

# 光学学报

## 基于小波-空间高阶特征聚合网络的遥感图像场景分类

倪康<sup>1,2\*</sup>, 翟明亮<sup>3</sup>, 王鹏<sup>4</sup>

<sup>1</sup>南京邮电大学计算机学院、软件学院、网络空间安全学院, 江苏 南京 210023;

<sup>2</sup>江苏省大数据安全与智能处理重点实验室, 江苏 南京 210023;

<sup>3</sup>南京邮电大学自动化学院、人工智能学院, 江苏 南京 210023;

<sup>4</sup>南京航空航天大学电子信息工程学院, 江苏 南京 211106

**摘要** 提出一种小波-空间高阶特征聚合网络(WHFA-Net),该网络可分为小波域特征提取和空间域特征提取两个分支。首先,将Harr小波变换嵌入卷积神经网络(CNNs),保留深度卷积特征的低频分量作为小波深度特征;其次,利用最大池化进行深度特征学习,并将其输出作为空间深度特征;将两分支的深度特征进行向量化后,获取其自相关和互相关高阶深度特征向量,并依次进行特征正规化、特征聚合和特征归一化操作;最后,引入交叉熵损失函数进行端到端网络训练。在NWPU45(NWPU-RESISC45 Dataset)和AID(Aerial Image Dataset)数据集上的实验结果表明:相较于基准网络(VGG-16),本文所提WHFA-Net的场景分类准确率有5.13%~12.12%的提升;与DCCNN、APDC-Net、GBNet、LCNN-BFF、MSCP和Wavelet CNN相比,WHFA-Net的场景分类准确率均有不同程度的提升;通过消融实验验证了各模块和分支的有效性及其性能差异。因此,WHFA-Net可有效且稳定地抽取遥感场景图像不同特征域的高阶聚合特征,并提升场景分类准确率。

**关键词** 遥感; 场景分类; 卷积神经网络; 特征可辨别性; 特征聚合

中图分类号 TP753

文献标志码 A

DOI: 10.3788/AOS202242.2428005

### Scene Classification of Remote Sensing Images Based on Wavelet-Spatial High-Order Feature Aggregation Network

Ni Kang<sup>1,2\*</sup>, Zhai Mingliang<sup>3</sup>, Wang Peng<sup>4</sup>

<sup>1</sup>School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, Jiangsu, China;

<sup>2</sup>Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing 210023, Jiangsu, China;

<sup>3</sup>College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, Jiangsu, China;

<sup>4</sup>College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, Jiangsu, China

**Abstract** This paper proposes a wavelet-spatial high-order feature aggregation network (WHFA-Net) which can be divided into two branches: wavelet domain feature extraction and spatial domain feature extraction. Firstly, a Harr wavelet transform is embedded into convolutional neural networks (CNNs), and low-frequency components of the depth-wise convolutional features are retained as wavelet depth features. Secondly, depth feature learning is performed by the max pooling, and then the output is used as spatial depth features. In addition, the wavelet domain and spatial domain depth features are vectored, and their auto-correlation and cross-correlation high-order depth feature vectors are obtained. Feature regularization, feature aggregation, and feature normalization are then performed in sequence. Finally, a cross-entropy loss function is utilized for end-to-end network training. The experimental results on NWPU45

收稿日期: 2022-04-26; 修回日期: 2022-05-24; 录用日期: 2022-06-16

基金项目: 国家自然科学基金(62101280,61801211)、江苏省自然科学基金(BK20210588,BK20210594)、江苏省高校自然科学基金(21KJB520016)、南京邮电大学引进人才科研启动基金(NY220135)

通信作者: \*tznikang@163.com

(NWPU-RESISC45 Dataset) and AID (Aerial Image Dataset) show that compared with that of the benchmark network (VGG-16), the accuracy of the proposed WHFA-Net in scene classification is improved by 5.13%–12.12%. Furthermore, compared with DCCNN, APDC-Net, GBNet, LCNN-BFF, MSCP, and Wavelet CNN, the accuracy of WHFA-Net in scene classification is higher. Additionally, the effectiveness and performance differences of each module and branch are verified through the ablation experiments. Therefore, WHFA-Net can effectively and stably extract the high-order aggregated features of different feature domains in remote sensing scene images and accurately perform scene classification.

**Key words** remote sensing; scene classification; convolutional neural network; feature discriminability; feature aggregation

## 1 引言

随着遥感传感器技术的飞速发展,遥感图像分辨率得到提高、数据量增加,遥感图像信息处理也相继进入大数据处理阶段,这为遥感图像解译相关工作的应用提供了可能<sup>[1]</sup>。遥感图像场景分类是遥感图像解译的重要部分,其主要目的是根据遥感场景图像中场景布局、目标等信息,利用相关算法提取该图像中所包含的高层语义信息,实现对场景内容的类别映射。但高分辨率遥感图像包含较为丰富的细节、结构、纹理等特征信息,使得高分辨率遥感图像场景分类极具挑战性<sup>[2-4]</sup>。

卷积神经网络(CNNs)能够更好地抽取图像的高层语义信息<sup>[5]</sup>,特征表述能力和特征泛化性较强,大大提升了遥感图像场景分类的准确率。近年来,专家和学者提出了一系列基于CNNs的遥感图像场景分类算法<sup>[6-9]</sup>,遥感图像场景分类算法大致可分为基于预训练模型的分分类算法和基于端到端网络训练的分分类算法。基于预训练模型的分分类算法是将已在ImageNet等大型数据集上训练好的网络模型作为特征提取器,将特征提取的结果作为分类器的输入,以此完成场景分类。He等<sup>[10]</sup>提出多层堆叠的协方差池化(MSCP)网络,Li等<sup>[11]</sup>提出基于多层卷积特征网络的遥感图像场景分类算法,Ni等<sup>[12]</sup>提出多层激励局部约束仿射子空间编码-卷积神经网络(MELASC-CNN)。上述算法大都在CNNs深度特征的基础上引入多层连接、视觉词袋模型(BoW)、稀疏低秩等理论,以此提高深度特征可辨别性。但由于上述算法中预训练模型均是利用自然场景图像数据集进行训练的,未利用遥感场景图像进行重新训练,这将导致该类算法的泛化性较弱,同时也限制了其在遥感图像场景分类精确度上的提升空间。

基于端到端网络训练的分分类算法在一定程度上提升了模型泛化性,基于多分支的深度网络亦可有效提高遥感图像场景分类性能。Bi等<sup>[13]</sup>提出了一种基于密集连接注意力池化网络(APDC-Net)的遥感图像场景分类算法,该算法利用注意力池化层提取密集连接网络中的多层深度特征,并将其作为最终的深度特征向量,该算法可以得到较高的场景分类准确性,但其最终的深度特征向量仍然为一阶深度特征,特征可辨别性仍有不足。为了嵌入高阶深度特征描述,He等<sup>[14]</sup>提出一种跳跃式连接协方差网络(SCCov),并将其应用

至遥感图像场景分类,该网络利用多层深度特征、协方差特征描述和特征归一化层抽取最终的深度特征向量;Li等<sup>[15]</sup>提出一种一阶-二阶融合深度特征学习网络(FSoI-Net),该网络利用全局池化层(GAP)获取一阶深度特征,二阶深度特征表述可通过基于注意力机制的协方差层获取。Sun等<sup>[16]</sup>将分层特征聚合与干扰聚合消除模块集成在CNNs中,提出一种门控双向遥感图像场景分类网络(GBNet),GBNet通过双向连接聚合多层深度卷积特征,门控函数用于消除双向连接的干扰信息。为了兼顾模型复杂度和场景分类准确率,Shi等<sup>[17]</sup>提出一种基于轻量多分支深度网络聚合网络的遥感图像场景分类方法(LCNN-BFF),该方法通过两分支深度网络进行深度特征抽取,并利用深度可分离卷积和常规卷积结合的方式降低模型复杂度。在综合多分支与高阶特征聚合方面,Li等<sup>[18]</sup>提出一种多尺度高阶特征残差网络(MRHNNet),该网络在多尺度残差网络基础之上,利用高阶特征表述刻画遥感图像场景信息,可有效提升遥感图像场景分类准确率。虽然上述算法嵌入了多分支网络架构和高阶特征提取层,但特征学习均限制在空间域,缺少对小波域等频域信息的考虑。图1给出利用最大池化、平均池化和小波变换进行特征提取的结果。从图1可以看出,最大池化和平均池化利用较为直接的取最大值和取平均的操作,当主要轮廓或者细节的灰度值(平均值)低于非主要轮廓或细节灰度值时,所提取的图像特征可辨别性欠佳;而小波变换所提取的低频向量可以较为有效地描述图像的轮廓等图像的大部分信息,其特征刻画能力更具针对性。

基于以上分析,本文立足于小波域与空间域特征学习,以高阶深度特征为基础,提出一种小波-空间高阶特征聚合网络(WHFA-Net)。将Harr小波变换嵌入CNNs,保留深度卷积特征的低频分量作为小波深度特征;同时利用最大池化层的输出特征向量作为空间深度特征;另外,引入互相关、自相关和特征聚合以获取具有可辨别性的高阶深度特征。所提网络可进行端到端的网络训练。

## 2 小波-空间高阶特征聚合网络

本文所提网络是一种基于端到端训练的深度神经网络,表1给出了其与现有遥感图像场景分类网络的区别,对比网络包括MSCP<sup>[10]</sup>、DCCNN<sup>[19]</sup>、APDC-Net<sup>[13]</sup>、

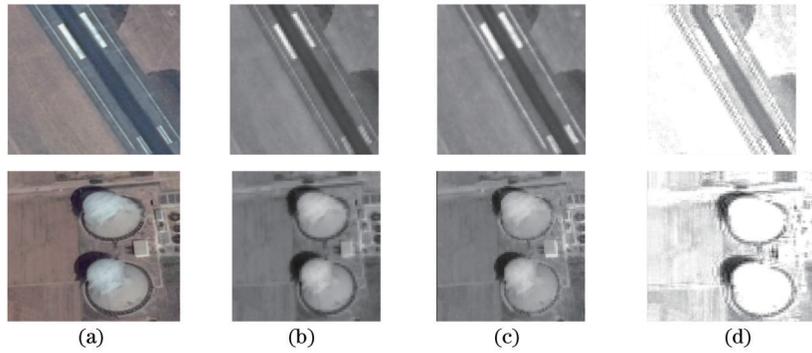


图 1 最大池化、平均池化和小波变换进行特征提取的结果。(a)遥感图像;(b)最大池化;(c)平均池化;(d)小波变换  
Fig. 1 Results of feature extraction by max-pooling, average-pooling, and wavelet transform. (a) Remote sensing images; (b) max pooling; (c) average pooling; (d) wavelet transform

Wavelet CNN<sup>[20]</sup>、SCCov<sup>[14]</sup>网络,主要从是否为空间域特征抽取、是否为小波域特征抽取、是否可抽取高阶特征和是否可进行端到端训练4个方面进行比较。

从表 1 可以看出:WHFA-Net 不仅在空间域进行

特征提取,且提取小波域的低频特征作为小波特征向量;另外,利用自积、互积获取不同域的高阶统计特征,并将其聚合为最终的深度特征向量。图 2 给出本文所提 WHFA-Net 的网络架构。

表 1 本文所提网络与现有遥感图像场景分类网络的区别

Table 1 Differences between proposed network and existing remote sensing image scene classification algorithms

Network	Spatial domain	Wavelet domain	High-order feature extraction	End-to-end training
MSCP <sup>[10]</sup>	✓	✗	✓	✗
DCCNN <sup>[19]</sup>	✓	✗	✗	✓
APDC-Net <sup>[13]</sup>	✓	✗	✗	✓
Wavelet CNN <sup>[20]</sup>	✗	✓	✗	✓
SCCov <sup>[14]</sup>	✓	✗	✓	✓
WHFA-Net	✓	✓	✓	✓

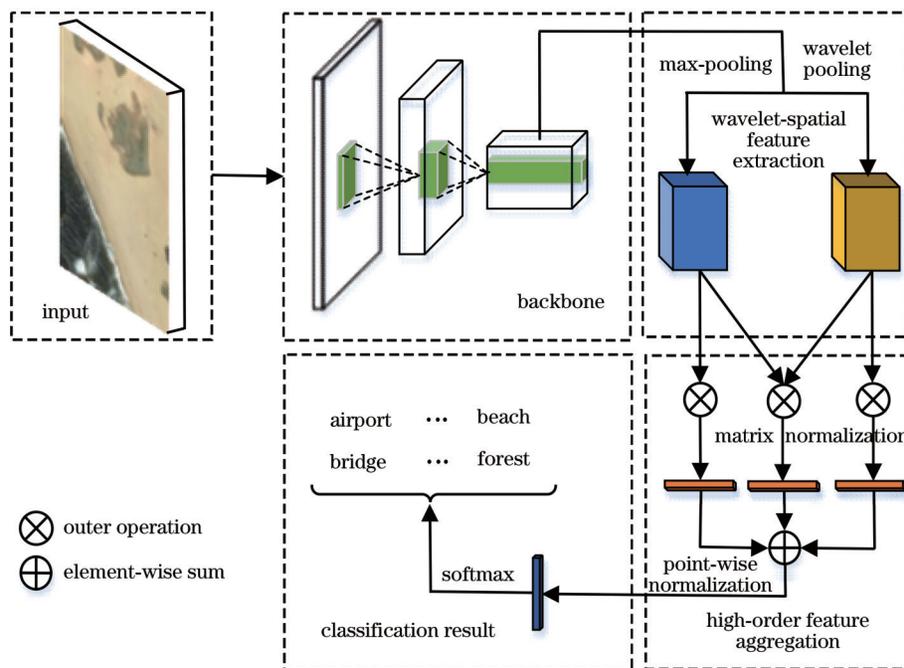


图 2 本文所提 WHFA-Net 的网络架构

Fig. 2 Network architecture of proposed WHFA-Net

WHFA-Net 主要由主干网、小波-空间特征提取模块、高阶特征聚合模块组成。本文利用 VGG-16 作为

主干网(主要是考虑到 VGG-16 最后一个卷积层特征图的维度为 512,相比于残差网特征维度(2048)较低,

以避免后续高阶特征聚合模块计算量的增加);小波-空间特征提取模块主要通过最大池化层和小波池化层抽取空间域和小波域深度特征表述;高阶特征聚合模块主要利用自积、互积提取高阶深度特征表述,利用矩阵平方根正规化(矩阵归一化)增强单分支高阶深度特征的特征表述能力,通过通道相加操作完成特征聚合,并利用特征归一化操作(逐像素归一化)提高深度特征的可辨别性。下文将详细阐述所提网络各模块的机理。

### 2.1 小波-空间特征提取模块

小波-空间特征提取可分为空间域特征提取和小波域特征提取。在小波-空间特征模块之前,首先利用 $1 \times 1$ 卷积层进行特征降维,假设主干网的深度特征输出图为 $X \in \mathbb{R}^{H \times W \times C}$ ( $H$ 为特征图的高, $W$ 为特征图的宽, $C$ 为特征图的通道数),降维后的特征图为 $X' \in \mathbb{R}^{H \times W \times C'}$ ( $C'$ 为降维后的特征图通道数)。在空间域特征提取分支,利用最大池化层进行深度特征提取,最大池化的窗口大小设置为2,步长为2,此时,空间域特征向量为 $S' \in \mathbb{R}^{(H/2) \times (W/2) \times C'}$ ,继而引入ReLU(Rectified Linear Units)层和BN(Batch Normalization)层完成对该特征的非线性激活和特征批归一化。

对于小波域深度特征提取分支,小波池化层基于正交Harr小波变换所设计的端到端特征提取层。 $X'$ 可以转化为二维信息,故可以利用二维离散Harr小波变换进行小波特征提取。这里,二维离散Harr小波变换可通过逐行逐列的一维离散Harr小波变换实现<sup>[20]</sup>。

设基向量为 $h = \frac{1}{\sqrt{2}}\{1, 1\}$ ,  $l = \frac{1}{\sqrt{2}}\{1, -1\}$ 。此时,一维离散Harr小波变换将会对 $X'$ 逐行进行计算,以获取水平方向的高频分量和低频分量;继而对已获取的高频、低频分量逐列进行计算,以获取一个低频特征分量和三个高频特征分量。前向运算的计算公式为

$$D_{a_0 a_1} = (1 \downarrow 2) \begin{bmatrix} y_{a_1} \otimes_c \\ (2 \downarrow 1)(y_{a_0} \otimes_r X') \end{bmatrix}, a_0, a_1 \in \{l, h\}, \quad (1)$$

式中: $(1 \downarrow 2)$ 和 $(2 \downarrow 1)$ 分别表示按行和按列的下采样; $\otimes_c$ 和 $\otimes_r$ 分别为列卷积和行卷积; $y_{a_0}$ 和 $y_{a_1}$ 表示由小波基 $l$ 和 $h$ 所确定的低通滤波器和高通滤波器<sup>[21]</sup>。通过式(1)的计算,可得到 $D_{ll}$ 、 $D_{lh}$ 、 $D_{hl}$ 和 $D_{hh}$ 4个特征分量。其中, $D_{ll} \in \mathbb{R}^{(H/2) \times (W/2) \times C'}$ 为低频分量,可以表示图像的轮廓等大部分特征信息, $D_{lh}$ 、 $D_{hl}$ 和 $D_{hh}$ 为高频分量,可以表示图像的噪声、细节等特征信息。小波池化层仅保留低频特征分量,这样既降低了特征冗余度,又充分顾及了图像的大部分特征信息。继而引入ReLU层和BN层,完成对小波域深度特征的抽取。

### 2.2 高阶特征聚合模块

高阶特征聚合模块主要完成对小波域和空间域深度特征的聚合。对于空间域深度特征 $S'$ 和小波域深度特征 $D_{ll}$ ,其聚合特征向量 $Z$ 的计算公式可表示为

$$Z = (D_{ll}^T D_{ll})^{1/2} + (D_{ll}^T S')^{1/2} + (S'^T S')^{1/2}, \quad (2)$$

式中: $Z \in \mathbb{R}^{C' \times C'}$ ;  $(\cdot)^{1/2}$ 表示矩阵平方根正规化操作。式(2)中包含 $S'$ 与 $S'$ 的自积、 $D_{ll}$ 与 $D_{ll}$ 的自积、 $D_{ll}$ 与 $S'$ 的互积,其结果可以有效刻画空间域深度特征和小波域深度特征的自相关性和互相关性<sup>[22-23]</sup>,特征表述能力较强。同时,矩阵平方根正规化层可以缩小矩阵最大、最小特征值的比率,以减弱矩阵缺秩现象,提高矩阵的鲁棒性。假设 $Z_1 = D_{ll}^T D_{ll}$ ,  $Z_1$ 的矩阵平方根可以通过EIG分解算法计算得到<sup>[24]</sup>,即 $Z_1 = U \Sigma U^T$ ,  $Z_1 \in \mathbb{R}^{D \times D}$ ,此时, $U^T U = I$ ,  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_D)$ 是对角矩阵, $\sigma_D$ 为 $\Sigma$ 的特征值。假设 $Z_1 = f(E) = U g(\Sigma) U^T$ ,  $f(\cdot)$ 为矩阵运算符, $g$ 为作用在 $\Sigma$ 对角元素上的平方根运算符,即 $g(\Sigma) = \text{diag}(\sigma_1^{1/2}, \sigma_2^{1/2}, \dots, \sigma_D^{1/2})$ 。此时,矩阵平方根正规化层前向运算可以表示为

$$Z_1^{1/2} = U \Sigma^{1/2} U^T. \quad (3)$$

为了将矩阵平方根正规化层嵌入深度神经网络端到端网络学习,可利用矩阵求导对其后向运算进行推导。此时,矩阵平方根正规化层<sup>[25]</sup>误差反向传播所需要的损失函数 $l$ 相对于 $E$ 的偏导数的具体计算公式为

$$\frac{\partial l}{\partial E} = U \left[ K^T \circ \left( U^T \frac{\partial l}{\partial U} \right) + \left( \frac{\partial l}{\partial \Sigma} \right)_{\text{diag}} \right] U^T, \quad (4)$$

式中: $\frac{\partial l}{\partial U}$ 和 $\frac{\partial l}{\partial \Sigma}$ 可通过链式求导法则求得; $\left( \frac{\partial l}{\partial \Sigma} \right)_{\text{diag}}$ 为

$\frac{\partial l}{\partial \Sigma}$ 构造矩阵的对角矩阵,即除主对角线以外的其他元素全部设置为0; $\circ$ 表示逐像素矩阵乘积运算符; $K$ 为一个斜对称矩阵<sup>[25]</sup>,对角线元素为0,其余元素定义为 $1/(\sigma_i - \sigma_j)$ 。

为了增强高阶聚合特征的可辨别性,引入符号平方根归一化层和 $l_2$ 归一化层,具体计算公式如下:

$$\begin{cases} Z'_{i,j} = s(Z_{i,j}) \sqrt{|Z_{i,j}| + \gamma} \\ Z''_{i,j} = \sqrt{\sum_{j=1}^{C'} (Z'_{i,j})^2 + \gamma} \end{cases}, \quad (5)$$

式中: $Z_{i,j}$ 为特征向量 $Z$ 的第 $i$ 行、第 $j$ 列特征描述子; $Z'_{i,j}$ 为符号平方根归一化层的输出向量; $Z''_{i,j}$ 为 $l_2$ 归一化层的输出向量;因 $Z_1 \in \mathbb{R}^{D \times D}$ ,则 $C' = D$ ;  $\gamma$ 为一个小整数,以保证算式有意义; $s(\cdot)$ 为符号函数,满足如下计算准则:

$$s(Z_{i,j}) = \begin{cases} 1, & |Z_{i,j}| > 0 \\ 0, & |Z_{i,j}| = 0. \\ -1, & |Z_{i,j}| < 0 \end{cases} \quad (6)$$

通过特征正规化层和归一化层可有效增强矩阵鲁棒性和特征可判别性,提升场景分类准确率。同时,高阶特征聚合模块聚合了空间域和小波域深度特征的高阶统计特征,更加有效地刻画了遥感场景图像的特征信息。

### 3 实验与分析

#### 3.1 遥感图像场景分类数据集

本文利用较为广泛的大规模遥感图像场景分类数据集进行实验。AID(Aerial Image Dataset)是由武汉大学和华中科技大学发布的遥感图像场景分类数据集<sup>[26]</sup>,该数据集共包含 10000 张图像、30 个场景类别,

如工业区、火车站、学校、森林等。图 3 给出该场景分类数据集的部分场景实例。

NWPU45(NWPU-RESISC45 Dataset)是由西北工业大学公布的包含 31500 张、45 个场景类别的大规模遥感图像场景分类数据集<sup>[27]</sup>,主要包含桥梁、火力发电站、公园、湿地等。图 4 给出该场景分类数据集的部分场景实例。

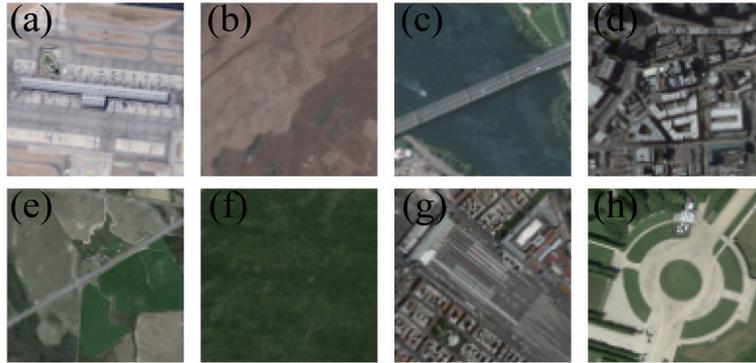


图 3 AID 数据集的若干类样本。(a)机场;(b)荒地;(c)桥梁;(d)密集住宅区;(e)农田;(f)森林;(g)火车站;(h)广场

Fig. 3 Images of several classes from AID dataset. (a) Airport; (b) bare land; (c) bridge; (d) dense residential area; (e) farmland; (f) forest; (g) railway station; (h) square

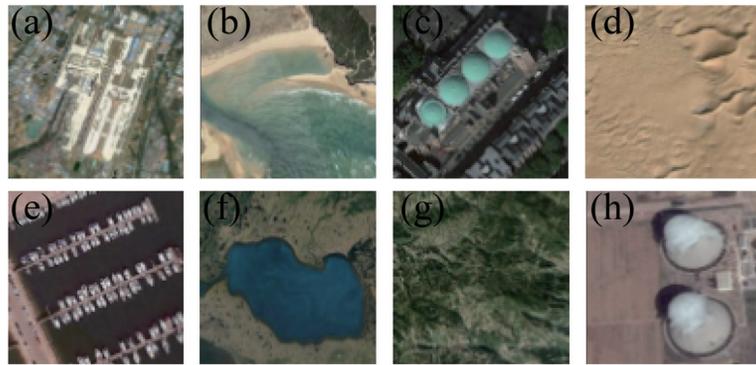


图 4 NWPU45 数据集的若干类样本。(a)机场;(b)海滩;(c)教会;(d)沙漠;(e)港口;(f)湖泊;(g)山脉;(h)热电站

Fig. 4 Images of several classes from NWPU45 dataset. (a) Airport; (b) beach; (c) church; (d) desert; (e) harbor; (f) lake; (g) mountain; (h) thermal power station

#### 3.2 实验参数设置

本文利用 PyTorch 深度学习框架进行实验与结果分析。本文所提 WHFA-Net 的主干网为 ImageNet 数据集预训练的 VGG-16 网络,神经网络优化算法为 Adam,初始学习率为 0.01,学习率衰减率为 0.1,学习率衰减步长为 20,权重衰减系数为 0.001,数据批次大小(batch size)为 128,动量为 0.9。其他硬件平台为:64 GB RAM 和 NVIDIA GeForce RTX 24G 3090 GPU 型号。

#### 3.3 算法性能对比

将本文所提 WHFA-Net 与基准网络 VGG-16,多分支网络 DCCNN<sup>[19]</sup>、GBNet<sup>[16]</sup>和 LCNN-BFF<sup>[17]</sup>,结合注意力机制的多分支网络 APDC-Net<sup>[13]</sup>,多分支二阶网络 MSCP<sup>[10]</sup>,以及小波域神经网络 Wavelet CNN<sup>[20]</sup>的算法性能进行对比。随机选取 AID 数据集中 20% 和 50% 的场景图像作为训练集,剩余场景图像

作为测试集;另外,随机选取 NWPU45 数据集中 10% 和 20% 的场景图像作为训练集,剩余其他场景图像作为测试集,利用总准确率(OA)作为客观评价指标,所有实验均进行 5 次。表 2 给出了算法性能的对比结果,表中±之前的数值表示场景分类准确率平均值,±之后的数值为场景分类准确率标准差,用以衡量模型的稳定性。

通过表 2 可以看出,与基准网络 VGG-16 相比,DCCNN 构建了一个具有 23 层密集连接、注意力机制以及嵌入具有特殊语义信息的 Loss 层,可以得到较好的遥感图像场景分类效果。类似地,APDC-Net 利用简化的密集连接以保留不同等级的深度特征,引入可训练的池化层增强局部特征,并嵌入多级监督策略完成网络的端到端训练,该网络可得到比 DCCNN、VGG-16 更好的场景分类效果。另外,在多分支 CNNs 方面,GBNet 嵌入双向连接聚合多层深度卷积特征,

表 2 算法的场景分类准确率对比

Table 2 Comparison of scene classification accuracy of algorithms

unit: %

Network	Classification accuracy of AID		Classification accuracy of NWPU45	
	20% scene images	50% scene images	10% scene images	20% scene images
VGG-16	86.59±0.29	89.64±0.36	76.47±0.18	79.79±0.15
DCCNN	87.37±0.41	91.49±0.22	83.97±0.19	85.63±0.18
MSCP	91.52±0.21	94.42±0.17	85.33±0.17	88.93±0.14
APDC-Net	88.56±0.29	92.15±0.29	85.94±0.22	87.84±0.26
Wavelet CNN	—	<b>96.65±0.24</b>	84.69±0.44	88.84±0.38
GBNet	90.16±0.24	93.72±0.34	—	—
LCNN-BFF	91.66±0.48	94.62±0.16	86.53±0.15	91.73±0.17
WHFA-Net	<b>91.72±0.19</b>	96.14±0.14	<b>88.59±0.21</b>	<b>91.84±0.13</b>

并利用门控函数消除双向连接的干扰信息,可得到比 DCCNN 和 ADPC-Net 更优的场景分类效果;同时, LCNN-BFF 所提出的双分支特征聚合网络从不同方面抽取场景信息,并嵌入深度可分离卷积模块,故得到了较高的场景分类准确率。上述 DCCNN、APDC-Net、GBNet、LCNN-BFF 可认为是一种多分支一阶深度特征网络,本文所提 WHFA-Net 为多分支高阶特征聚合网络,进行对比实验。MSCP 网络在增加多分支结构之后,引入高阶深度统计特征作为最终的深度特征表述,其能够更好地刻画场景图像,与基准网络 VGG-16 相比,MSCP 网络的场景分类效果得到大幅提升。Wavelet CNN 构建了基于小波域的深度特征表述,虽然该网络是基于单分支的,与基于多分支空间域的其他相关网络相比,Wavelet CNN 的特征表述能力更强,得到了具有竞争力的场景分类效果。本文所提出的 WHFA-Net 融合了空间域和小波域深度特征表述,且利用高阶特征融合策略和特征归一化增强特征可判别性,故 WHFA-Net 在大部分情况下能够得到最优的场景分类效果。与 VGG-16、DCCNN 和 APDC-Net 相比,WHFA-Net 的场景分类准确率得到提升;与 GBNet、LCNN-BFF 相比,WHFA-Net 的场景分类效果也有不同程度的改善。这主要是因为 WHFA-Net 嵌入了小波域特征抽取模块和高阶特征聚合模块,而 DCCNN、APDC-Net、GBNet、LCNN-BFF 均属于空间域多分支网络,其多采用密集连接、注意力机制、双向连接等增强深度特征表述能力。与 MSCP 网络相比,WHFA-Net 的场景分类准确率提升了 0.20%~3.26%,虽然 MSCP 网络中嵌入多分支二阶特征统计量作为特征描述,但该网络仍属于空间域深度特征学习网络,缺少对不同域深度特征抽取的考虑。与小波域神经网络 Wavelet CNN 相比,在 NWPU45 数据集上,WHFA-Net 的场景分类准确率得到提高,虽然在 50% 训练集比例的 AID 数据集上,其分类效果有小幅度的降低,但 WHFA-Net 的稳定性较好,另外,在更大规模的 NWPU45 数据集上,本文所提 WHFA-Net 的场景分类效果提升明显,网络稳定性更优,这是由于 Wavelet CNN 仅利用小波域深度特征刻画遥感图像场景信息,缺少对遥感场景图像高阶特征的抽取,这也从

另一方面表明高阶特征聚合模块的有效性。

### 3.4 混淆矩阵

图 5 和图 6 分别为 NWPU45 数据集的 10% 和 20% 训练集比例下的场景分类混淆矩阵,图中 1~45 代表 45 类场景,混淆矩阵中的坐标数字为数据集中按照场景类别英文首字母顺序排序所得到的场景类别,详细信息请参考文献[27]。

从图 6 可以看出:所提 WHFA-Net 对大部分场景类别的分类准确率均达到了 80% 以上,仅有少部分场景类别的分类准确率在 80% 以下,如对教堂场景类别的分类准确率为 74%、对工业区的分类准确率为 73%、对宫殿的分类准确率为 63%,这些场景类别被错分的情况比较严重,主要是因为其类间差异性较小,如教堂和宫殿场景类别较为相似,因此:在 10% 的训练集比例下,有 12% 的教堂场景被错分为宫殿场景,在 20% 的训练集比例下也有着相似的分类结果。所提 WHFA-Net 对部分场景类别如 10% 训练集比例下的森林场景类别的分类准确率达到 97%,对 20% 训练集比例下的丛林和海冰场景的分类准确率分别达到了 98% 和 100%。

### 3.5 网络曲线

为了较好地分析本文所提 WHFA-Net 的训练与测试稳定性,图 7 和图 8 给出当训练集比例为 10% 和 20%,WHFA-Net 在 NWPU45 数据集的训练、测试和 Loss 曲线。WHFA-Net 训练过程中采用学习率衰减策略,故其学习率会根据训练期数(epoch)而发生变化。从图 7 和图 8 可以得出:在两种训练集比例下,WHFA-Net 所得到的网络模型均处于稳定状态。具体来讲:当学习率为 0.01 时,模型 Loss 下降较快,且模型 Top1 错误率(Top1 error)和 Top5 错误率(Top5 error)也随之下降;当学习率下降至 0.001 时,三种曲线具有不同程度的下降;当学习率为 0.0001 时,模型趋于稳定,此时说明 WHFA-Net 已学习到较为稳定的深度特征。

### 3.6 消融实验

WHFA-Net 包含小波域特征提取、空间域特征提取、高阶特征聚合模块。本节采用消融实验验证各分支和模块的有效性,数据集选定为 20% 训练集比例下

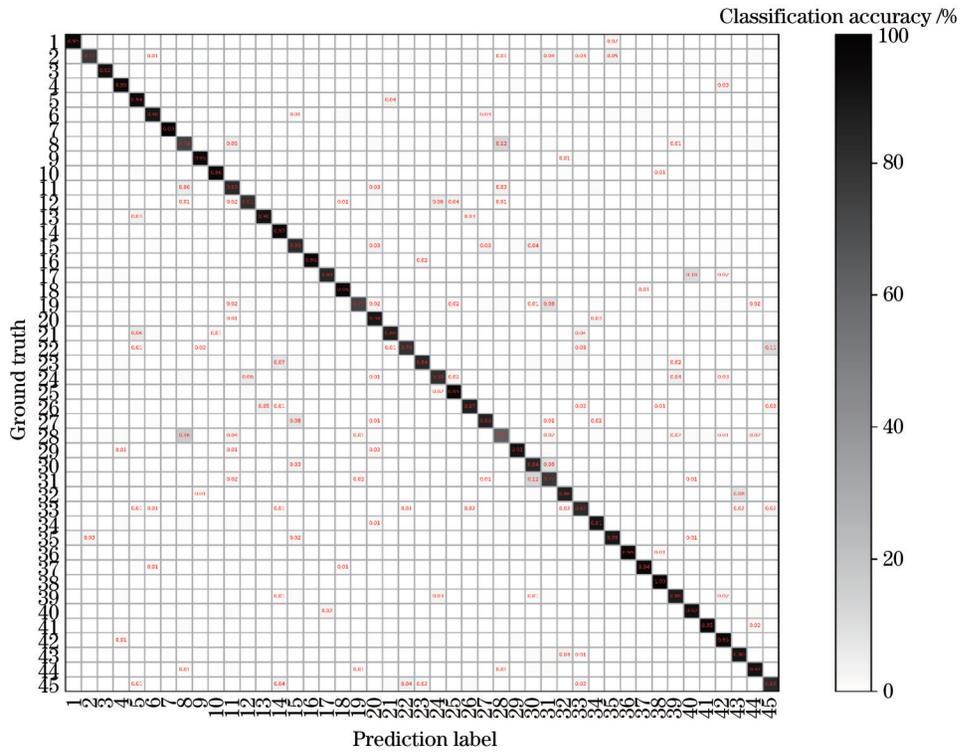


图 5 NWPU45数据集的10%训练集比例下的场景分类混淆矩阵

Fig. 5 Confusion matrix of scene classification on NWPU45 dataset under 10% training set ratio

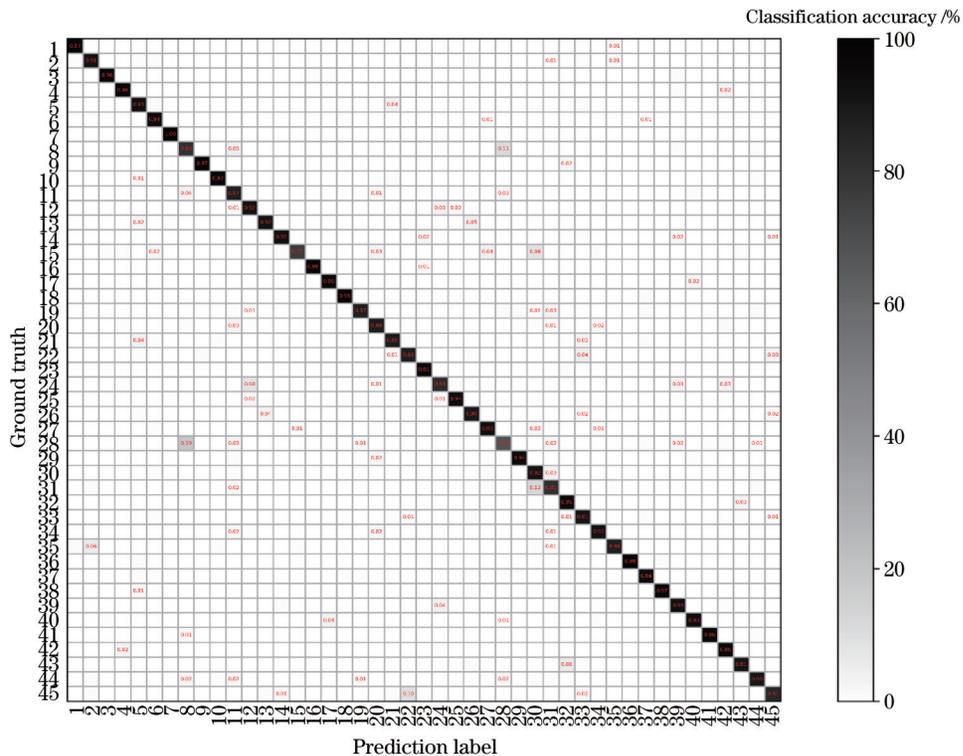


图 6 NWPU45数据集的20%训练集比例下的场景分类混淆矩阵

Fig. 6 Confusion matrix of scene classification on NWPU45 dataset under 20% training set ratio

的NWPU45,表3给出消融实验结果。

与空间域深度特征相比,小波域深度特征更具有可辨别性,场景分类准确率有约为0.57%的提升,这主要是因为小波域特征提取分支是根据正交Harr小

波变换所设计的。空间域+小波域、空间域+高阶特征聚合、小波域+高阶特征聚合的场景分类准确率均有不同程度的提高,值得注意的是:与其他两种方法相比,小波域+高阶特征聚合方法得到了较高的场景分

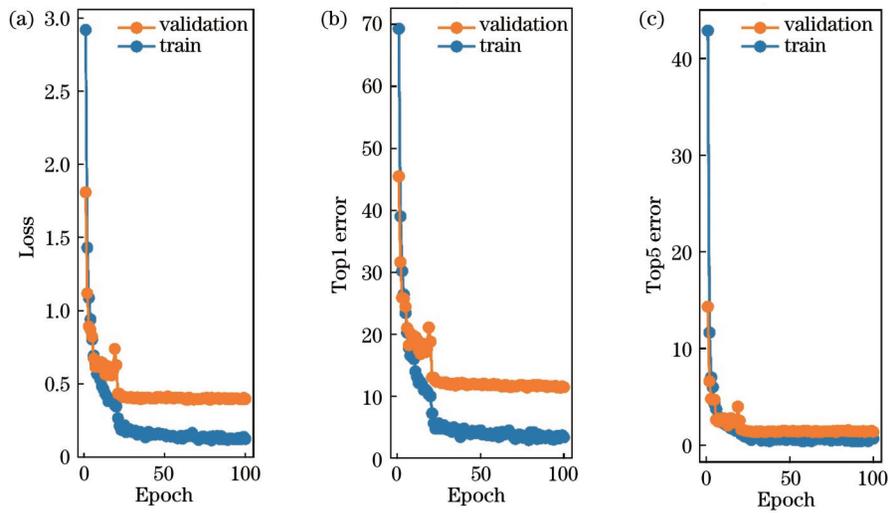


图 7 NWPU45 数据集的 10% 训练集比例下的网络曲线。(a) 损失-期数曲线; (b) Top1 错误率-期数曲线; (c) Top5 错误率-期数曲线

Fig. 7 Network curves on NWPU45 dataset under 10% training set ratio. (a) Loss-epoch curves; (b) Top1 error-epoch curves; (c) Top5 error-epoch curves

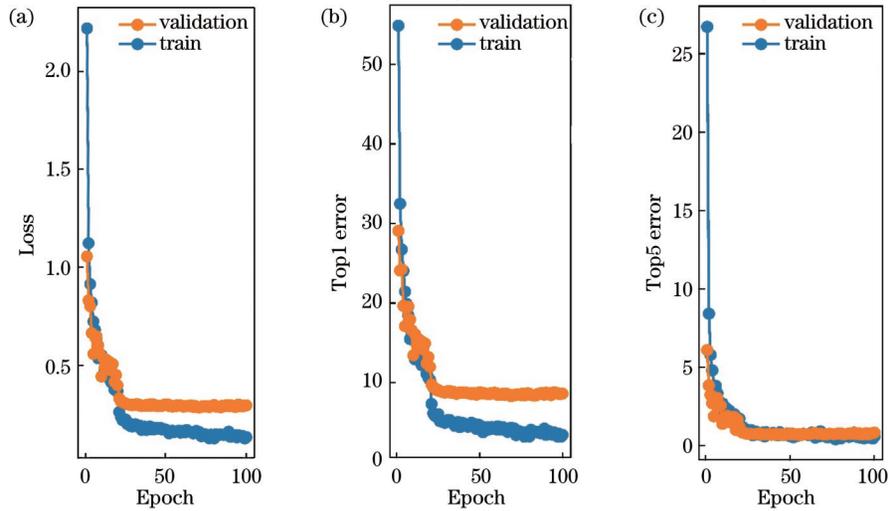


图 8 NWPU45 数据集的 20% 训练集比例下的网络曲线。(a) 损失-期数曲线; (b) Top1 错误率-期数曲线; (c) Top5 错误率-期数曲线

Fig. 8 Network curve on NWPU45 dataset under 20% training set ratio. (a) Loss-epoch curves; (b) Top1 error-epoch curves; (c) Top5 error-epoch curves

表 3 消融实验结果

Table 3 Ablation experimental results

Spatial domain	Wavelet domain	Higher-order feature aggregation	OA / %
√	×	×	83.41
×	√	×	83.98
√	×	√	87.18
×	√	√	87.61
√	√	×	87.10
√	√	√	88.59

类准确率。因此,小波域特征提取分支、空间域特征提取分支、高阶特征聚合模块均对遥感图像场景分类效

果有不同程度的提升。

另外,高阶特征聚合模块中包含三个分支:小波域-小波域(WWD)、空间域-空间域(SSD)、小波域-空间域(WSD)。为了验证三支高阶特征的有效性,图 9 给出了实验结果对比。

图 9 中 W 代表 WWD, S 代表 SSD, WS 代表 WSD。W、S、WS 三支均对场景分类有效,且 W 分支的分类效果最佳,可达到 87.61%,这也从另一方面说明本文所提的小波域特征分支的有效性。另外, W+S、W+WS、S+WS 两分支融合学习也可称为不同分支深度特征的自积和互积,与 W、S、WS 三支相比,其所生成的高阶深度特征更具有可辨别性,三个融合学习分支的场景分类效果相当,但从更为细致的角

度观察,  $W+WS$  分支和  $S+WS$  分支相较于  $W+S$  分支, 场景分类准确率有小幅(约 0.1%)提升, 这说明在不同分支互相关性的引导下, 分别嵌入小波域和空间域深度特征在一定程度上提升了场景分类准确率, 这主要是因为小波域特征分支可提取特征图边缘、轮廓等低频特征信息, 其与空间特征信息的高阶特征统计量可更加有效地刻画遥感图像场景信息, 但仅与小波域低频特征信息进行互相关特征统计存在一定的特征缺失, 需要与  $W$  和  $S$  分支进行融合, 形成特征互补, 才可得到较优的场景分类效果, 这也是  $WS$  分支场景分类效果略差于  $S$  和  $W$  分支的主要原因。另外,  $W+WS$  和  $S+WS$  分支的场景分类效果相当, 分类准确率浮动在 0.05% 左右, 因此, 在  $WS$  互相关分支存在的情况下,  $W$  和  $S$  自相关分支的特征表述能力基本相当。值得注意的是,  $W+S+WS$  方法可得到最优分类效果 (88.59%), 这主要是因为多分支的高阶融合深度特征更具可辨别性。因此, 三支高阶特征均对场景分类效果的提升有促进作用。

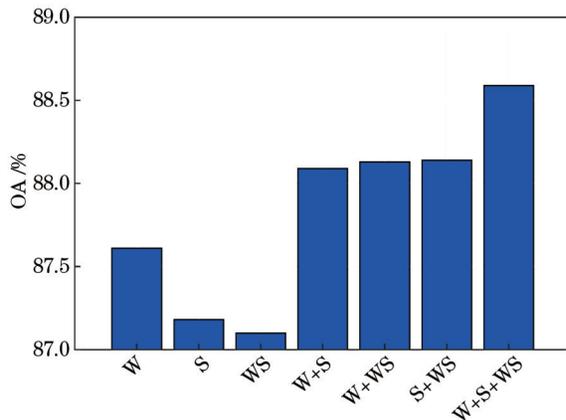


图9 三支实验结果对比

Fig. 9 Comparison of results of three branches

## 4 结 论

提出一种基于小波-空间高阶特征聚合网络的遥感图像场景分类算法。所提网络以 VGG-16 为主干网, 嵌入小波-空间特征提取模块, 其中空间域特征提取分支和小波域特征提取分支可在空间域和小波域有效抽取遥感场景图像的深度特征表述; 继而引入高阶特征聚合模块, 通过自积、互积获取不同分支的高阶深度统计特征, 并利用特征归一化层增强特征可辨别性。实验结果表明所提网络在两个具有竞争力的遥感图像场景分类数据集上的分类精度最高可达到 96.14% 和 91.84%, 有效提升遥感图像场景分类的准确率, 同时消融实验结果和模型训练曲线验证了本文所提模型中所有模块及分支的性能差异与有效性。另外, 本文所提模型仅利用主干网最后一个网络层的深度特征, 且小波域特征分支的引入和高阶深度特征的计算均会造成计算量的增加。因此, 后续的研究工作将会针对多层深度特征的合理构建和计算量的有效减少有序

开展。

## 参 考 文 献

- [1] 杨州, 慕晓冬, 王舒洋, 等. 基于多尺度特征融合的遥感图像场景分类[J]. 光学精密工程, 2018, 26(12): 3099-3107.  
Yang Z, Mu X D, Wang S Y, et al. Scene classification of remote sensing images based on multiscale features fusion[J]. Optics and Precision Engineering, 2018, 26(12): 3099-3107.
- [2] Xu Y H, Du B, Zhang L P. Assessing the threat of adversarial examples on deep neural networks for remote sensing scene classification: attacks and defenses[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(2): 1604-1617.
- [3] 龚希, 陈占龙, 吴亮, 等. 用于高分辨遥感影像场景分类的迁移学习混合专家分类模型[J]. 光学学报, 2021, 41(23): 2301003.  
Gong X, Chen Z L, Wu L, et al. Transfer learning based mixture of experts classification model for high-resolution remote sensing scene classification[J]. Acta Optica Sinica, 2021, 41(23): 2301003.
- [4] Peng C, Li Y Y, Jiao L C, et al. Efficient convolutional neural architecture search for remote sensing image scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(7): 6092-6105.
- [5] 汪鹏, 刘瑞, 辛雪静, 等. 基于残差网络的遥感图像场景分类算法[J]. 激光与光电子学进展, 2021, 58(2): 0210001.  
Wang P, Liu R, Xin X J, et al. Scene classification of optical remote sensing images based on residual networks [J]. Laser & Optoelectronics Progress, 2021, 58(2): 0210001.
- [6] Li Y S, Zhu Z H, Yu J G, et al. Learning deep cross-modal embedding networks for zero-shot remote sensing image scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(12): 10590-10603.
- [7] 朱淑鑫, 周子俊, 顾兴健, 等. 基于 RCF 网络的遥感图像场景分类研究[J]. 激光与光电子学进展, 2021, 58(14): 1401001.  
Zhu S X, Zhou Z J, Gu X J, et al. Scene classification of remote sensing images based on RCF network[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1401001.
- [8] 吴晨, 王宏伟, 袁昱纬, 等. 基于图像特征融合的遥感场景零样本分类算法[J]. 光学学报, 2019, 39(6): 0610002.  
Wu C, Wang H W, Yuan Y W, et al. Image feature fusion based remote sensing scene zero-shot classification algorithm[J]. Acta Optica Sinica, 2019, 39(6): 0610002.
- [9] Ma A L, Yu N, Zheng Z, et al. A supervised progressive growing generative adversarial network for remote sensing image scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5618818.

- [10] 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.
- [11] Li E Z, Xia J S, Du P J, 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.
- [12] Ni K, Wu Y Q. Scene classification from remote sensing images using mid-level deep feature learning[J]. *International Journal of Remote Sensing*, 2020, 41(4): 1415-1436.
- [13] Bi Q, Qin K, Zhang H, et al. APDC-net: attention pooling-based convolutional network for aerial scene classification[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(9): 1603-1607.
- [14] He N J, Fang L Y, Li S T, et al. Skip-connected covariance network for remote sensing scene classification [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(5): 1461-1474.
- [15] Li E Z, Samat A, Zhang C, et al. First and second-order information fusion networks for remote sensing scene classification[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 8014405.
- [16] Sun H, Li S Y, Zheng X T, et al. Remote sensing scene classification by gated bidirectional network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(1): 82-96.
- [17] Shi C P, Wang T, Wang L G. Branch feature fusion convolution network for remote sensing scene classification[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 5194-5210.
- [18] Li C, Zhuang Y, Liu W C, et al. Effective multiscale residual network with high-order feature representation for optical remote sensing scene classification[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 6003105.
- [19] Bi Q, Qin K, Li Z L, et al. Multiple instance dense connected convolution neural network for aerial image scene classification[C]//2019 IEEE International Conference on Image Processing, September 22-25, 2019, Taipei, China. New York: IEEE Press, 2019: 2501-2505.
- [20] Williams T, Li R. Wavelet pooling for convolutional neural networks[C]// 6th International Conference on Learning Representations, April 30-May 3, 2018, Vancouver, BC, Canada. [S.l.]: [s.n.], 2018.
- [21] Fujieda S, Takayama K, Hachisuka T. Wavelet convolutional neural networks[EB/OL]. (2018-05-20) [2021-05-04]. <https://arxiv.org/abs/1805.08620>.
- [22] Kong S, Fowlkes C. Low-rank bilinear pooling for fine-grained classification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 7025-7034.
- [23] Ni K, Wu Y Q, Wang P. Scene classification from synthetic aperture radar images using generalized compact channel-boosted high-order orderless pooling network[J]. *Remote Sensing*, 2019, 11(9): 1079.
- [24] Li P H, Xie J T, Wang Q L, et al. Towards faster training of global covariance pooling networks by iterative matrix square root normalization[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 947-955.
- [25] Li P H, Xie J T, Wang Q L, et al. Is second-order information helpful for large-scale visual recognition? [C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2089-2097.
- [26] Xia G S, Hu J W, Hu F, et al. AID: a benchmark data set for performance evaluation of aerial scene classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(7): 3965-3981.
- [27] 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.