

基于双重注意力机制的遥感图像场景分类特征表示方法

徐从安^{①②} 吕亚飞^{*①③} 张筱晗^{①④} 刘瑜^① 崔晨浩^⑤ 顾祥岐^①

^①(海军航空大学信息融合研究所 烟台 264000)

^②(清华大学电子工程系 北京 100084)

^③(91977部队 北京 100089)

^④(61646部队 北京 100089)

^⑤(32144部队 渭南 714000)

摘要: 针对遥感图像场景分类面临的类内差异性大、类间相似性高导致的部分场景出现分类混淆的问题, 该文提出了一种基于双重注意力机制的强鉴别性特征表示方法。针对不同通道所代表特征的重要性程度以及不同局部区域的显著性程度不同, 在卷积神经网络提取的高层特征基础上, 分别设计了一个通道维和空间维注意力模块, 利用循环神经网络的上下文信息提取能力, 依次学习、输出不同通道和不同局部区域的重要性权重, 更加关注图像中的显著性特征和显著性区域, 而忽略非显著性特征和区域, 以提高特征表示的鉴别能力。所提双重注意力模块可以与任意卷积神经网络相连, 整个网络结构可以端到端训练。通过在两个公开数据集AID和NWPU45上进行大量的对比实验, 验证了所提方法的有效性, 与现有方法对比, 分类准确率取得了明显的提升。

关键词: 遥感图像处理; 场景分类; 注意力机制; 特征表示

中图分类号: TN911.73; TP183

文献标识码: A

文章编号: 1009-5896(2021)03-0683-09

DOI: [10.11999/JEIT200568](https://doi.org/10.11999/JEIT200568)

A Discriminative Feature Representation Method Based on Dual Attention Mechanism for Remote Sensing Image Scene Classification

XU Cong'an^{①②} LÜ Yafei^{①③} ZHANG Xiaohan^{①④} LIU Yu^①
CUI Chenhao^⑤ GU Xiangqi^①

^①(Information Fusion Institute, Naval Aviation University, Yantai 264000, China)

^②(Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

^③(91977 Troop, Beijing 100089, China)

^④(61646 Troop, Beijing 100089, China)

^⑤(32144 Troop, Weinan, 714000, China)

Abstract: Considering the problem of low classification accuracy caused by large intra-class differences and high inter-class similarity in remote sensing image scene classification, a discriminative feature representation method based on dual attention mechanism is proposed. Due to the difference in the importance of the features contained in different channels and the significance of different local regions, the channel-wise and spatial-wise attention module are designed, based on the high-level features extracted by the Convolutional Neural Networks. Relying on the ability to extract contextual information, the Recurrent Neural Network is adopted to learn and output the importance weights of different channels and different local regions, paying more attention to the salient features and salient regions, while ignoring non-salience features and regions, to enhance the discriminative ability of feature representation. The proposed dual attention module can be connected to the last convolutional layer of any convolutional neural network, and the network structure can be trained end-to-end.

收稿日期: 2020-07-10; 改回日期: 2020-12-07; 网络出版: 2020-12-15

*通信作者: 吕亚飞 YFei_Lv@163.com, xcatougao@163.com

基金项目: 国家自然科学基金(61790550, 61790554, 61531020, 61671463)

Foundation Items: The National Natural Science Foundation of China (61790550, 61790554, 61531020, 61671463)

Comparative experiments are conducted on the two public data sets AID and NWPU45. Compared with the existing methods, the classification accuracy has been significantly improved, and the effectiveness of the proposed method can be verified.

Key words: Remote sensing image process; Scene classification; Attention mechanism; Feature representation

1 引言

随着卫星与无人机等遥感观测技术的飞速发展, 高分辨率遥感图像的数据总量和数据类型都不断增加, 遥感信息处理正在进入“遥感大数据时代”^[1]。高分辨率遥感图像相比于中低分辨率遥感图像包含着更丰富的纹理、细节和地物特征, 在为遥感图像的自动解译工作提供了丰富信息的同时, 也提出了更多的挑战和要求。

作为遥感图像解译工作中的重要组成部分, 遥感图像场景分类(Remote Sensing Image Scene Classification, RSISC)是根据图像中包含的高层语义信息, 将遥感图像映射到预定的类别标签中, 实现对图像场景内容的自动判别。根据特征表示方法的不同, 现有的研究方法主要可以分为基于手工特征的方法^[2,3]和基于学习特征的方法^[4,5]。尤其自AlexNet^[6]的提出, 以卷积神经网络(Convolutional Neural Networks, CNNs)为代表的基于学习特征的方法凭借较好的高层语义特征提取能力较大提高了算法的性能表现, 逐渐取代了基于手工特征的相关方法。

基于CNNs的网络结构在计算机视觉领域不断提高着分类、识别等任务的性能表现, 但在RSISC中, 由于自然场景图像与遥感图像在成像内容和成像特点上的差异, 用于自然场景图像的深度网络结构难以直接适用于RSISC问题。这主要是由遥感图像俯视成像带来的成像范围更广、尺度变化大和语义内容更复杂等问题导致的^[7]。成像目标的繁杂性和语义内容的复杂性导致遥感图像场景分类面临着类内差异性大、类间相似性高的难点。为了解决该问题, 已有较多研究方法提出, 文献^[8]提出了鉴别性卷积神经网络(Discriminative-Convolutional Neural Networks, D-CNN)的网络结构, 将度量学习和CNNs相结合, 以拉近同类场景、推远异类场景图像特征表示间的距离, 较好地提高了特征表示的鉴别性能力(discriminative ability), 但该方法忽略了对遥感图像局部特征的关注; 文献^[9]提出了基于区域的深度特征提取网络, 通过先检测后特征表示的方式, 对图像中显著的局部区域进行特征提取, 虽然关注了局部区域特征, 但丢失了全局特征; 文献^[10]提出了多尺度CNN结构, 利用两个不同尺度的网络结构分别提取图像的特征表示, 以学习遥感

图像中的尺度不变性特征; 文献^[11]通过先聚类再重排的方式挑选遥感图像中重要的局部区域进行特征表示, 方法无法端到端训练, 算法复杂度高; 文献^[12]利用注意力机制通过长短时记忆网络生成了高层特征图的重要性权重, 对高层特征中每一个元素的重要性进行修正, 虽然该方法可以端到端训练, 但对每一个元素的修正没有实现对局部区域重要性的修正。文献^[13]中针对遥感图像成像范围广、语义复杂的特点提出了适用于遥感图像特征表示的注意力模块, 在高层特征的基础上, 分别从通道维和空间维两个维度对图像的特征表示进行修正, 以使得图像中显著区域和显著性特征更被关注, 提高特征表示的鉴别性能力, 该方法在遥感图像检索问题上得到了有效性验证。不同于文献^[13]中注意力模块学习的无先验知识, 文献^[14]提出了一种局部和全局特征融合的特征表示方法, 利用循环神经网络分别读取和捕获不同区域特征表示间的上下文关系, 以生成不同区域的重要性权重, 实现对局部重要区域的重点关注, 整个方法可端到端地实现局部和全局特征的提取与融合训练, 在多个公开数据集上达到了最佳实现(State-Of-The-Art, SOTA), 但该方法只是在空间维对不同区域的重要性进行了修正, 忽略了通道维重要性的作用。

基于以上讨论, 针对遥感图像成像范围广、语义信息复杂的特点, 而现有的特征表示方法又难以同时对遥感图像中的局部显著区域和显著性特征进行关注并有效地表达, 本文提出了一种双重注意力机制的遥感图像场景分类特征表示方法, 以CNNs提取的高层特征为先验信息, 利用循环神经网络对上下文信息的提取能力, 分别设计了一个通道维和空间维相结合的双重注意力模块, 通过生成不同通道和不同区域的重要性权重, 对原有高层特征表示中显著性特征和显著性区域两个维度的特征表示进行增强, 减弱对非重要区域(背景区域)和非显著性特征的关注, 提高特征表示的鉴别性能力。

2 本文方法

2.1 方法整体框架

卷积神经网络的强特征表示主要归因于其对高层语义特征的提取能力, 如图1所示ResNet50所提取的高层特征($7 \times 7 \times 2048$), 其3维的特征结构可以

分解为空间维(7×7)和特征维(2048)两个维度。一方面,空间维(7×7)与输入图像(一般为224×224)的空间位置存在着对应关系,即高层特征空间维中每一个像素(1×1)对应着输入图像中大小为32×32的图像块,因此,空间维(7×7)中每一个像素就可以看作输入图像中不同区域的特征代表;另一方面,卷积神经网络中不同的卷积层代表着不同的特征,每一层卷积层可能关注着不同类型的特征,因此,通道维的物理意义可以看作2048中不同类型的特征,对不同类别的遥感图像来说,决定其身份信息特征也必定不相同。因此综上分析,为了进一步提高CNNs特征表示的鉴别性能力,需要从空间维和通道维两个维度进一步对高层特征进行修正,对重要区域和特征进行增加,非重要区域和特征进行减弱。

以CNNs提取的高层特征为基础,所提的双重注意力模块在通道维和空间维进一步对高层特征进行再权重化,方法的框架图如图1所示。以ResNet50为例,以其最后一个大小为7×7×2048的卷积层为高层特征,依次连接通道维注意力模块和空间维注意力模块;通道维注意力模块以高层特征中各通道展开后得到的49维向量为输入,共2048个49维的通道向量依次输入到以门限循环单元(Gated Recurrent Units, GRU)^[15]为基本单元的循环神经网络中,以提取各通道间的上下文关系,生成各通道的重要性权重,对高层特征的通道重要性进行加权修正;经通道注意力模块修正后的高层特征进一步输入到空间维注意力模块中,以高层特征中不同区域的特征表示为先验以此输入到以GRU为基本单元的循环神经网络中,输出得到不同区域的重要性权重,以此实现对显著性区域的增强、背景区域的抑制。最后,将经过通道注意力和空间注意力修正后的高层特征以全局平均池化的方式得到维度为2048维的特征向量,并与两个节点为2048的全连接层连接后,通过softmax函数预测输入图像类别。

2.2 通道维注意力模块

经CNNs提取获得的特征图中,不同的通道代表着不同种类的特征,以ResNet50为例,高层特征图中包含2048个通道,可以看作2048种视觉特征。不同特征对遥感图像语义内容的表达起着不同的作用,通道维注意力模块的目的就是在CNNs提取的特征图基础上,根据不同通道的重要性程度,进一步生成各通道的重要性权重,实现对显著性特征的加强,对非显著特征的忽略。通道维注意力模块的具体网络结构如图2所示。

将得到的高层特征 $F(7 \times 7 \times 2048)$ 按照空间维展开得到2048个维度为49的通道特征表示: $F_1 = \{f_1, f_2, \dots, f_{2048}\}, f_i \in d^{49}$ 。文献[13]对通道注意力模块的生成是以随机初始化的方式,学习生成各通道的重要性,虽然起到了对不同通道重要性修正的目的,但对各通道间的相互关系和通道自身先验信息的利用不足。而循环神经网络作为处理序列信息的重要手段,能有效提取序列信息中的上下文关系,对进一步确定不同通道间的相互关系和重要性权重提供了解决方法。因此,为了准确生成不同通道间的重要性权重,本文以各通道的特征表示为先验信息,依次输入到以GRU为基本单元的循环神经网络中,GRU是一种包含参数更少,但在很多序列关系提取任务中都有较好性能表现的一种循环神经网络。通过对不同通道间的上下文关系的提取,生成不同通道修正后的特征表示 $F_2 = \{f'_1, f'_2, \dots, f'_{2048}\}, f'_i \in d^{49}$;修正后的2048个特征表示通过一个权重共享的全连接层(49×1)和relu激活函数,将各通道的特征表示映射到维度为1的向量;然后,将2048个维度为1的向量以相连(concatenate)的方式连接,并进一步与一个全连接层和sigmoid激活函数相连,以sigmoid函数输出的值作为各通道的重要性权重,该过程可见式(1)

$$W_{C_attention} = \text{sigmoid} \cdot (\text{relu}(\text{Concatenate}(F_2 \times W_1)) \times W_2) \quad (1)$$

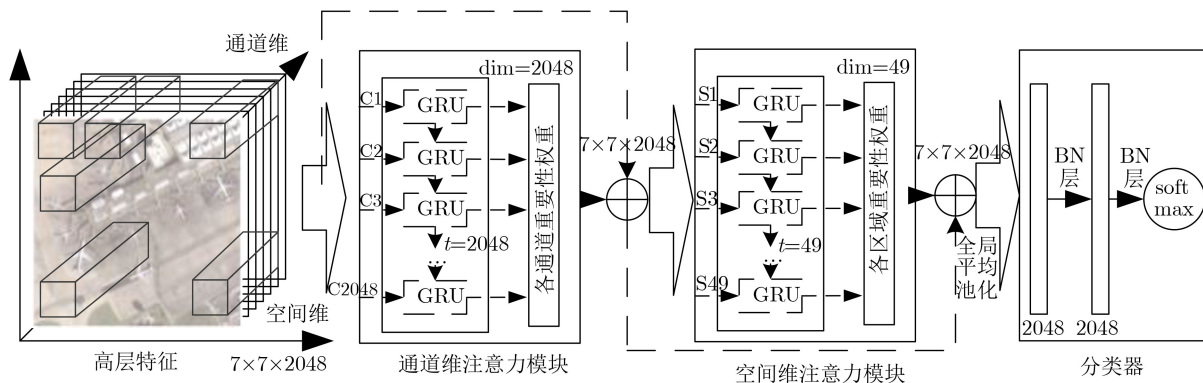


图1 本文算法框架图

其中，两个全连接层的权重 $W_1 \in d^{49 \times 1}$, $W_2 \in d^{2048 \times 2048}$, $W_{C_attention} \in d^{2048}$ 。

在得到各通道的重要性权重 $W_{C_attention}$ 后，将权重与原高层特征相乘，实现对各通道重要性的修正；此外，借鉴ResNet中的跳跃连接，将修正后的高层特征与原高层特征之间相加，以确保如果在通道注意力模块不起作用的情况下，整个网络能够保持原有的特征信息不被减少，整个通道注意力模块的计算过程可以表示为

$$F_3 = F + F \cdot W_{C_attention} \tag{2}$$

其中， $W_{C_attention}$ 经调整尺寸(resize)为 $1 \times 1 \times 2048$ ，·代表按元素相乘。

2.3 空间维注意力模块

遥感图像广阔的成像范围导致遥感图像成像内容繁杂、目标众多，如何更好地关注显著性区域，忽略背景区域对于准确地实现遥感图像特征表示起着关键作用。空间注意力模块的目的就是在CNNs提取的高层特征基础上，进一步捕获不同区域特征的上下文联系，以生成各区域的重要性权重，实现

对显著区域的重点关注，对背景区域的忽略。与通道维注意力模块的网络结构相似，空间维注意力模块的网络结构图如图3所示。

将经通道维注意力模块修正后的高层特征 $F_3(7 \times 7 \times 2048)$ 输入到空间维注意力模块中。由文献[14]可知，高层特征 F_3 空间维中每个元素的特征表示 $(1 \times 1 \times 2048)$ 可以看作原图像中局部区域的特征表示，因此，可以得到图像中 $49(7 \times 7)$ 个区域的高层特征表示 $F_4 = \{v_1, v_2, \dots, v_{49}\}, v_i \in d^{2048}$ 。与通道维注意力模块的结构类似，为了充分探索不同图像区域间的相互关系和重要性权重，利用循环神经网络作为不同区域序列信息的上下文关系提取器。将49个不同区域的特征表示依次输入到由GRU构成的循环神经网络中，并与共享全连接层和最后的分类全连接层相连，通过softmax函数输出不同区域的重要性权重 $W_{S_attention}$ ；与通道维注意力模块相同，将空间维注意力模块的重要性权重与输入的高层特征按元素相乘，并与原输入的高层特征相加，得到最终的经通道维和空间维双重注意力模块修正

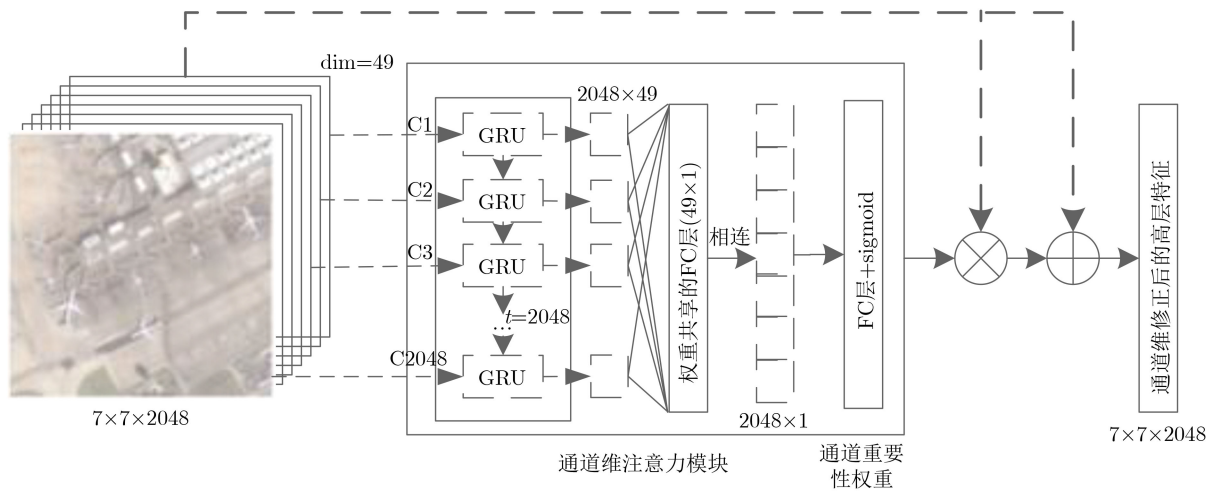


图2 通道维注意力模块网络结构图

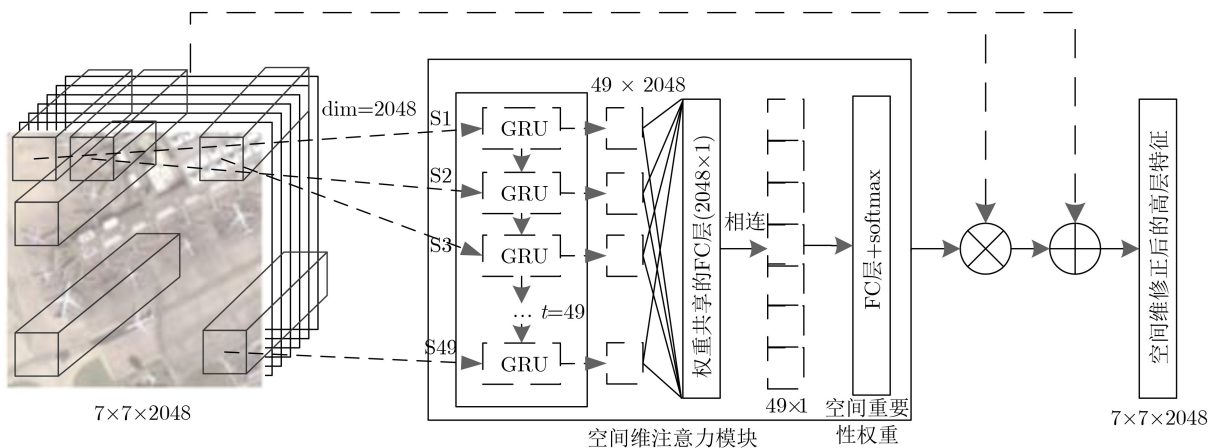


图3 空间维注意力模块网络结构图

后的高层特征 F_4 。整个过程的计算如式(3)和式(4)所示

$$W_{S_attention} = \text{softmax} \cdot (\text{relu}(\text{Concatenate}(F_3 \times W_3)) \times W_4) \quad (3)$$

其中, 两个全连接层的权重 $W_3 \in d^{2048 \times 1}$, $W_4 \in d^{49 \times 49}$, $W_{S_attention} \in d^{49}$

$$F_4 = F_3 + F_3 \cdot W_{S_attention} \quad (4)$$

其中, $W_{S_attention}$ 经调整尺寸(resize)为 $7 \times 7 \times 1$, \cdot 代表按元素相乘。

3 实验验证与结果分析

3.1 数据集

为验证所提双重注意力模块的有效性, 选取两个代表性数据集AID (Aerial Image Dataset)^[16]和NWPU-RESISC45 dataset (NWPU45)^[17]进行实验验证。数据集AID和NWPU45是遥感图像场景分类领域数据量最大、使用最广泛的两个数据集之一。AID中的数据来自谷歌地球, 共包含10000张大小为 600×600 像素的航空场景图像, 空间分辨率为1~8 m, 数据集共被划分为30个类别, 每个类别包含220~420张图像不等。数据集NWPU45也是收集自谷歌地球, 覆盖率100多个国家的遥感图像, 共包含31500张图像和45个场景类别, 每类各包含700张大小为 256×256 像素的图像, 分辨率为0.2~30 m。

为便于与相关方法进行比较, 与相关文献^[11-14, 16, 17]中的数据设置保持一致, 随机从两个数据集中挑选部分数据进行训练和测试, 数据集AID分别挑选20%和50%进行训练, 剩余的作为测试集; 数据集NWPU45中随机挑选10%和20%的数据作为训练集, 剩余的90%和80%作为测试集。

3.2 实验设置

在实验中, 整个网络采用随机梯度下降(Stochastic Gradient Descent, SGD)优化器进行训练, 数据批次大小(batch size)为64, 数据集训练迭代30个循环(epochs), 学习率设置为 $1e-6$ 。

对于高层特征的获取, 本实验主要采用VGG16和ResNet50作为基准网络, 与所提的双重注意力模块相结合; 在测试阶段, 以整个网络结构倒数第2个全连接层作为遥感图像最终的特征表示; 根据相关文献^[11, 14]中的验证, 利用支持向量机作为测试阶段分类器的性能要优于训练阶段使用softmax的分类性能, 因此, 采用线性支持向量机作为测试阶段的分类器。

此外, 算法的评价指标采用总体分类准确率

(Overall Accuracy, OA)和混淆矩阵(Confusion Matrix, CM)两个综合性指标。OA的定义为测试集中分类正确的样本数占总测试集的百分比, 计算方法见式(5)

$$OA = \frac{T}{N} \times 100\% \quad (5)$$

其中, T 为测试集中分类正确的样本数, N 为测试集的样本总数。本实验中对每次测试结果重复10次, 以10次的平均准确率和标准差作为最终的实验结果。

CM是分类任务中能更直观地表现算法性能表现的表达方式, 其对角线元素表示各类别的分类准确率, 其余元素 a_{mn} 代表第 m 类被误分为第 n 类所占的比例, 通过CM能更好地看出相近、混淆类别间的分类情况。

3.3 实验结果与分析

3.3.1 模型简化测试

为了验证双重注意力模块中各组成部分的作用, 通过模型简化测试的方法对两种注意力模块的有效性进行实验验证, 分别用CA(Channel Attention module)和SA(Spatial Attention module)代表通道注意力模块和空间注意力模块, 将两类注意力模块分别与基准网络VGG16^[18]和ResNet50相连, 在两个数据集上的实验结果对比如表1所示。

从实验结果中可以看到两个注意力模块的有效性, 具体分析如下: 首先, 在两个基准网络VGG16和ResNet50上, 两种注意力模块都能提升深度卷积网络所提取特征的特征能力, 在两个数据集上的整体分类准确率都取得了较明显的提升, 但空间注意力模块的性能提升要更优于通道注意力模块, 本文认为这主要是由于受遥感图像在成像范围上广阔性的影响, 通过空间注意力模块对局部显著区域进行重点的关注更能提高对遥感图像语义内容的理解; 然后, 通过表1可以发现两种注意力模块的结合能更进一步提高特征表示的鉴别性能力, 使得基准网络的分类准确率大幅提升, 尤其在数据量更大、更具挑战性的NWPU45数据集上, 所提方法将原基准网络的准确率提升到了99%, 证实了两种注意力模块针对的侧重点不同, 可以相互补充、相互促进; 最后, 可以发现, 两个基准网络的性能表现还存在着明显差距, ResNet50的特征表征能力明显超出VGG16, 但所提的两个注意力模块都能较好适用于两个基准网络, 证明了所提注意力模块的适用性是可以与任意卷积神经网络相结合。

3.3.2 与其他方法的对比及分析

为了进一步验证所提方法的有效性, 在数据集

AID和NWPU45上与近几年的相关基准方法进行比较,实验对比结果如表2和表3所示。

数据集AID下的对比结果如表2所示,本文所提方法在20%和50%两种训练比例下都实现了最佳性能表现,准确率超过了现有基准方法,分别达到了91%和95%左右的整体分类准确率。其中,2019年提出的方法ResNet_LGFFE将两种训练比例下原有最佳分类表现分别提高了2%和1.3%左右,在此基础上,本文方法又将整体准确率分别提高了0.5%和0.8%,达到了最佳实现(SOTA)。数据集NWPU45下的实验对比结果如表3所示,本文所提方法在两种训练比例下的准确率达到98.55%和99.07%,接近于100%,较大提高了该数据集下的最佳实现。由此,在以上两个数据集下的表现可以看出所提方法的有效性。此外,将所提的双注意力模块与ResNet50相结合在数据集AID下所取得的最佳混淆矩阵展示如图4所示。

本文所提方法虽然能较明显地提升特征表示的鉴别性能力,在两个具有挑战性的数据集上取得了较好的总体分类准确率,但是对局部特征和局部区

域的过分关注也反而导致对相似类别图像的混淆。如图4的混淆矩阵所示,所提方法对各个类别的分类准确率都比较高,尤其在50%的训练比例下,各个类别的准确率都接近1,但错误最明显的是将真值为类别22(度假村, resort)的图像大量误判为类别16(公园, park)。以图5中两个类别的代表性图像为例,可以发现两类图像在部分局部特征上存在较高的相似性,如游泳池、树木和建筑等。而所提方法是在原有特征表示的基础上对局部区域和局部特征的进一步关注,这就导致了所提方法在该数据集下训练学习后,对度假村和公园的局部相似特征过分关注,导致了对两个类别信息的混淆。

4 结束语

针对遥感图像场景分类存在的类内差异性大、类间相似性高导致的分类准确率不高的问题,本文提出了一种双重注意力模块的特征表示方法,在CNNs提取的高层特征基础上,分别从通道维和空间维各设计了一个注意力模块,利用循环神经网络的上下文信息提取能力,捕捉不同特征和不同区域

表1 数据集AID和NWPU45下的模型简化测试OA(%)结果对比表

方法	AID		NWPU45	
	20%	50%	10%	20%
VGG16	86.59±0.29	89.64±0.30	87.15±0.45	90.36±0.18
VGG16+CA	87.73±0.19	89.98±0.25	88.54±0.39	90.89±0.23
VGG16+SA	89.36±0.21	94.06±0.19	93.23±0.21	95.05±0.18
VGG16+CA+SA	89.87±0.30	94.58±0.23	97.89±0.12	98.82±0.20
ResNet50	86.48±0.49	89.22±0.34	89.88±0.26	92.35±0.19
ResNet50+CA	88.23±0.34	91.45±0.30	91.52±0.19	93.48±0.21
ResNet50+SA	90.83±0.55	94.46±0.48	97.56±0.08	98.79±0.04
ResNet50+CA+SA	91.34±0.38	95.22±0.36	98.55±0.11	99.07±0.23

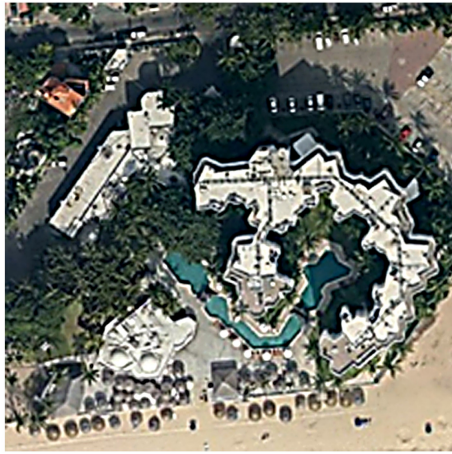
表2 数据集AID下所提方法与其他基准方法的OA(%)结果对比表

方法	年份	AID	
		20%	50%
VGG16 ^[16]	2017	86.59±0.29	89.64±0.30
CaffeNet ^[16]	2017	86.86±0.47	89.53±0.31
GoogLeNet ^[16]	2017	83.44±0.40	86.39±0.55
Fusion-by-add ^[19]	2017		91.87±0.36
MCNN ^[11]	2018		91.80±0.22
ARCNet ^[12]	2019	88.75±0.40	93.10±0.55
Finetune_ResNet50 ^[14]	2019	86.48±0.49	89.22±0.34
ResNet_LGFFE ^[14]	2019	90.83±0.55	94.46±0.48
VGG16+CA+SA	本文方法	89.87±0.30	94.58±0.23
ResNet50+CA+SA	本文方法	91.34±0.38	95.22±0.36

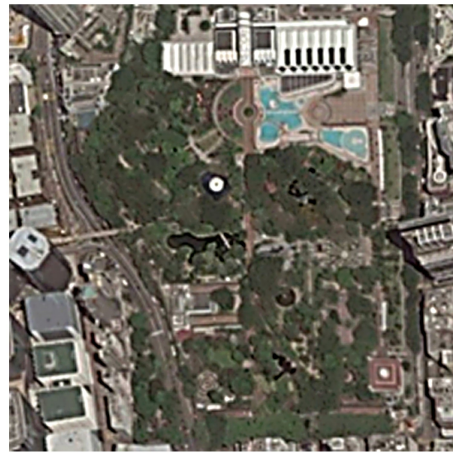
表3 数据集NWPU45下所提方法与其他基准方法的

OA(%)结果对比表

方法	年份	NWPU45	
		10%	20%
AlexNet ^[17]	2017	81.22±0.19	85.16±0.18
VGG_16 ^[17]	2017	87.15±0.45	90.36±0.18
GoogleNet ^[17]	2017	86.02±0.18	86.02±0.18
D_CNN ^[11]	2018	89.22±0.5	91.89±0.22
LGFFE ^[20]	2018	93.61±0.1	96.37±0.05
文献[21]	2019	91.73±0.21	93.47±0.30
Finetune_ResNet50 ^[14]	2019	89.88±0.26	92.35±0.19
ResNet_LGFFE ^[14]	2019	97.56±0.08	98.79±0.04
VGG16+CA+SA	本文方法	97.89±0.12	98.82±0.20
ResNet50+CA+SA	本文方法	98.55±0.11	99.07±0.23



(a) 度假村类别22



(b) 公园类别16

图5 数据集AID在所提方法中的误判实例

间的重要性程度, 以实现对重点区域和显著特征的重点关注, 对背景区域和非显著特征进行忽略, 以此提高特征表示的鉴别性能力。在两个公开数据集上, 本文方法均超过了现有基准方法, 验证了所提方法的有效性。但是, 双注意力模块的引入不可避免地增加了一定的训练时间, 对训练数据量有一定的要求, 且通过实验结果可以发现对局部特征相似的部分类别容易出现混淆的现象, 下一步重点从自监督学习的角度出发, 减少模型对训练数据量的依赖。

参考文献

- [1] CHI Mingmin, PLAZA A, BENEDIKTSSON J A, *et al.* Big data for remote sensing: Challenges and opportunities[J]. *Proceedings of the IEEE*, 2016, 104(11): 2207–2219. doi: [10.1109/JPROC.2016.2598228](https://doi.org/10.1109/JPROC.2016.2598228).
- [2] ZHANG Liangpei, ZHANG Lefei, and DU Bo. Deep learning for remote sensing data: A technical tutorial on the state of the art[J]. *IEEE Geoscience and Remote Sensing Magazine*, 2016, 6(4): 22–40. doi: [10.1109/MGRS.2016.2540798](https://doi.org/10.1109/MGRS.2016.2540798).
- [3] CHENG Gong, MA Chengcheng, ZHOU Peicheng, *et al.* Scene classification of high resolution remote sensing images using convolutional neural networks[C]. 2016 IEEE International Geoscience and Remote Sensing Symposium, Beijing, China, 2016: 767–770. doi: [10.1109/IGARSS.2016.7729193](https://doi.org/10.1109/IGARSS.2016.7729193).
- [4] SZEGEDY C, LIU Wei, JIA Yangqing, *et al.* Going deeper with convolutions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 1–9. doi: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
- [5] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [6] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[C]. The 25th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2012: 1097–1105.
- [7] HU Fan, XIA Guisong, YANG Wen, *et al.* Recent advances and opportunities in scene classification of aerial images with deep models[C]. 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018: 4371–4374. doi: [10.1109/IGARSS.2018.8518336](https://doi.org/10.1109/IGARSS.2018.8518336).
- [8] CHENG Gong, YANG Ceyuan, YAO Xiwen, *et al.* When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(5): 2811–2821. doi: [10.1109/TGRS.2017.2783902](https://doi.org/10.1109/TGRS.2017.2783902).
- [9] LI Peng, REN Peng, ZHANG Xiaoyu, *et al.* Region-wise deep feature representation for remote sensing images[J]. *Remote Sensing*, 2018, 10(6): 871. doi: [10.3390/rs10060871](https://doi.org/10.3390/rs10060871).
- [10] LIU Yanfei, ZHONG Yanfei, and QIN Qianqing. Scene classification based on multiscale convolutional neural network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(12): 7109–7121. doi: [10.1109/TGRS.2018.2848473](https://doi.org/10.1109/TGRS.2018.2848473).
- [11] YUAN Yuan, FANG Jie, LU Xiaoqiang, *et al.* Remote sensing image scene classification using rearranged local features[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(3): 1779–1792. doi: [10.1109/TGRS.2018.2869101](https://doi.org/10.1109/TGRS.2018.2869101).
- [12] WANG Qi, LIU Shaoteng, CHANUSSOT J, *et al.* Scene classification with recurrent attention of VHR remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(2): 1155–1167. doi: [10.1109/TGRS.2018.2864987](https://doi.org/10.1109/TGRS.2018.2864987).

- [13] XIONG Wei, LV Yafei, CUI Yaqi, *et al.* A discriminative feature learning approach for remote sensing image retrieval[J]. *Remote Sensing*, 2019, 11(3): 281. doi: [10.3390/rs11030281](https://doi.org/10.3390/rs11030281).
- [14] LV Yafei, ZHANG Xiaohan, XIONG Wei, *et al.* An end-to-end local-global-fusion feature extraction network for remote sensing image scene classification[J]. *Remote Sensing*, 2019, 11(24): 3006. doi: [10.3390/rs11243006](https://doi.org/10.3390/rs11243006).
- [15] CHO K, VAN MERRIËNBOER B, GULCEHRE C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]. 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 2014: 1724–1734. doi: [10.3115/v1/D14-1179](https://doi.org/10.3115/v1/D14-1179).
- [16] XIA Guisong, HU Jingwen, HU Fan, *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. doi: [10.1109/TGRS.2017.2685945](https://doi.org/10.1109/TGRS.2017.2685945).
- [17] CHENG Gong, HAN Junwei, and LU Xiaoqiang. Remote sensing image scene classification: Benchmark and state of the art[J]. *Proceedings of the IEEE*, 2017, 105(10): 1865–1883. doi: [10.1109/JPROC.2017.2675998](https://doi.org/10.1109/JPROC.2017.2675998).
- [18] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]. The 3rd International Conference on Learning Representations, San Diego, USA, 2015: 7–12.
- [19] CHAIB S, LIU Huan, GU Yanfeng, *et al.* Deep feature fusion for VHR remote sensing scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(8): 4775–4784. doi: [10.1109/TGRS.2017.2700322](https://doi.org/10.1109/TGRS.2017.2700322).
- [20] ZHU Qiqi, ZHONG Yanfei, LIU Yanfei, *et al.* A deep-local-global feature fusion framework for high spatial resolution imagery scene classification[J]. *Remote Sensing*, 2018, 10(4): 568. doi: [10.3390/rs10040568](https://doi.org/10.3390/rs10040568).
- [21] 叶利华, 王磊, 张文文, 等. 高分辨率光学遥感场景分类的深度度量学习方法[J]. *测绘学报*, 2019, 48(6): 698–707. doi: [10.11947/j.AGCS.2019.20180434](https://doi.org/10.11947/j.AGCS.2019.20180434).
- YE Lihua, WANG Lei, ZHANG Wenwen, *et al.* Deep metric learning method for high resolution remote sensing image scene classification[J]. *Acta Geodaetica et Cartographica Sinica*, 2019, 48(6): 698–707. doi: [10.11947/j.AGCS.2019.20180434](https://doi.org/10.11947/j.AGCS.2019.20180434).
- 徐从安：男，1987年生，博士，研究方向为遥感图像智能处理、多目标跟踪。
- 吕亚飞：男，1992年生，博士，研究方向为遥感图像智能处理、跨模态检索。
- 张筱晗：女，1992年生，博士，研究方向为遥感图像智能处理、目标检测。
- 刘瑜：男，1986年生，副教授，研究方向为智能数据处理。
- 崔晨浩：男，1991年生，研究方向为雷达数据处理。
- 顾祥岐：男，1995年生，博士生，研究方向为雷达数据处理、信息融合。

责任编辑：余蓉