

文章编号 1004-924X(2018)12-3099-09

## 基于多尺度特征融合的遥感图像场景分类

杨 州\*, 慕晓冬, 王舒洋, 马晨晖

(火箭军工程大学 信息工程系, 陕西 西安 710025)

**摘要:**为了解决遥感图像场景分类中因样本量小而分类精度不高的问题,提出了一种基于多尺度特征融合(MSFF)的分类方法。首先,对遥感图像进行尺度变换,得到同一遥感源图像的多个不同尺度图像。接着,将其分别输入深度卷积神经网络(DCNN)中进行卷积操作。然后,将各卷积层和全连接层提取出的不同尺度特征进行降维和编码/平均池化操作。最后,将各尺度特征进行编码融合并利用多核支持向量机(MKSVM)进行场景分类。在两个公开遥感图像数据集UCM Land-Use和NWPU-RESISC45中进行试验,分类精度最高分别达到98.91%和99.33%。本文方法能够利用不同尺度的图像特征,结合低、中、高层语义表示,使融合特征的可辨识性更高,同时使用多核支持向量机提高了深度网络学习的泛化能力,因此分类效果更好。

**关键词:**遥感图像;场景分类;深度卷积神经网络;特征融合;多核支持向量机

**中图分类号:**TP751 **文献标识码:**A **doi:**10.3788/OPE.20182612.3099

### Scene classification of remote sensing images based on multiscale features fusion

YANG Zhou\*, MU Xiao-dong, WANG Shu-yang, MA Chen-hui

(Department of Information Engineering, Rocket Force Engineering  
University, Xi'an 710025, China)

\* Corresponding author, E-mail: yzmailbox2015@163.com

**Abstract:** To solve the low accuracy problem of remote sensing image scene classification due to small sample sizes, a classification method was proposed based on Multiscale Features Fusion (MSFF). First, the remote sensing images were scaled to obtain several different scale images of the same remote sensing image. Thereafter, they were inputted into a Deep Convolutional Neural Network (DCNN) for convolutional operation, and the different scale features of the convolutional and the fully connected layers were reduced and coded or average pooled. Finally, the different scale features were coded and fused, and a multikernel support vector machine was used to classify the scenes. In the two public remote sensing image data sets UCM Land-Use and NWPU-RESISC45, the highest classification accuracy of the experiment are 98.91% and 99.33%, respectively. This method can use image features of different scales and low, middle and high-level semantic representations combined, thus the fusion feature is more recognizable. Furthermore, the use of a multikernel support vector machine

收稿日期:2018-03-13;修订日期:2018-05-06.

基金项目:国家自然科学基金资助项目(No. 61601475);航空科学基金资助项目(No. 201555U8010)

improves the generalization of the deep network learning ability, so the classification effect is better.

**Key words:** remote sensing image; scene classification; deep convolutional neural network; features fusion; multi-kernel support vector machine

## 1 引言

随着遥感技术的不断发展进步,人们能够获得的遥感图像越来越多,分辨率也越来越高,图像中包含的信息更丰富、结构更复杂,在城市规划、自然灾害探测等领域发挥重要作用的同时,也因其同物异谱、异物同谱等问题的存在对场景分类等带来了更大的挑战。虽然已有许多分类方法被不同的研究者提出<sup>[1-10]</sup>,但大部分方法是基于图像的 low 中层特征,分类效果并不理想。其中,比较有名的方法之一就是视觉词袋法<sup>[4]</sup>,它是一种基于图像中层特征的视觉编码分类方法,后来的许多方法都是对视觉词袋方法的改进或扩展。

近几年,深度学习的方法因其强大的学习能力被成功地应用于许多领域<sup>[11-16]</sup>,不少学者将其应用在遥感场景分类中<sup>[17-23]</sup>,同样取得了不错的表现。但是,在这些研究过程中,只有部分研究者意识到深度学习过程中不同尺度的特征均具有较强的表示能力<sup>[20,24-26]</sup>。同时,在应用 DCNN 进行遥感图像场景分类时,通常需要大量有标签的数据进行训练,然而由于遥感图像内所含对象的复杂多样性,致使具有单一标签的图像较少,而人工标注又是一件费时费力的事,导致在遥感图像场景分类时因标签样本量少而分类精度不高,文献<sup>[20,24-26]</sup>利用多尺度特征进行遥感图像场景分类正好缓解了这种情况。其中,文献<sup>[20]</sup>提出的多尺度特征方法是通过不同尺度的输入图像产生稠密特征,然后用不同的编码方式进行特征编码并分类。文献<sup>[24]</sup>主要是利用非下采样变换方法对遥感图像进行多尺度分解,然后利用深度卷积网络训练得到不同尺度的图像特征进行场景分类。文献<sup>[25]</sup>提出的另外一种方法是对各卷积层特征采用 VLAD 编码,对各全连接层特征进行平均池化和标准化,最后连接所有处理后的特征进行场景分类。文献<sup>[26]</sup>将多尺度遥感图像各卷积层特征进行多尺度改进费舍尔编码,再将其与固定尺度遥感图像的全连接层特征进行主成分分析

降维,然后应用光谱回归核区分分析法处理后进行融合并分类。

在以上场景分类方法中,只是利用了局部多尺度,并没有充分利用图像的多尺度和深度网络各级的不同尺度特征。因此,本文提出一种新的基于多尺度特征融合的遥感图像场景分类方法——MSFF,即在增加遥感图像多尺度样本数量的同时,利用深度网络各层多尺度特征较强的语义表示,提高小样本遥感图像场景分类的精度。

## 2 基于 MSFF 的遥感图像分类

本文提出的多尺度特征融合的分类方法主要分为三个步骤:第一步为源图像的多尺度变换,即对同一遥感图像进行不同尺度变换,得到多个尺度图像,增加遥感图像的总样本数量,有效解决部分遥感图像场景分类中样本量小的问题;第二步结合深度网络各层不同尺度特征较强的语义表示,尤其是中高级语义特征进行分类,丰富样本数量的同时,也进一步丰富提取特征的多样性和互补性,增强融合特征的可辨识性;第三步用 MKSVM 分类器代替 DCNN 网络的输出层 Softmax 进行场景分类,提高深度网络学习的泛化能力。

### 2.1 对 DCNN 中各卷积层和全连接层特征的编码

从文献<sup>[27]</sup>可知,一般的 DCNN 中因全连接层的存在要求输入固定大小的图像,为了能够输入不同尺度的遥感图像,我们采用文献<sup>[25]</sup>中的方法,对各全连接层特征进行降维和平均池化。在对各卷积层特征编码时,分别采用常用的四种编码方法进行对比,即词袋法(BOW)<sup>[28]</sup>、局部合成描述符向量法(VLAD)<sup>[29]</sup>、局部限制线性编码法(LLC)<sup>[30]</sup>和改进费舍尔核编码法(IFK)<sup>[31]</sup>。

### 2.2 多核支持向量机

从文献<sup>[24]</sup>可知,在利用多尺度特征和深度网络进行遥感图像分类时,由于多尺度特征间的差异性大,用多核支持向量机(MKSVM)分类器<sup>[32]</sup>代替 DCNN 网络的输出层 Softmax,可以提

高图像分类精度。本文采用  $k$  型核函数<sup>[33]</sup> 作为多核支持向量机的基核函数,因为它避免了径向基核函数(RBF)复杂的指数运算,同时具有多项式核函数计算量少和 RBF 核函数逼近精度高与泛化能力强的优点,性能更加优越。 $k$  型核函数和多核函数分别如式(1)和式(2)所示:

$$k(x_i, x_j) = \prod_t \frac{1}{1 + l^2(x_i^{(t)} - x_j^{(t)})^2}, \quad (1)$$

$$\begin{cases} K(x_i, x_j) = \sum_{m=1}^M d_m K_m(x_i - x_j) \\ \sum_{m=1}^M d_m = 1 \quad d_m \geq 0 \end{cases}, \quad (2)$$

式中,  $0 < 1$  反映  $k$  型核函数的宽度;  $M$  为基核函数的个数,  $K_m$  为基核函数,即式(1)表示的  $k$  型核函数。可以通过两步交替优化的方式进行多核支持向量机的最优化,即首先固定核函数权重  $d_m$ , 求解基本的支持向量机问题,然后固定超平面权重,再用梯度下降法对关于  $d_m$  的目标函数进行求解。

### 2.3 MSFF 模型

本文提出的 MSFF 模型结构如图 1 所示,分为三个部分:第一部分为源图像的多尺度变换,即对同一遥感图像进行不同尺度变换,得到多个尺度图像;第二部分是预训练的深度卷积神经网络,这是 MSFF 模型中最重要的部分,通过对各卷积层特征降维并进行编码,由于编码后的特征维数仍然较高,再次进行降维处理,对各全连接层特征进行降维和平均池化,消除输入图像尺度多样性带来的影响,然后经过标准化处理后再与前面卷积层特征进行融合;第三部分为多核支持向量机分类器。对于模型中使用的各预训练 DCNN,其具体网络结构、各层参数设置以及特征编码可参考本文第三部分或参考文献中的内容。其中,  $C_i$  ( $i=1 \cdots n$ ) 和  $F_j$  ( $j=1 \cdots n$ ) 分别表示各卷积层和各全连接层,  $CF_i$  ( $i=1, \cdots, n$ ) 和  $FCF_j$  ( $j=1 \cdots n$ ) 分别表示处理后的各卷积层特征和各全连接层特征。

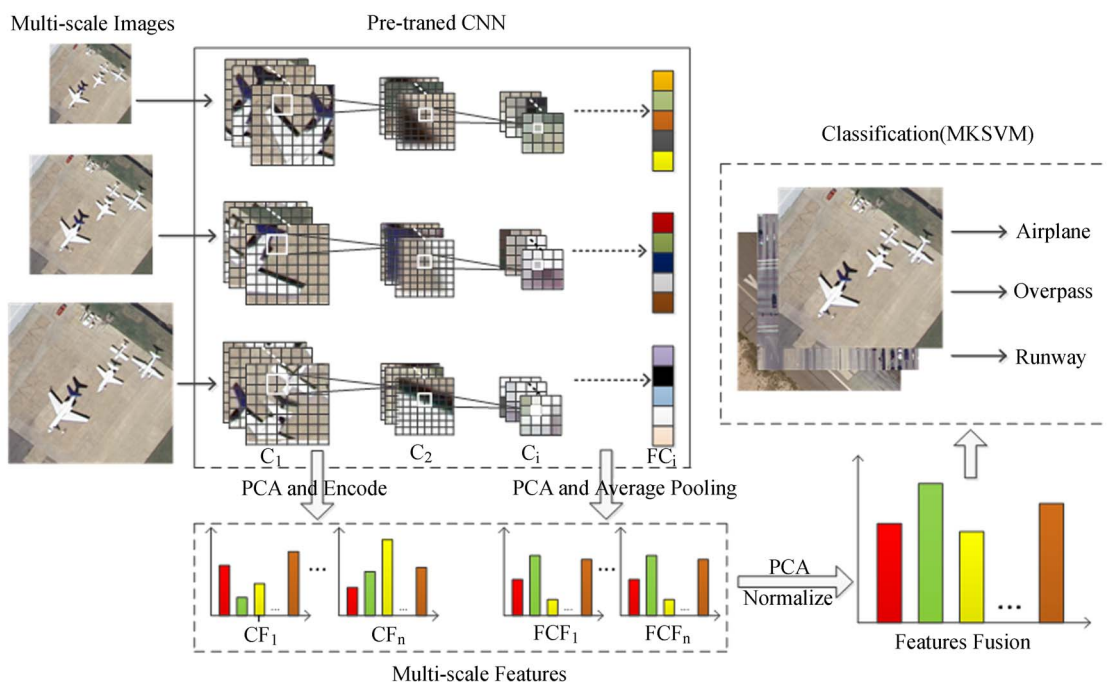


图 1 MSFF 模型结构图

Fig. 1 Structure of MSFF

## 3 试验结果与分析

### 3.1 试验数据集

试验数据集 1 为 UC Merced Land-Use data

set (University of California Merced Land-Use data set), 包括飞机、森林、高速公路等共 21 类遥感图像,每类包含 100 幅  $256 \times 256$  像素的图像,空间分辨率高达  $0.3 \text{ m/pixel}$ ,图 2 给出了各类图像的一个示例。在本文试验中均随机选择每类图

像中的 80 幅作为训练数据,其余 20 幅图像作为 测试数据。



图 2 UC Merced Land-Use data set 场景示例

Fig. 2 Sample charts of UC Merced Land-Use data set

本文采用的试验数据集 2 为 NWPU-RESISC45 data set (Remote Sensing Image Scene Classification\_Benchmark and State of the Art), 它是由程功、韩俊伟等人在已有各种数据集中挑选出 30 个被广泛使用的场景类别,再加上从科学网和搜狗学术网上甄选出的 15 个场景类别,得到的一个包含教堂、沙漠、篮球场等 45 个场景类别的新数据集,每类共有 700 幅  $256 \times 256$  pixel 的图像,大部分图像的空间分辨率为  $0.2 \sim 30$  m/pixel,图 3 给出了部分类别图像的一个示例。在本文试验中均随机选择每类图像中的 600 幅作为训练数据,其余 100 幅图像作为测试数据。

### 3.2 试验参数设置

在输入图像尺度变换时,太小或太大的尺度图像都不能产生较好的结果,因为太小的尺度图像在经过卷积和池化操作后,产生的特征图像会越来越小,使得后面的特征编码效率不高,而尺度太大的图像会产生大量多余的特征,而且使提取图像特征的过程变得很慢,耗时较多,不利于分类的整体表现<sup>[20]</sup>;在特征降维过程中,也存在类似的情况,即维数降低太多或太少都不利于最终的结果。经过多次试验,本文各参数设置分别为:在图像多尺度变换时,采取  $128 \times 128$ 、 $256 \times 256$  和  $512 \times 512$  三个尺度作为同一源图像的多尺

度变换进行输入;同时,在卷积层特征编码前后将特征向量维数分别降至 128 维和 512 维,将全连接层特征向量维数降至 512 维;虽然已有文献提出了各种融合方法,本文为了简化处理,采用编码方法 BOW<sup>[28]</sup>进行特征融合,融合后的特征为  $FF = [CF_1', \dots, CF_n', FCF_1', \dots, FCF_n']$ 。在多核支持向量机中,采用 3 个  $k$  型核函数  $l$ , 的取值分别为 0.65、0.73 和 1,相应的核函数权重  $d$  分别为 0.3、0.5 和 0.2。本文使用开源码库 Caffe 和 VLFeat<sup>[34]</sup>来提取卷积神经网络的特征并进行编码,使用的预训练模型可以在 Caffe 模型库中找到。

### 3.3 分类结果分析

对于每种特征编码方法,本文均取 100 次试验结果的平均值作为最终的分类结果(每次试验都随机选取训练数据和测试数据),4 种特征编码方法的分类结果如表 1 所示:从分类结果来看,本文方法应用不同的卷积神经网络均取得了较好的表现,不管编码方法简单与否,分类精度都在 97% 以上,其中,分类效果表现突出的编码方法为 IFK,在 UCM Land-Use 数据集上的分类精度高于其他 3 种编码方法,最高为 98.91% (使用 OverFeatL 深度网络),在 NWPU-RESISC45 数据集上的分类,除了比 VLAD 编码方法的

99.33%(使用 VGG-M 深度网络)低之外,使用其他深度网络时分类精度均高于其他 3 种方法,可见,IFK 编码方法对深度网络中各卷积层特征具有更好的编译分类效果。

从表中的试验结果还可以看出,对 NWPU-

RESISC45 数据集的分类正确率比 UCM Land-Use 数据集的普遍要高,那是因为前者比后者在同类场景中具有更多的图像,丰富的图像数据有助于对深度网络进行更充分地训练,较好的发挥了深度网络的潜在优势。

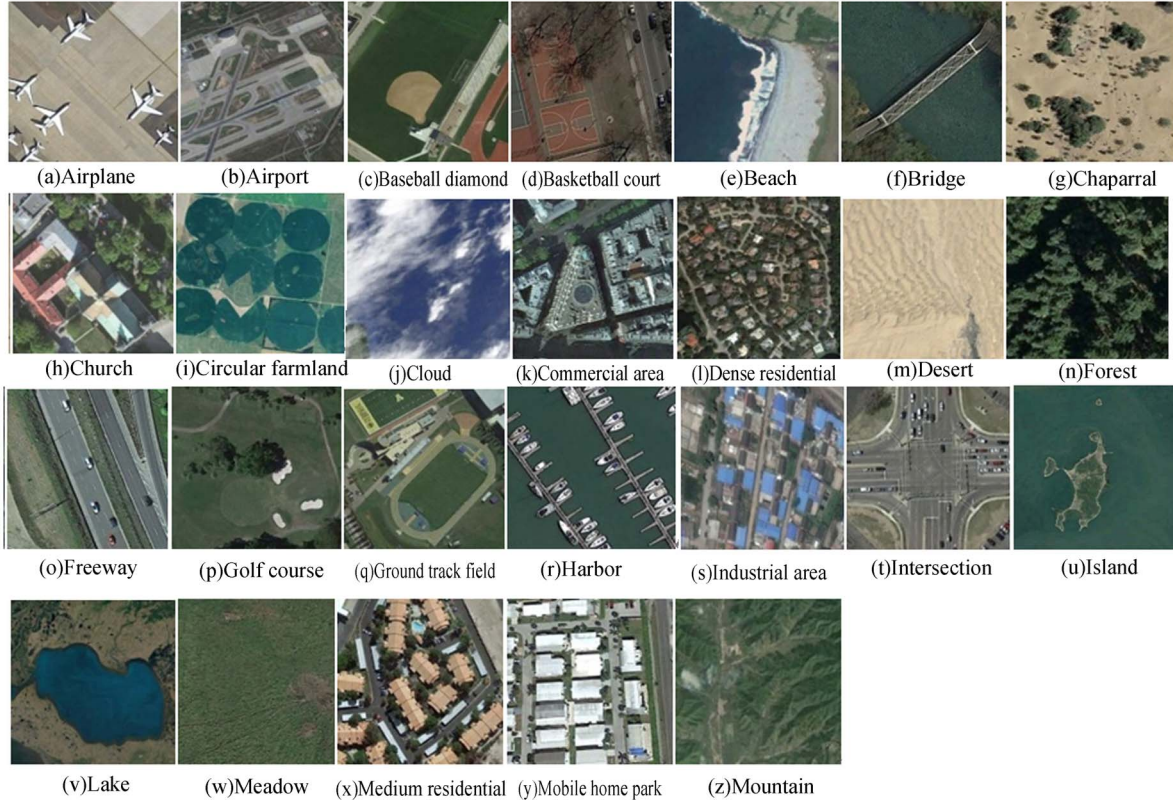


图 3 NWPU-RESISC45 data set 部分场景示例

Fig. 3 Some sample charts of NWPU-RESISC45 data set

表 1 不同预训练卷积神经网络和 4 种特征编码方法的分类精度

Tab. 1 Classification accuracy of different pre-trained convolutional neural networks and four feature encoding methods

特征编码方法	UCM Land-Use			NWPU-RESISC45		
	CaffeNet <sup>[35]</sup>	VGG-M <sup>[36]</sup>	OverFeatL <sup>[37]</sup>	CaffeNet	VGG-M	OverFeatL
LLC	97.05	97.66	98.01	97.81	98.12	98.53
BOW	97.56	98.14	98.45	98.33	98.51	98.95
VLAD	97.81	98.39	98.63	98.41	99.33	99.22
IFK	98.06	98.53	98.91	98.50	99.02	99.31

### 3.4 MKSVM 和 Softmax 分类能力分析

在本文提出的网络模型第三部分中,应用预训练深度网络时,用 MKSVM 代替原网络中的输出层 Softmax 进行图像特征的最终分类,为了比较两个分类器的优劣,对源图像和多尺度图像及特征用两个分类器分别进行分类,出于简单,本文只应用 CaffeNet 预训练模型进行试验,在多尺度

图像和特征分类时,分类结果如表 2 所示。从表中的试验结果可以看出,两个分类器在源图像分类时,分类效果相差不大,但是在多尺度图像和特征分类时,Softmax 分类精度显著降低,而 MKSVM 的分类精度则显著提高。由此可以看出 Softmax 分类器对差异性较大的图像特征分类效果不是很好,而 MKSVM 由于能够利用多核的组

合空间对特征参数进行映射,从而有助于多尺度 图像特征的正确分类。

表 2 MKSVM 和 Softmax 对不同遥感图像的分类精度

Tab. 2 Classification accuracy of different remote sensing images by MKSVM and Softmax

分类精度	UCM Land-Use		NWPU-RESISC45	
	源图像	多尺度图像和特征(LLC)	源图像	多尺度图像和特征(LLC)
Softmax	88.12	72.54	88.78	73.48
MKSVM	88.35	97.05	89.13	97.81

### 3.5 不同分类方法分类能力对比

为了验证本文方法的优势,对文中遥感图像数据集 UCM Land-Use,应用本文方法与现有文献中的方法对比,各方法的分类精度如表 3 所示。从表中的试验结果可以看出,基于深度网络的方法(如CaffeNet等)分类精度普遍高于基于低中

层特征的方法(如 SVM-LDA 等),因为深度网络能够训练得到图像高层次的语义特征,比只能获得较低层特征的方法具有更好的可分性,特别是基于多尺度特征的深度网络方法分类精度高达 98%以上(如文献[25]中的 98.14%和本文的 98.91%),可见基于多尺度特征的方法更能充分发挥深度网络的自身优势,加上本文应用 MKSVM 代替原深度网络中的 Softmax 分类器,对多尺度特征进行了更好地分类。

表 3 不同方法对 UCM Land-Use 数据集的分类精度

Tab. 3 Classification accuracy of UCM Land-Use data set by different methods

分类方法	分类精度
SVM-LDA <sup>[38]</sup>	80.33
SIFT+SC <sup>[7]</sup>	81.67
FK-S <sup>[39]</sup>	91.63
VLAD <sup>[40]</sup>	92.50
CaffeNet <sup>[41]</sup>	93.42
GoogLeNet <sup>[42]</sup>	97.10
VGG-M+IFK <sup>[20]</sup>	96.90
MS-DCNN <sup>[24]</sup>	91.34
VGG16_EMN <sup>[25]</sup>	98.14
VGG-VD19 <sup>[26]</sup>	97.67
本文方法(OverFeatL+IFK)	98.91

使用本文方法(OverFeatL+IFK)对 UCM Land-Use 遥感场景的分类识别率和混淆矩阵分别如图 1 和图 2 所示,同时在图 1 中还列出了 SVM-LDA<sup>[38]</sup>和 CaffeNet<sup>[41]</sup>两种方法的分类识别率用作对比。从两图中的试验结果来看,在对不同场景分类中,对纹理差异性小的场景分类准确度普遍较高,如森林、农田、海滩等,而对含有多个对象的复杂场景分类准确率则普遍较低,如建筑物、居民区、立交桥等,尤其是对于存在二义性的场景容易出现误分,如稠密居住区和中等居住区等等。本文提出的方法在各类场景分类中均得到了比另外两种方法要好的分类结果。

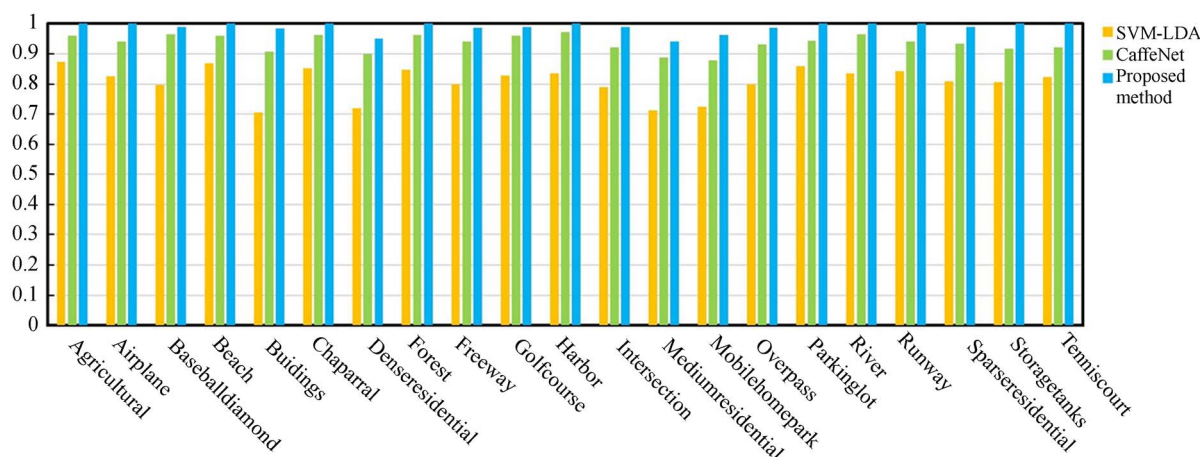


图 4 UCM Land-Use 数据集场景分类准确率

Fig. 4 Scene classification accuracy on UCM Land-Use data set

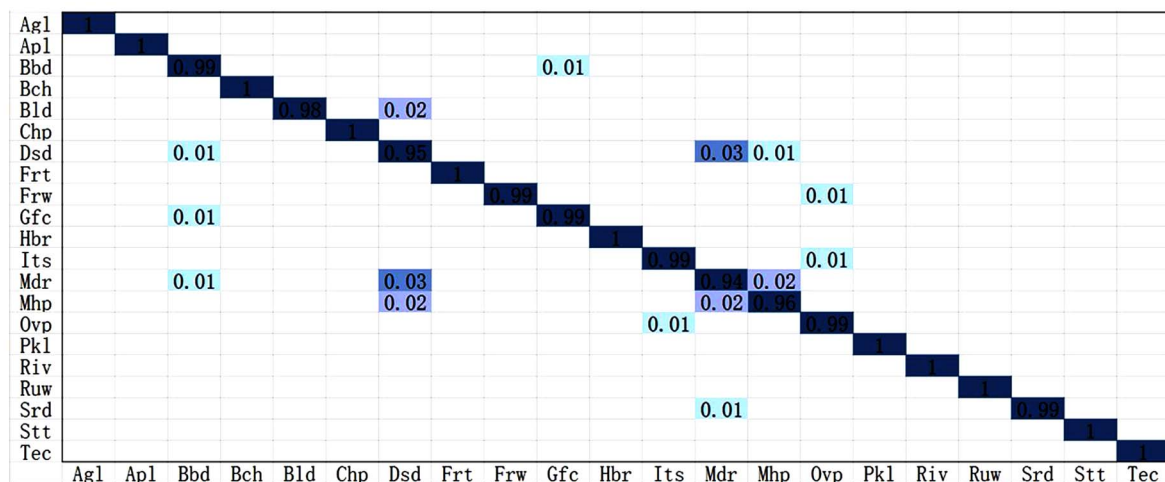


图 5 本文方法对 UCM Land-Use 数据集场景分类的混淆矩阵

Fig. 5 Confusion matrix of scene classification on UCM Land-Use data set

## 4 结 论

本文提出的基于深度网络和多尺度特征融合的遥感图像场景分类方法,不仅能为深度网络输入较丰富的遥感图像,既增加了图像的标签样本数量,也减少了原本因固定尺寸要求缩小图像而带来的信息损失,同时还利用了深度网络各中间

层和全连接层特征较强的表示能力,最后多核支持向量机的使用提高了网络学习的泛化能力。通过在常用遥感数据集 UCM Land-Use 和 NWPU-RESISC45 中的试验结果表明,本文方法使遥感图像场景分类整体精度较以往方法有了一定的提高,分类精度最高分别达到 98.91% (OverFeatL+IFK) 和 99.33% (VGG-M+VLAD)。

## 参考文献:

- [1] PERRONNIN F, DANCE C. Fisher kernels on visual vocabularies for image categorization[C]. *Computer Vision and Pattern Recognition*, 2007. CVPR 07. IEEE Conference on. IEEE, 2007; 1-8.
- [2] JIANG Y. Texture description based on multiresolution moments of image histograms[J]. *Optical Engineering*, 2008, 47(3): 037005.
- [3] VAN DE SANDE KE, GEVERS T, SNOEK CG. Evaluating color descriptors for object and scene recognition[J]. *IEEE Trans Pattern Anal Mach Intell*, 2010, 32(9): 1582-1596.
- [4] YANG Y, NEWSAM S. Bag-of-visual-words and spatial extensions for land-use classification[C]. *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. ACM, 2010; 270-279.
- [5] CHENG G, GUO L, ZHAO T, et al.. Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA[J]. *International Journal of Remote Sensing*, 2013, 34(1): 45-59.
- [6] RAJA R, MANSOOR ROOMI S M, DHARMA-LAKSHMI D. Outdoor scene classification using invariant features[J]. 2013 *Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, 2013;1-4.
- [7] CHERIYADAT A M. Unsupervised feature learning for aerial scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(1): 439-451.
- [8] CHEN S, TIAN Y L. Pyramid of spatial relations for scene-level land use classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(4): 1947-1957.
- [9] HU F, XIA G S, WANG Z, ET AL. Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*,

- 2015, 8(5): 2015-2030.
- [10] ZHAO B, ZHONG Y, XIA G S, *et al.*. Dirichlet-derived multiple topic scene classification model fusing heterogeneous features for high spatial resolution remote sensing imagery[J]. *IEEE Trans. Geosci. Remote Sens.*, 2016, 54(4): 2108-2123.
- [11] ABDEL H O, MOHAMED A, JIANG H, *et al.*. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition [C]. *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE*, 2012:4277-4280.
- [12] KARPATHY A, TODERICI G, SHETTY S, *et al.*. Large-scale video classification with convolutional neural networks [C]. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2014: 1725-1732.
- [13] BALABAN S. Deep learning and face recognition: the state of the art[C]. *Biometric and Surveillance Technology for Human and Activity Identification XII. International Society for Optics and Photonics*, 2015, 9457: 94570B.
- [14] GIRSHICK R, DONAHUE J, DARRELL T, *et al.*. Region-based convolutional networks for accurate object detection and segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(1): 142-158.
- [15] ZHENG S, JAYASUMANA S, ROMERA P B, *et al.*. Conditional random fields as recurrent neural networks[C]. *Proceedings of the IEEE International Conference on Computer Vision*. 2015: 1529-1537.
- [16] 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*. 2016: 770-778.
- [17] VAKALOPOULOU M, KARANTZALOS K, KOMODAKIS N, *et al.*. Building detection in very high resolution multispectral data with deep learning features[C]. *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International. IEEE*, 2015: 1873-1876.
- [18] SZEGEDY C, LIU W, JIA Y, *et al.*. Going deeper with convolutions[J]. 2014:1-9.
- [19] NOGUEIRA K, PENATTI O A B, DOS S J A. Towards better exploiting convolutional neural networks for remote sensing scene classification[J]. *Pattern Recognition*, 2017, 61: 539-556.
- [20] HU F, XIA G S, HU J, *et al.*. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery [J]. *Remote Sensing*, 2015, 7(11): 14680-14707.
- [21] ZHANG F, DU B, ZHANG L. Scene classification via a gradient boosting random convolutional network framework [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(3): 1793-1802.
- [22] ZHAO W, DU S. Scene classification using multi-scale deeply described visual words[J]. *International Journal of Remote Sensing*, 2016, 37(17): 4119-4131.
- [23] OTHMAN E, BAZI Y, ALAJLAN N, *et al.*. Using convolutional features and a sparse autoencoder for land-use scene classification [J]. *International Journal of Remote Sensing*, 2016, 37(10): 2149-2167.
- [24] 许凤晖, 慕晓冬, 赵鹏, 等. 利用多尺度特征与深度网络对遥感影像进行场景分类[J]. *测绘学报*, 2016, 45(7): 834-840.
- XU S H, MU X D, ZHAO P, *et al.*. Scene classification of remote sensing image based on multi-scale feature and deep neural network[J]. *Acta Geodaetica et Cartographica Sinica*, 2016, 45(7): 834-840. (in Chinese)
- [25] WANG G, FAN B, XIANG S, *et al.*. Aggregating rich hierarchical features for scene classification in remote sensing imagery[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 10(9): 4104-4115.
- [26] LI E, XIA J, DU P, *et al.*. Integrating multilayer features of convolutional neural networks for remote sensing scene classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(10): 5653-5665.
- [27] HE K, ZHANG X, REN S, *et al.*. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE Trans Pattern Anal Mach Intell*, 2015, 37(9): 1904-1916.
- [28] SIVIC J, ZISSERMAN A. Video Google: A text retrieval approach to object matching in videos [C]. *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003)*, 2003: 1470-1477.

- [29] JEGOU H, DOUZE M, SCHMID C, *et al.*. Aggregating local descriptors into a compact image representation[J]. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010:3304-3311.
- [30] WANG J, YANG J, YU K, *et al.*. Locality-constrained Linear Coding for image classification[J]. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010: 3360-3367.
- [31] PERRONNIN F, SANCHEZ J, MENSINK T. Improving the fisher kernel for large-scale image classification[J]. *Computer Vision - ECCV 2010*, 2010: 143-156.
- [32] CHEN Z, LI J, WEI L, *et al.*. Multiple-kernel SVM based multiple-task oriented data mining system for gene expression data analysis[J]. *Expert Systems with Applications*, 2011, 38 ( 10 ): 12151-12159.
- [33] 孙翠娟. 基于 K 型核函数的支持向量机[J]. *淮海工学院学报(自然科学版)*, 2006, 15(4): 4-7.  
SUN C J. Support vector machine based K-type kernel function[J]. *Journal of Huaihai Institute of Technology (Natural Sciences Edition)*, 2006, 15(4):4-7. (in Chinese)
- [34] VEDALDI A, FULKERSON B. VLFeat: An open and portable library of computer vision algorithms[C]. *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010: 1469-1472.
- [35] 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*. ACM, 2014: 675-678.
- [36] CHATFIELD K, SIMONYAN K, VEDALDI A, *et al.*. Return of the devil in the details: delving deep into convolutional nets [J]. *Proceedings of the British Machine Vision Conference*, 2014.
- [37] SERMANET P, EIGEN D, ZHANG X, *et al.*. OverFeat: integrated recognition localization and detection using convolutional networks [Z]. *Eprint Arxiv*, 2013.
- [38] ZHANG F, DU B, ZHANG L. Saliency-guided unsupervised feature learning for scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(4): 2175-2184.
- [39] ZHAO B, ZHONG Y, ZHANG L, *et al.*. The Fisher kernel coding framework for high spatial resolution scene classification[J]. *Remote Sensing*, 2016, 8(2): 157.
- [40] NEGREL R, PICARD D, GOSELIN P H. Evaluation of second-order visual features for land-use classification[C]. *Content-Based Multimedia Indexing (CBMI)*, 2014 *12th International Workshop on*. IEEE, 2014: 1-5.
- [41] PENATTI O A B, NOGUEIRA K, DOS S J A. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015: 44-51.
- [42] CASTELLUCCIO M, POGGI G, SANSONE C, *et al.*. Land use classification in remote sensing images by convolutional neural networks [Z]. *arXiv preprint arXiv:1508.00092*, 2015.

#### 作者简介:



杨 州(1982—),男,陕西西安人,博士研究生,工程师,分别于 2005 年和 2010 年于国防科技大学获得学士和硕士学位,主要从事遥感图像分类和目标检测方面的研究。E-mail: yzmail-box2015@163.com

#### 导师简介:



慕晓冬(1965—),男,山东栖霞人,教授,博士生导师,分别于 1987 年、1990 年和 1994 年于第二炮兵工程学院获得学士、硕士和博士学位,主要从事目标识别、智能算法领域的研究。E-mail: Mu\_403@163.com