合, 以此训练 CNN 模型。实验表明, 多源多特征相融合能够加快模型收敛速度, 有效提高遥感图像的分类精度; 与其他分类方法相比, CNN 能够取得更高的分类精度, 获得更优的分类效果。

关键词: 遥感图像; 地物分类; 卷积神经网络; 特征融合

中图分类号: TP301 **文献标志码:** A **文章编号:** 1673-4785(2018)04-0550-07

中文引用格式: 李亚飞, 董红斌. 基于卷积神经网络的遥感图像分类研究[J]. 智能系统学报, 2018, 13(4): 550-556.

英文引用格式: LI Yafei, DONG Hongbin. Classification of remote-sensing image based on convolutional neural network[J]. CAAI transactions on intelligent systems, 2018, 13(4): 550-556.

Classification of remote-sensing image based on convolutional neural network

LI Yafei, DONG Hongbin

(College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)

Abstract: The classification of remote-sensing images is a specific application of pattern recognition technology in the remote-sensing domain. In this paper, we propose a method for the classification of remote-sensing images based on convolutional neural networks (CNN). In addition, to address the difficulty of providing effective information regarding a single-source feature in convolutional neural networks, we propose a multi-source and multi-feature fusion method. We combine the spectral, texture, and spatial-structure features of remote-sensing images in the form of vectors or matrices according to their spatial dimensions, and train the CNN model using these combined features. The experimental results show that multi-source and multi-feature fusion can effectively improve the model convergence speed and classification accuracy, in comparison with traditional classification methods, and that the CNN method achieves higher classification accuracy and classification effect.

Keywords: remote-sensing image; classification of land cover; convolutional neural networks; feature fusion

遥感图像分类就是依据遥感数据的各种信息, 通过采用某种算法挖掘每类地物的独有特征并将其分割为互不相交的子空间, 进而将各个像素划分到对应的子空间。但对于高分辨率的遥感图像, 地物的光谱特征越来越丰富, “同物异谱”

和“异物同谱”现象更加明显^[1]。因此传统的参数化方法如最小距离分类法 (minimum distance classification, MDC)^[2]、极大似然分类法 (maximum likelihood classification, MLC)^[3]等分类准确度降低。而非参数化方法如支持向量机^[4]、人工神经网络 (artificial neural network, ANN)^[5]、决策树 (decision tree, DT)^[6]等在高分辨率遥感影像分类中得

收稿日期: 2017-06-26. 网络出版日期: 2018-03-28.

基金项目: 国家自然科学基金项目 (61472095).

通信作者: 董红斌. E-mail: donghongbin@hrbeu.edu.cn.

到广泛的应用。然而这些算法都属于浅层学习算法^[7], 很难有效地表达复杂函数, 缺少对复杂样本的适应性, 并且分类器模型的参数无法估计或估计不准确, 导致分类效果不理想^[8]。

2006年, 加拿大多伦多大学教授, 机器学习领域的泰斗 Hinton^[9]和他的学生 Salakhutdinov 在《科学》上发表了一篇文章, 掀起了深度学习在学术界和工业界的浪潮。深度学习是通过建立一种类似人脑分层的模型结构, 对输入信息逐层进行特征提取, 层级越深, 提取的特征越抽象复杂, 称为深度神经网络 (deep neural networks, DNN)^[10]。如今, 深度学习作为机器学习的一个重要分支, 已在图像识别、语音识别和自然语言处理等领域取得了巨大的成功^[11-13]。

目前, 国内将深度神经网络应用到遥感图像分类中的研究工作相对较少。因此, 本文提出一

种基于卷积神经网络 (CNN) 的遥感图像分类方法。针对单源特征无法提供更多有效信息的问题, 本文设计了一种多源多特征融合方法, 该方法将遥感图像的光谱特征、纹理特征、空间结构特征等按空间维度以向量或矩阵的形式进行融合, 并以融合的特征训练卷积神经网络模型。实验结果表明: 该融合方法能使 CNN 模型学习到更抽象、更具代表性的高层特征, 有效提高了分类精度, 获得最优的分类效果。

1 卷积神经网络简介

卷积神经网络作为深度学习的一种典型模型, 是一种多层神经网络结构。它主要由输入层、卷积层、下采样层 (池化层)、全连接层和输出层组成。如图 1 所示。

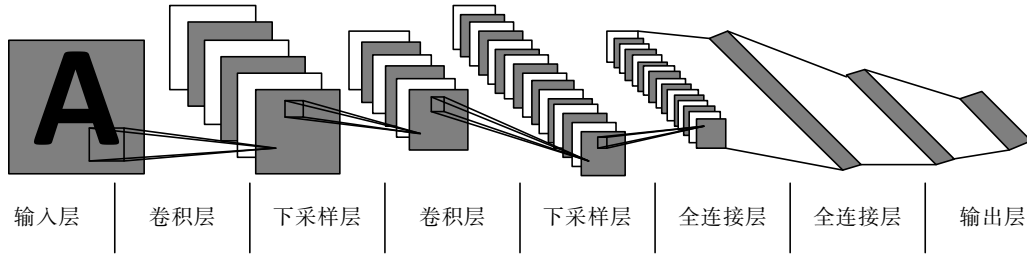


图 1 卷积神经网络示意

Fig. 1 Schematic diagram of CNN

输入层用于接收原始图像, 卷积层用于提取图像的特征并减少噪声的影响。假设输入的原始图像为 X , Y_i 表示第 i 层的特征图, 则 $Y_0 = X$ 。假设 Y_i 是第 i 层卷积层, 则

$$Y_i = f(W_i \cdot Y_{i-1} + b_i) \quad (1)$$

式中: W_i 表示第 i 层卷积核的权重; 运算符 \cdot 表示 W_i 与第 $i-1$ 层的特征图进行卷积操作; b_i 表示第 i 层的偏置向量; f 为非线性激活函数, 通常采用 ReLU 函数, ReLU 表达式为

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (2)$$

下采样层紧跟卷积层之后, 基于图像的局部相关性对特征图进行降维, 同时保持特征的尺度不变性。假设 Y_i 为下采样层特征图, 则

$$Y_i = \text{subsample}(Y_{i-1}) \quad (3)$$

池化方式一般有两种, 最大池化 (max pooling) 和平均池化 (mean pooling)。

经过多个卷积层和下采样层的交替连接, 全连接层对提取的特征进一步降维, 最后输出层依据全连接层提取的特征向量输出样本对应的标签。

卷积神经网络分类过程主要是网络的训练过程, 相当于人脑的学习过程。分为两个阶段, 1)

前向传播, 实现样本从输入层到输出层的特征学习; 2) 反向传播, 依据损失函数 $L(W, b)$ 计算出输出值与期望值之间的误差, 又称为“残差”^[14], 并根据梯度下降法调整网络参数。目前 CNN 中使用广泛的损失函数为交叉熵 (cross-entropy, CE) 损失函数, 其计算公式为

$$L(W, b) = \text{CE}(W, b) = - \sum_{i=1}^N \sum_{j=1}^C 1\{\hat{y}_i = j\} \log p_i^j \quad (4)$$

式中: \hat{y}_i 为第 i 个训练样本的期望值, p_i^j 为第 i 个训练样本第 j 个类别的预测概率, C 为训练样本的总类别数, N 为训练样本的总数。

卷积神经网络的训练目标是通过梯度下降法最小化网络的损失函数 $L(W, b)$ 。在整个训练过程中, 经过前向传播计算损失值, 然后通过梯度下降进行反向传播, 逐层更新每一层的训练参数 W 和 b 。参数更新公式定义为

$$W_i = W_i - \eta \frac{\partial L(W, b)}{\partial W_i} \quad (5)$$

$$b_i = b_i - \eta \frac{\partial L(W, b)}{\partial b_i} \quad (6)$$

式中: η 为网络学习率, 用来控制损失值反向传播的强度。

2 基于 CNN 的遥感图像分类方法

2.1 CNN 分类模型

AlexNet 网络模型是由 Alex Krizhevsky 于

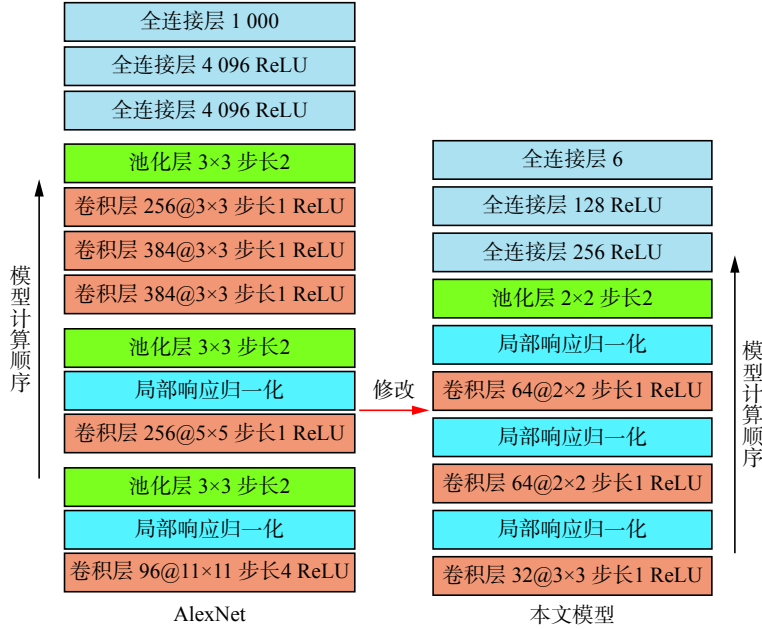


图 2 本文模型示意

Fig. 2 Schematic diagram of our model

与 AlexNet 的不同之处在于,设计的 CNN 模型去掉了一部分卷积层和池化层,原因在于池化层的主要作用是用来降维,而本文是对遥感图像进行像素级的分类,为了防止训练样本在高层次的尺寸过小,只保留最后一层的池化层。3 个卷积层参数设置为 $32@3 \times 3$ 、 $64@2 \times 2$ 和 $64@2 \times 2$,步长均为 1。池化层采用大小为 2×2 ,步长为 2 的最大池化。两个全连接层的节点个数分别设置为 256 和 128,最后以 Softmax 层作为网络的输出层。整个网络的学习率初始化为 0.1,采用指数衰减法,衰减因子为 0.1,权重衰减因子为 0.98,Dropout 参数训练时设置为 0.5,测试时设置为 1,采用 Adam 优化算法。

2.2 多源多特征融合

目前大多数分类方法都是依据人工设计的单源特征来进行分类,由于单源特征往往不能很好地反映出所有地物类别之间的差异,从而导致分类算法的泛化性能较差,针对此问题,本文设计了一种多源多特征融合的方法。为了更形象地说明本文的融合方法,给出了如图 3 所示的流程图。对于每一个像素点,考虑了其周围大小为 $winsize \times winsize$ 的邻域像素,这有助于消除图像的斑点噪声。

首先利用 PCA 对原始数据进行变换,然后选择几乎包含所有波段 95% 以上信息的前 3 个主成

2012 年设计的一种深度卷积神经网络模型^[15],鉴于该模型层数不是很深,并且有很好的分类性能,因此本文以 AlexNet 模型为基础,构建了适用于遥感图像分类的 CNN 模型,其模型结构如图 2 所示。

分 (PCA_1 、 PCA_2 、 PCA_3) 作为变换后的原始图像。

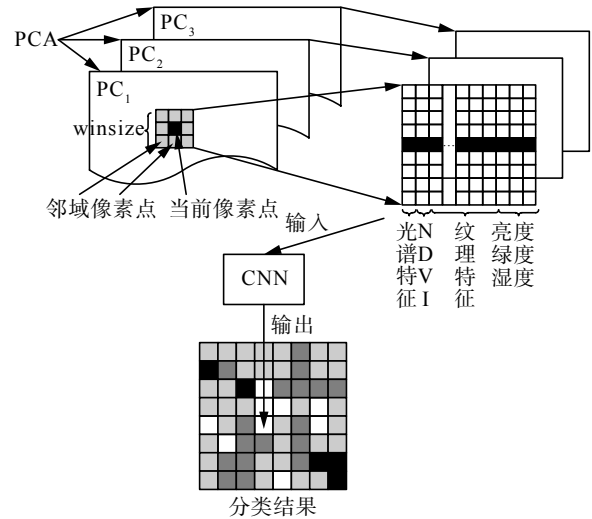


图 3 基于 CNN 模型的分分类方法流程

Fig. 3 Flowchart of the CNN-based classification approach

接着提取训练样本对应的光谱值组成一维光谱特征向量 A ,同时计算 NDVI,组成一维特征向量 B 。

其次对每幅图像计算灰度共生矩阵 (gray level co-occurrence matrix, GLCM)^[16],并基于 GLCM 提取均值、方差、熵、角二阶距、相关性、相异性、对比度和协同性共 8 种二阶概率统计的纹理滤波,按照提取顺序将其组成纹理特征矩阵 C 。

最后对图像进行 K-T 变换,提取亮度、绿度和湿度 3 个分量的数据组成特征矩阵 D 。

按照图 3 所描述的多源多特征融合方法,将 A 、 B 、 C 、 D 按照 $[A \ B \ C \ D]$ 组成一个大小为 $9 \times 13 \times 3$ 的特征融合矩阵,并将此矩阵输入 CNN 中进行特征学习,最后进行分类处理。

3 实验与分析

3.1 实验环境

本文采用 TensorFlow1.1.0 开源框架,搭建环境为个人 PC,操作系统为 Ubuntu16.04,处理器为 Intel(R) Core(TM) i5-4440 CPU@3.10 GHz,运行内存为 8 GB。

3.2 实验数据及样本选取

研究区域选择位于黑龙江省松嫩平原西部,乌裕尔河下游的扎龙湿地自然保护区 ($46^{\circ}52' \sim 47^{\circ}32'N$, $123^{\circ}47' \sim 124^{\circ}37'E$)。根据该地区实际地物分布情况,分为耕地、草地、湿地、水域、居民区以及裸地共 6 类。本文使用 2016 年 9 月的 Landsat-8 卫星遥感图像,共选择 16 160 个像素级训练样本,其中每类地物中随机选择 1/4 的数据样本作为验证集,剩余数据样本作为训练集。

3.3 实验结果及分析

为了验证本文方法的有效性,实验将与其他文献中的方法进行对比,如 SVM、NN、RF、DBN 和 CNN(Patch)。其中 CNN(Patch) 方法是一种基于区域块的 CNN 模型,该模型是以像素点周围 5×5 大小的邻域块作为单个样本的输入,可看作是为样本加入了邻域信息。除此之外,在本文方法的基础上,又设计了另外两种对比实验模型,一种是只将光谱特征作为 CNN 的输入 (CNN(ST)),该模型是为了验证本文的多源多特征融合方法的有效性;另外一种是不采用 PCA 的多源多特征融合模型 (CNN(SST)),该模型主要是为了验证 PCA 是否能够缩短模型的训练时间,加快收敛速度。

表 1 展示了本文方法分类结果的混淆矩阵,由此可知, CNN(PCA) 最终取得了 97.83% 的总体正确率, Kappa 系数达到了 0.973 6。对于单类地物, PA 和 UA 都超过了 90%, 某些地物的分类精度达到了最优,例如水域和裸地,两者的 PA 和 UA 都达到了 99% 以上。但是对于耕地类别来说,分类精度相对差一些,本文模型将其错分为湿地类别的样本数量较多,分析原因是农作物的光谱值范围和湿地的光谱值范围存在重叠,即所谓的“异物同谱”现象严重,导致模型难以有效区分耕地和湿地。

表 1 CNN(PCA) 分类结果的混淆矩阵

Table 1 Confusion matrix of CNN(PCA) classification results

类别	耕地	草地	湿地	水域	居民区	裸地	总和	PA/%	UA/%
耕地	647	11	39	0	3	0	700	92.43	98.33
草地	3	596	1	0	0	0	600	99.33	97.70
湿地	1	3	576	0	0	0	580	99.31	93.51
水域	0	0	0	520	0	0	520	100.00	100.00
居民区	6	0	0	0	593	1	600	98.83	99.33
裸地	1	0	0	0	1	230	232	99.14	99.57
总和	658	610	616	520	597	231	—	—	—

表 2 展示了本文方法与其他分类方法的实验结果对比,从表中可知,本文方法取得了最优的分类效果。相比于 SVM、NN 和 RF, CNN(PCA) 总体精度分别提高了约 13.61%、9.34% 和 7.3%, Kappa 系数提高了约 20.42%、13.16% 和 10.02%,这说明 CNN 的分类效果要远远优于浅层分类算法,这得益于 CNN 独特的结构,如局部连接、权值共享、池化等。这些特点使得 CNN 拥有某种尺度的位移、尺度和形变不变性,其强大的学习能力和容错能力使得 CNN 能够自动学习到更抽象、更具代表性的特征,从而获得更高的分类精

度。而浅层分类算法却不能在信息不充分或信息量少时从原始样本中获得更多更有用的信息。

同样,相比于其他深度学习方法, CNN(PCA) 的分类精度也要高于 DBN 和 CNN(Patch)。分析原因是 DBN 采用的是无监督的方式逐层训练网络,最后采用有监督的方式进行微调,这种逐层训练方式使得网络参数存在更大的随机性,不利于网络的整体优化。而对于 CNN(PCA),该方法虽然也是采用 CNN 模型,但是输入的信息只考虑了样本周围的光谱信息,而没有考虑图像的纹理特征,恰恰纹理特征最能反映不同类别之间的差

异, 所以导致分类效果较差。从另一个方面也能说明这一点, 如本文设计的对比实验模型 CNN(ST) 和 CNN(SST), 前者只考虑了样本的光谱特征, 而后者将光谱、纹理特征相融合, 这使得 CNN(SST)

比 CNN(ST) 总体精度提高了约 7.2%, Kappa 系数提高了约 5.29%。而且, CNN(SST) 方法能有效改善 CNN(ST) 中严重的“椒盐”现象, 对比效果如图 4 所示。

表 2 不同分类方法分类效果比较

Table 2 Comparison of different classification methods

方法	类别/%						OA/%	Kappa
	耕地	草地	湿地	水域	居民区	裸地		
SVM ^[17]	50.29	79.84	97.76	100.00	85.33	99.57	84.22	0.808 5
NN ^[18]	63.00	92.00	98.45	100.00	91.33	98.26	88.49	0.860 4
RF ^[19]	66.29	91.17	99.83	100.00	99.00	95.69	90.53	0.884 9
DBN ^[1]	81.57	93.83	94.83	100.00	94.83	99.57	92.95	0.917 1
CNN(Patch) ^[20]	67.00	98.00	97.59	100.00	92.17	95.26	90.25	0.881 8
CNN(ST)	58.71	96.83	95.86	100.00	88.50	93.97	87.14	0.884 3
CNN(SST)	83.14	96.17	97.93	100.00	95.17	99.57	94.34	0.931 1
CNN(PCA)	92.43	99.33	99.31	100.00	98.83	99.14	97.83	0.973 6

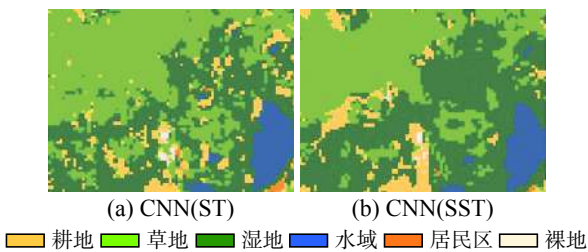


图 4 分类效果对比

Fig. 4 Comparison of classification results

图 5 和图 6 分别显示了本文设计的对比模型关于训练时间和训练精度的对比结果。由图 5 可知, 采用 PCA 变换后, 相比不使用 PCA 的 CNN (SST) 方法训练时间减少了约 80%, 不仅训练时间

大幅度缩短, 同时总体精度也相对提高了约 3.49%。这与 CNN 模型本身的网络结构有关, 输入的样本信息维度越高, 模型需要训练的参数个数就会呈几何式增长, 从而导致训练速度慢, 训练时间长。而由图 6 可知, CNN(PCA) 大约在迭代 300 步时模型性能趋于稳定, 而 CNN(ST) 和 CNN (SST) 分别在 2 000 步和 1 100 步以后才趋于稳定, 这说明 PCA 降维以后, 模型收敛的速度更快, 性能更好。原因是经过 PCA 变换以后, 使得类间差距增大, 类内差距缩小, 同时能够在一定程度上消除样本中的噪声信息, 从而加快了模型收敛速度, 提高了总体分类精度。图 7 为 CNN(PCA) 方法的最终分类效果图。

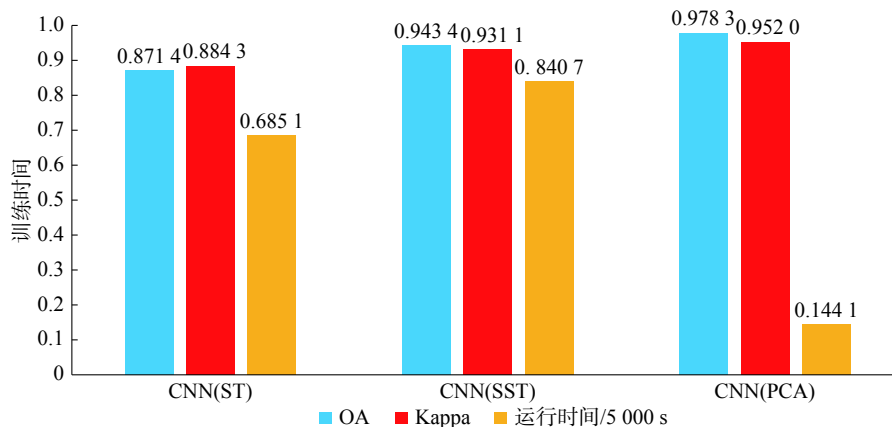


图 5 不同方法性能对比

Fig. 5 Performance comparison of different methods

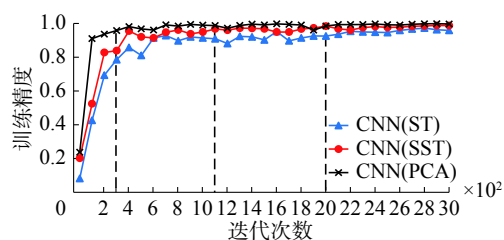


图6 不同方法训练精度对比

Fig. 6 Training accuracy comparison of different methods

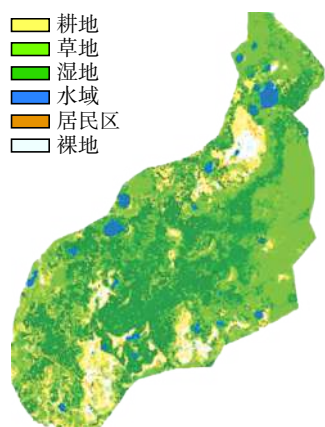


图7 CNN(PCA)分类效果

Fig. 7 Classification results of CNN(PCA)

4 结束语

本文提出了一种基于卷积神经网络的遥感图像分类方法,并针对单源特征无法提供更多有用信息的问题,设计了一种多源多特征融合方法,该方法能够有效融合遥感图像的光谱特征、纹理特征和空间结构特征。通过与SVM、NN、RF、DBN和CNN(Patch)等分类方法进行比较,说明CNN的局部连接、权值共享和池化等独特结构使得它能够自动挖掘遥感图像的空间分布规律,学习到遥感图像更多更本质的特征,从而获得最优的分类准确率。此外,通过本文设计的对比模型CNN(ST)和CNN(SST),验证了本文设计的多源多特征融合方法能为CNN提供更多的有用信息,在加快模型收敛速度的同时进一步提高了模型的分精度。

虽然通过本文研究可以发现CNN分类方法能够有效提高遥感图像的分类精度,但是作为一种新的机器学习方法,仍然存在很多不完善的地方,主要体现在网络模型的参数多、训练时间长;网络模型的结构设计没有完善的理论支撑,只能通过反复实验来选取最优参数。下一步,将以如何提高模型训练速度和寻找最优参数作为研究的方向。

参考文献:

- [1] 刘大伟, 韩玲, 韩晓勇. 基于深度学习的高分辨率遥感影像分类研究[J]. 光学学报, 2016, 36(4): 0428001.
- LIU Dawei, HAN Ling, HAN Xiaoyong. High spatial resolution remote sensing image classification based on deep learning[J]. Acta optica sinica, 2016, 36(4): 0428001.
- [2] ALBERGA V. A study of land cover classification using polarimetric SAR parameters[J]. International journal of remote sensing, 2007, 28(17): 3851-3870.
- [3] HAGNER O, REESE H. A method for calibrated maximum likelihood classification of forest types[J]. Remote sensing of environment, 2007, 110(4): 438-444.
- [4] NIU Xin, BAN Yifang. Multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using an object-based support vector machine and a rule-based approach[J]. International journal of remote sensing, 2013, 34(1): 1-26.
- [5] HEERMANN P D, KHAZENIE N. Classification of multispectral remote sensing data using a back-propagation neural network[J]. IEEE transactions on geoscience and remote sensing, 1992, 30(1): 81-88.
- [6] PAL M, MATHER P M. An assessment of the effectiveness of decision tree methods for land cover classification [J]. Remote sensing of environment, 2003, 86(4): 554-565.
- [7] BENGIO Y. Learning deep architectures for AI[J]. Foundations and trends in machine learning, 2009, 2(1): 1-127.
- [8] 杜培军, 夏俊士, 薛朝辉, 等. 高光谱遥感影像分类研究进展[J]. 遥感学报, 2016, 20(2): 236-256.
- DU Peijun, XIA Junshi, XUE Zhaohui, et al. Review of hyperspectral remote sensing image classification[J]. Journal of remote sensing, 2016, 20(2): 236-256.
- [9] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [10] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.
- [11] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: a unified embedding for face recognition and clustering[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA, 2015: 815-823.
- [12] ZHANG Yu, BAI Xiangzhi, WANG Tao. Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure[J]. Information fusion, 2017, 35: 81-101.
- [13] TATULLI E, HUEBER T. Feature extraction using multimodal convolutional neural networks for visual speech

- recognition[C]//Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing. New Orleans, USA, 2017: 2971–2975.
- [14] 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述[J]. 计算机应用, 2016, 36(9): 2508–2515.
- LI Yandong, HAO Zongbo, LEI Hang. Survey of convolutional neural network[J]. Journal of computer applications, 2016, 36(9): 2508–2515.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84–90.
- [16] HARALICK R M, SHANMUGAM K, DINSTEN I. Texture features for image classification[J]. IEEE transactions on systems, man and cybernetics, 1973, 3(6): 610–621.
- [17] 张策, 臧淑英, 金竺, 等. 基于支持向量机的扎龙湿地遥感分类研究[J]. 湿地科学, 2011, 9(3): 263–269.
- ZHANG Ce, ZANG Shuying, JIN Zhu, et al. Remote sensing classification for Zhalong Wetlands based on support vector machine[J]. Wetland science, 2011, 9(3): 263–269.
- [18] YUAN H, VAN DER WIELE C F, KHORRAM S. An Automated artificial neural network system for land use/land cover classification from Landsat TM imagery[J]. Remote sensing, 2009, 1(3): 243–265.
- [19] NA Xiaodong, ZANG Shuying, LIU Lei, et al. Wetland mapping in the Zhalong National Natural Reserve, China, using optical and radar imagery and topographical data[J]. Journal of applied remote sensing, 2013, 7(1): 073554.
- [20] SHARMA A, LIU Xiuwen, YANG Xiaojun, et al. A patch-based convolutional neural network for remote sensing image classification[J]. Neural networks, 2017, 95: 19–28.

作者简介:



李亚飞, 男, 1992 年生, 硕士研究生, 主要研究方向为深度学习。



董红斌, 男, 1963 年生, 教授, 博士生导师, 主要研究方向计算智能、机器学习、数据挖掘和多 Agent 系统。