第 58 卷 第 14 期/2021 年 7 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于 RCF 网络的遥感图像场景分类研究

朱淑鑫[1]，周子俊[1]，顾兴健[1]，任守纲[1*]，徐焕良[1,2]

[1]南京农业大学人工智能学院，江苏 南京 210095；

[2]国家信息农业工程技术中心，江苏 南京 210095

摘要 为了提高 ResNet50 网络对遥感场景图像中目标特征的提取能力和场景分类的可解释性，提出一种基于 ResNet50-CBAM-FCAM(RCF)网络的遥感图像场景分类方法。该方法在 ResNet50 网络中增加卷积注意力模块和全卷积类激活映射分支，利用注意力机制将分支特征分别与提取的通道注意力特征和空间注意力特征融合，生成各类场景的类激活映射图。实验结果证明，所提方法在数据集 AID 上的总体分类准确率达到 96%以上，在实验数据集 NWPU-RESISC45 上的总体分类准确率达到 93%以上，且类激活映射图可视化结果可以准确聚焦遥感场景图像的目标对象。

关键词 大气光学；遥感图像；场景分类；深度学习；网络解释性；注意力机制

中图分类号 TP753 **文献标志码** A **doi**：10.3788/LOP202158.1401001

Scene Classification of Remote Sensing Images Based on RCF Network

Zhu Shuxin[1], Zhou Zijun[1], Gu Xingjian[1], Ren Shougang[1*], Xu Huanliang[1,2]

[1] *College of Artificial Intelligence, Nanjing Agricultural University, Nanjing, Jiangsu* 210095, *China*;

[2] *National Engineering and Technology Center for Information Agriculture, Nanjing, Jiangsu* 210095, *China*

Abstract To improve the ability of ResNet50 to extract target object features of remote sensing scene images and interpretability of scene classification, a Resnet50-CBAM-FCAM(RCF) network-based method of remote sensing image scene classification is proposed in this paper. This method increases the convolution attention module and full convolution-class activation mapping branch in the ResNet50 network. With the help of an attention mechanism, the branch features are fused with the extracted channel attention features and spatial attention features, respectively, and the class activation maps of various scenes are generated. The experimental results show that the overall classification accuracy of the proposed method in AID and NWPU-REISC45 datasets is more than 96% and 93%, respectively, and the visual results of the class activation maps can focus the target objects of remote sensing scene image accurately.

Key words atmospheric optics; remote sensing image; scene classification; deep learning; network interpretation; attention mechanism

OCIS codes 010.0280; 100.3008; 100.4996

1 引 言

随着遥感技术的发展，遥感图像分辨率日益提高，所携带的空间特征和语义信息不断增加。然而高分辨遥感图像场景复杂，在精度不断提高的同时也造成不同类别场景相似度增加、同一类别场景差异性不够显著等问题。因此，在遥感图像场景分类研究中，有效地提取图像特征、显著区分遥感图像场景信息和背景信息显得尤为重要[1]。

进行传统遥感图像场景分类时，通常采用“基

收稿日期：2020-09-15；**修回日期**：2020-10-29；**录用日期**：2020-11-14

基金项目：国家自然科学基金青年项目(61806097)、国家级大学生创新创业训练计划(202010307063Z)

通信作者：*rensg@njau.edu.cn

于底层特征”[2]和“基于中层特征”[3]两类特征提取方法。前者利用工程技能和领域知识通过人工设计各种特征，如颜色、纹理、形状、空间和光谱信息等。后者针对底层特征进行编码，基于统计分布建立底层特征与语义的联系。但传统方法的特征描述能力有限，模型的分类效果和泛化能力仍需进一步优化[4-8]。

相比传统方法，以卷积神经网络(CNN)为代表的深度学习方法在图像分类、目标识别、语义分割等领域表现突出[9-10]。随着大规模遥感图像场景数据的出现，深度学习方法被广泛地应用于遥感图像场景分类中[11-13]。Krizhevsky 等[14]凭借 AlexNet 网络，在 ImageNet 图像分类竞赛中获得冠军。Szegedy 等[15]提出 GoogleNet 模型，加深网络的同时加宽网络，但增加了模型训练时间和占用内存。He 等[16]提出 ResNet 网络，解决了因网络层数过深而难以训练的问题。Hu 等[17]提出 Squeeze and Excitation(SE)模块，重新校准通道的特征响应，改进图像通道间的相互依赖关系。Woo 等[18]提出轻量型 Convolutional Block Attention Module(CBAM)，关注通道和空间维度的特征信息，增加了特征的判别力。Zhou 等[19]提出类激活映射(CAM)机制，利用卷积层的响应可视化类激活映射图，但修改了原始网络结构而影响图像分类性能。Selvaraju 等[20]提出加权类激活映射(Grad-CAM)机制，依据梯度和卷积层的响应可视化类激活映射图，保持原始网络结构但需要反向传播获得梯度。张桐等[21]提出多分支卷积神经网络，该网络融合遥感图像高、中、低三个层次的特征信息，解决单一模型特征提取不足的问题。Wang 等[22-23]通过加权卷积神经网络的高层特征，增强遥感图像目标对象的特征表示能力。汪鹏等[24]在残差网络中引入跳跃连接和协方差池化模块，提取遥感图像不同层次的信息。然而，以上方法在特征提取时未能显著区分遥感场景图像的目标对象，模型的关键特征提取能力还需进一步提高。

为增强目标对象特征表示能力，显著区分遥感场景图像中的对象信息和背景信息，本文提出 ResNet50-CBAM-FCAM(RCF)网络。在 ResNet50 网络中引入改进的卷积注意力模块(CBAM)和全卷积类激活映射分支(FCAM)，以增强 ResNet50 网络特征提取能力及场景分类的可解释性，并在数据集 AID 和 NWPU-RESISC45 上进行验证。

2　RCF 网络结构

用于遥感图像场景分类的 RCF 网络结构如图 1 所示。该网络选用 ResNet50 为主干网络，主要步骤为：利用预训练的 ResNet50 网络提取输入图像的特征信息；利用改进的 CBAM 提取遥感图像通道和空间维度深层次特征，分别得到通道注意力特征图和空间注意力特征图；利用基于响应的 FCAM 生成类激活映射图和概率分数；利用注意力机制，将类激活映射图分别与通道注意力特征图和空间注意力特征图融合。

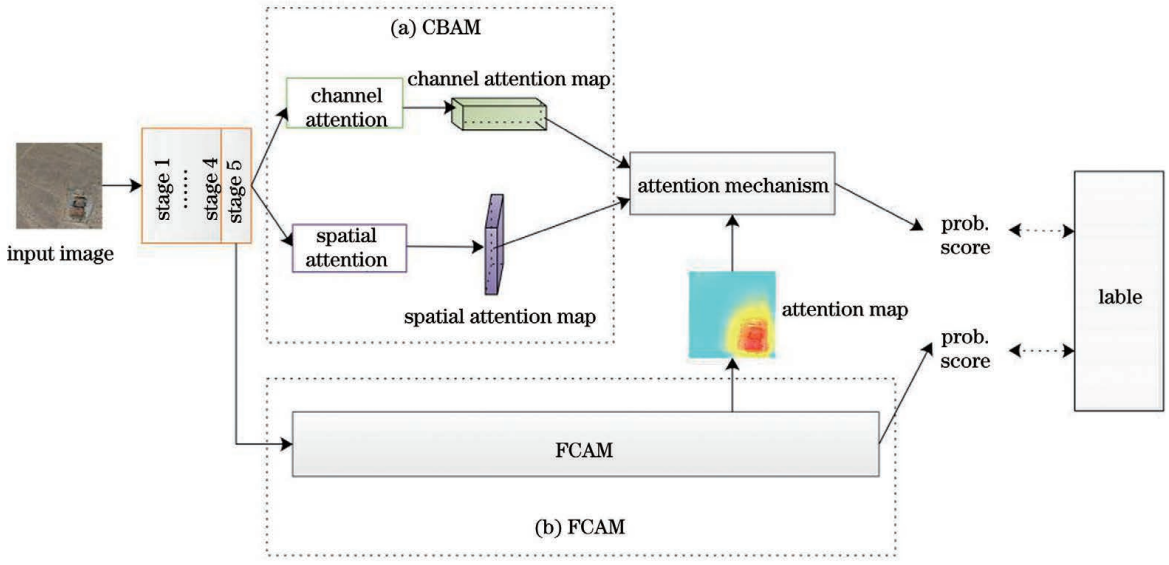

图 1　RCF 结构

Fig. 1　Architecture of RCF

2.1　ResNet50 网络

ResNet50 网络有 5 个特征提取阶段。第 1 阶段仅包含一层卷积层，提取初步特征，其余 4 个阶段分别由含有 3、4、6、3 个残差块的卷积层组成。每个残差块由卷积核大小分别为 1×1、3×3、1×1 的三个卷积层串联在一起。

$$y_i = F(x_i, \omega_i) + x_i, \tag{1}$$

式中：x_i 为残差块的输入；ω_i 为 x_i 经过残差块后的权重；$F(\cdot)$为学习的残差映射；y_i 为经过残差块后的输出。与普通卷积层不同的是，残差块通过跳跃连接将输入信息直接传到输出，减少了特征提取过程中损失的信息量，保护了信息的完整性。

2.2　改进的 CBAM

CBAM 是一种轻量、通用的卷积注意力模块。在 ResNet50 网络中引入 CBAM，可基于注意力从通道和空间两个维度对目标特征进行自适应细化。但 ResNet50 网络中低层特征的判别性较差，若在每个阶段增加 CBAM，不仅增加了参数量，还无法强调目标对象的关键特征，会对下一步的特征提取产生误导。因此，在 ResNet50 第五个阶段提取的高层特征后引入 CBAM，从通道和空间两个维度突出目标对象的高层特征。

输入遥感图像，通过 ResNet50 网络提取高层特征。再通过最大池化与平均池化对阶段 5 输出的高层特征图空间维度上的特征进行压缩，生成两种不同的空间上下文描述符。两个空间背景描述符共享一个多层感知器，得到有效权重，以逐元素求和的方式合并输出特征向量。使用 Sigmoid 函数进行归一化，以产生通道注意力特征图$C(x_i)$，关注遥感图像中有意义的目标对象，如图 2(a)所示。通道注意力特征图计算公式为

$$C(x_i) = \sigma\{\mathrm{MLP}\{\mathrm{AvgPool}[F'(x_i)]\} + \mathrm{MLP}\{\mathrm{MaxPool}[F'(x_i)]\}\} = \sigma\{W_1[W_0(F_{\mathrm{avg}})] + W_1[W_0(F_{\mathrm{max}})]\}, \tag{2}$$

式中：x_i 为输入图像样本；$F'(x_i)$为阶段 5 输出的特征图；σ 为 Sigmoid 函数；W_0 为经过 ReLU 激活函数的权重；W_1 为经过共享网络的权重；F_{avg} 为空间上平均池化特征；F_{max} 为空间上最大池化特征。

与此同时，在阶段 5 输出的高层特征图中，首先在通道维度上使用最大池化和平均池化对通道维度上的映射进行压缩，分别得到两个通道上下文描述符；然后连接两个特征映射，生成有效的特征描述符，将两个通道特征图串联起来；最后使用 7×7 的卷积操作生成空间注意力特征图 $S(x_i)$，关注目标对象在遥感图像上的位置信息，如图 2(b)所示。空间注意力特征图计算公式为

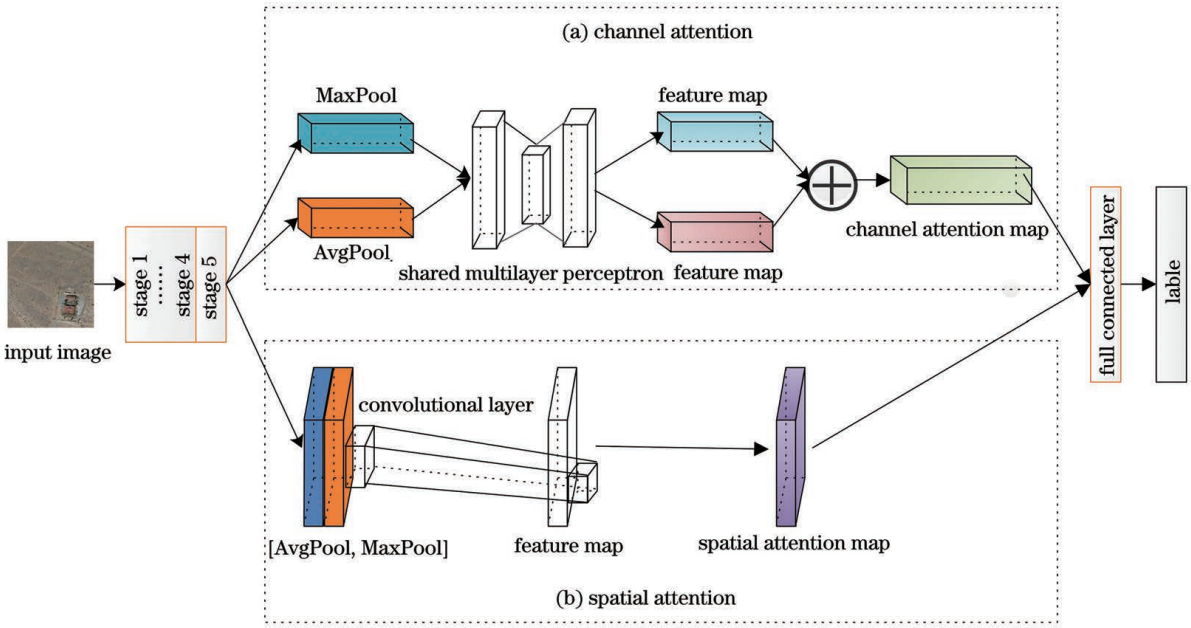

图 2　ResNet50-CBAM 结构

Fig. 2　Architecture of ResNet50-CBAM

$$S(x_i)=\sigma\{f^{7\times7}\{\mathrm{AvgPool}[F'(x_i)];\mathrm{MaxPool}[F'(x_i)]\}\}=\sigma\{f^{7\times7}[F_{\mathrm{avg}};F_{\mathrm{max}}]\},\tag{3}$$

式中：$f^{7\times7}$ 为滤波器大小为 7×7 的卷积运算。

2.3 基于响应的 FCAM

文献[25]已验证特征重要性度量与注意力权重相关性很弱，目标对象的注意力分布对模型输出没有影响，注意力模块不具有视觉解释性。因此，仅在 ResNet50 网络中引入基于注意力的 CBAM 无法解释遥感图像场景分类的依据。

CAM 是一种基于响应的视觉解释类激活映射网络结构，通过卷积层的前馈响应值和全连接层的权重可视化类激活映射图。它在无像素级对象标注的情况下，利用特征图权重叠加实现目标对象区域的定位，为图像分类任务提供视觉解释依据。但 CAM 在使用时必须修改原始网络重新训练，无法在训练过程中生成类激活映射图。本文改进 CAM，提出一种基于响应的视觉解释 FCAM。

FCAM 结构如图 3 所示，接收主干分支阶段 5 输出的特征图，通过归一化、卷积层、ReLU 激活函数提取深层次特征，接着基于 CAM 构建输出层。输出层由卷积层和全局平均池化(GAP)[26]组成，取代了分类任务中的全连接层，防止过拟合并减少了大量参数。由卷积层降低特征维度，通过全局平均池化操作得到每个特征图的均值，通过加权和得到输出，使用 Softmax 函数输出类别概率。

$$F^k=\sum_{x,y}f_k(x,y),\tag{4}$$

式中：$f_k(x,y)$ 为卷积层在空间坐标 (x,y) 中单元 k 的激活值。对于每个类 c，Softmax 层的输入为

$$S_c=\sum_k w_k^c F_k,\tag{5}$$

式中：w_k^c 为单元 k 对应的类 c 的权重，即 F_k 对类 c 的重要性。最后类 c 的 Softmax 输出概率分数为

$$P_c=\frac{\exp(S_c)}{\sum_c\exp(S_c)},\tag{6}$$

式中：将 Softmax 的偏差项设置为 0，由于偏差项对图像分类几乎没有影响，所以忽略偏差项。把(5)式代入(6)式中，得

$$S_c=\sum_k w_k^c\sum_{x,y}f_k(x,y)=\sum_{x,y}\sum_k w_k^c f_k(x,y)。\tag{7}$$

每个类别的类激活映射为

$$M_c(x,y)=\sum_k w_k^c f_k(x,y)。\tag{8}$$

将(8)式代入(7)式，得

$$S_c=\sum_{x,y}M_c(x,y)。\tag{9}$$

通过线性加权不同空间区域，类激活映射可视化每个被激活的单元，按照原始图像大小输出类激活映射图，清楚地显示特定类最相关的区域。

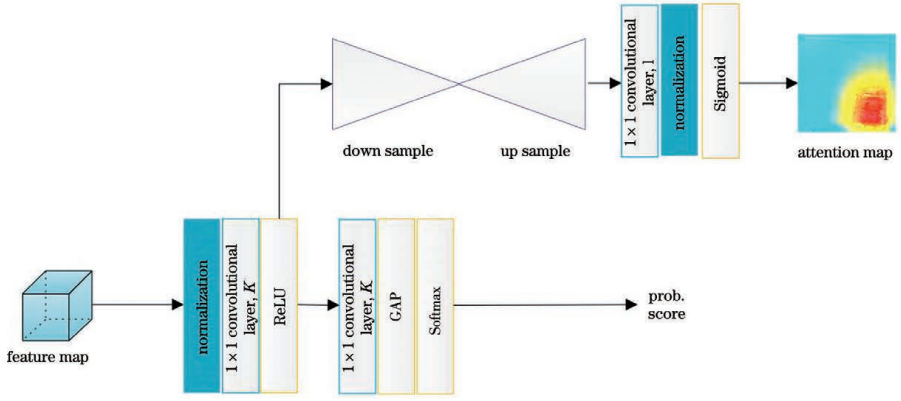

图 3　FCAM 结构

Fig. 3　Architecture of FCAM

与此同时，借鉴语义分割任务中全卷积网络(FCN)对特征图进行编码-解码的思想[27]，在 ReLU 层输出的特征图后增加自底向上、自上而下的全卷积结构。通过自底向上的架构扩展全局信息，以指导每个位置输入特征图，有意识地聚焦于分类对象的主要特征；通过自上而下的下采样结构，保证输入特征图与输出特征图大小相同。即先执行两次最大池化操作，以快速增加感受野，使特征图达到最小分辨率，再执行两次上采样，使输出的特征图与输入的特征图具有相同的大小。在全卷积结构之后，用 1×1 的卷积层对这些特征图进行卷积，以聚合特征图，得到 $M(x_i)$。

$$M(x_i)=\sigma\{f^{1\times1}\{\mathrm{UpSample}\{\mathrm{UpSample}\{\mathrm{MaxPool}\{\mathrm{MaxPool}[F''(x_i)]\}\}\}\}\}=\sigma[f^{1\times1}(F_{\mathrm{Up}})],\quad(10)$$

式中：$F''(x_i)$为注意分支 ReLU 输出的特征图；MaxPool(•)为池化操作；UpSample(•)为上采样操作；F_{Up} 为全卷积结构输出的特征。

2.4　特征融合

在 ResNet50 阶段 5 后增加改进的 CBAM，分别得到通道注意力特征图和空间注意力特征图，增加改进的 FCAM 得到基于响应的类激活映射图。为了更加增强目标对象在通道和空间两个维度中的特征，利用注意力机制将 FCAM 输出的类激活映射图应用于通道注意力特征图，得到加权的通道注意力特征图 $H_1(x_i)$。

$$H_1(x_i)=C(x_i)*[M(x_i)+1]。\quad(11)$$

同时利用注意力机制将 FCAM 输出的类激活映射图应用于空间注意力特征图，得到加权的空间注意力特征图$H_2(x_i)$。

$$H_2(x_i)=S(x_i)*[M(x_i)+1]。\quad(12)$$

在(12)式中，FCAM 输入注意图增加 1 目的是突出目标对象特征图的峰值，同时防止特征图的低值区域退化为零。以逐元素求和的方式融合加权的通道注意力特征图和空间注意力特征图，得到融合的特征图 $H(x_i)$。

$$H(x_i)=H_1(x_i)+H_2(x_i)。\quad(13)$$

3　实验与分析

3.1　数据集

实验数据集采用遥感图像场景分类领域中的两个大规模数据集 AID[28] 和 NWPU-RESISC45[29]。AID 数据集包含 30 个类别，每类样本有 220 到 420 张，共有 10000 张图像。该数据集具有 8～0.5 m 的不同分辨率，每个图像的大小为 600×600，代表图像如图 4 所示。NWPU-RESISC45 数据集包含 45 个类别，每类有 700 张图像，每张图像的大小为 256×256，共有 31500 张图像，且空间分辨率为每像素 30～0.2 m，代表图像如图 5 所示。

图 4　AID 数据集场景图

Fig. 4　Scene images of AID dataset

图 5　NWPU-RESISC45 数据集场景图

Fig. 5　Scene images of NWPU-RESISC45 dataset

3.2　实验设置

为了验证所提方法的有效性，与其他论文成果进行比较。参考文献[23]和[30]中数据集划分方法，分别在 AID 和 NWPU-RESISC45 数据集上设置两类实验。在 AID 数据集上，第一类实验，从每类场景图像中随机取 20%作为训练集，剩余 80%作为测试集；第二类实验，从每类场景图像中随机取 50%作为训练集，剩余 50%作为测试集。在 NWPU-RESISC45 数据集上，第一类实验，从每类场景图像中随机取 10%作为训练集，剩余 90%作为测试集；第二类实验，从每类场景图像中随机取 20%作为训练集，剩余 80%作为测试集。

为提高模型的泛化能力，在训练前对训练集进行随机缩放、裁剪等数据增强处理操作。利用 ImageNet1K 数据集上的 ResNet50 预训练模型初始化网络参数，并使用训练集对模型进行微调。利用微调后的网络对测试集进行分类预测，得到最终的分类结果。模型的批处理大小为 32，完全训练 400 次后停止训练，确保网络损失值基本平稳以得到模型的可靠性。在模型训练过程中使用 SGD 优化器，设置阶梯下降的学习率，以提高模型的训练效率和稳定性，初始学习率为 0.01，在 121 轮到 250 轮学习率衰减为 0.001，在 251 轮到 400 轮学习率衰减为 0.0001，如图 6 所示。在网络反向传播过程中，主干分支和注意分支均使用交叉熵损失函数，最终以两分支损失函数加和的方式优化网络参数。重复 5 次，随机划分数据集进行实验，得到平均分类准确率。

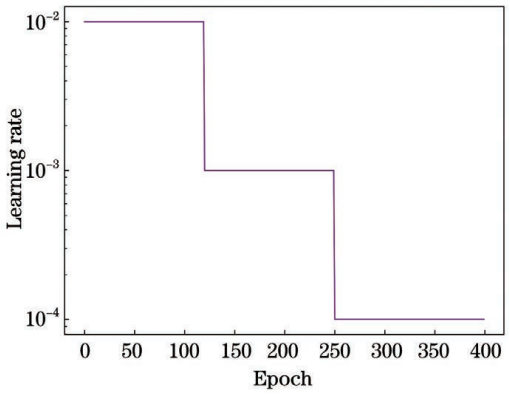

图 6　学习率随循环次数的变化

Fig. 6　Learning rate versus number of cycles

实验设置参数量(Params)、FLOPs、运行时间(Time)和分类准确率(Accuracy)4 个评价指标。其

中参数量决定模型文件的大小，影响模型预测时内存的占用量；FLOPs 用来衡量网络模型的复杂度，该值越大代表网络的计算复杂度越高，占用计算资源越高；运行时间包括平均每轮训练时间和测试时间；分类准确率是测试集分类正确的样本数与总样本数之比，衡量模型的分类效果。

3.3 实验结果及分析

为了验证 RCF 网络性能，在 AID 和 NWPU-RESISC45 数据集上分别对 ResNet50 网络模型（第一组）、ResNet50-CBAM 网络模型（第二组）、ResNet50-FCAM 网络模型（第三组）、RCF 网络模型（第四组）4 组模型的参数量、FLOPs、分类准确率进行对比。

实验结果如表 1 所示。对比第一组和第二组实验结果发现：所提方法在 ResNet50 网络引入 CBAM，参数量仅增加 0.52×10^6，模型复杂度增加 0.01GFLOPs；在文献[20]中，ResNet50-CBAM 的参数量增加 2.53×10^6，模型复杂度增加 0.01GFLOPs；ResNet50-CBAM 降低 2.01×10^6 参数量，模型复杂度几乎没有增加，且相比 ResNet50 在 AID 和 NWPU-RESISC45 数据集上第一类实验准确率分别提升 1.22 个百分点、1.09 个百分点，第二类实验准确率分别提升 0.64 个百分点、1.02 个百分点。对比第一组和第三组实验结果发现：在 ResNet50 网络中增加改进的 FCAM，参数量增加 1.01×10^6，模型复杂度仅增加 2.74GFLOPs，相比 ResNet50 在 AID 和 NWPU-RESISC45 数据集上第一类实验准确率分别提升 0.86 个百分点、1.67 个百分点，第二类实验准确率分别提升 0.69 个百分点、1.22 个百分点。对比第一组和第四组实验结果发现：在 ResNet50 网络引入 CBAM 并结合改进的 FCAM，参数量增加 1.53×10^6，模型复杂度增加 2.74GFLOPs，相比 ResNet50 在 AID 和 NWPU-RESISC45 数据集上第一类实验准确率分别提升 1.41 个百分点、2.06 个百分点，第二类实验准确率分别提升 1.13 个百分点、1.97 个百分点。实验证明，提出的 RCF 网络参数量与计算量较小，能够有效提高遥感图像场景分类准确率。

表 1 4 种网络在 AID 和 NWPU-RESISC45 数据集上的测试集的精度和参数量对比

Table 1 Comparison of test accuracy and number of parameters using four networks on test set in the AID and NWPU-RESISC45 datasets

Group	Model	Params /10^6	FLOPs/GFLOPs	Accuracy /%			
				AID		NWPU-RESISC45	
				Experiment I	Experiment II	Experiment I	Experiment II
1	ResNet50	25.56	4.10	92.35±0.33	95.04±0.21	88.44±0.25	91.30±0.50
2	ResNet50-CBAM	26.08	4.11	93.57±0.41	95.68±0.25	89.53±0.23	92.32±0.42
3	ResNet50-FCAM	26.57	6.84	93.21±0.21	95.73±0.15	90.11±0.20	92.52±0.43
4	RCF	27.09	6.84	93.76±0.32	96.17±0.47	90.50±0.13	93.27±0.11

为了进一步分析 ResNet50 网络分别引入改进的 CBAM 和 FCAM 模型的速度，对 4 组模型的平均每轮训练时间和测试时间进行了对比实验，结果如表 2 所示。发现引入改进的 CBAM 后，网络对模

表 2 4 种网络在 AID 和 NWPU-RESISC45 数据集上的平均每轮训练时间和平均每轮测试时间对比

Table 2 Comparison of average training time and average test time per epoch using four networks on AID and NWPU-RESISC45 datasets

Group	Model	AID				NWPU-RESISC45			
		Experiment I		Experiment II		Experiment I		Experiment II	
		Train time /s	Test time /s	Train time /s	Test time /s	Train time /s	Test time /s	Train time /s	Test time /s
1	ResNet50	30.65	17.33	43.23	11.34	78.07	57.26	87.61	49.23
2	ResNet50-CBAM	32.70	18.15	46.10	12.21	80.27	59.48	93.51	52.75
3	ResNet50-FCAM	48.34	28.97	65.12	18.96	126.81	97.70	144.35	87.78
4	RCF	51.62	31.15	68.76	19.86	134.61	104.35	153.27	92.95

型速度的影响较小，而引入 FCAM 后，网络在提高模型准确率的同时对模型速度有一定的影响。

为了更好地比较所提方法的优势，对 RCF 网络在 AID 数据集和 NWPU-RESISC45 数据集上的分类结果分别与传统方法（salM3LBP-CL[5]、BoCF[8]）、经典 CNN 方法（GoogleNet、VGG-VD16[31]）、改进的 CNN 方法（ARCNet、scale-attention network、Integrated CNN[30]）进行对比，结果如表 3、4 所示。

1）AID 数据集：从表 3 数据可知，在两类实验中，所提方法的分类准确率分别为 93.76%（第一类实验）和 96.17%（第二类实验），与 salM3LBP-CLM 相比，分类准确率分别提高 6.84 个百分点（第一类实验）和 6.41 个百分点（第二类实验）；与 GoogleNet、VGG-VD16 相比，第一类实验中分类准确率分别提高 2.39 个百分点、1.70 个百分点，第二类实验中分类准确率分别提高 2.18 个百分点、1.43 个百分点；与 ARCNet、scale-attention network 相比，第一类实验中分类准确率分别提高 5.01 个百分点、1.23 个百分点，第二类实验中分类准确率分别提高 3.07 个百分点、0.45 个百分点。

表 3 在 AID 数据集上的分类准确率

Table 3 Classification accuracy on AID dataset unit: %

Algorithm	Experiment I	Experiment II
salM3LBP-CLM	86.92±0.35	89.76±0.45
GoogleNet	91.37±0.35	93.99±0.37
VGG-VD16	92.06±0.29	94.74±0.51
ARCNet	88.75±0.40	93.10±0.55
scale-attention network	92.53±0.33	95.72±0.27
RCF	93.76±0.32	96.17±0.47

2）NWPU-RESISC45 数据集：从表 4 数据可知，在两类实验中，所提方法的分类准确率分别为 90.50%（第一类实验）和 93.27%（第二类实验），与 BoCF 相比，分类准确率分别提高 7.85 个百分点（第一类实验）和 8.95 个百分点（第二类实验）；与 GoogleNet、VGG-VD16 相比，第一类实验中分类准确率分别提高 3.76 个百分点、2.48 个百分点，第二类实验中分类准确率分别提高 3.60 个百分点、2.20 个百分点；与 Integrated CNN、scale-attention network 相比，第一类实验中分类准确率分别提高 2.03 个百分点、1.58 个百分点，第二类实验中分类准确率分别提高 0.74 个百分点、1.02 个百分点。

表 4 在 NWPU-RESISC45 数据集上的分类准确率

Table 4 Classification accuracy on NWPU-RESISC45 dataset unit: %

Algorithm	Experiment I	Experiment II
BoCF	82.65±0.31	84.32±0.17
GoogleNet	86.74±0.39	89.67±0.27
VGG-VD16	88.02±0.14	91.07±0.12
Integrated CNN	88.47±0.00	92.53±0.00
scale-attention network	88.92±0.29	92.25±0.18
RCF	90.50±0.13	93.27±0.11

以上对比结果说明，RCF 网络在遥感图像场景分类中具有一定优势，能够增强目标对象的特征表示，弱化复杂背景对场景分类的影响。

另外为了更好地分析所提方法在遥感图像场景分类中的有效性，图 7 分别展示在两个实验数据集上准确率和损失值随循环次数的变化。从图 7(a)～(d)得出，RCF 网络在 AID 数据集上通过约 120 次迭代后，分类精度趋于稳定，在两类实验中准确率分别保持在 93%、96%以上，同时训练损失函数也降低至 0.001 左右并保持稳定。从图 7(e)～(h)中得出，RCF 网络在 NWPU-RESISC45 数据集上通过约 120 次迭代后，分类精度趋于稳定，在两类实验中准确率分别保持在 90%、93%以上，同时训练损失函数也降低至 0.001 左右并保持稳定。RCF 网络在两个数据集中，测试数据集的准确率均能取得很好的效果，由此可见 RCF 网络在遥感图像场景分类中能够达到较好分类效果且算法稳定，能够快速收敛。

3.4 类激活映射图可视化结果与分析

为验证所提方法可视化类激活映射图的有效性，采用常用的 Grad-CAM 可视化方法和所提改进的可视化方法在 NWPU-RESISC45 数据集中进行实验，可视化每类的类激活映射图。图 8 分别展示分布于图像全局的“棒球场”、分布于图像局部的“空闲住宅”和目标小而多的“储油罐”这 3 种具有代表性的场景。

类激活映射图通过颜色变化程度，直观地反映图像关注的位置，即红色越深代表此位置越受关注，蓝色越深代表此位置越不受关注。从图 8(a)可以看出：Grad-CAM 方法只关注“棒球场”的部分边缘区域，将棒球场的内外野显示为深蓝色不受关注区域；RCF方法能够关注棒球场的内外野和边缘区

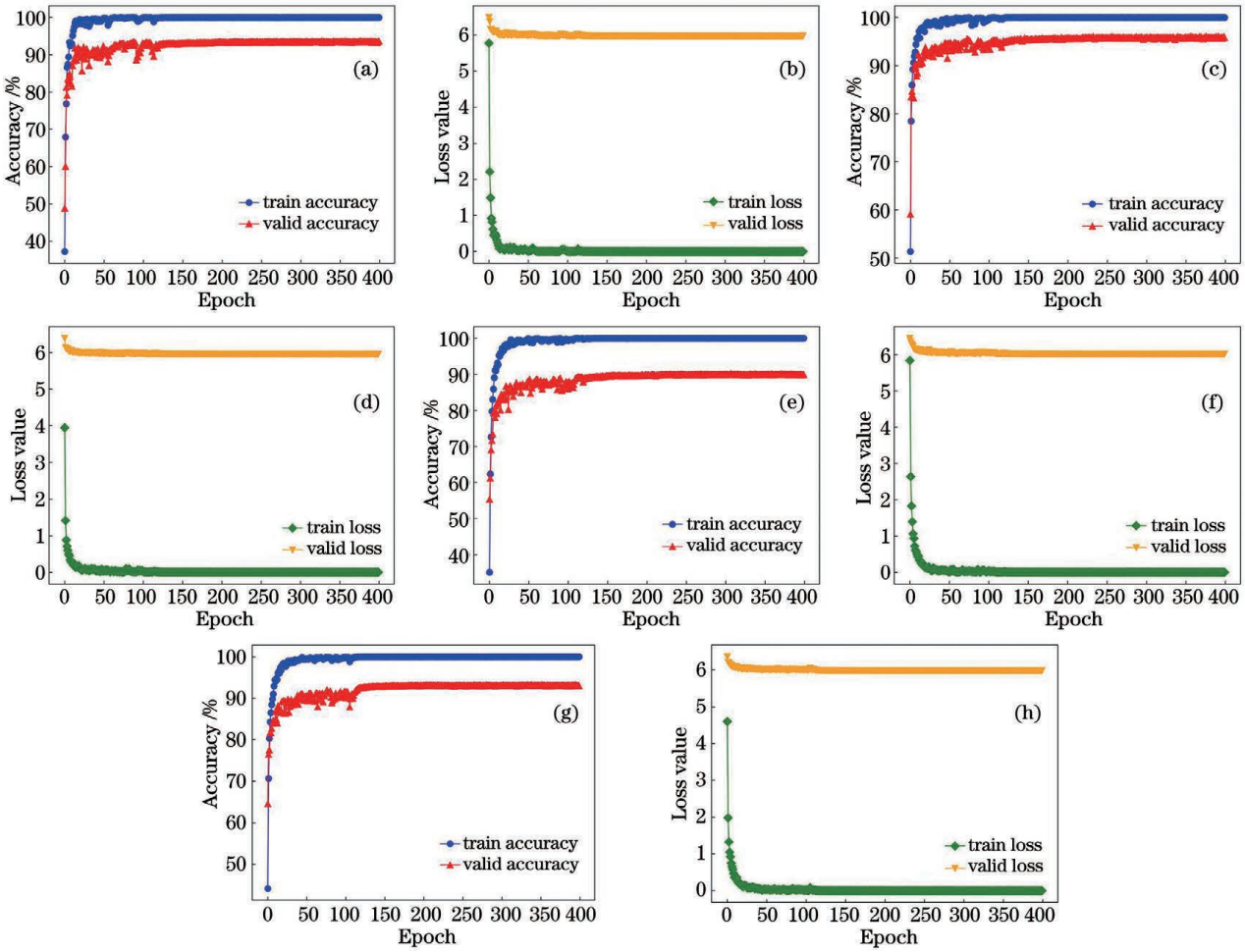

图 7　准确率和损失值随循环次数的变化结果。(a)(b)AID 数据集中训练比例为 20%;(c)(d)AID 数据集中训练比例为 50%;(e)(f)NWPU-RESISC45 数据集中训练比例为 10%;(g)(h)NWPU-RESISC45 数据集中训练比例为 20%

Fig. 7　Accuracy and loss value versus number of cycles. (a)(b) Training proportion is 20% in AID dataset; (c)(d) training proportion is 50% in AID dataset; (e)(f) training proportion is 10% in NWPU-RESISC45 dataset; (g)(h) training proportion is 20% in NWPU-RESISC45 dataset

域。从图 8(b)可以看出:Grad-CAM 方法关注区域偏离“空闲住宅”;RCF 方法能显示空闲住宅的全部区域。从图 8(c)可以看出:Grad-CAM 方法关注到部分储油罐;RCF 方法能够显示所有的大、小型储油罐区域。

从实验结果发现,RCF 方法应用于遥感图像不同场景时都能够较好地聚焦于场景中的目标区域,突显目标对象,分类解释性好。所提方法一方面增强遥感图像中目标对象的特征表示能力,提高了分类准确率;另一方面有利于我们了解每类遥感场景信息在网络中的流动,增加了场景分类的可解释性。

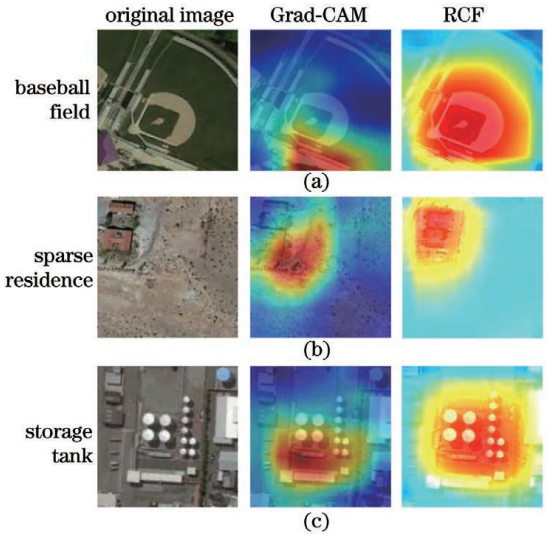

图 8　类激活映射图可视化结果对比

Fig. 8　Comparison of visualization results of class activation map

4　结　　论

提出一种基于 RCF 网络的遥感图像场景分类方法,在 ResNet50 网络中引入改进的 CBAM 和基

于响应的 FCAM，利用注意力机制自适应增强目标对象的特征，同时可视化类激活映射图。实验结果表明，相比传统方法和其他深度学习方法，所提方法能够有效地提升分类精度，较好地解释场景分类依据。

由于高分辨率遥感场景图像目标对象多而小，RCF 网络可视化的类激活映射图结果无法精细到每个单独小对象的区域，且增加 FCAM 后对训练时间和测试时间有一定的影响。因此精细可视化类激活映射图和提升网络整体运行速度将是未来的研究重点。

参 考 文 献

[1] Yu D H, Zhang B M, Zhao C, et al. Scene classification of remote sensing image using ensemble convolutional neural network[J]. Journal of Remote Sensing, 2020, 24(6): 717-727.
余东行, 张保明, 赵传, 等. 联合卷积神经网络与集成学习的遥感影像场景分类[J]. 遥感学报, 2020, 24(6): 717-727.

[2] Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope [J]. International Journal of Computer Vision, 2001, 42(3): 145-175.

[3] Sivic J, Zisserman A. Video Google: a text retrieval approach to object matching in videos[C]// Proceedings Ninth IEEE International Conference on Computer Vision, October 13-16, 2003, Nice, France. New York: IEEE Press, 2003: 1470-1477.

[4] Gong X, Wu L, Xie Z, et al. Classification method of high-resolution remote sensing scenes based on fusion of global and local deep features [J]. Acta Optica Sinica, 2019, 39(3): 0301002.
龚希, 吴亮, 谢忠, 等. 融合全局和局部深度特征的高分辨率遥感影像场景分类方法[J]. 光学学报, 2019, 39(3): 0301002.

[5] 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.

[6] Zhao L J, Tang P. Improved visual vocabularies for scene classification of high resolution remote sensing imagery in urban areas[C]//2019 Joint Urban Remote Sensing Event (JURSE), May 22-24, 2019, Vannes, France. New York: IEEE Press, 2019.

[7] Wu C, Wang H W, Wang Z Q, et al. Zero-shot classification for remote sensing scenes based on locality preservation[J]. Acta Optica Sinica, 2019, 39(7): 0728001.
吴晨, 王宏伟, 王志强, 等. 基于局部保持的遥感场景零样本分类算法[J]. 光学学报, 2019, 39(7): 0728001.

[8] 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.

[9] Yang Q L, Zhou B H, Zheng W, et al. Dim and small target detection based on fully convolutional recursive network[J]. Acta Optica Sinica, 2020, 40 (13): 1310002.
杨其利, 周炳红, 郑伟, 等. 基于全卷积递归网络的弱小目标检测方法[J]. 光学学报, 2020, 40(13): 1310002.

[10] Yuan L, Yuan J S, Zhang D Z. Remote sensing image classification based on DeepLab-v3 + [J]. Laser & Optoelectronics Progress, 2019, 56(15): 152801.
袁立, 袁吉收, 张德政. 基于 DeepLab-v3+的遥感影像分类[J]. 激光与光电子学进展, 2019, 56(15): 152801.

[11] 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.

[12] Cheng G, Ma C C, Zhou P C, et al. Scene classification of high resolution remote sensing images using convolutional neural networks[C]//2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July 10-15, 2016, Beijing, China. New York: IEEE Press, 2016: 767-770.

[13] Wang F, Jiang M Q, Qian C, et al. Residual attention network for image classification[EB/OL]. (2017-04-23)[2020-09-14]. https://arxiv.org/abs/1704.06904.

[14] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[15] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015.

[16] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[17] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[18] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.

[19] Zhou B L, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2921-2929.

[20] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 618-626.

[21] Zhang T, Zheng E R, Shen J G, et al. Remote sensing image scene classification based on deep multi-branch feature fusion network[J]. Acta Photonica Sinica, 2020, 49(5): 0510002.
张桐, 郑恩让, 沈钧戈, 等. 基于深度多分支特征融合网络的光学遥感场景分类[J]. 光子学报, 2020, 49(5): 0510002.

[22] Wang Q, Liu S T, 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.

[23] Bian X Y, Fei X J, Mu N. Remote sensing image scene classification based on scale-attention network[J]. Journal of Computer Applications, 2020, 40(3): 872-877.
边小勇, 费雄君, 穆楠. 基于尺度注意力网络的遥感图像场景分类[J]. 计算机应用, 2020, 40(3): 872-877.

[24] 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.
汪鹏, 刘瑞, 辛雪静, 等. 基于残差网络的光学遥感图像场景分类算法[J]. 激光与光电子学进展, 2021, 58(2): 0210001.

[25] Jain S, Wallace B C. Attention is not explanation[EB/OL]. (2019-05-08)[2020-09-14]. https://arxiv.org/abs/1902.10186.

[26] Lin M, Chen Q, Yan S C. Network in network[EB/OL]. (2014-03-04)[2020-09-14]. https://arxiv.org/abs/1312.4400.

[27] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.

[28] 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.

[29] 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.

[30] Zhang X N, Zhong X, Zhu R F, et al. Scene classification of remote sensing images based on integrated convolutional neural networks[J]. Acta Optica Sinica, 2018, 38(11): 1128001.
张晓男, 钟兴, 朱瑞飞, 等. 基于集成卷积神经网络的遥感影像场景分类[J]. 光学学报, 2018, 38(11): 1128001.

[31] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2020-09-14]. https://arxiv.org/abs/1409.1556v6.