ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 40,No. 6,Nov. 2025,pp. 1637－1649
DOI:10. 16337/j. 1004-9037. 2025. 06. 020

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于FACNNCN的高分遥感影像场景分类方法

张　婧[1]，杨宇浩[2]，曹　峰[2]，张　超[2]，李德玉[2]

(1. 太原学院数学系,太原 030032;2. 山西大学计算机与信息技术学院,太原 030006)

摘　要：高分遥感影像场景分类旨在对复杂的地表场景影像进行精确认知,对于高分遥感影像的理解和信息提取具有重要的意义。本文提出了一种高分遥感影像场景方法,该方法基于特征聚合卷积神经网络(Feature aggregated convolution neural network, FACNN)和向量胶囊网络(Capsule network, CapsNet),即FACNNCN网络。通过增加聚合特征提升场景分类中影像特征的区分力和鲁棒性,并基于向量胶囊网络表征场景影像中地物与场景的空间关系,有效弥补了当前基于卷积神经网络的高分遥感影像场景分类方法中普遍存在的场景影像特征提取不充分、地物空间特征欠考虑的不足。本文提出的方法在2个公共高分遥感影像场景分类数据集(UC Merced Land-Use和NWPU-RESISC45)上进行了测试,实验结果表明该方法的分类精度优于相关的对比方法。

关键词：高分遥感影像;场景分类;特征聚合;卷积神经网络;胶囊网络

中图分类号：TP391　　**文献标志码**:A

Scene Classification Method of High-Resolution Remote Sensing Images Based on FACNNCN

ZHANG Jing[1], YANG Yuhao[2], CAO Feng[2], ZHANG Chao[2], LI Deyu[2]

(1. Department of Math, Taiyuan University, Taiyuan 030032, China;2. School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)

Abstract: High-resolution remote sensing image scene classification aims to accurately perceive complex surface scenes, which is significant for the understanding and information extraction of high-resolution remote sensing images. A new scene classification method based on feature aggregated convolution neural network (FACNN) and capsule network(CapsNet), named FACNNCN, is proposed in this paper. For the proposed method, the distinguish ability and robustness of convolutional features for scene classification are enhanced by adding aggregated features. Meanwhile, the spatial relationship between geographic entity and scene is represented based on CapsNet. Therefore, the proposed method can overcome some drawbacks usually found in existing high-resolution remote sensing image scene classification methods based on CNNs. For example, the extracted representative features of scene images are insufficient and the spatial features of geographical objects are lack of consideration. The method proposed in this paper is tested on two public high-resolution remote sensing image scene classification datasets (UC Merced Land-Use and NWPU-RESISC45). Experimental results show that the classification accuracy of FACNNCN is better than those of comparison methods.

Key words: high-resolution remote sensing image; scene classification; feature aggregation; convolutional neural network (CNN); capsule network (CapsNet)

基金项目:国家自然科学基金(62072291; 62472269; 62272284; 62072294);山西省科技创新青年人才团队项目(202204051001015)。
收稿日期:2024-03-30;**修订日期**:2024-10-16

引　言

随着遥感技术的飞速发展，遥感影像分辨率快速提升，数据量成几何级数增长。针对海量的高分辨率遥感影像，传统的基于像元的分类解译已转变为基于场景的语义理解。高分遥感影像场景分类旨在通过影像的主要内容辨识影像的场景类别，实现精准的高分遥感影像高层语义理解。目前，高分遥感影像场景分类已广泛应用于城市规划、土地利用、环境污染检测以及军事目标检测等众多领域，体现出重要的实用价值[1-2]。然而，由于高分遥感影像场景构成非常复杂，具有类内多样性、类间相似性、尺度差异以及多类地物目标共存等特点，因此场景分类仍然极具挑战性[3]。

在过去的数十年间，研究者已经提出了多种高分遥感影像场景分类方法。根据分类过程中使用的影像特征，可以分为基于低层特征、中层特征和深度特征的方法。基于低层特征的方法依赖于手工设计的各种特征提取算子，如颜色直方图(Color histogram，CH)、局部二值模式(Local binary pattern，LBP)、方向梯度直方图(Histogram of oriented gradients，HOG)、灰度共生矩阵(Gray level co-occurrence matrix，GLCM)和尺度不变特征变换(Scale-invariant feature transform，SIFT)等，获取颜色、纹理、形状和空间结构等低层特征，进而基于提取的低层特征进行分类。该类方法广泛应用于高分遥感影像场景分类研究的早期[4-7]。基于中层特征的方法在低层特征提取的基础上，采用视觉词袋模型(Bag of visual words，BoVW)、空间金字塔匹配(Spatial pyramid matching，SPM)以及局部特征聚合描述符(Vector of locally aggregated descriptors，VLAD)等方法对低层特征进行编码，进一步提取更具判别能力的中层特征，并用于场景分类[8-10]。基于深度特征的方法主要基于深度学习模型，尤其是可以利用卷积神经网络(Convolutional neural network，CNN)强大的特征表征能力提取场景影像的深层次抽象特征。同基于低层和中层特征的分类方法相比，基于深度特征的分类方法可以显著提升场景分类的性能。因此，基于深度特征的场景分类受到众多学者的广泛关注，并取得了一系列的科研成果。

2015年，巴西三星研究院Penatti等[11]利用CNN来精确辨识高分遥感影像场景的类别，这是深度学习在高分遥感影像场景分类中的首次应用，从此场景分类的研究进入了一个新的阶段。He等[12]提出了一种基于多层堆叠协方差池化网络(Multilayer stacked covariance pooling，MSCP)的场景分类方法，该方法组合CNN网络的多个特征层形成堆叠特征，在此基础上，计算堆叠特征协方差矩阵。该协方差矩阵是一种新的二阶特征，可以有效地区分复杂的场景。Liu等[13]针对场景分类中存在的变尺度问题，提出了多尺度卷积神经网络(Multi-scale CNN，MCNN)。MCNN包含固定尺度和可变尺度两个网络，通过网络之间的参数共享可以获取尺度不变的分类特征。Lu等[14]提出了一种特征聚合卷积神经网络(Feature aggregated CNN，FACNN)，并应用于高分遥感影像场景分类。FACNN将特征学习、特征聚合和分类器三者进行一体化联合训练。实验结果表明，特征聚合可以较好地提高场景分类的精度。

CNN在高分遥感影像场景分类中体现出强大的优势，但是CNN的池化操作带来的空间不变性导致CNN难以从场景影像中识别出地物的姿势、纹理和变化等空间信息。Sabour等[15]于2017年提出了向量胶囊网络(Capsule network，CapsNet)，该网络可以对实体属性进行编码，并通过识别一个实体的部分属性来识别整体。因此，向量胶囊网络可以弥补CNN在场景分类应用中的不足。但是，在实际的应用过程中发现，仅使用向量胶囊网络进行场景分类很难取得显著优于CNN的应用效果。因此，研究者尝试将CNN与向量胶囊网络进行有机结合，通过融合两个网络的优势提升场景分类的精度。Zhang等[16]提出了基于CNN-CapsNet的高分遥感影像场景分类方法，实验结果表明，该方法可以取得较好的分类性能。Asif等[17]提出了一种D-CapsNet网络，该网络通过融合CNN、注意力机制和向量胶囊网络，学习更加丰富和更具鲁棒性的场景影像特征，分类结果良好。

融合卷积神经网络和向量胶囊网络的遥感影像场景分类方法已经取得了一些研究成果。但是随着遥感影像场景分类应用的发展，遥感影像场景的构成越来越复杂，导致存在相同类别场景影像之间

特征分异性大,不同类别场景影像之间特征相似性强,以及多种地物共存引起的特征混淆度大等诸多问题。为了更加有效地区分复杂、易混的场景影像,设计特征表达和区分能力更强以及鲁棒性更好的高分遥感影像场景分类方法具有重要的研究价值。本文提出了一种新的基于FACNN和CapsNet(FACNNCN网络)的高分遥感影像场景分类方法。该方法通过特征聚合的方式获取增强的场景影像分类特征,并基于增强的特征更加准确地表达地物的位置、角度、旋转、倾斜度和尺寸等空间信息,更好地学习地物与场景之间的空间关系,进而提升复杂背景下场景分类的精度。

1　卷积神经网络与向量胶囊网络融合

1.1　卷积神经网络

CNN是受生物学上感受野机制的启发提出的神经网络模型。CNN自提出以来在图像处理领域的各种任务上取得了巨大的成功,如图像分类、人脸识别和图像分割等。随着CNN研究的不断发展,出现了一系列优秀的CNN模型。VGGNet是牛津大学计算机视觉组和Google DeepMind公司一起研发的新型深度CNN模型。该模型使用的较小的感受野和深层次的网络结构可以显著地提升图像分类的精度。VGGNet常用的两种结构分别为VGG-16和VGG-19,其中VGG-16应用更为广泛。

VGG-16共包含5个卷积池化组,其中每个卷积池化组包含的卷积层个数分别为2、2、3、3和3。卷积层是VGG-16最核心的部分。通过不同大小的卷积核得到的卷积层可以有效地提取图像的多层次复杂特征。卷积核的大小通常为3×3或者5×5。在每个卷积池化组的最后,都包含1个池化层。池化层主要是将卷积层得到的特征图进行压缩,减小特征的维度。经常使用的池化操作主要包括最大池化和平均池化。在5个卷积池化组之后,VGG-16包含3个全连接层和1个Softmax输出层。全连接层将卷积、池化操作后得到的特征进行高度综合并输入到分类器中进行分类。一般情况下,全连接层最后一层的输出维数为分类任务的类别数。为了增加各层之间的非线性关系,VGG-16的卷积层和全连接层都使用ReLU函数来增加网络的非线性表达能力。

1.2　向量胶囊网络

胶囊网络是由Hinton等[18]在2011年提出的一种新的深度学习模型,是CNN的一种变体。2017年Sabour等[15]提出了胶囊网络的改进模型向量胶囊网络和胶囊间的动态路由算法。向量胶囊中的胶囊是由1组神经元组成的向量。该向量的参数代表实体的不同属性,如位置、尺寸、方向、形变、速度、反射率、色度和纹理等。向量的长度表示实体存在的概率。

向量胶囊网络由卷积层、低层胶囊层和高层胶囊层构成。卷积层进行图像特征的抽取,并作为低层胶囊层的输入。低层胶囊层对图像的卷积特征进行向量化表示,形成多个向量胶囊,实现实体属性信息的编码。高层胶囊层也称类胶囊层,接受低层胶囊层的向量输入,形成新的向量胶囊,并通过向量胶囊的长度计算胶囊所属类别的概率,进而完成分类任务。为了实现不同胶囊层之间的信息传递和动态连接,向量胶囊网络采用动态路由算法,自动筛选更有效的胶囊,进而提高模型的性能。动态路由算法采用迭代的方法连接不同胶囊层之间的胶囊。由于胶囊是1组用于表示不同特征的神经元的集合,能够在路由过程中建立不同实体间的位置关系,因此,向量胶囊网络对于实体的位置和角度等空间信息的变化更具鲁棒性。

1.3　融合网络

卷积神经网络具有强大的表征学习能力,在大规模、复杂任务中表现出优异的性能。伴随着数值计算设备的快速发展,特别是GPU计算集群的强有力支持,卷积神经网络的性能得到了更加充分的发挥。但是,卷积神经网络的池化操作会丢失大量有用信息,而且不能识别图像中部分与整体之间的空间结构。向量胶囊网络作为卷积神经网络的变体,其设计的初衷旨在弥补卷积神经网络的以上不足。

然而,向量胶囊网络在绝大部分应用任务中的性能却无法超越卷积神经网络。

融合卷积神经网络和向量胶囊网络的CNNCN网络,充分发挥两个网络的优势,既可以利用卷积神经网络获取深层次卷积特征,充分发挥卷积神经网络的强大表征能力,同时又可以通过向量胶囊网络良好的空间特征表达能力,获取丰富的空间信息,构建的网络模型具有更强的特征表达能力和更好的鲁棒性。

CNNCN已成功应用于高分遥感影像场景分类中,并取得了较好的分类性能。然而,随着高分遥感影像场景分类应用的不断发展,场景影像分类任务的复杂度不断提升,同时人们对场景影像分类精度的要求也不断提高。因此,如何对CNNCN网络进行改进,进一步提升CNNCN的分类性能值得深入研究。

2 基于FACNNCN的高分遥感影像场景分类

本文在CNNCN网络的基础上提出了新的FACNNCN网络,并应用于高分遥感影像场景分类中。FACNNCN在传统的CNNCN网络中加入新的聚合特征,通过聚合特征进一步增强场景分类中特征的表达能力。基于FACNNCN的场景影像分类方法首先利用CNN中最具代表性的VGG-16提取场景影像的卷积特征,并将其第3、4、5卷积池化组最后1个特征层所提取的特征进行池化后聚合,得到聚合特征AF_1;接着,将AF_1与VGG-16第5个卷积池化组最后1个卷积层的特征进行聚合,得到聚合特征AF_2;最后将区分能力增强的聚合特征AF_2输入向量胶囊网络,增强向量胶囊网络的地物空间信息表达能力和地物与场景的空间关系学习能力。

FACNNCN主要由输入层、特征聚合模块、胶囊模块和输出层4个部分构成。图1详细描述了FACNNCN的网络结构。输入层输入的是带有类标记的多张高分遥感场景影像;特征聚合模块首先通过卷积层对影像的卷积特征进行提取,然后对多个卷积层提取的卷积特征进行聚合,增强原卷积特征的表现力和鲁棒性;胶囊模块将特征聚合模块获取的卷积特征作为向量胶囊网络的输入,通过卷积特征的矢量化实现场景影像地物空间信息的编码,学习场景影像的局部与整体之间的空间关系;输出层基于训练的网络结构输出测试场景影像所属的类别。图中H、W分别表示特征图的高度和宽度;T表示胶囊个数;S_1和S_2分别表示胶囊神经元个数和胶囊维数;L表示卷积核个数N及通道数D_4之和。

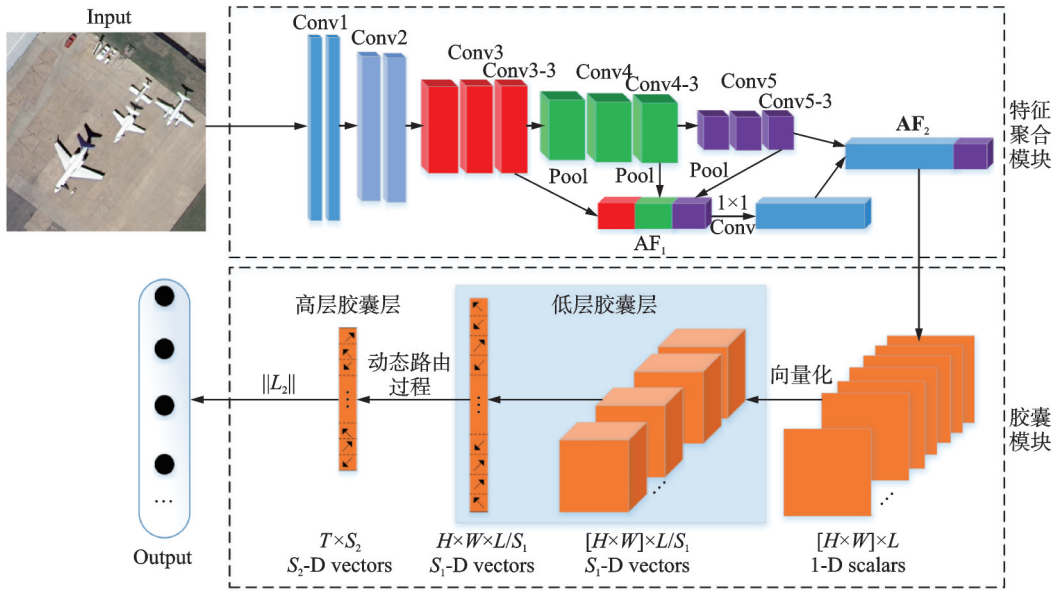

图1 FACNNCN网络结构

Fig.1 Architecture of the FACNNCN network

下面将结合FACNNCN的4个组成部分对场景分类的步骤进行详细描述。

2.1　输入层

步骤1　将具有RGB三个颜色通道的带有类标记的高分遥感场景影像进行裁剪等预处理,得到统一尺寸大小的输入影像。

2.2　特征聚合模块

特征聚合模块是FACNNCN的重要组成部分。特征聚合模块的主要计算步骤如下:

步骤2　针对输入的高分遥感场景影像,利用VGG-16进行卷积特征提取。将VGG-16的第3、4、5卷积池化组的最后1个卷积层Conv3_3、Conv4_3、Conv5_3进行平均池化,得到具有相同尺寸的新的中间特征,这些特征是原有特征的补充,分别记作:$\boldsymbol{X}_1\in\mathbf{R}^{H\times W\times D_1}$、$\boldsymbol{X}_2\in\mathbf{R}^{H\times W\times D_2}$和$\boldsymbol{X}_3\in\mathbf{R}^{H\times W\times D_3}$。依次将得到的特征进行聚合,得到新的卷积特征,记作

$$\mathrm{AF}_1=[\boldsymbol{X}_1;\boldsymbol{X}_2;\boldsymbol{X}_3]\in\mathbf{R}^{H\times W\times(D_1+D_2+D_3)} \tag{1}$$

式中:D_1、D_2、D_3分别为$\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$的通道数。

步骤3　对聚合后的特征AF_1进行1×1的卷积和ReLU操作,实现聚合特征跨通道之间的信息融合,增加不同通道之间的非线性交互作用。将卷积核的个数设为N,此时AF_1的维度由$D_1+D_2+D_3$变为N。

步骤4　将AF_1与卷积层Conv5-3最大池化后的特征$\boldsymbol{X}_4$进行聚合,记$\boldsymbol{X}_4\in\mathbf{R}^{H\times W\times D4}$,$L=N+D_4$,特征聚合模块最终输出的聚合卷积特征为

$$\mathrm{AF}_2=[\mathrm{AF}_1;\boldsymbol{X}_4]\in\mathbf{R}^{H\times W\times L} \tag{2}$$

2.3　胶囊模块

胶囊模块主要由卷积层、低层胶囊层和高层胶囊层3部分构成。胶囊模块的主要计算步骤如下:

步骤5　构建低层胶囊层和高层胶囊层。低层胶囊层用于描述场景影像中较小的地物,并编码地物的属性,提供地物属于某个场景类型的概率。高层胶囊层用于描述整个场景,并基于编码的属性判断场景所属的类别。例如,对于火车站场景影像,低层胶囊用于描述站台、火车、铁轨和建筑物等场景的组成实体,并编码实体的属性。高层胶囊用于描述整个场景,并编码场景的属性。通过低层胶囊层和高层胶囊层之间的信息传递,胶囊模块可以学习场景影像中的较小地物与整体场景之间的关系。

设定低层胶囊层每个胶囊包含的神经元个数S_1的取值,构建包含$H\times W\times L/S_1$个胶囊的低层胶囊层,其中任一胶囊i对应卷积特征AF_2中长度为S_1维的特征向量,即胶囊i的输出向量,记作$\boldsymbol{u}_i$。设定高层胶囊层每个胶囊的维数S_2的取值和胶囊的个数T。T的取值为场景影像的预测类别的个数。构建包含T个胶囊的高层胶囊层,其中任意一个胶囊j对应S_2维的特征向量。

步骤6　计算低层胶囊对高层胶囊的预测向量,即高层胶囊的输入向量。图2给出了胶囊层之间的连接方式。低层胶囊i对高层胶囊j的预测向量记作$\boldsymbol{u}_{j|i}$[15],表达式为

$$\boldsymbol{u}_{j|i}=\boldsymbol{W}_{ij}\boldsymbol{u}_i \tag{3}$$

式中$\boldsymbol{W}_{ij}$为权重矩阵,可以通过反向传播进行优化。所有低层胶囊都要对高层胶囊j进行预测,因此,高层胶囊j的输入向量$\boldsymbol{s}_j$是所有低层胶囊预测向量的加权和,即

$$\boldsymbol{s}_j=\sum_i c_{ij}\boldsymbol{u}_{j|i} \tag{4}$$

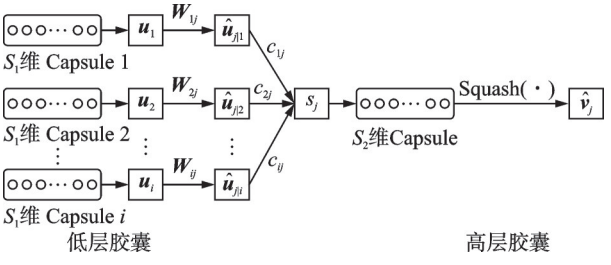

图2　低层胶囊与高层胶囊之间的连接图

Fig.2　Connection between primary capsules and final capsule

式中c_{ij}为耦合系数,在动态路由迭代过程中进行优化。c_{ij}可通过式(5)计算得到,且胶囊i与高层中所有胶囊的耦合系数之和为1。

$$c_{ij}=\frac{\exp(b_{ij})}{\sum_{k}\exp(b_{ik})} \tag{5}$$

式中b_{ij}表示低层胶囊i与高层胶囊j耦合的对数先验概率,初始值设为0。

步骤7 对高层胶囊的输入向量进行压缩,得到高层胶囊的输出向量。高层胶囊的输出向量的长度代表场景影像类别的预测概率,因此为了保证向量的长度不超过1,需要用非线性挤压函数Squash,使得短向量几乎收缩到0,长向量收缩到略小于1。高层胶囊j通过对输入向量$\boldsymbol{s}_j$进行挤压,得到输出向量$\boldsymbol{v}_j$[16]为

$$\boldsymbol{v}_j=\mathrm{Squash}(\boldsymbol{s}_j)=\frac{\|\boldsymbol{s}_j\|^2}{1+\|\boldsymbol{s}_j\|^2}\frac{\boldsymbol{s}_j}{\|\boldsymbol{s}_j\|} \tag{6}$$

步骤8 对胶囊模块进行参数更新。胶囊模块中,对数概率b_{ij}的更新依赖于向量$\boldsymbol{v}_j$和向量$\boldsymbol{u}_{j|i}$的一致性,即两个向量具有较大的内积。因此,b_{ij}的更新公式如下

$$b_{ij}=b_{ij}+\boldsymbol{u}_{j|i}\cdot\boldsymbol{v}_j \tag{7}$$

式(3~7)构成了计算$\boldsymbol{v}_j$的一个完整的动态路由过程。更新了对数概率b_{ij}后,耦合系数c_{ij}根据式(5)进行更新。步骤8之后,转步骤6,进行多次动态路由过程的迭代,实现胶囊模块参数的更新,完成地物与整体场景之间空间信息的传递。

步骤9 计算损失函数。高层胶囊层中每个胶囊k的损失函数L_k计算为[16]

$$L_k=T_k\cdot\max(0,m^+-\|\boldsymbol{v}_k\|)^2+\lambda(1-T_k)\cdot\max(0,\|\boldsymbol{v}_k\|-m^-)^2 \tag{8}$$

式中:$\boldsymbol{v}_k$为高层胶囊k的输出向量,$\|\boldsymbol{v}_k\|$为其长度;m^+、m^-、λ为超参数,需要根据输入的场景影像进行设置。当对应的类别k存在时,T_k的值为1。训练过程中,总的损失为该层所有胶囊的损失之和。

2.4 输出层

步骤10 输出层通过高层胶囊层输出的向量$\boldsymbol{v}_j$(式(9))的长度$\|L_2\|_{\hat{v}_j}$(式(10))来判断预测的场景影像所属的类别C_k。

$$\hat{\boldsymbol{v}}_j=(x_1,x_2,\cdots,x_{S_2}) \tag{9}$$

$$\|L_2\|_{\hat{v}_j}=\left(\sum_{i=1}^{S_2}{x_i}^2\right)^{1/2} \tag{10}$$

比较多个输出向量长度的大小,长度最大的向量对应的类别为场景影像的预测类别C_k。

3 实验与分析

本文基于一个安装3.6 GHz 8核E5-1650 v4 CPU、64 GB内存和Linux操作系统的实验平台进行实验,同时利用GPU进行加速。实验采用UC Merced Land-Use和NWPU-RESISC45两个高分遥感影像场景分类数据集进行算法性能验证。

3.1 数据集介绍

UC Merced Land-Use(如图3所示)是2010年美国UC Merced大学Yang等[10]从USGS(United States Geological Survey)National Map整理收集的包含21类典型场景的高分辨率遥感数据集。该数据集是第一个公开的高分遥感影像场景分类数据集,现已被国内外学者广泛使用。每个场景类别包含100

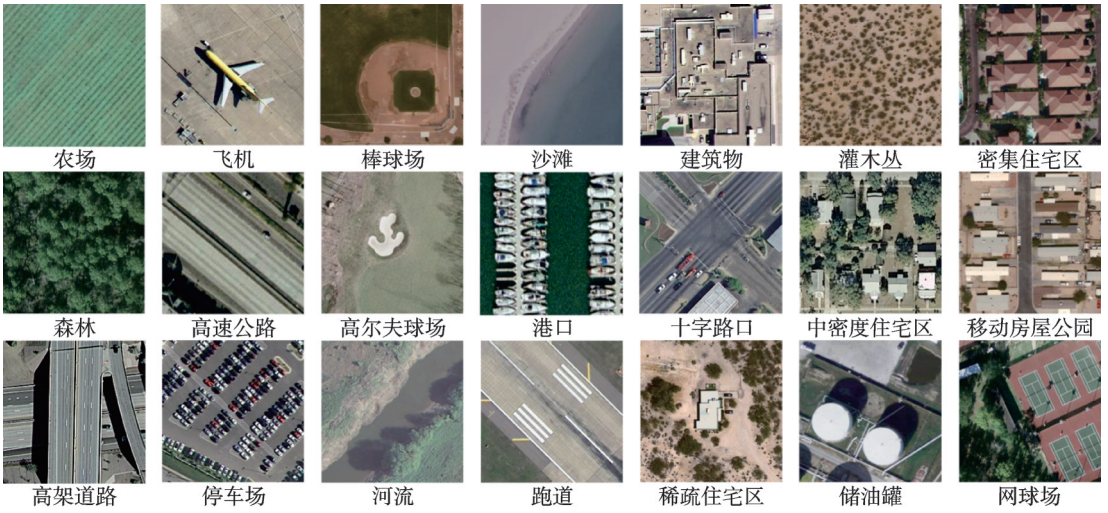

图3　UC Merced Land-Use数据集场景类别图

Fig.3　Some samples of the UC Merced Land-Use dataset

幅大小为256像素×256像素、空间分辨率为0.3 m的影像。21个场景类别分别为:农场、飞机、棒球场、海滩、建筑物、丛林、密集住宅区、森林、高速公路、高尔夫球场、港口、十字路口、中密度住宅区、移动房屋公园、高架道路、停车场、河流、跑道、稀疏住宅区、储油罐和网球场。

NWPU-RESISC45(如图4所示)是西北工业大学研究团队于2017年构建的高分遥感影像场景分类数据集[19]。该数据集场景类别丰富,是目前规模最大的高分遥感影像场景分类数据集,且具有较高

图4　NWPU-RESISC45数据集场景类别图

Fig.4　Some samples of the NWPU-RESISC45 dataset

的类内多样性和类间相似性,对于高分遥感影像场景分类方法具有较高的挑战。该数据集包含45个场景类别,每个场景类别包含700幅大小为256像素×256像素,空间分辨率30～0.2 m不等的影像。45个场景类别分别为:飞机、机场、棒球场、篮球场、沙滩、桥、树丛、教堂、圆形农场、云、商业区、密集住宅区、沙漠、森林、高速公路、高尔夫球场、田径场、港口、工业区、路口、岛、湖、草地、中密度住宅区、移动式家庭公园、山地、立交桥、宫殿、停车场、铁路、火车站、矩形农田、河流、环岛、跑道、海冰、船舶、雪峰、稀疏住宅区、体育场、储油罐、网球场、梯田、热电站和湿地。

3.2 实验的参数设置

本文方法的输入是具有RGB三个颜色通道的高分遥感场景影像,为了方便模型对比,场景影像经过随机裁剪,输入大小变为224 × 224 × 3。在特征聚合模块中,VGG-16网络的卷积核的大小均为3 × 3,每个卷积池化组中均包含1个最大池化层,池化操作窗口的大小为2 × 2,步长为2。VGG-16的权重参数由基于ImageNet数据集预训练的参数进行初始化,并利用随机梯度下降法进行权重参数优化。学习率初始值为0.001,每经过30个epoch,学习率除以10。训练阶段的权重衰减参数和动量分别为0.000 5和0.9。2个数据集训练的批次大小分别为32和8。在胶囊模块中,低层胶囊层和高层胶囊层胶囊的维度分别为8和16;损失函数中参数m^+、m^-、λ的值分别设置为0.9、0.1、0.5。

3.3 评价指标

通过与目前最新的多种基于CNN模型的高分场景影像分类方法的分类精度及其误差进行对比分析,验证了本文方法的有效性。为了实现评价的客观性,针对每个数据集,重复进行10次实验,以减少随机性对实验结果的影响。

3.4 实验结果与分析

3.4.1 UC Merced Land-Use数据集

实验从UC Merced Land-Use数据集的每个场景类别中随机选择80%的影像作为训练集,其余20%的影像作为测试集。本文将提出的FACNNCN方法与10种场景分类方法进行了对比。对比方法中,GoogLeNet、CaffeNet、VGG-VD-16、Fine-tuned VGG-16属于经典的基于CNN的方法;LGFBOVW、Fusion by addition、Two-Stream Fusion、MSCP和FACNN属于基于特征聚合的方法;CNN-CapsNet和D-CapsNet属于结合CNN和向量胶囊网络的方法。基于不同场景分类方法的分类精度对比结果如表1所示。由表1可以看出,经典的基于CNN的场景分类方法的分类精度相对较低,但是对比4种方法,Fine-tuned VGG-16的分类精度比其他3种模型高约2%,体现出微调方式的有效性。FACNNCN特征融合模块同样使用Fine-tuned VGG-16进行特征提取,且分类精度比Fine-tuned VGG-16提高了约2%,证明了FACNNCN中特征聚合模块和胶囊模块的有效性。

表1 UC Merced Land-Use数据集80%训练比例下分类精度对比表

Table 1 Comparison of classification accuracies and errors with 80% training ratio for UC Merced Land-Use dataset %

方法	分类精度
GoogLeNet[4]	94.31±0.89
CaffeNet[4]	95.02+0.81
VGG-VD-16[4]	95.21±1.20
Fine-tuned VGG-16[19]	97.14±0.22
LGFBOVW[20]	96.88±1.32
Fusion by addition[21]	97.42±1.79
Two-Stream Fusion[22]	98.02+1.03
MSCP[12]	98.36±0.58
FACNN[14]	98.81±0.24
CNN-CapsNet[16]	98.81±0.22
D-CapsNet[17]	99.03±0.23
FACNNCN	**99.25±0.25**

与4种经典的基于CNN的场景分类方法相比,5种基于特征聚合的分类方法的分类精度相对较高。结合CNN和向量胶囊网络的CNN-CapsNet方法同样取得了比经典的基于

CNN分类方法更高的分类精度。这表明特征聚合以及向量胶囊网络都是提升场景影像分类精度的有效途径。特征聚合通过对多个中间特征进行聚合获取区分性更强的场景影像的卷积特征,而向量胶囊网络通过胶囊的向量化表示能够获取场景影像地理实体的多种空间特征,这些特征的融合丰富了模型分类的输入信息,更加有利于区分复杂的场景类别,进而提高场景影像的整体分类精度。

FACNNCN在进行遥感场景分类时,同时利用了聚合特征和多种空间特征,模型的输入信息更加丰富,分类性能必将得到改善。实验结果也验证了FACNNCN的有效性。与性能最好的基于特征聚合的分类方法FACNN相比,FACNNCN的平均精度、最大精度和最小精度分别提高了0.44%、0.45%和0.43%,而与结合CNN和向量胶囊网络的CNN-CapsNet方法相比,FACNNCN的平均精度、最大精度和最小精度分别提高了0.44%、0.47%和0.41%。总之,在UC Merced Land-Use数据上,本文所提FACNNCN方法的平均精度、最高精度和最低精度分别为99.25%、99.50%和99.00%,均高于所有对比方法。

图5直观显示了不同分类方法之间的分类精度误差的对比情况。可以看出,FACNNCN的分类误差较小,明显优于7种对比方法。而与其余3种方法Fine-tuned VGG-16、FACNN和CNN-CapsNet相比,分类误差的大小非常接近,这表明FACNNCN方法不仅能够取得较高的分类精度,而且具有较好的稳定性。

图6显示了FACNNCN在UC Merced Land-Use数据集上最好分类结果的混淆矩阵。由图6可以看出,在21个场景类别中,有19个场景类别的分类精度为100%,其余2个场景类别建筑物和中密度住宅区的预测精度为95%。模型将5%的类别为建筑物的场景影像预测为密集住宅区,将5%的类别为中密度住宅区的场景影像预测为密集住宅区。预测错误的原因在于建筑物、中密度住宅区和密集住宅区这3类场景影像的构成实体以房屋为主,且光学影像和空间结构特征都非常相似,极易造成预测结果的混淆。

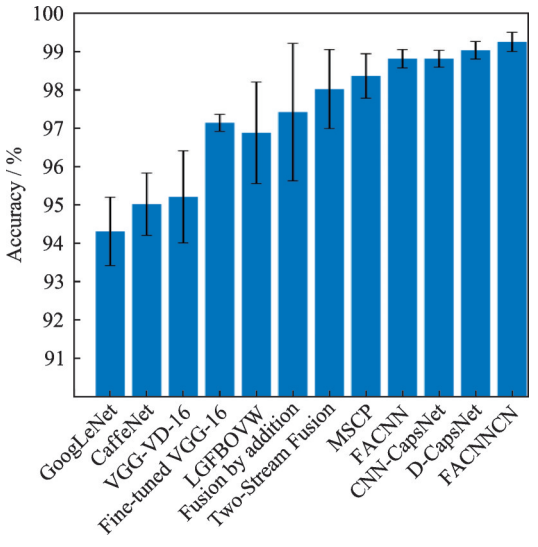

图5　UC Merced Land-Use数据集80%训练比例下不同模型分类精度及误差对比图

Fig.5　Comparison of classification accuracies and errors with 80% training ratio for UC Merced Land-Use dataset

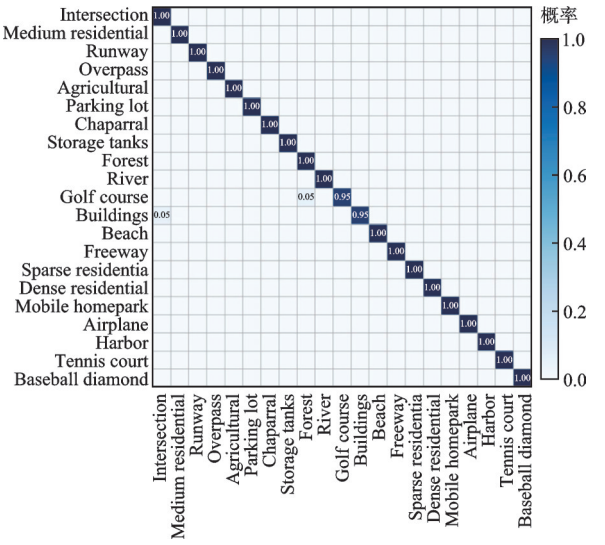

图6　UC Merced Land-Use数据集80%训练比例下FACNNCN方法分类结果混淆矩阵

Fig.6　Confusion matrix of FACNNCN method with 80% training ratio for UC Merced Land-Use dataset

3.4.2 NWPU-RESISC45数据集

UC Merced Land-Use数据集的数据规模相对较小,场景类别数也相对较少。为了更客观地评价方法的性能,在场景类别数较多、规模更大的NWPU-RESISC45数据集上进行了对比实验。实验分为两组:(1)对每个场景类别随机选择10%的场景影像作为训练集,其余的影像作为测试集;(2)对每个场景类别随机选择20%的场景影像作为训练集,其余的影像作为测试集。在NWPU-RESISC45数据集上,同样选择了10种场景分类方法进行了对比实验。不同方法的场景分类精度对比结果如表2所示。

表2 NWPU-RESISC45数据集在10%和20%训练比例下分类精度对比表

Table 2 Comparison of classification accuracies under 10% and 20% training ratios on NWPU-RESISC45 dataset %

方法	10%训练比例	20%训练比例
GoogLeNet[19]	76.19±0.38	78.48±0.26
AlexNet[19]	76.69±0.21	79.85±0.13
VGG-16[19]	76.47±0.18	79.79±0.15
Fine-tuned GoogLeNet[19]	82.57±0.12	86.02±0.18
Fine-tuned AlexNet[19]	81.22±0.19	85.16±0.18
Fine-tuned VGG-16[19]	87.15±0.45	90.36±0.18
BoVW[23]	82.65±0.31	84.32±0.17
Two-Stream Fusion[22]	80.22±0.22	83.16±0.18
MSCP[12]	85.33±0.21	88.93±0.14
CNN-CapsNet[16]	85.08±0.13	89.18±0.14
D-CapsNet[17]	87.91±0.18	90.25±0.13
FACNNCN	**88.60±0.15**	**91.56±0.11**

从表2中可以看出,由于NWPU-RESISC45数据集的规模和场景类别的数量显著增加,因此,所有方法的分类精度都显著低于UC Merced Land-Use数据集。但是,同10种对比方法相比,在10%和20%的训练比例下,FACNNCN的平均精度、最大精度和最小精度仍然都是最高的,表明在大规模数据集上FACNNCN仍然可以取得较优的分类性能。

图7显示了在10%和20%的训练比例下不同方法的分类误差之间的差异,可以看出,当训练比例为10%时,各方法的分类误差比较接近,而当训练比例由10%增加到20%时,分类误差的差异更小。在20%的训练比例下,FACNNCN的分类误差低于所有对比方法,表明在大规模和多类别的遥感影像场景数据上,本文方法仍能保持良好的稳定性。

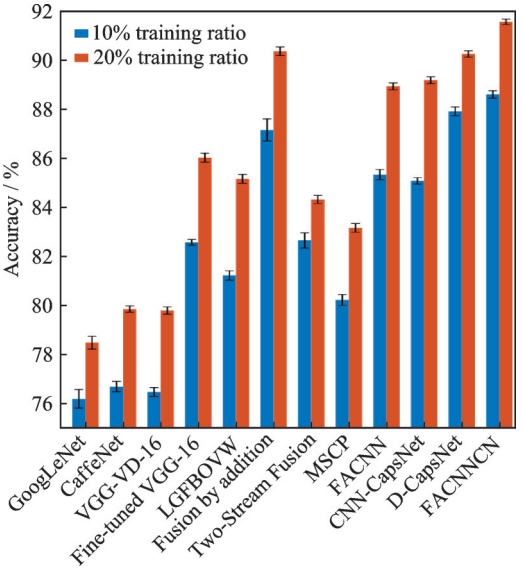

图7 NWPU-RESISC45在数据集20%比例下不同模型分类精度与误差对比图

Fig.7 Comparison of classification accuracies and errors under 10% and 20% training ratios on NWPU-RESISC45 dataset

图8和图9分别给出了在10%和20%的训练比例下,FACNNCN分类结果的混淆矩阵。图8中有29个场景类别的预测精度超过90%,图9中有34个场景类别的预测精度超过了90%。两个混淆矩阵中,预测精度最差且最易混淆的两个场景类别是宫殿和教堂,主要原因

在于它们的建筑风格非常相似。

3.4.3　消融实验

为了验证FACNNCN方法中特征聚合模块和胶囊模块的性能,分别在UC Merced Land-Use和NWPU-RESISC45两个数据集上进行了消融实验。对于UC Merced Land-Use数据集,当FACNNCN仅保留特征聚合模块或胶囊模块时,FACNNCN的平均分类精度均为98.81%。当同时保留特征聚合模块和胶囊模块时,FACNNCN的平均分类精度提高到99.25%。对于NWPU-RESISC45数据集,进行同样的消融实验。在10%和20%的训练比例下,当保留特征聚合模块时,FACNNCN的平均分类精度分别为86.41%和90.68%。当保留胶囊模块时,FACNNCN的平均分类精度分别为85.08%和89.18%。当同时保留特征聚合模块和胶囊模块时,FACNNCN的平均分类精度提高到87.34%和91.12%。

由消融实验可以发现,本文所提FACNNCN方法中特征聚合模块和胶囊模块在场景分类过程中均体现出良好的性能。在场景影像规模较大的NWPU-RESISC45数据集上,特征聚合模块发挥的作用优于向量胶囊模块,而二者的有机结合可以获得优于任一模块的分类性能。

4　结束语

高分遥感影像场景分类是航空和遥感卫星影像分析领域的重要科学问题,众多研究者已经开展了广泛的研究,并取得了丰富的研究成果。近年来,人工智能技术快速发展,其中深度学习最具代表性,已在计算机视觉、语音识别和自然语言处理等领域取得了远远超过其他相关技术的优越性能。深度学习同样被广泛应用于高分遥感影像场景分类研究

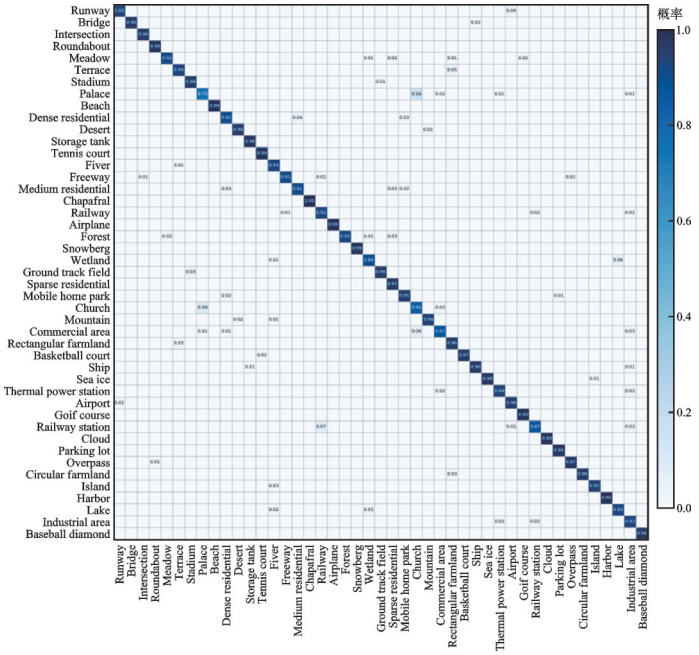

图8　NWPU-RESISC45数据集在10%训练比例下FACNNCN方法分类结果混淆矩阵

Fig.8　Confusion matrix of the IFACNNCN method under 10% training ratio on NWPU-RESISC45 dataset

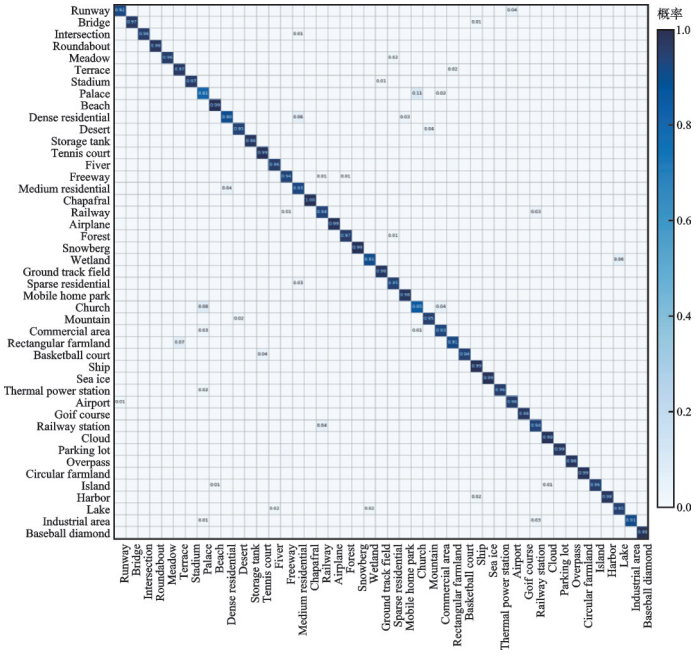

图9　NWPU-RESISC45数据集在20%训练比例下FACNNCN方法分类结果混淆矩阵

Fig.9　Confusion matrix of the FACNNCN method under 20% training ratio on NWPU-RESISC45 dataset

中,并成为高分遥感影像场景分类研究的利器。本文提出了一种新的基于FACNNCN的高分遥感影像场景分类方法。与传统的基于CNN的高分遥感影像场景分类方法相比,FACNNCN的特征聚合模块通过获取场景影像的聚合卷积特征,有效提升了网络对较复杂场景的区分能力,而其胶囊模块利用向量胶囊对场景的整体和部分之间的空间关系进行表达和学习,能够深度挖掘场景影像的空间特征。因此,FACNNCN充分利用了场景影像中蕴含的特征信息,具有更优的复杂场景分类性能。未来将从两个方面开展研究工作:(1)结合高分遥感影像场景数据自身的特点,针对易混淆场景影像设计辨识度强的特征信息,提高方法对易混淆类别的区分能力;(2)优化场景影像的空间特征表达模式,通过对网络结构的改进,进一步提升方法的分类性能。

参考文献:

[1] WANG X, XIONG X N, NING C, et al. Integration of heterogeneous features for remote sensing scene classification[J]. Journal of Applied Remote Sensing, 2018, 12(1): 15-23.

[2] 殷慧, 曹永锋, 孙洪. 基于多维金字塔表达和AdaBoost的高分辨率SAR图像城区场景分类算法 [J]. 自动化学报, 2010, 36(8): 1099-1106.
YIN Hui, CAO Yongfeng, SUN Hong. Urban scene classification based on multi-dimensional pyramid representation and AdaBoost using high resolution SAR Images[J]. Acta Automatica Sinica, 2010, 36(8): 1099-1106.

[3] CHENG G, XIE X X, HAN J W, et al. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 3735-3756.

[4] XIA G S, HU J, HU F, et al. AID: A benchmark data det for performance evaluation of aerial scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3965-3981.

[5] NOGUEIRA K, PENATTI O A B, DOS SANTOS J. Towards better exploiting convolutional neural networks for remote sensing scene classification[J]. Pattern Recognition, 2016, 61: 539-556.

[6] BIAN X Y, CHEN C, TIAN L, et al. Fusing local and global features for high-resolution scene classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(6): 2889-2901.

[7] RISOJEVIĆ V, BABIĆ Z. Unsupervised quaternion feature learning for remote sensing image classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(4): 1521-1531.

[8] 黄鸿, 徐科杰, 石光耀. 联合多尺度多特征的高分遥感图像场景分类[J]. 电子学报, 2020, 48(9): 1824-1833.
HUANG Hong, XU Kejie, SHI Guangyao. Scene classification of high-resolution remote sensing image by multi-scale and multi-feature fusion[J]. Acta Electronica Sinica, 2020, 48(9): 1824-1833.

[9] LOW D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[10] YANG Y, NEWSAM S. Bag-of-visual-words and spatial extensions for land-use classification[C]//Proceedings of the SIGSPATIAL International Conference on Advances in Geographic Information Systems. San Jose, CA, USA: ACM, 2010: 270-279.

[11] PENATTI O, NOGUEIRA K, SANTOS J. 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. [S.l.]: IEEE, 2015: 44-51.

[12] HE N J, FANG L Y, LI S T, et al. Remote sensing scene classification using multilayer stacked covariance pooling[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56 (12): 6899-6910.

[13] LIU Y F, ZHONG Y F, QIN Q Q. Scene classification based on multiscale convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(12): 7109-7121.

[14] LU X Q, SUN H, ZHENG X T. A feature aggregation convolutional neural network for remote sensing scene classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(10): 7894-7906.

[15] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[C]//Proceedings of the Conference on Neural

Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc., 2017: 3856-3866.

[16] ZHANG W, TANG P, ZHAO L. Remote sensing image scene classification using CNN-CapsNet[J]. Remote Sensing, 2019, 11: 494.

[17] ASIF R, HONG H, SALAYIDIN S, et al. Diverse capsules network combining multiconvolutional layers for remote sensing image scene classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 5297-5313.

[18] HINTON G E, KRIZHEVSKY A, WANG S D. Transforming auto-encoders[C]//Proceedings of the International Conference on Artificial Neural Networks. Espoo, Finland: Springer, 2011: 44-51.

[19] CHENG G, HAN J W, LU X Q. Remote sensing image scene classification: Benchmark and state of the art[J]. Proceedings of the IEEE, 2017, 105(10): 1865-1883.

[20] ZHU Q Q, ZHONG Y F, ZHAO B, et al. Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(6): 747-751.

[21] CHAIB S, LIU H, GU Y F, et al. Deep feature fusion for VHR remote sensing scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(8): 4775-4784.

[22] YU Y L, LIU F X. A two-stream deep fusion framework for high-resolution aerial scene classification[J/OL]. Computational Intelligence and Neuroscience, 2018. https://doi.org/10.1155/2018/8639367.

[23] CHENG G, LI Z P, YAO X W, et al. Remote sensing image scene classification using bag of convolutional features[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(10): 1735-1739.

作者简介:

张婧(1982-),女,副教授,研究方向:数据挖掘、人工智能,E-mail:zj6amanda@163.com。

杨宇浩(1996-),男,硕士研究生,研究方向:遥感信息处理,E-mail: 1459338016@qq.com。

曹峰(1980-),**通信作者**,男,教授,研究方向:人工智能、空间数据挖掘、遥感信息处理,E-mail: caof@sxu.edu.cn。

张超(1980-),男,教授,研究方向:粒计算、数据挖掘、智能决策,E-mail: czhang@sxu.edu.cn。

李德玉(1965-),男,教授,研究方向:粒计算、数据挖掘、人工智能,E-mail:lidy@sxu.edu.cn。

(编辑:刘彦东)