第 60 卷 第 4 期/2023 年 2 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

细粒度显著区域引导的遥感图像场景分类

李飞扬，王江涛*，王子阳

淮北师范大学物理与电子信息学院，安徽 淮北 235000

摘要 近年来，遥感图像场景在监测环境、勘探地球资源及预测自然灾害等方面有着越来越广泛的应用，大量的数据需求推动了遥感图像场景分类的快速发展。尽管基于深度学习的方法已经在场景分类方面取得了比较好的性能，但如何对背景复杂、尺度变化剧烈的遥感场景进行有效识别仍然是分类任务中的一个巨大挑战。为了解决这一问题，提出一种细粒度方法来检测显著区域，并使用全局分支和局部分支将整体和局部联合起来，分别从整幅图像和关键区域提取全局特征和局部关键信息。为了验证所提方法的有效性，基于ResNet18模型在三个公共遥感图像场景分类数据集上对不同方法进行对比实验，实验结果表明所提方法的准确率优于大多数先进方法。

关键词 深度学习；遥感；细粒度；显著区域检测

中图分类号 TP751 **文献标志码** A **DOI**：10.3788/LOP212616

Scene Classification of Remote Sensing Images Guided by Fine-Grained Salient Region

Li Feiyang, Wang Jiangtao*, Wang Ziyang

School of Physics and Electronic Information, Huaibei Normal University, Huaibei 235000, *Anhui, China*

Abstract Recently, remote sensing image scenes have been increasingly widely used in monitoring the environment, exploring earth resources, and predicting natural disasters. Numerous data requirements aid in the rapid development of remote sensing image scene classification. Although the deep learning-based method has achieved decent performance in scene classification, how to effectively classify remote sensing scenes with complex backgrounds and drastic scale changes remains a great challenge in the classification task. To address this issue, this paper proposes a fine-grained approach to detect the salient region, uses the global and local branches to combine the global and local parts, and extracts the global features and local key information from the whole image and key region, respectively. To verify the effectiveness of the proposed method, comparative experiments are conducted using ResNet18 on three public remote sensing image scene classification datasets, and the experimental results show that the accuracy of the proposed method is better than that of most advanced methods.

Key words deep learning; remote sensing; fine-grained; salient region detection

1 引 言

随着遥感技术的快速发展和成像设备的更新换代，遥感图像已经广泛应用于土地覆盖分类[1]、目标检测[2]、交通监控[3]及自然灾害监测[4]等多个领域。这些应用领域对遥感数据的需求不断增加，使得遥感图像的数量也呈现大幅度的增长，而遥感图像场景分类作为这些任务的基础，也成为计算机视觉领域的一个研究热点。

遥感图像覆盖范围广，背景复杂，同一类的图像之间可能有着非常大的差异，而不同类的图像之间则可能含有许多相似的、容易混淆的特征和对象，这些都给遥感图像场景分类带来了非常大的挑战。为了能够提高对遥感图像场景分类的精确度，人们在研究各种分类方法上付出了巨大的努力，而图像特征提取作为其中的核心问题，经历了手工制作到深度学习的转变。

收稿日期：2021-09-27；**修回日期**：2021-11-23；**录用日期**：2021-12-21；**网络首发日期**：2021-12-30

基金项目：国家自然科学基金(61976101)、安徽省高校自然科学研究重大项目(KJ2018ZD038)、安徽省高校优秀青年骨干教师国内访问研修项目(gxgnfx2021175)

通信作者：*jiangtaoking@126.com

传统的方法主要依靠人们的先验知识，然后手工设计一些低级特征的描述符，例如直方图[5]、纹理描述符[6]和尺度不变特征变换(SIFT)[7]等。为了能够有进一步的提高，研究人员以这些手工构建的特征描述符为基础提出了无监督的学习方法，即视觉词袋[8]、稀疏编码[9]、自动编码器[10]等。然而这些手工制作的特征有着很大的局限性，它们往往带着大量人为的想法，并不能有效地应对遥感图像场景分类中所出现的问题。之后，卷积神经网络的提出大大改善了这个问题。与传统的方法相比，基于卷积神经网络的方法可以自动地从原始图像中学习更深层次、更具有鉴别能力的语义级特征。

随着卷积神经网络的发展，近年来基于深度学习的细粒度图像分类算法研究也得到了飞速发展。针对细粒度图像的检索问题，Wei 等[11]提出了选择性卷积描述符聚合(SCDA)对卷积层和池化层的激活特征进行选择和保留，最后将保留下来的多层特征融合后进行分类。而遥感图像和细粒度图像之间有着诸多相似的特点，但是目前关于基于遥感图像的细粒度分类方法还没有太多的文献研究。为了验证基于细粒度的遥感图像场景分类方法的有效性，在 SCDA 的启发下，本文提出以细粒度显著区域引导的方法来联合全局和局部，从不同尺度的输入图像中分别提取全局和局部特征，并用于对遥感图像场景的分类。

2　网络结构与方法

图 1 为所提方法的流程。使用两个 ResNet18 模型分别作为全局分支和局部分支的基线网络来提取图像的特征并完成分类。针对不同的分支，两个模型分别有着不同的作用。全局分支在整幅图像上提取图像的全局特征，包括轮廓、纹理等信息。对于局部分支，输入图像经过全局分支的所有卷积层之后得到一系列特征图。然后利用生成的特征图检测局部关键区域，并使用局部分支提取更多细粒度特征。最后融合两个分支输出的分类分数。

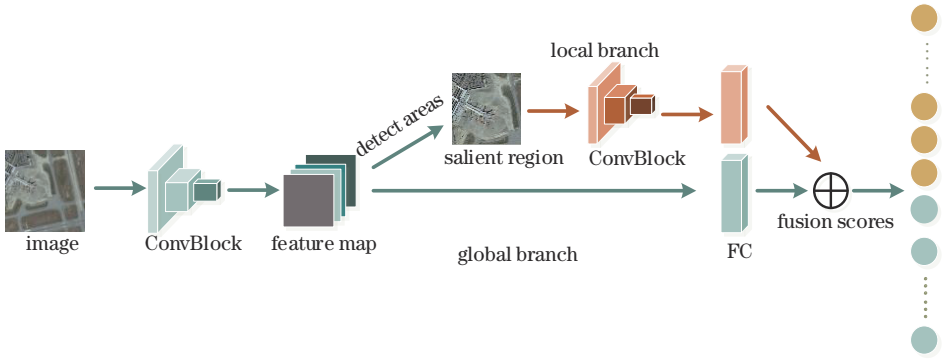

图 1　所提方法的流程

Fig. 1　Flow chart of the proposed method

2.1　基线网络

深度卷积神经网络的问世给计算机视觉领域的研究带来了极大的便利。2012 年以来 AlexNet 和 VggNet 已成功应用于各种各样的计算机视觉任务，例如目标检测[12]、图像分类[13]、对象跟踪[14]等。但是，随着网络层数的不断加深，梯度在网络的反向传播过程中逐渐减少和消失，导致训练精度快速下降。为了解决这一问题，He 等[15]在 VggNet 的基础上提出了一种残差网络(ResNet)，一定程度上解决了梯度消失的问题。

如图 2 所示，ResNet 主要使用残差单元来优化整个模型，解决网络的退化问题。每一个残差单元都由两层卷积层堆叠而成，然后通过跳跃链接将低层特征直接映射到高层特征，可以定义为

$$Y(x)=F\left(x,\{\theta_i\}\right)+x, \tag{1}$$

式中：x 和 $Y(x)$ 分别是输入和输出；$F\left(x,\{\theta_i\}\right)$ 是学习到的残差映射。图 2 中，$F\left(x,\{\theta_i\}\right)=\theta_2 g(\theta_1 x)$，$g$ 为激活函数，θ_1 和 θ_2 分别为第一层和第二层的权重参数。

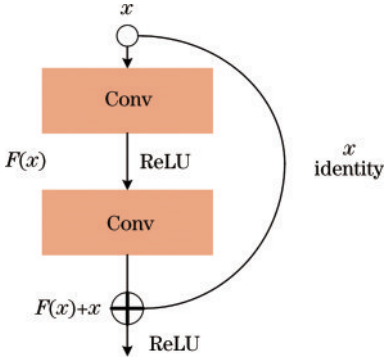

图 2　残差单元结构

Fig. 2　Structure of residual unit

2.2　显著区域检测

显著区域检测是整个模型中最关键的部分。如图 3 所示，输入图像经过全局分支后生成一系列特征图，如图 3(b)所示，在特征图上检测显著区域并得到显著区域的外接矩形框，如图 3(d)所示。然后采用如

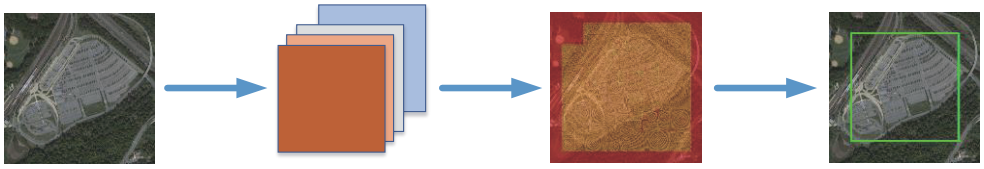

图 3　局部关键区域检测

Fig. 3　Local critical area detecting

图 1 所示的网络框架将局部的显著性区域特征和全局特征联合起来，最终完成遥感图像的分类。整个检测过程主要由以下几个步骤组成。

1）聚合映射

输入图像经过 ResNet18 网络的所有卷积层后生成一系列激活的特征图，将其表示成大小为 $H\times W\times C$ 的三维张量 $\boldsymbol{T}$。然后沿深度的方向进行加法运算，将 $\boldsymbol{T}$ 聚合为二维的张量 $\boldsymbol{M}$：

$$M_i=\sum_{i=1}^{C}T_i(H,W)。\tag{2}$$

得到聚合映射 $\boldsymbol{M}$ 后，将空间位置 (m,n) 上的激活值记为 δ。激活响应越高，就证明该区域包含有关键信息的可能性越大。

2）选择区域

为了能够准确找到并将可能含有关键信息的重要区域保留下来，首先计算 $\boldsymbol{M}$ 中所有位置的平均值作为阈值，计算方式为

$$\bar{a}=\frac{\sum_{m=0}^{H}\sum_{n=0}^{W}\delta}{H\times W}。\tag{3}$$

如果位置 (m,n) 处的激活响应值大于所求均值 $\bar{a}$，就认为该位置是需要关注的区域，并通过运算得到一幅和 $\boldsymbol{M}$ 大小相同的掩模 M'：

$$M'_{m,n}=\begin{cases}1, & M_{m,n}\geqslant\bar{a}\\ 0, & M_{m,n}<\bar{a}\end{cases}。\tag{4}$$

最后，将得到的掩模图覆盖到输入图像上，尽可能地保留模型感兴趣的区域。

3）去除噪声

经过前面的计算，网络基本已经检测到关键局部区域，但是遥感图像背景复杂，其中会有一些不相关的噪声区域被激活。为了减少噪声的影响，只保留其最大的连通区域，如图 4 所示。

通过以上几个步骤，最终得到如图 3(c)所示的描述符。同时得到关键区域的整个边界 $[x_1,x_2,y_1,y_2]$，如图 3(d)所示，并用它来对整幅图像进行采样。

4）提取显著区域引导的局部特征

使用双线性插值的方法，利用前面得到的边界框对原始输入图像 I_2 进行局部区域的采样，记为

$$I_2'=\text{bilinear}\left[I_2,(x_1,x_2,y_1,y_2)\right]。\tag{5}$$

将采样后的图像 I_2' 输入局部分支网络，提取局部的特征。

Algorithm 1: finding the largest connected region in binary image

1 convert to binary image;

2 mark all connected areas in the image;

3 create an empty list Pix;

4 calculate the area size of each connected domain and store it in the list;

5 get the index of the maximum value in the list, and return the coordinate attribute of the external frame of the connected domain.

图 4　在二值图像中寻找最大连通区域的算法流程

Fig. 4　Algorithm flow of finding the largest connected region in binary image

5）融合分类评分

全局和局部特征经过全连接层后分别得到全局和局部的分类评分 S_g 和 S_l。为了能够调节两个分类评分在最终输出中的占比，引入尺度调节参数 λ。最后融合两个分类评分的表达式为

$$S=\lambda S_g+(1-\lambda)S_l。\tag{6}$$

当输入样本为 j，分类类别数为 Z，每批样本数量为 N 时，损失函数为

$$L_{\text{cross}}=-\frac{1}{N}\sum_{j}\sum_{c=1}^{Z}\left[p_{jc}\log y_{jc}+(1-p_{jc})\log(1-y_{jc})\right],\tag{7}$$

式中：y_{jc} 表示样本 j 属于类别 c 的概率；p_{jc} 表示真实类别为 c 的概率，取值为 0 或 1。

3　实验结果与分析

3.1　数据集

1）RSSCN7 数据集

该数据集一共有 2800 张遥感图像，它们分别来源于 7 个场景：草地、农田、森林、河湖、停车场、工业区和住宅区。其中每一个场景都是根据 4 个不同的尺度进行采样的，最后得到 400 张像素大小为 400×400 的图像。图 5(a)展示了该数据集中的部分图像，由于图像来源于不同的季节、不同的天气、不同尺度的变换，场景内容的变化较大，RSSCN7 数据集的分类具有较大挑战性。

2）Aerial Image 数据集

该数据集一共有 10000 张航空遥感图像，其中包含 30 个类别，每一个类别都有数百张尺度不一的图像。图 5(b)展示了该数据集中的部分图像，由于航空遥感图像的视野比较大，其中可能包含各种各样复杂

的背景，有时图像中的关键目标过小（几十个像素），这些都给分类任务造成了比较大的困难。

3）NWPU-RESISC45数据集

该数据集一共有31500张遥感图像，包含45个场景类别，每一类都有700张像素大小为256×256的图像。图5(c)展示了该数据集中的部分图像，由于该数据集所含图像和类别数众多，且图像之间的相似程度较高，分类精度很难有大幅度的提升。

(a) RSSCN7　　(b) Aerial Image　　(c) NWPU-RESISC45

图5　部分数据集展示

Fig. 5　Partial dataset display

3.2　评估指标

1）准确率（accuracy）：准确率是指模型分类正确的图像数量与图像总数量的比值，可以用来评估整个模型的性能。总体准确率（OA）表示为

$$A_{\mathrm{O}}=\frac{T}{T+F},\tag{8}$$

式中：T表示分类正确的样本数量；F表示分类错误的样本数量。

2）混淆矩阵（confusion matrix）：混淆矩阵用n行n列的矩阵形式来表示精度评价。每一列代表的是预测的样本类别，每一行代表的是实际的样本类别。它能够进一步分析模型性能，然后对算法做出总结。

3.3　实验设置

1）环境配置：本实验在Windows 10系统下进行，以开源的深度学习框架Pytorch[16]作为模型的训练平台，在NVIDIA GeForce GTX 1660Ti GPU的计算环境下完成训练。

2）参数设置：在全局分支中，采用中心裁剪的方法将输入图像的大小变换为224×24；在局部分支中，整体图像先放大为448×448，然后对局部关键区域进行采样，并设置大小为224×224；在训练和测试的过程中选择Adam算法作为优化器，损失函数使用交叉熵损失；整个模型一共训练50个epoch，将批处理的大小设置为32；此外，模型的初始学习率都为1×10^{-4}，且每20个epoch学习率衰减为原来的0.1。

3.4　分类结果与分析

为了验证所提方法的性能，以ResNet18作为基线，在三个广泛使用的遥感数据集上进行训练和测试，并将所提方法的实验结果与其他一些先进的方法进行比较。同时，为了方便和公平，将RSSCN7和Aerial Image两个数据集的训练样本比例都设置为20%和50%分别进行测试，而NWPU-RESISC45数据集的训练样本比例设置为10%和20%。

1）RSSCN7数据集：由于受到季节、天气和尺度变化等原因，RSSCN7数据集的分类具有一定的难度。首先，利用所提方法在该数据集上对尺度调节参数进行实验研究，找到一个合理的λ。实验结果如表1所示，当全局评分和局部评分的权重相等时，模型的准确率最高，因此将λ都设置为0.5。

表1　在不同调节尺度下所提方法对RSSCN7数据集的OA

Table 1　OA of the proposed method under different regulation scales on RSSCN7 dataset　　unit:%

Parameter	20% training	50% training
$\lambda=0.3$	92.66	94.63
$\lambda=0.5$	93.17	95.11
$\lambda=0.7$	92.99	94.93
$\lambda=0.9$	92.30	94.60

为了进一步分析改进模型的空间复杂度和时间复杂度，在训练样本比例为50%的RSSCN7数据集基础上，对所提方法和其他模型的大小及运行时间进行比较，结果如表2所示。虽然基于全局和局部分支的模

表2　不同模型的大小和运行时间

Table 2　Size and running time of different models

Method	Model size /MB	Test time /s
Vgg16	112	0.0121
ResNet18	89	0.0096
ResNet50	195	0.0107
Proposed method	89×2	0.0117

型大小是全局分支的2倍，但是测试时间却比2倍小很多。

此外，不同方法在RSSCN7数据集上完成分类，对所提方法与其他方法的总体分类准确率进行对比，如表3所示。从表3可以看到：当训练样本的比例为20%时，所提方法的准确率比ResNet18模型的准确率高1.60个百分点左右；当训练样本的比例为50%时，所提方法的准确率要比EfficientNetB3[17]高0.90个百分点左右；综合来看，所提方法的性能比Resnet50[18]要好，与其他方法相比，所提方法在准确率上占据了很大的优势。

表3　不同方法对RSSCN7数据集的OA

Table 3　OA of different methods on RSSCN7 dataset

unit:%

Method	20% training	50% training
ResNet18	91.76±0.25	94.36±0.20
Proposed method	93.11±0.20	95.23±0.20
Resnet50[18]	90.23±0.43	93.12±0.55
Resnet50-TEX-Net-LF[18]	92.45±0.45	94.00±0.57
EfficientNetB3[17]	92.06±0.39	94.39±0.10
VGG-M-TEX-Net-EF-4ch[18]	86.77±0.76	89.61±0.54
VGG-M-TEX-Net-EF-6ch[18]	85.65±0.79	88.70±0.78
Deep filter banks[19]		90.40±0.60
Gan-full pipeline[20]	83.47±0.63	87.32±0.54
FV+HCV[21]		86.40±0.70
CaffeNet[22]	85.57±0.95	88.25±0.62
VGG-VD-16[22]	83.98±0.87	87.18±0.94

当训练样本的比例为50%时，制作了如图6所示的混淆矩阵。从混淆矩阵中可以看到：草地和田野之间非常容易出现错误分类；工业园区由于复杂的场景会被错误分类为停车场、河湖和居民区；森林则会被错误分类为河湖和停车场，原因可能是这三个类别之间会出现非常相似的场景，从而导致分类准确度受到影响。

2）Aerial Image数据集：Aerial Image数据集中的图像视野大、背景复杂、目标物体较小等情况给分类任务带来了非常大的困难。同样分别用两种比例的样本来训练、测试模型的性能，并对所提方法与其他方法进行比较，结果如表4所示。从表4可以清楚看到：在两种不同训练样本比例的数据集上，所提方法的准确率比ResNet18模型高大约1.40个百分点；当训练样本的比例为20%时，所提方法的准确率比Resnet50-TEX-Net-LF模型[21]高0.10个百分点左右；当训练样本的比例为50%时，所提方法的准确率比Resnet101-FSL方法[23]提高了0.27个百分点左右。相比其他方法，所提方法的准确率具有优势。

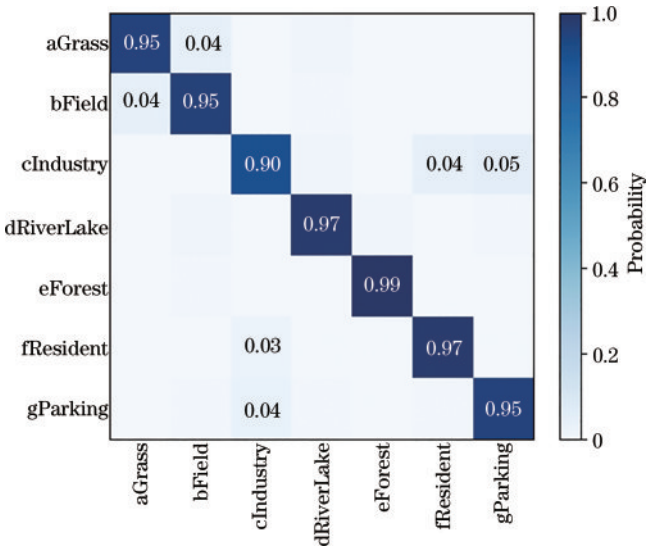

图6　当训练样本的比例为50%时RSSCN7数据集上的混淆矩阵

Fig. 6　Confusion matrix on RSSCN7 dataset when proportion of training samples is 50%

表4　不同方法对Aerial Image数据集的OA

Table 4　OA of different methods on Aerial Image dataset

unit:%

Method	20% training	50% training
ResNet18	92.31±0.20	95.35±0.22
Proposed method	93.90±0.25	96.06±0.20
Resnet101-FSL[23]		95.88
Resnet50-TEX-Net-LF[21]	93.81±0.12	95.73±0.16
EfficientNetB3[17]	93.43±0.33	94.45±0.76
Two-Stream Fusion[24]	92.32±0.41	94.58±0.41
Fusion by Addition[25]		91.87±0.36
CaffeNet[22]	86.86±0.47	89.53±0.31
TEX-TS-Net (addition)[26]	88.56±0.25	90.29±0.19
RADC-Net[27]	88.12±0.43	92.35±0.19
SalM3LBPCLM[28]	86.92±0.35	89.76±0.45
VGG-VD-16[22]	86.59±0.29	89.64±0.36

同时在训练样本的比例为50%时，绘制了如图7所示的混淆矩阵。从图7可以观察到：中心（center）、公园（park）、度假村（resort）、学校（school）、广场（square）非常容易被错误分类。这些分类错误的样本之间有着非常相似的场景，极大地影响了模型的分类精度。

3）NWPU-RESISC45数据集：该数据集是目前最大的数据集，同时也是分类任务中的一个巨大挑战。因为数据集所含图像数量众多，所以将训练样本的比例设为10%和20%分别进行训练和测试，结果如表5

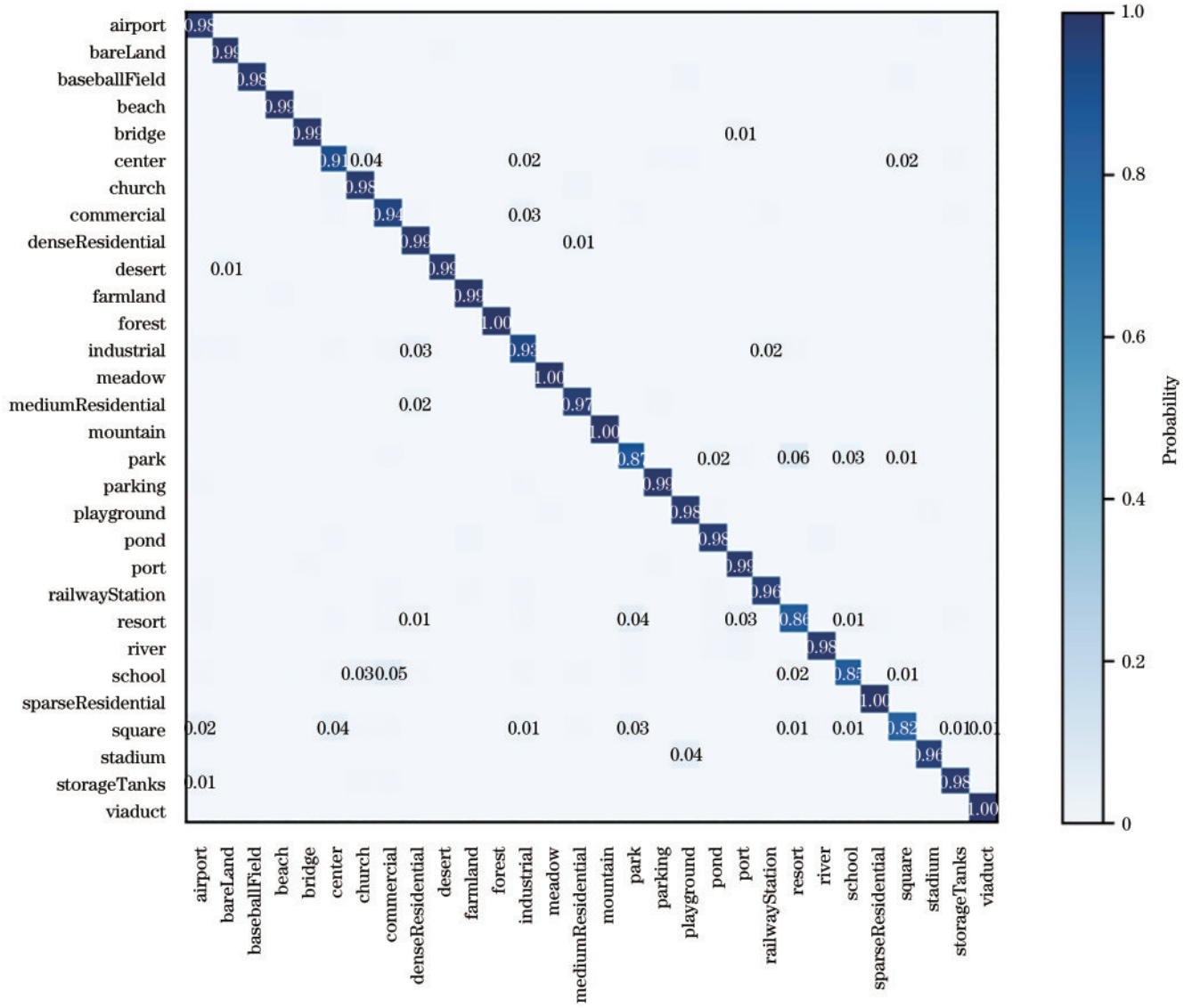

图 7　当训练样本的比例为 50% 时 Aerial Image 数据集上的混淆矩阵

Fig. 7　Confusion matrix on Aerial Image dataset when proportion of training samples is 50%

表 5　不同方法对 NWPU-RESISC45 数据集的 OA

Table 5　OA of different methods on NWPU-RESISC45 dataset

unit: %

Method	10% training	20% training
ResNet18	88.51±0.12	91.72±0.22
Proposed method	89.73±0.10	92.70±0.09
Multi-scale triplet loss[29]	88.30±0.24	91.62±0.35
Two-Stream Fusion[24]	80.22±0.22	83.16±0.18
RADC-Net[27]	85.72±0.25	87.63±0.28
Fine-tuned VGGNet-16[30]	87.15±0.45	90.36±0.18
Fine-tuned GoogLeNet[30]	82.57±0.12	86.02±0.18
SAL-TS-Net (addition)[26]	79.75±0.41	81.52±0.28
TEX-TS-Net (addition)[26]	79.63±0.30	81.22±0.27
Gan-full pipeline[20]	72.21±0.21	77.99±0.19

所示。可以清晰地看到：所提方法在 NWPU-RESISC45 数据集的分类任务上取得了较大的进步；当训练样本的比例为 10% 时，所提方法的准确率比 Multi-scale triplet loss[29] 提高了大约 1.40 个百分点；当训练样本的比例为 20% 时，所提方法的准确率比其高 1.00 个百分点左右。实验结果表明，所提方法的准确率比其他方法都要好得多。

训练样本比例为 20% 时，绘制了如图 8 所示的混淆矩阵。从中可以清楚地看到教堂（church）、商业区（commercial_area）、高速公路（freeway）、中型住宅（medium_residential）、宫（palace）、火车站（railway_station）、河流（river）等场景分类更加容易出错。与前两个数据集相同，遥感图像场景中背景复杂、相似目标过多的问题仍然需要更好的方法去解决。

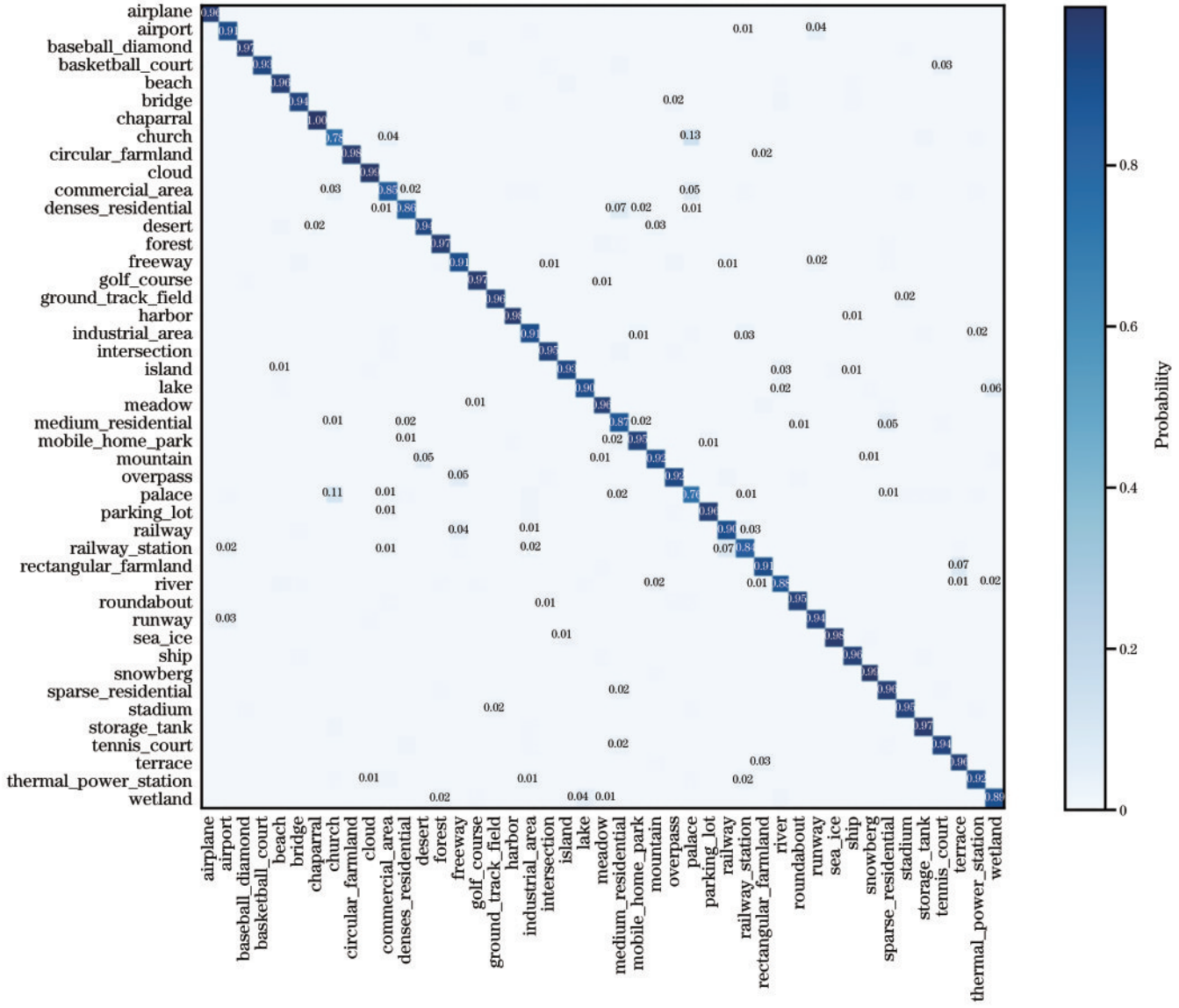

图 8　当训练样本的比例为 20% 时 NWPU-RESISC45 数据集上的混淆矩阵

Fig. 8　Confusion matrix on NWPU-RESISC45 dataset when proportion of training samples is 20%

4　结　　论

使用一种基于 CNN 模型的双分支结构，从细粒度的角度出发，联合两个 ResNet18 模型分别提取整体和局部的特征。其中关键是检测图像中的显著性区域，对增强后的局部图像再进行特征提取和分类。两个分支协同作用，从多个视角、多个尺度提取特征进行学习并完成分类评分。最后，在 3 个大尺度变化的数据集上进行了大量实验并对不同方法进行了对比，结果表明所提方法能够在一定程度上解决遥感图像场景分类的难题。在未来的学习中，图像显著区域的精确定位或将成为研究重点。

参　考　文　献

[1] Franklin S E, Wulder M A. Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas[J]. Progress in Physical Geography: Earth and Environment, 2002, 26(2): 173-205.

[2] Larsen S Ø, Koren H, Solberg R. Traffic monitoring using very high resolution satellite imagery[J]. Photogrammetric Engineering & Remote Sensing, 2009, 75(7): 859-869.

[3] Sunar F, Özkan C. Forest fire analysis with remote sensing data[J]. International Journal of Remote Sensing, 2001, 22(12): 2265-2277.

[4] Yan L, Zhu R X, Mo N, et al. Improved class-specific codebook with two-step classification for scene-level classification of high resolution remote sensing images[J]. Remote Sensing, 2017, 9(3): 223.

[5] Walton N S, Sheppard J W, Shaw J A. Using a genetic algorithm with histogram-based feature selection in hyperspectral image classification[C]//Proceedings of the Genetic and Evolutionary Computation Conference, July 13-17, 2019, Prague, Czech Republic. New York: ACM Press, 2019: 1364-1372.

[6] Yang Y, Newsam S. Geographic image retrieval using

local invariant features[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(2): 818-832.

[7] Cheng G, Han J W, Zhou P C, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98: 119-132.

[8] Csurka G, Dance C, Fan L, et al. Visual categorization with bags of keypoints[EB/OL]. [2021-09-26]. https://www.researchgate.net/publication/228602850_Visual_categorization_with_bags_of_keypoints.

[9] Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: a strategy employed by V1? [J]. Vision Research, 1997, 37(23): 3311-3325.

[10] Zhou W X, Shao Z F, Diao C Y, et al. High-resolution remote-sensing imagery retrieval using sparse features by auto-encoder[J]. Remote Sensing Letters, 2015, 6(10): 775-783.

[11] Wei X S, Luo J H, Wu J X, et al. Selective convolutional descriptor aggregation for fine-grained image retrieval[J]. IEEE Transactions on Image Processing, 2017, 26(6): 2868-2881.

[12] 程叶群, 王艳, 范裕莹, 等. 基于卷积神经网络的轻量化目标检测网络[J]. 激光与光电子学进展, 2021, 58(16): 1610023.
Cheng Y Q, Wang Y, Fan Y Y, et al. Lightweight object detection network based on convolutional neural network[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1610023.

[13] 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.
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.

[14] 刘美菊, 曹永战, 朱树云, 等. 基于卷积神经网络的特征融合视频目标跟踪方法[J]. 激光与光电子学进展, 2020, 57(4): 041502.
Liu M J, Cao Y Z, Zhu S Y, et al. Feature fusion video target tracking method based on convolutional neural network[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041502.

[15] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[16] Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library [C]//Advances in Neural Information Processing Systems 32, December 8-14, 2019, Vancouver, BC, Canada. New York: Curran Associates, 2019: 8024-8035.

[17] Alhichri H, Alswayed A S, Bazi Y, et al. Classification of remote sensing images using EfficientNet-B3 CNN model with attention[J]. IEEE Access, 2021, 9: 14078-14094.

[18] Anwer R M, Khan F S, van de Weijer J, et al. Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 138: 74-85.

[19] Wu H, Liu B Z, Su W H, et al. Deep filter banks for land-use scene classification[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(12): 1895-1899.

[20] Yu Y L, Li X Z, Liu F X. Attention GANs: unsupervised deep feature learning for aerial scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(1): 519-531.

[21] Wu H, Liu B Z, Su W H, et al. Hierarchical coding vectors for scene level land-use classification[J]. Remote Sensing, 2016, 8(5): 436.

[22] 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.

[23] Huang W, Wang Q, Li X L. Feature sparsity in convolutional neural networks for scene classification of remote sensing image[C]//2019 IEEE International Geoscience and Remote Sensing Symposium, July 28-August 2, 2019, Yokohama, Japan. New York: IEEE Press, 2019: 3017-3020.

[24] Yu Y L, Liu F X. A two-stream deep fusion framework for high-resolution aerial scene classification[J]. Computational Intelligence and Neuroscience, 2018, 2018: 8639367.

[25] 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.

[26] Yu Y L, Liu F X. Dense connectivity based two-stream deep feature fusion framework for aerial scene classification [J]. Remote Sensing, 2018, 10(7): 1158.

[27] Bi Q, Qin K, Zhang H, et al. RADC-Net: a residual attention based convolution network for aerial scene classification[J]. Neurocomputing, 2020, 377: 345-359.

[28] 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.

[29] Zhang J M, Lu C Q, Wang J, et al. Training convolutional neural networks with multi-size images and triplet loss for remote sensing scene classification[J]. Sensors, 2020, 20(4): 1188.

[30] 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.