

DOI: 10.13382/j.jemi.B2205792

基于阶段聚焦损失和并行增广策略的 遥感图像场景分类*

陈 燕 杨 艳 杨春兰 邓运生 李 壮
(蚌埠学院电子与电气工程学院 蚌埠 233030)

摘 要:随着深度学习的不断普及,卷积神经网络已成为遥感图像场景分类的主要手段,然而当前的研究主要集中于多网络主干的信息融合以及注意力机制等领域,在提高分类精度的同时也带来更高的计算复杂度。针对上述问题,分别从改进卷积神经网络损失函数和设计新的样本训练策略两个角度出发,在不增加计算复杂度的前提下,提升卷积神经网络的分类性能。首先,在对传统交叉熵和 Focal loss 损失函数进行分析的基础上,提出一种阶段聚焦损失函数,该损失函数可以在训练阶段对卷积网络进行有侧重的性能挖掘。其次,设计了一种并行样本训练策略,将采用 Gridmask 算法增广后的样本图像和原始样本图像,分为两路输入卷积网络进行并行训练,进一步提升卷积网络的分类性能。实验结果表明,所提出的算法分别在 AID 和 NWPU-RESISC45 两个大规模数据库上取得了 96.72% 和 93.95% 的检测精度,可以显著提升遥感图像场景分类的性能。

关键词: 遥感图像场景分类;阶段聚焦损失;并行 Gridmask 样本增广

中图分类号: TP751 **文献标识码:** A **国家标准学科分类代码:** 510.40

Remote sensing image scene classification via stage-based focal loss and parallel data augmentation strategy

Chen Yan Yang Yan Yang Chunlan Deng Yunsheng Li Zhuang
(School of Electronic and Electrical Engineering, Bengbu University, Bengbu 233030, China)

Abstract: With the continuing popularity of deep learning techniques, convolutional neural network (CNN) has become the main tool to solve the remote sensing image scene classification tasks. However, current research interests are highly focused on the topic of how to fuse multi-branch-based CNN and how to apply attention models. Despite that these approaches enhance the classification accuracy markedly; it leads to high computational complexity. In this paper, the above problems are addressed by means of introducing a modified loss function and designing a novel data augmentation strategy, which can significantly improve the classification performance of CNN without increasing the computational complexity. First, a stage-based focal loss function is presented to adaptively mining the hard sample during the training process. Second, a parallel training strategy is conducted to feed the original image samples and samples after Gridmask operation into the sharing CNN separately. Experimental results show that the proposed algorithm achieves 96.72% and 93.95% detection accuracy on two large-scale databases of AID and NWPU-RESISC45, respectively, and can significantly improve the performance of remote sensing image scene classification.

Keywords: remote sensing image scene classification; stage-based Focal loss; parallel Gridmask data augmentation

0 引 言

近年来遥感技术发展迅速,相关遥感平台的增加和观测能力的提高,也使得获取大规模的地球观测遥感图像数据变得不再困难。对遥感图像进行高效的分析在自

然灾害监测、农业普查、区域规划、自然资源检测等领域都发挥着重要作用。而遥感图像场景分类则是对遥感图像进行解译分析的一项核心内容。自从遥感图像场景分类这一任务提出以来,获得了越来越多研究人员的关注,目前已成为遥感信息处理领域的一个研究热点。

与传统的图像相比,遥感图像纹理复杂且颜色信息

不够丰富,呈现出明显的类间差异小,类内差异大的特点,这给遥感场景的精确分类带来较大的困难。为了解决这一问题,各种分类方法层出不穷。近年来,得益于深度学习的不断普及,特别是卷积神经网络的强大学习能力,基于深度学习的遥感场景分类方法已成为当前研究的主流^[1-6]。Minetto 等^[7]提出一种对不同的卷积网络进行多次微调训练,然后将多个网络进行集成的方法来提高遥感场景分类的精度。同样采用多个网络集成的策略,Xue 等^[8]将 CaffeNet、VGG-VD16 和 GoogLeNet 这 3 种常用的卷积网络特征融合在一起,共同完成遥感图像场景分类任务。文献[9]则通过设计一种改进的注意力机制模块,将均值池化和最大值池化进行融合,从而提高网络的分类鉴别能力。Bi 等^[10]在密集连接卷积网络的基础上,加入残差注意力模块来改进网络的分类性能。文献[11]则设计了一种循环注意力结构来挖掘卷积网络的高层语义信息和空间特征,从而提升系统的分类性能。然而无论通过注意力机制还是使用多源信息融合手段来提升遥感场景分类的性能,都需要在原始的卷积网络基础上增加相应的任务单元,从而不可避免的增加整个分

类网络的参数量和计算复杂度。

针对上述问题,本文从改进网络损失函数和设计新的训练机制两个角度出发,设计了一种新的遥感图像场景分类算法,在不改变骨干卷积网络、不增加新的网络结构的基础上,显著提高了遥感场景分类的性能。

1 模型的构建

1.1 主干网络的选择

本文所提出的整体模型框架如图 1 所示。不失一般性的,这里选择被广泛应用的残差网络 Resnet50 作为框架的主干网络。自从 Resnet 系列卷积网络被提出以来,其已被成功应用分类、目标检测等不同的任务场景中,并展现出优异的性能。如图 1 所示,Resnet50 的卷积结构可以分为 5 个部分,在待分类图像进入网络后,首先经过 1 个卷积层 Conv1,然后进入 4 个由残差模块组成的卷积结构 Layer1 至 Layer4,每一个结构包含残差模块的数量分别为 3、4、6、3,最后经过全连接层得到最终的分类输出。

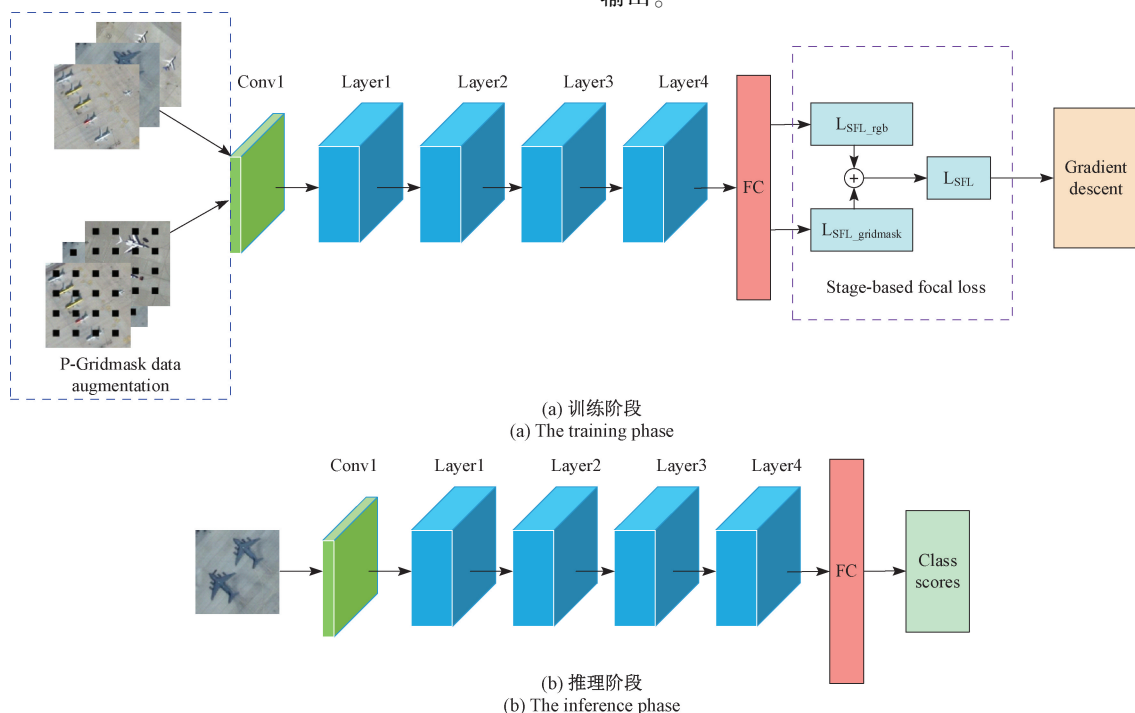


图1 本文提出的模型框架

Fig. 1 The proposed model

1.2 阶段聚焦损失

在基于卷积神经网络的分类任务中损失函数是关键的一环,不同的损失函数对网络的训练性能和收敛速度有着不同的影响。

目前,在众多的损失函数中,交叉熵损失函数是应

用最广泛的一种。以二分类为例,令训练样本为 (x_i, y_i) , $i=1, 2, \dots, n$ 。其中 x_i 为样本特征值, y_i 则为类别标签, y_i 的取值为 $\{0, 1\}$ 。假定样本 x_i 经过卷积神经网络后的预测值为 $p(x_i)$ 。则其对应的交叉熵损失可以定义为:

$$L_{CE}(x_i) = \begin{cases} -\log(1 - p(x_i)), & y_i = 0 \\ -\log(p(x_i)), & y_i = 1 \end{cases} \quad (1)$$

进一步的令:

$$p_i(x) = \begin{cases} p(x), & y = 1 \\ 1 - p(x), & \text{其他} \end{cases} \quad (2)$$

则上式可以整合为:

$$L_{CE}(x_i) = -y_i \log(p_i(x_i)) \quad (3)$$

观察上式可以看出,预测值 $p_i(x)$ 的大小在某种程度上反映了样本的难易程度。实践证明对困难样本进行正确的处理更有利于目标的分类。基于此 Lin 等^[12]提出了聚焦损失(focal loss, FL)这一概念,并将其成功应用在目标检测任务中。聚焦损失的工作原理是通过给传统的交叉熵损失增加一个加权项,从而对 $p_i(x)$ 值不同的样本进行区别对待,如式(4)所示:

$$L_{FL}(x_i) = -(1 - p_i(x_i))^\gamma \log(p_i(x_i)) \quad (4)$$

其中, γ 为幂次,在使用中一般取 2,用于对难易样本进行区别性对待。

然而通过实践发现在遥感图像场景分类任务中,直接使用 FL 损失函数并不能提高网络的分类性能。经过分析可知,在训练初期卷积网络的分类性能较弱,难以正确的对易分类样本和困难样本进行区分,此时若直接使用 FL 损失则可能将被误判的易分类样本当做困难样本进行着重学习,从而导致最终的分类性能不佳。基于上述分析,本文提出一种按照卷积网络所处的训练阶段对 CE 和 FL 损失进行自适应融合的损失函数,称之为阶段性聚焦损失(stage-based focal loss, SFL)。给出 SFL 的表达公式如式(5)所示:

$$L_{SFL}(x_i) = \alpha L_{FL}(x_i) + \beta L_{CE}(x_i) \quad (5)$$

其中, α 和 β 分别为 FL 损失和 CE 损失的权重因子,且有 $\alpha + \beta = 1$ 。 α 和 β 的值应随着训练 epoch 的增加而改变,为了更好的体现阶段性聚焦损失的思路,将 α 参数表达为 sigmoid 函数形式,使 α 和 β 的值可以平滑的变化,如式(6)~(7)所示:

$$\alpha = \frac{1}{1 + \exp((SP - s) \cdot ep)} \quad (6)$$

$$\beta = 1 - \alpha = \frac{\exp((SP - s) \cdot ep)}{1 + \exp((SP - s) \cdot ep)} \quad (7)$$

其中, s 为当前训练所处的阶段比值,假定一共训练 ep 个 epoch,且当前已经训练到第 c 个 epoch,则有 $s = c/ep$ 。而 SP 则为阶段性因子,其取值范围为 $[0, 1]$ 。 SP 的取值可以控制 CE 和 FL 的作用阶段,从而达到两者的有效融合。

1.3 基于并行 Gridmask 数据增广的样本训练策略

Gridmask 是 Chen 等^[13]提出的一种性能优越的数据增广方法,与 cutout, mixup 等流行的数据增广方法相比,其在 ImageNet 等数据集上取得了更好的分类性能。给

定输入图像 x , 则采用 Gridmask 算法增广后的图像可以表示为:

$$\tilde{x} = x \times M \quad (8)$$

其中, M 为预先生成的二值掩膜,其可以采用 $\{r, d, t_x, t_y\}$ 4 个参数来唯一的确定,如图 2 所示。其中 d 为单个掩膜的边长, r 为去除区域边长与 d 的比值, t_x, t_y 则为掩膜在以左上角为原点的原始图像中的坐标位置,上述参数给定了掩膜的初始位置,以该掩膜为起始,重复去除对应位置的像素,得到 Gridmask 处理后的最终图像,且原始图像的保留比例为:

$$R = 1 - \left(\frac{r}{d}\right)^2 \quad (9)$$

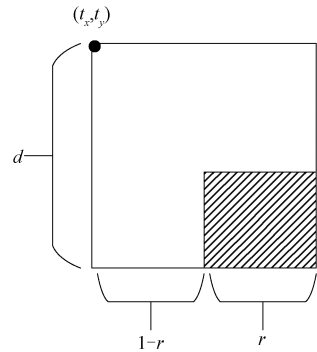


图 2 二值掩膜示意图

Fig. 2 Illustration of the binary mask

传统的 Gridmask 数据增广如图 3(a) 所示,然而在对遥感图像进行场景分类时,发现上述方法并不能带来性能的提升。针对该问题,本文设计了一种并行的样本增广训练方法 P-Gridmask,如图 3(b) 所示。该方法将原始样本和增广后的样本作为两类独立的数据,分别进行输入到主干网络进行训练。在训练过程中两种样本信息分别对应了不同的损失,如图 1(a) 所示,令原始的 RGB 图像经过主干网络后得到的损失为 L_{SFL_RGB} ,而 Gridmask 图像对应的损失为 $L_{SFL_Gridmask}$,则最终得到的损失为:

$$L_{SFL} = L_{SFL_RGB} + L_{SFL_Gridmask} \quad (10)$$

得到 L_{SFL} 后即可采用梯度下降方法对网络参数进行优化,从而完成网络的训练。

2 实验

2.1 实验设置

在实验中采用 AID 和 NWPU-RESISC45 两个被研究人员广泛使用的大规模数据集作为实验数据来对算法进行评估。表 1 给出了这两个数据集的细节信息,其中 AID 数据集包含 30 类不同场景下的 10 000 个样本数据,样本图像的分辨率为 600×600 ,而 NWPU-RESISC45 数据

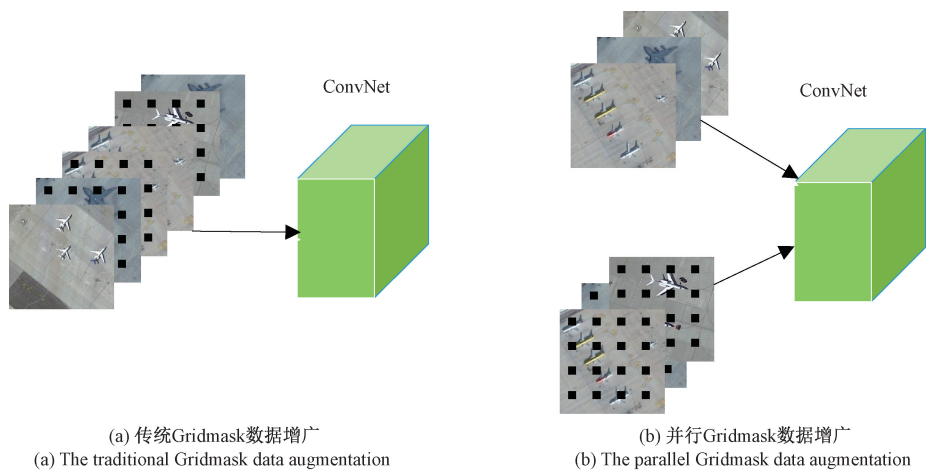


图 3 P-Gridmask 数据增广
Fig. 3 The P-Gridmask data augmentation

集则规模更大,其包含了 45 类不同的场景,样本总数则为 31 500。图 4 给出了两个数据集中部分不同场景类别的示例图像,从图中看出不同类别的图像往往具备较为类似的景物信息和纹理分布,这为对场景进行正确分类带来较大的困难。

表 1 AID 和 NWPU-RESISC45 数据集详细信息

Table 1 detail information of AID and NWPU-RESISC45 datasets				
数据集名称	图像分辨率	分辨率/m	样本数	场景类别
AID	600×600	0.5~8	10 000	30
NWPU-RESISC45	256×256	0.2~30	31 500	45

所有实验均在一台 CPU 为 i7-8700, GPU 为 GTX 1080Ti, 内存为 16 G, 显存为 11 G 的机器上进行, 所采用的操作系统为 Ubuntu16. 04, 实验中所用的深度学习框架为 Pytorch。采用 ImageNet 预训练模型作为 Resnet50 的初始参数, 每次实验均对模型训练 100 个 epoch, 初始学习率为 0. 001。训练的批量大小为 32, 优化器采用 SGD 方法。模型评估时, 采用总体分类正确率作为评价指标, 该指标给出了样本集中所有样本被正确检测的比率。

对于训练集和测试集的划分, 沿用基线算法的划分设置, 其中 AID 数据集在实验时分别从每类样本中随机抽取 20% 和 50% 作为训练集, 剩余的样本则作为测试集。由于 NWPU-RESISC45 数据集的规模更大, 在实验时则分别从每类样本中随机抽取 10% 和 20% 作为训练集, 剩余的样本则作为测试集。实验时随机的对数据集进行 5 次划分和训练测试, 以测试集上 5 次分类结果的平均值和标准差作为最终的分类正确率指标。

2.2 参数选择

在本文提出的分类框架中, 除通用的训练设置外, 涉及的参数主要有 Focal loss 中的 γ , Gridmask 生成二值掩

膜的控制参数和 SFL 中的阶段性因子 SP , 其中前两类参数沿用了文献[12]和[13]中的设置, 而在本节中着重分析 SP 参数对分类性能的影响。如表 2 所示, 以 NWPU-RESISC45 数据集采用 20% 的训练比为例, 给出了 SP 取不同值时所提出算法的分类总体精度。从表中可以看出, SP 取值不同, 分类总体精度随之发生较大变化, 通过对比可以发现, 当 SP 取 0. 6 时取得了最优的分类性能, 因此在实验中将 SP 设为 0. 6。

表 2 SP 取不同值时对 NWPU-RESISC45 数据集分类性能的影响

Table 2 the influence of classification performance on NWPU-RESISC45 dataset with different value of SP	
SP	OA (20%)
0. 0	92. 87±0. 36
0. 2	93. 39±0. 32
0. 4	93. 90±0. 25
0. 6	93. 95±0. 22
0. 8	93. 75±0. 29
1. 0	92. 72±0. 34

2.3 和其他算法的对比

为了进一步评估本文所提出算法的综合性能, 分别在 AID 和 NWPU-RESISC45 两个大型数据集上对本文算法的分类结果进行评估。在两个数据集上进行实验时, 除了本文算法外, 还与近年提出的一系列先进算法进行对比, 这些算法包括 Resnet50 + EAM^[9]、RADNet-Net^[10]、ADFF^[14]、MLCBF^[15]、TFADNN^[16]、MF²Net^[17] 以及 MG-CAP^[18] 等。

表 3 给出了在 AID 数据集上采用不同算法时的整体准确率。当采用 20% 数据集进行训练时, 本文算法取得

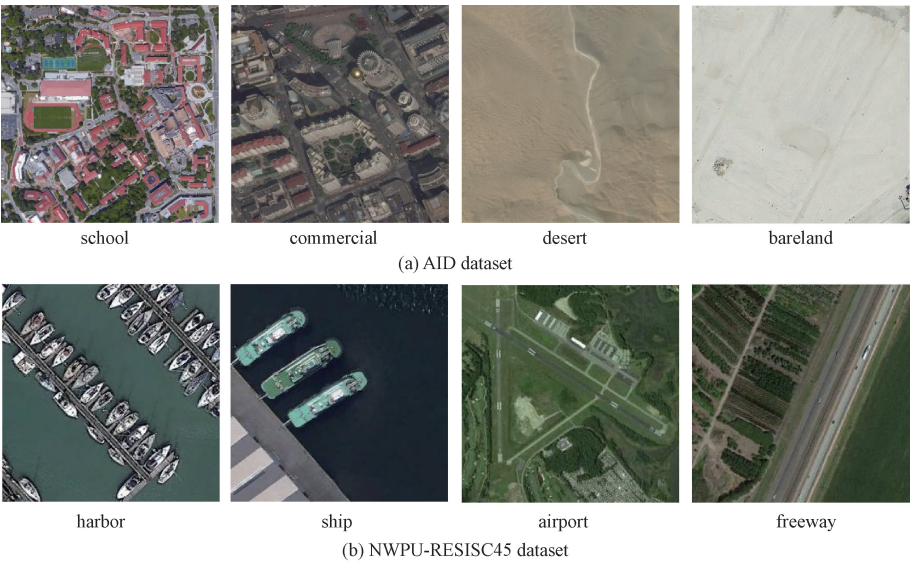


图 4 数据集遥感图像示例

Fig. 4 Remote sensing image samples in the datasets harbor

了 95.10% 的整体准确率,要高出注意力机制类算法^[19]中最好的 Resnet50+EAM 1.5%,即使与次优算法 MF²Net 相比,本文算法依然得到了 1.3% 的提升。当采用 50% 数据集进行训练时,由于训练样本的增多,相对采用 20% 的数据进行训练,所有的算法性能都有所提升,其中本文算法取得了 96.72% 的整体准确率,要高出细粒度算法^[20-21]中的 MG-CAP 0.6%。整体而言,在 AID 数据集中,不管采用 20% 还是采用 50% 的数据集作为训练集时,本文算法都取得了最高的整体准确率,对检测精度有着显著的提升。

表 3 不同方法在 AID 数据集上的分类性能比较

Table 3 Overall accuracy and standard deviation of the proposed method and other methods on AID dataset

方法	OA (20%)	OA (50%)
Resnet50+EAM ^[9]	93.64±0.25	96.62±0.13
RADC-Net ^[10]	88.12±0.43	92.35±0.19
MLCBF ^[15]	93.70±0.11	95.84±0.26
TFADNN ^[16]	93.21±0.32	95.64±0.16
MF ² Net ^[17]	93.82±0.26	95.93±0.23
MG-CAP ^[18]	93.34±0.18	96.12±0.12
ours	95.10±0.21	96.72±0.15

NWPU-RESISC45 数据集的实验结果由表 4 给出,由于 NWPU-RESISC45 数据集的规模远大于 AID 数据集,因此采用了更小的比例的数据集来构建训练集,实验时分别采用了 10% 和 20% 的数据集进行训练,剩余样本则为测试集。如表 4 中所示,当采用 10% 的数据集时,本文算法取得了 92.27% 的整体检测率,高出特征融合类算法 2% 以上,即使与基于注意力机制的次优算法 Resnet50+

EAM 相比,本文算法依然得到了 1.4% 的提升。当采用 50% 数据集进行训练时,由于训练样本的增多,相对采用 20% 的数据进行训练,所有的算法性能都有所提升,其中本文算法取得了 93.95% 的整体准确率,高出次优算法 Resnet50+EAM 0.44%。总的来说,在 NWPU-RESISC45 数据集中,不管采用 20% 还是采用 50% 的数据集作为训练集时,本文算法都取得了最高的整体准确率,这说明本文所提出的算法模块是有效的。

表 4 不同方法在 NWPU-RESISC45 数据集上的分类性能比较

Table 4 Overall accuracy and standard deviation of the proposed method and other methods on NWPU-RESISC45 dataset

方法	OA (10%)	OA (20%)
Resnet50+EAM ^[9]	90.87±0.15	93.51±0.12
RADC-Net ^[10]	85.72±0.25	87.63±0.28
ADFP ^[14]	86.01±0.15	88.79±0.17
TFADNN ^[16]	87.78±0.11	90.86±0.24
MF ² Net ^[17]	90.17±0.25	92.73±0.21
MG-CAP ^[18]	90.83±0.12	92.95±0.13
ours	92.49±0.18	93.95±0.22

2.4 消融实验

如表 5 所示,通过在原始的 Resnet50 主干网络中选择是否加入 SFL 和 P-Gridmask,一共可以得到 4 种不同的方案。首先只采用 Resnet50 主干网络进行分类,此时在 10% 和 20% 的 NWPU-RESISC45 数据集上进行训练,分别取得了 91.42% 和 92.72% 的总体准确率。然后将 SFL 损失函数引入到主干网络中,此时两种训练设置下

的总体准确率均得到显著的提升,分别达到 92.02% 和 93.68%。接下来分别将 Gridmask 和 P-Gridmask 两种不同的数据增广策略引入到训练中考察其对数据集分类性能的影响,从表中可以看出引入 Gridmask 后,分类性能不但没有提升反而严重衰减,相对而言,P-Gridmask 策略则可以进一步的提升主干网络的分类性能,在两种不同的训练设置下分别取得了 92.49% 和 93.95% 的最优准确率,分别比 fine-tuning Resnet50 提升 1.07% 和 1.23%,整体性能得到显著提升,表明了 SFL 和 P-Gridmask 两种策略的有效性。

表 5 各模块对 NWPU-RESISC45 数据集分类性能的影响
Table 5 different module's influence of classification performance on NWPU-RESISC45 dataset

网络结构	OA (10%)	OA (20%)
fine-tuning Resnet50	91.42±0.32	92.72±0.34
Resnet50+SFL	92.02±0.17	93.68±0.29
Resnet50+SFL+Gridmask	91.26±0.21	92.81±0.31
Resnet50+SFL+P-Gridmask	92.49±0.18	93.95±0.22

为了更加清晰的对原始的 Gridmask 和本文提出的 P-Gridmask 策略进行直观性对比,在图 5 中分别给出了采用 10% 和 20% 的比例设置训练集时,NWPU-RESISC45 数据集上的 OA 曲线。其中图 5(a) 为采用 10% 的样本集进行训练时在测试集上的 OA 曲线。从图中可以看出,随着训练的进行整体准确率逐步提升,采用 P-Gridmask 策略要显著优于原始的 Gridmask。同样的,如图 5(b) 所示,在采用 20% 的样本作为训练集时,P-Gridmask 的整体准确率依然要显著高于 Gridmask,充分表明 P-Gridmask 策略对系统的分类性能有着显著提高。

3 结 论

本文从设计阶段性侧重点不同的损失函数和并行样本数据增广策略两个角度出发提出了一种新的卷积网络分类框架来解决遥感图像场景分类问题。整体分类框架以 Resnet50 为主干网络,通过设计阶段性聚焦损失,使得该网络前期集中于提升网络对所有样本的整体分类性能。此外,通过对 Gridmask 数据增广方法进行改进,对原始图样和增广后的图像进行并行训练,进一步的提高对遥感图像的识别能力。实验表明本文所提出的方法显示出较为明显的优势。然而,本方法中的阶段性因子这一参数需要手动设定,如何对该参数进行自学习并确定最优值,将是未来一个值得关注的研究方向。

参考文献

[1] 高原,陈爱斌,周国雄,等. 基于 OctConv 的 DCNN 在遥感图像场景分类中的应用[J]. 电子测量与仪器学

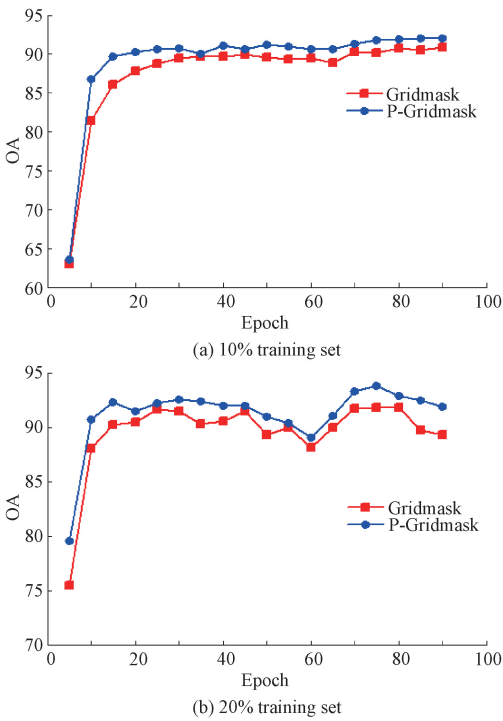


图 5 P-Gridmask 与 Gridmask 的检测率曲线
Fig. 5 The OA curves of P-Gridmask and Gridmask

报, 2020, 34(1):61-67.
GAO Y, CHEN AI B, ZHOU G X, et al. Application of DCNN based on OctConv in scene classification of remote sensing images [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(1):61-67.
[2] YU Y, LIU F. A two-stream deep fusion framework for high-resolution aerial scene classification [J]. Computational Intelligence and Neuroscience, 2018, 2018(4-5):1-13.
[3] ALHICHRI H, ALSUWAYED A, BAZI Y, et al. Classification of remote sensing images using efficientNet-B3 CNN model with attention[J]. IEEE Access, 2021, 9: 14078-14094.
[4] LI L, LIANG P, MA J, et al. A multiscale self-adaptive attention network for remote sensing scene classification[J]. Remote Sensing, 2020, 12(14): 2209.
[5] ZENG Z, CHEN X, SONG Z, et al. MGFN: A multi-granularity fusion convolutional neural network for remote sensing scene classification[J]. IEEE Access, 2021, 9: 76038-76046.
[6] 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):76-86.
ZHU SH 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):76-86.
- [7] MINETTO R, SEGUNDO M P, SARKAR S, et al. Hydra: An ensemble of convolutional neural networks for geospatial land classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(9): 6530-6541.
- [8] XUE W, DAI X, LIU L. Remote sensing scene classification based on multi-structure deep features fusion[J]. *IEEE Access*, 2020, 8: 28746-28755.
- [9] ZHAO ZH, LI J, LUO Z, et al. Remote sensing image scene classification based on an enhanced attention module[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(11): 1926-1930.
- [10] 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.
- [11] WANG Q, HUANG W, XIONG Z, et al. Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(4): 1414-1428.
- [12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2020, 42(2): 318-327.
- [13] CHEN P, LIU S, ZHAO H, et al. GridMask data augmentation [J]. *arXiv preprint*; 2020, arXiv: 2001.04086.
- [14] LI B, SU W, WU H, et al. Aggregated deep fisher feature for VHR remote sensing scene classification[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019, 12(9): 3508-3523.
- [15] 余甜微, 郑恩让, 沈钧戈, 等. 基于多级别跨层双线性融合的光学遥感图像场景分类[J]. *光子学报*, 2022, 51(2): 260-273.
- YU T W, ZHENG EN R, SHEN J G, et al. Optical remote sensing image scene classification based on multi-level cross-layer bilinear fusion [J]. *Acta Photonica Sinica*, 2022, 51(2): 260-273.
- [16] XU K, HUANG H, DENG P, et al. Two-stream feature aggregation deep neural network for scene classification of remote sensing images[J]. *Information Sciences*, 2020, 539: 250-268.
- [17] XU K, HUANG H, LI Y, et al. Multilayer feature fusion network for scene classification in remote sensing [J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(11): 1894-1898.
- [18] WANG S, GUAN Y, SHAO L. Multi-granularity canonical appearance pooling for remote sensing scene classification [J]. *IEEE Transactions on Image Processing*, 2020, 29: 5396-5407.
- [19] LI J, LIU X, ZHANG W, et al. Spatio-temporal attention networks for action recognition and detection [J]. *IEEE Transactions on Multimedia*, 2020, 22(11): 2990-3001.
- [20] DCAB C, WEI T, RUI C D, et al. A hierarchical approach for fine-grained urban villages recognition fusing remote and social sensing data[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2022, 106: 102661.
- [21] ARAUJO V M, BRITTO A S, OLIVEIRA L, et al. Two-view fine-grained classification of plant species [J]. *Neurocomputing*, 2022, 467: 427-441.

作者简介



陈燕(通信作者), 2019 年于淮北师范大学获得硕士学位, 现为蚌埠学院电子与电气工程学院助教, 主要研究方向为计算机视觉与图像处理。

E-mail: ychen_bbc@126.com

Chen Yan (Corresponding author)

received her M. Sc. degree from Huaibei Normal University in 2019. Now she is an assistant of the School of Electronic and Electrical Engineering of Bengbu University. Her main research interests include image processing and computer vision.