

残差网络分层融合的高光谱地物分类

张怡卓, 徐苗苗, 王小虎, 王克奇*

东北林业大学机电工程工程学院, 黑龙江 哈尔滨 150040

摘要 高光谱图像具有较高的空间分辨率, 蕴含着丰富的空间光谱信息, 近年来被广泛用于城市地物分类中。在高光谱图像分类过程中, 空间光谱特征的提取直接影响着分类精度; 传统的高光谱图像特征提取方法只利用了4或8邻域的像素进行简单卷积处理, 因而丢失了大量的复杂、有效信息; 卷积神经网络(CNN)虽然可以自动提取空间光谱特征, 在保留图像空间信息的同时, 简化网络模型, 但是, 随着网络深度增加, 网络分类产生退化现象, 而且网络间缺乏相关信息的互补性, 从而影响分类精度。该工作引入CNN自动提取空间光谱特征, 并且针对CNN深度增加所导致的退化问题, 设计了面向地物分类的高光谱特征融合残差网络。首先, 为了降低高光谱图像的光谱冗余度, 利用PCA提取主要光谱波段; 然后, 为了逐级提取光谱图像的空间光谱特征, 定义了卷积核为16, 32, 64的低、中、高三层残差网络模块, 并利用64个 1×1 的卷积核对3层特征输出进行卷积, 完成维度匹配与特征图融合; 接着, 对融合后的特征图进行全局平均池化(GAP)生成用于分类的特征向量; 最后, 引入具有可调节机制的Large-Margin Softmax损失函数, 监督模型完成训练过程, 实现高光谱图像分类。实验采用Indian Pines, University of Pavia和Salinas地区的高光谱图像来验证方法有效性, 设置批次训练的样本集为100, 网络训练的初始学习率为0.1, 当损失函数稳定后学习率降低为0.001, 动量为0.9, 权重延迟为0.0001, 最大训练迭代次数为 2×10^4 , 当3个数据集的样本块像素分别设置为 25×25 , 23×23 , 27×27 , 网络深度分别为28, 32和28时, 3个数据集的分类准确率最高, 其平均总体准确率(OA)为98.75%、平均准确率(AA)的评价值为98.1%, 平均Kappa系数为0.98。实验结果表明, 基于残差网络的分类方法能够自动学习更丰富的空间光谱特征, 残差网络层数的增加和不同网络层融合可以提高高光谱分类精度; Large-Margin Softmax实现了类内紧凑和类间分离, 可以进一步提高高光谱图像分类精度。

关键词 高光谱图像分类; 深度学习; 残差网络; Large-Margin Softmax

中图分类号: P407.8 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)11-3501-07

引言

高光谱图像含有丰富的光谱信息, 这些信息是由可见光谱到红外光谱的数百个光谱带组成。高光谱图像的每个像素点通过高维向量来表示, 对应于特定波长的光谱反射率。近年来, 国内外学者提出了多种高光谱图像分类算法。Jia等^[1]将SVM用于高光谱图像分类, 但是, SVM只利用了高光谱图像的光谱信息, 没有考虑图像的空间信息。随着稀疏理论的发展, 稀疏表达在高光谱图像分类中得到广泛研究。2011年, Chen等^[2]提出联合稀疏表示模型来进行高光谱图像分类, 该模型通过空间信息确定待测像元的类别, 有效地融合

了高光谱图像的空间光谱特征, 但是, 当地物种类丰富时, 容易出现错分现象。2014年, Zhang等^[3]提出了非局部加权联合稀疏分类模型, 针对邻域像元对中心像元贡献值存在的差异进行了研究。传统机器学习方法在高光谱图像分类中得到了很大的进展, 但是需要复杂的特征提取过程, 因而分类精度有待于进一步提高。

深度学习通过分层网络获取分层次的特征信息, 简化了复杂的人工设计特征过程, 成为高光谱图像分类的研究热点。卷积神经网络(CNN)是深度学习的一个重要分支, 在高光谱图像分类中的应用越来越广泛^[4-5]。2015年, Hu等^[6]利用CNN网络模型对高光谱图像进行分类, 利用CNN的局部连接、权值共享等特性, 减少了模型参数, 降低了训练难度,

收稿日期: 2019-01-13, 修订日期: 2019-04-20

基金项目: 林业公益性行业科研专项(21504307)资助

作者简介: 张怡卓, 1978年生, 东北林业大学机电工程工程学院教授 e-mail: nefuzyz@163.com

* 通讯联系人 e-mail: zdhwkq@163.com

提高了分类性能。2016年, Zhao等^[7]通过 CNN 和平衡局部判别嵌入算法提取高光谱图像的空间和光谱特征来进行高光谱图像分类;但是,基于 CNN 模型分类算法仅包含单一卷积层,降低了其特征学习的能力。通过增加网络深度可以提高网络的性能,尤其处理具有非常复杂的空间光谱特性的高光谱图像。但是,网络层数过多导致训练误差增大,即退化现象,使准确率达到一定值后迅速下降。2018年, Zhong等^[8]提出了基于残差网络的高光谱图像分类方法,网络中的残差块通过恒等映射连接其他卷积层,并且在每个卷积层上进行批量归一化,解决了网络层数增加带来的负担,在高光谱图像分类中具有很好的效果。残差网络利用残差块学习有判别力的特征,但是没有充分利用浅层和深层的互补和相关信息;而且,传统的 softmax 分类器没有明确地鼓励类内紧凑性和类间可分离性^[9],在进行复杂地物的分类时精确度不高。基于深度学习的融合策略在图像分类中得到广泛研究^[10-12]。本工作引入残差网络进行特征提取,并设计了残差网络中不同层次特征融合策略,通过 Large-Margin Softmax 分类器实现高光谱图像的有效分类。

1 实验部分

1.1 基于 CNN 的高光谱图像分类

图 1 为高光谱图像的 CNN 特征提取和分类流程图,为充分利用空间信息,提取以待分类像素为中心的样本作为输入,经过卷积、池化和全连接一系列的操作,输入到 softmax 分类器得到分类结果。定义样本块经过卷积和池化的输出特征向量为 x , softmax 输出该样本块在各个类别的概率 p_j 如式(1)所示

$$p_j = \frac{e^{f_j}}{\sum_{c=1}^C e^{f_c}}, j = 1, \dots, C \quad (1)$$

其中, C 为类别总数; p_j 的最大值为该样本块的所属类别; f_j 表示全连接层输出向量的第 j 个元素,其表达式如式(2)所示

$$f_j = W_j^T x = \|W_j\| \|x\| \cos(\theta_j) \quad (2)$$

其中, W 表示网络权值。

网络的损失函数表示为

$$L = -\log \left(\frac{e^{\|W_j\| \|x\| \cos(\theta_j)}}{\sum_{c=1}^C e^{\|W_c\| \|x\| \cos(\theta_c)}} \right), c = 1, \dots, C \quad (3)$$

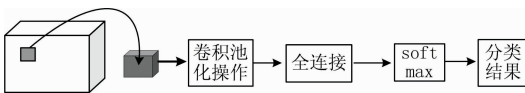


图 1 基于 CNN 的高光谱图像分类流程图

Fig. 1 The flow chart of hyperspectral image classification based on CNN

虽然 CNN 在图像分类中有着广泛的应用,但是,仅利用几个简单的卷积和池化层,不能充分提取高光谱图像深层的特征以进行准确分类。虽然可以通过增加网络深度提高 CNN 的性能,但是过度增加网络深度将会使网络产生退化现象,会导致 CNN 出现过拟合、梯度消失和梯度爆炸等问题。

1.2 残差网络

残差网络为解决深层 CNN 出现的退化问题将深层网络后面的卷积层变为恒等函数。让一些层去拟合一个恒等函数很困难,所以把网络设计为

$$H(x) = F(x) + x \quad (4)$$

其中, x 和 $H(x)$ 分别表示一个残差块的输入和输出, $F(x)$ 是残差块学习的函数。

学习一个恒等映射函数转换为学习一个残差函数

$$F(x) = H(x) - x \quad (5)$$

当 $F(x) = 0$ 时,就构成了恒等映射 $H(x) = x$ 。由于恒等映射没有引入额外的参数,不会给网络增加额外的参数和计算量,同时能够加快模型训练速度、提高训练效果。

残差网络使用 shortcut 回路实现两个卷积层的优化^[10],即在两个卷积层的基础上叠加 $y = x$ 的恒等映射,将输入和输出相加。加入 shortcut 后,训练过程中误差可以通过恒等映射向上层传播,减弱层数过多造成的梯度消失现象,残差块的基本结构如图 2 所示。但是,该网络忽视了不同网络层之间信息的互补性与相关性,影响分类精度。

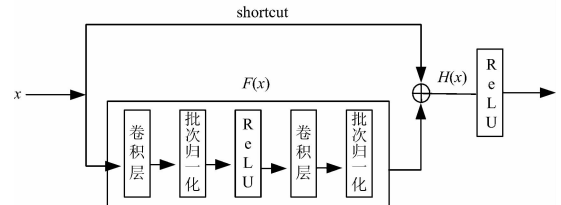


图 2 残差块的基本结构

Fig. 2 The structure of the residual block

1.3 基于残差网络分层融合的 Large-Margin Softmax 高光谱图像分类

CNN 中不同卷积层可以学习不同的特征,浅层具有图像细节信息,深层含有更多抽象信息。为了充分利用不同层次特征信息,将浅层特征与深层特征进行融合,可以增强模型对复杂图像的表现能力,图 3 为残差融合网络设计结构。

根据卷积核的数量将残差网络分为 3 个模块,其中 O_L , O_M 和 O_H 分别表示 3 个模块的输出;在 3 个模块中分别具有 16, 32, 64 个卷积核,输出不同维度的特征图,实现特征信息的逐层提取。为确保特征融合前具有相同的维度,经过数量为 64,大小为 1×1 的卷积核进行卷积完成维度匹配。最后,将 3 个模块的输出相加融合。激活之后的融合特征图进行全局平均池化(GAP),生成最终的特征图,融合过程如图 4 所示。

最后,将融合的特征图经过全连接网络转换成输出的特征向量,输入到分类器进行分类。传统的 softmax 分类器不能很好地学习到使类内紧凑、类间分散的特征。所以,引入 Large-Margin Softmax 分类器。

以二分类问题为例,假设一个样本块属于类别 1,由式(1)和式(2)可知,softmax 是通过让 $p_1 > p_2$,即

$$f_1 = W_1^T x > f_2 = W_2^T x \quad (6)$$

来进行分类,可以写为

$$\|W_1\| \|x\| \cos(\theta_1) > \|W_2\| \|x\| \cos(\theta_2) \quad (7)$$

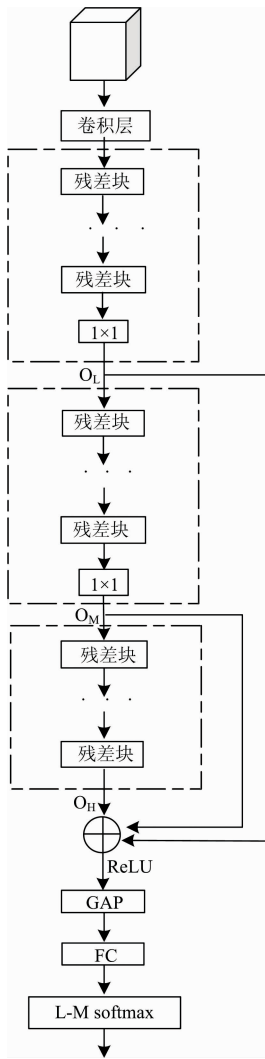


图 3 基于残差网络多层融合分类算法流程图

Fig. 3 Flow chart of multi-layer fusion classification algorithm based on residual network

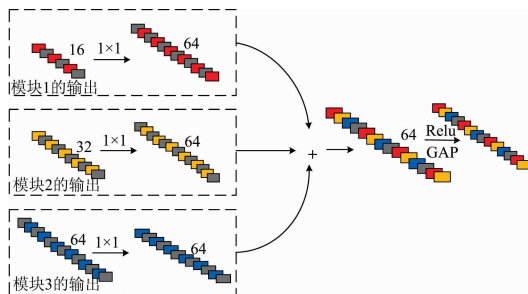


图 4 多层特征融合过程

Fig. 4 Multi-layer feature fusion process

而 Large-Margin softmax 是将其变为

$$\begin{aligned} \|W_1\| \|x\| \cos(\theta_1) &\geq \|W_1\| \|x\| \cos(m\theta_1) > \\ &\|W_2\| \|x\| \cos(\theta_2) \end{aligned} \quad (8)$$

其中, m 是正整数, $0 \leq \theta_1 \leq \frac{\pi}{m}$ 。

通过增加一个正整数变量 m , 产生一个决策余量, 能够更加严格地约束式(7)。

Large-Margin Softmax 损失函数表示为

$$L = -\log\left(\frac{e^{\|W_j\| \|x\| \Psi(\theta_j)}}{e^{\|W_j\| \|x\| \Psi(\theta_j)} + \sum_{c \neq j} e^{\|W_c\| \|x\| \cos(\theta_c)}}\right) \quad (9)$$

其中 $\Psi(\theta)$ 为

$$\Psi(\theta) = \begin{cases} \cos(m\theta), & 0 \leq \theta \leq \frac{\pi}{m} \\ \cos\left(\frac{\pi}{m}\right), & \frac{\pi}{m} < \theta \leq \pi \end{cases} \quad (10)$$

通过调节间隔 m 的值来调节分类边界, 式(8)对学习 W_1 和 W_2 提出了更高的要求, 使模型学习到类内距离更小、类间距离更大的可区分性的特征。

2 结果与讨论

2.1 实验数据

实验选用的高光谱数据为 Indian Pines, University of Pavia 和 Salinas 3 个典型的数据源。其中, Indian Pines 是美国印第安纳州西北部农区的高光谱图像, 该数据有 10 249 个地物像素, 包括苜蓿 (Alfalfa)、玉米 (Corn)、燕麦 (Oats)、小麦 (Wheat)、木材 (Woods) 等 16 个地物类别; University of Pavia 是意大利帕维亚城的一部分高光谱数据, 该数据有 42 776 个用于分类的地物像素, 包含树、沥青道路 (asphalt)、牧场 (meadows)、砾石 (gravel)、裸土 (bare soil) 等 9 类地物; Salinas 是美国加利福尼亚州的 Salinas 山谷的高光谱图像, 该图像有 16 类 54 129 个像素, 包括休耕 (fallow)、残株 (stubble)、葡萄 (grapes)、芹菜 (celery)、生菜 (lettuce) 等。

图 5 是 3 个地区分别抽取 3 个波段叠加所成的伪彩色图像。实验所用计算机配置为 Intel(R) Core™ i5-4460 处理器, NVIDIA GeForce GTX 1070 Ti 显卡; 在 Linux 操作系统下基于 caffe 框架运行实验。在进行数据降维时, 将 3 个高光谱图像经过 PCA 处理, 提取主要的光谱波段。将 3 个数据集中被标记的像素分成训练集和测试集, 分别选择 10%, 2% 和 0.5% 的样本进行训练, 其余作为测试样本。设置批次训练的样本集为 100, 网络训练的初始学习率为 0.1, 当损失函数稳定后学习率降低为 0.001, 动量为 0.9, 权重延迟为 0.000 1,

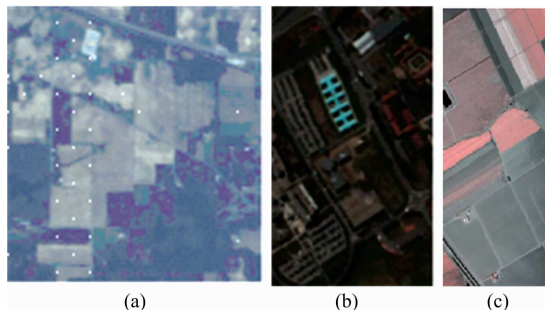


图 5 三类样本伪彩色图像

(a): 印第安农场; (b): 帕维亚大学; (c): 萨利纳斯

Fig. 5 Pseudocolored images

(a): Indian Pines; (b): University of Pavia; (c): Salinas

最大训练迭代次数为 2×10^4 。

2.2 结果与分析

图 6 为 3 个数据集基于残差网络的特征融合算法训练过程中, 样本块大小分别在 25×25 , 23×23 和 27×27 , 网络深度分别为 28, 32 和 28 时, 随迭代次数增加损失函数值和准确率随迭代次数变化的曲线图。

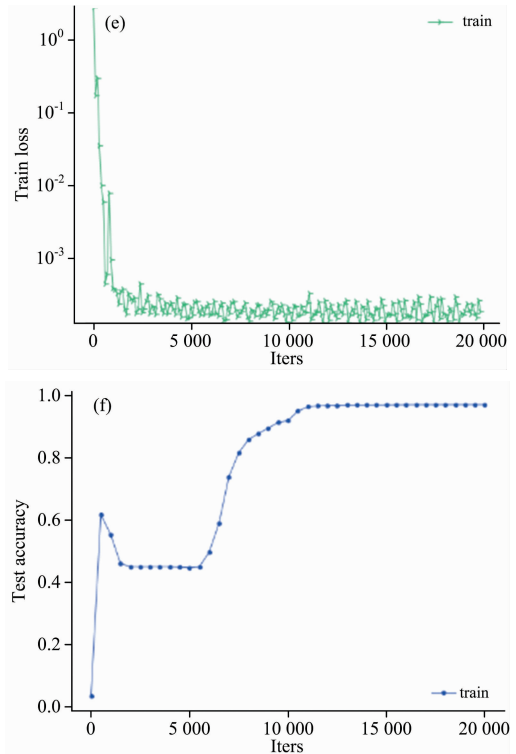
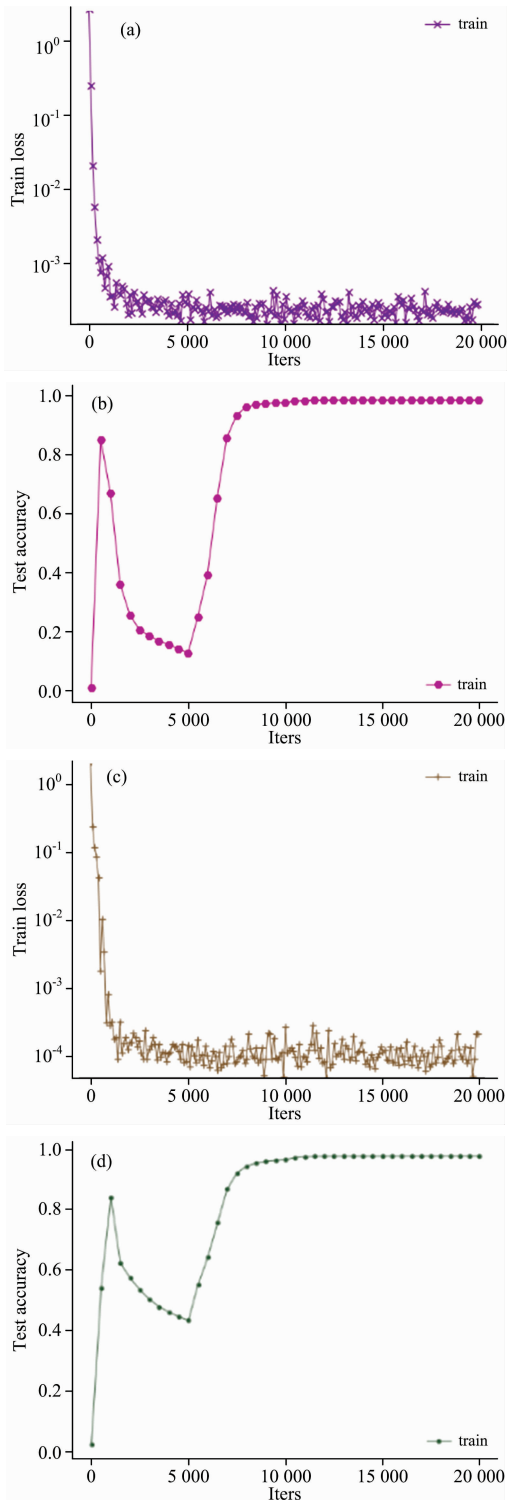


图 6 损失函数(a, c, e)和准确率(b, d, f)变化曲线图

(a), (b): Indian Pines 样本大小为 25×25 , 深度为 28;

(c), (d): PaviaU 在样本大小为 23×23 , 深度为 32;

(e), (f): Salinas 在样本大小为 27×27 , 深度为 28

Fig. 6 Loss function (a, c, e) and accuracy curve (b, d, f)

(a), (b): Indian Pines with a sample size of 25×25 and a depth of 28;

(c), (d): PaviaU with a sample size of 23×23 and a depth of 32;

(e), (f): Salinas with a sample size of 27×27 and a depth of 28

由图 6(a), (c)和(e)可以看出, 当迭代次数达到 2 500 时, 损失函数开始趋于稳定, 比较图 6(b), (d)和(f), 准确率呈现不同程度的下降, 说明随着网络层数的增加, 出现了退化问题。当迭代次数超过 10 000 时, 由于残差块的学习, 将输入直接传到输出, 保护了信息的完整性, 解决了退化问题, 准确率达到最大并趋于稳定。

2.3 网络参数分析

样本块大小和网络层数重要的网络参数。样本块大小会影响分类精度, 样本块过大, 会忽视一些细节信息; 样本块过小, 也会放大冗余信息, 降低分类精度。增加网络深度在某种程度上能够提高分类精度; 但是, 层数过多会使网络产生过拟合、梯度消失和梯度爆炸等问题。在此, 通过交替迭代实验对这两个参数的确定进行分析。实验过程中, 样本块大小设置范围为 17×17 到 31×31 , 网络深度设置在 8~36 之间。为定量评估不同参数的表现, 使用总体准确率(OA)、平均准确率(AA)和 Kappa 系数三个评价指标对分类精度进行评价, 其中, OA 是对分类结果质量的总体评价, AA 是所有类别分类精度的平均值, Kappa 系数反映了分类结果与参考数据之间的吻合程度。

图 7 为 3 个数据集的网络深度分别设为 28, 32 和 28 时,

样本块大小对分类精度的影响曲线。由图可以看出,随着样本块的增大,Indian Pines 分类的 OA 值变化较为平缓,Salinas 逐渐增大,而 University of Pavia 先增大后减小。对于不同的数据集,类内和类间的光滑度差异使得样本块大小对分类精度呈现不同的变化趋势,University of Pavia 的类别区域更加复杂。3 个数据集的样本块分别设为 25×25 , 23×23 和 27×27 时分类精度最高。

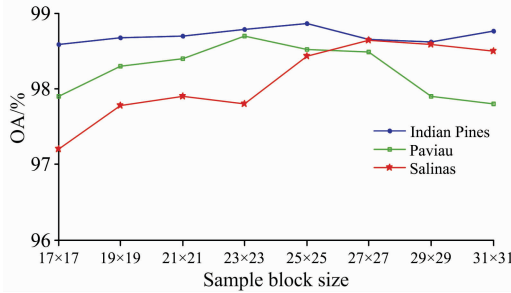


图 7 样本块大小对分类精度的影响曲线
Fig. 7 The effect of sample block size on classification rate

图 8 所示为 3 个数据集的样本块大小分别设为 25×25 , 23×23 , 27×27 时,网络深度对分类精度的影响曲线图。由图 8 可以看出,在网络层数 8~36 之间时,分类精度都达到 96% 以上,证明本工作提出的网络能有效平衡网络深度和分类的表现。3 个数据集的网络深度分别设为 28, 32 和 28 时,分类精度达到最高。

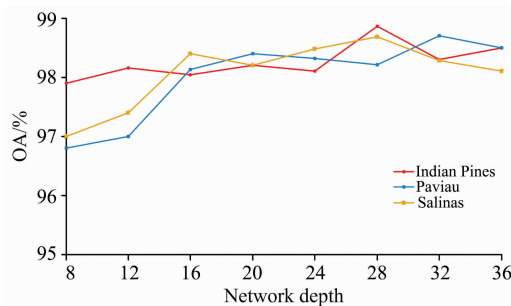


图 8 网络深度对分类精度的影响曲线
Fig. 8 The effect of network depth on classification accuracy

2.4 方法比较

为验证基于残差网络分层融合分类算法的性能,将该算法与常用的 SVM 算法、卷积神经网络 CNN 和深度残差网络 DRN 高光谱分类方法进行比较。其中,SVM 算法利用了高光谱的像素作为输入,使用具有高斯核的支持向量机库来实现分类;CNN 是经过浅层的卷积层后输入到分类器进行分类;而 DRN 采用了残差网络。比较结果如表 1 所示,相应的网络训练和测试时间如图 9 所示。

通过比较表 1 中的 OA 可知,与传统的机器学习算法 SVM 相比,CNN 通过卷积层获取分层次的空间光谱特征,减少了人工提取特征的经验性误差,OA 值分别提高了 4.9%, 7.8% 和 2.5%,具有更好的分类效果;但是,卷积神

经网络得到的特征数目多,处理时间长,大大增加了训练时间。残差网络的引入,解决了网络层数增加带来的负担,同时提取到深层次的空间光谱特征,将分类精度提高到 95% 左右;但是网络深度的增加,使训练时间明显增大。通过比较 DRN 和本算法可以看出,分层融合机制融合了不同卷积层输出的互补与相关信息,得到更有利于分类的特征,使分类精度达到 98% 以上,证明了特征融合的有效性。实验结果表明,使用 Large-Margin Softmax,能够提取更有区分性的特征,使分类精度得到进一步提高。从图 9 可以看出,网络结构的改变,对训练时间影响较大,但是对测试时间影响较小。

表 1 SVM, CNN, DRN softmax 和 large-margin softmax 的平均分类精度

Table 1 The average classification rates of SVM, CNN, DRN, softmax and large-margin softmax

精度	OA/%	AA/%	Kappa
SVM	79.82	76.89	0.79
CNN	91.53	91.02	0.88
DRN	94.07	94.76	0.93
Softmax	98.34	98.00	0.98
Large-margin Softmax	98.74	98.10	0.98

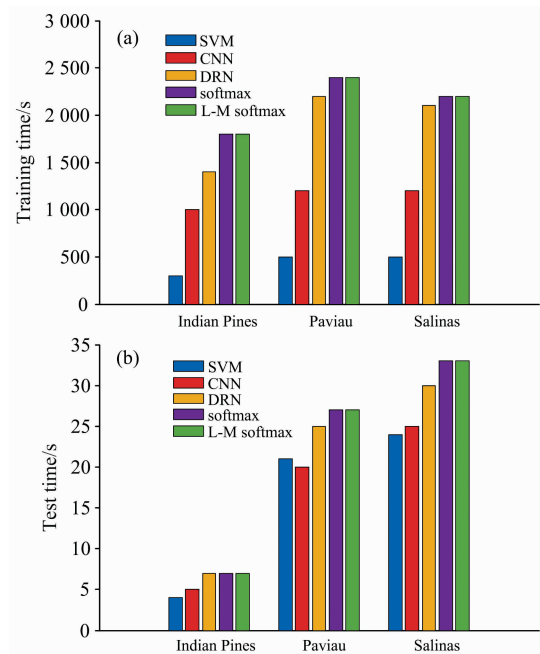


图 9 不同网络结构在 3 个数据集上训练与测试时间的对比
Fig. 9 Comparison of training and test time on 3 data sets with different network structures

3 结论

高光谱图像含有丰富的空间光谱信息,但是在分类过程中复杂特征通常不能被充分利用。在高光谱图像分类的方法

中, SVM、多元逻辑回归等传统的机器学习算法特征提取过程较为复杂, 而简单的 CNN 不能充分提取高光谱图像的深层光谱与空间特征, 并且没有充分利用多个网络层的互补与相关信息。为解决高光谱数据特征丰富难以提取和特征信息利用不充分的问题, 设计了一种基于残差网络融合的 Large-Margin Softmax 进行高光谱图像分类的新方法。通过对各种分类方法比较分析发现, 3 个数据集的平均总体准确率(OA)为 98.75%、平均准确率(AA)为 98.1%, 平均 Kappa 系数为

0.98, 实验结果证明基于残差网络融合分类算法可以获得较好的效果。

基于残差网络分层融合的 Large-Margin Softmax 分类方法可以提取高光谱图像的有效空间光谱特征, 解决了网络层数增加带来的负担, 为高光谱图像分类提供了新思路, 但是该方法仍然存在耗时较长的问题。所以, 在高光谱图像分类的应用中, 残差网络和分类器的设计存在巨大的潜力, 如何提高分类的速度还有待进一步地研究。

References

- [1] Fauvel M, Tarabalka Y, Benediktsson J A, et al. Proceedings of the IEEE, 2013, 101(3): 652.
- [2] Chen Y, Nasrabadi N M, Tran T D. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(10): 3973.
- [3] Zhang H, Li J, Huang Y, et al. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(6): 2056.
- [4] Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012: Neuram Information Processing Systems, 2012.
- [5] Chen Yushi, Jiang Hanlu, Li Chunyang, et al. IEEE Trans Geoscience and Remote Sensing, 2016, 54(10): 6232.
- [6] Hu W, Huang Y, Wei L, et al. Journal of Sensors, 2015(2): 1.
- [7] Zhao W Z and Du S H. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54 (8): 4544.
- [8] Zhong Z, Li J, Luo Z, et al. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(2): 847.
- [9] Liu W, Wen Y, Yu Z, et al. Large-Margin Softmax Loss for Convolutional Neural Networks. Proceedings of 33rd International Conference on Machine Learning, 2016: 507.
- [10] Chai S, Liu H, Gu Y, et al. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(8): 4775.
- [11] Zhang X, Zhang H, Zhang Y, et al. IEEE Transactions on Image Processing, 2016, 25(3): 1033.
- [12] Song Weiwei, Li Shutao, Fang Leyuan, et al. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(6): 3173.

Hyperspectral Image Classification Based on Hierarchical Fusion of Residual Networks

ZHANG Yi-zhuo, XU Miao-miao, WANG Xiao-hu, WANG Ke-qi*

College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China

Abstract Hyperspectral images contain a wealth of feature information, and they have been widely used in urban feature classification in recent years. In the process of hyperspectral image classification, the extraction of spatial spectral features directly affects the classification accuracy. Traditional hyperspectral image feature extraction methods only use 4 or 8 neighborhood pixels for simple convolution processing, thus losing a lot of complex and effective information. Convolution neural network (CNN) can automatically extract spatial spectral features and retain the same spatial information of the image, and the network model is simplified. However, with the increase of network depth, the network classification will degenerate, and the network lacks complementarity of relevant information, which will affect the classification accuracy. In this paper, a hyperspectral residual network for feature classification is designed for the degradation problem. Firstly, define the residual network module of the low, medium and high three-layer structure with convolution kernels of 16, 32, and 64. Then, convolve the 3-layer output features with 64 1×1 convolution kernels to complete the dimension matching and feature map. Next, the global average pooling (GAP) of the feature map is generated to generate the feature vector for classification. Finally, the Large-Margin Softmax objective function is introduced to achieve hyperspectral image classification. The experiments were performed using hyperspectral images from the Indian Pines, University of Pavia, and Salinas regions. The primary bands of the hyperspectral image were extracted by PCA. With the sample set of batch training being 100, the initial learning rate being 0.1, the momentum being 0.9, the weight delay being 0.000 1, and the maximum number of training iterations being 2×10^4 , when the sample sizes of the three data sets are set to be 25×25 , 23×23 and 27×27 , the network depth is 28, 32 and 28, the classification accuracy of the three data sets is the highest, and the average overall accuracy OA is 98.75%, the average accuracy AA is 98.1% and the average Kappa coefficient is 0.98. The experimental results show that the classification method based on residual network can get more affective features.

It can improve the classification accuracy with the increase of the number of residual network layers and the fusion of complementary information of different network layer outputs; Large-Margin Softmax achieves intra-class compactness. Separation between classes further improves classification accuracy.

Keywords Hyperspectral image classification; Deep learning; Residual network; Large-Margin Softmax

(Received Jan. 13, 2019; accepted Apr. 20, 2019)

* Corresponding author