

神经网络训练策略对高分辨率遥感图像场景分类性能影响的评估

郑海颖¹, 王 峰¹, 姜 维¹, 王志强², 姚西文²

(1. 华北水利水电大学信息工程学院, 河南郑州 450046; 2. 西北工业大学自动化学院, 陕西西安 710129)

摘 要: 机器学习方法在高分辨率遥感图像场景分类任务中已经得到大规模应用, 但当前研究主要围绕数据特征和神经网络结构展开, 极少提及神经网络训练策略对遥感图像分类性能的影响. 因此, 本文选取7种自然图像分类中常用的神经网络训练策略进行实验, 根据其在3个规模较大的遥感图像数据集和4个广泛使用的神经网络模型上的实验表现, 筛选出适用于遥感图像场景分类的神经网络训练策略. 通过消融研究详细评估多个神经网络训练策略对遥感图像场景分类性能的影响, 通过分析总体分类精度、混淆矩阵、Kappa系数得到有效的神经网络训练策略, 并证明神经网络训练策略对遥感图像场景分类性能的有效性; 根据叠加实验的结果分析, 7种训练策略的组合可以在不同网络模型和数据集上表现出良好的适用性.

关键词: 机器学习; 高分辨率; 遥感场景分类; 训练策略; 神经网络

中图分类号: TP753 文献标识码: A 文章编号: 0372-2112(2021)08-1599-16

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20200961

Evaluation of the Effect of Neural Network Training Tricks on the Performance of High-Resolution Remote Sensing Image Scene Classification

ZHENG Hai-ying¹, WANG Feng¹, JIANG Wei¹, WANG Zhi-qiang², YAO Xi-wen²

(1. School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou, Henan 450046, China;

2. School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi 710129, China)

Abstract: Machine learning have been widely used in high-resolution remote sensing image scene classification task. However, the current research mainly focuses on data features and neural network structure, and the effect of neural network training tricks on remote sensing image classification performance is rarely mentioned. Therefore, this paper selects 7 neural network training tricks commonly used in natural image classification for experiments. According to their experimental performance in 3 large remote sensing image data sets and 4 widely used neural network models, neural network training tricks suitable for remote sensing image scene classification are selected. The effect of multiple neural network training tricks on the scene classification performance of remote sensing images was evaluated in detail through ablation experiment. An effective neural network training strategy was obtained by analyzing the overall accuracy, confusion matrix and Kappa coefficient, and the effectiveness of the neural network training strategy on the scene classification performance of remote sensing images was proved. According to the results of the stacking experiment, the combination of 7 training tricks can show good applicability in different network models and data sets.

Key words: machine learning; high-resolution; remote sensing scene classification; training tricks; neural network

1 引言

近年来,随着遥感成像技术逐渐成熟,具有更高分辨率的遥感图像场景数据不断出现,关于遥感图像场

景分类方法的研究得到广泛关注. 由于遥感场景图像具有类内多样性大和类间相似度高的特点,使高分辨率遥感图像场景分类任务面临巨大挑战^[1]. 此外,高分

收稿日期: 2020-08-31; 修回日期: 2020-12-14; 责任编辑: 孙瑶

基金项目: 国家自然科学基金(No.61601184); 河南省科技攻关计划(No.192102210265, No.202102210141); 河南省教育厅科学技术研究重点项目(No.13A520713); 河南省重点科技攻关计划(No.152102210112)

分辨率遥感图像场景分类方法在城市规划、军事目标识别、灾害评估等方面具有重要意义^[2,3]。因此,遥感图像场景分类具有重要的理论研究意义和应用价值。

深度学习(Deep Learning, DL)方法,特别是卷积神经网络(Convolutional Neural Network, CNN)依靠强大的特征学习和推理能力提高遥感图像场景分类性能。目前,DL和CNN在计算机视觉领域得到广泛应用^[4-6],使得目标检测、语义分割和图像分类的精度得到极大

提高。因此,将CNN引入高分辨率遥感图像场景分类中是土地场景分类研究的热点。根据现有研究成果^[3,7-13],对比传统方法,基于DL和CNN的场景分类方法在总体分类精度上更具优越性。在近五年影响因子较高的期刊中,遥感场景分类方法的研究主要集中在数据特征^[9,14-25]、网络结构^[8,15,17,19,26-38]和损失函数^[8,39]三个方面,很少有研究提及训练策略对分类性能的影响和训练策略的优化方法,如图1所示。

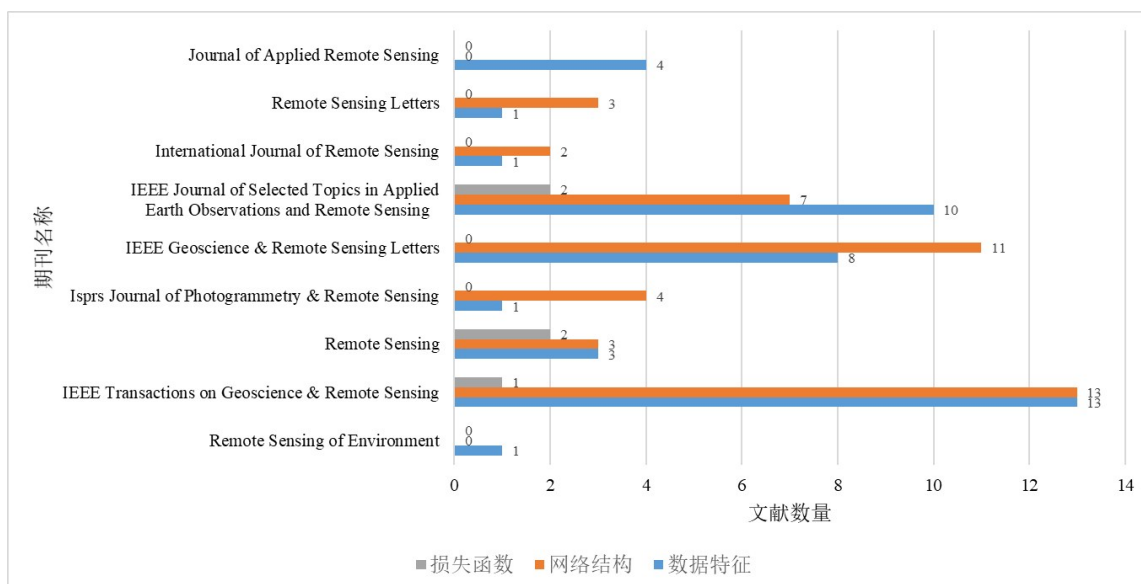


图1 2015—2020年高分辨率遥感场景分类方法主要研究方向

尽管新场景分类方法不断涌现,但是现有研究主要针对某一种特定网络或者特定使用场景和对象的优化,对于不同研究对象的适用性不高,并且对基于CNN场景分类方法中神经网络训练策略上的研究缺乏总结和针对性^[40-47]。从研究成果来看,基于CNN基线模型的遥感图像场景分类结果,在总体分类精度、混淆程度和稳定性方面要明显优于基于传统手工特征的分类结果,此外,基于度量学习的CNN模型优化方案使分类性能进一步提高,为解决类内多样性和类间相似性问题提供了研究思路^[3,8,40]。Cheng等人^[8]应用度量学习方法,在3种数据集上基于VGG-16模型的最优精度分别达到98.93%、96.89%、91.89%;边小勇等人^[1]应用注意力机制和多尺度特征变换方法在AID、NWPU-RESISC45数据集上的最优精度分别达到95.72%、92.25%;钱晓亮等人^[3]对CNN方法与传统方法进行了详细评估,证明了CNN方法在遥感场景分类任务中的优越性。上述研究证明,CNN应用于遥感图像场景分类的方法具有普遍性。神经网络训练策略是一种CNN训练程序优化方法,可以显著改善各种CNN模型,在ImageNet数据集上,ResNet-50的精度由75.3%提升到79.29%,但此种方法在遥感数据方面的研究涉及较

少。随着DL方法广泛应用于遥感图像场景分类任务,缺乏神经网络训练策略对场景分类方法性能影响的相关研究,不利于遥感图像场景分类方法的进一步发展。

因此,本文将神经网络训练策略引入遥感图像场景分类方法中,从优化神经网络学习过程,提高网络计算有效性方向着手,依据理论分析和实验对比,研究神经网络训练策略对遥感场景分类任务的有效性和适用性。本研究将主要解决以下问题:①对常用的神经网络训练策略进行总结,并从理论上对各类神经网络训练策略进行定性评估;②使用3个高分辨率遥感图像数据集(包含1个经典遥感图像数据集和2个场景类别规模不低于30类的大型遥感图像数据集),基于4个应用最广泛的神经网络的预训练模型(AlexNet、VGG-16、ResNet-50、GoogLeNet)进行实验,评估神经网络训练策略对高分辨率遥感图像场景分类性能的影响;③通过消融实验和叠加实验进行对比分析,验证单一策略和叠加策略对遥感影像分类精度的影响。最后,综合分析实验结果,证明神经网络训练策略在遥感图像场景分类任务中对不同网络模型和数据集组合的适用性和有效性。

2 训练策略

本研究中,CNN 参与的遥感图像场景分类基础实验流程如图2所示. 遥感图像为输入数据,在CNN中实现特征提取和分类任务,得到最终分类结果. 根据自然图像分类中常用神经网络训练策略^[48-55]分析,可迁移至高分辨率遥感图像场景分类任务的神经网络训练策略主要包括:学习率热身训练、线性比例学习率、无偏置衰减、余弦学习率衰减、标签平滑、混合训练、混合精度训练.

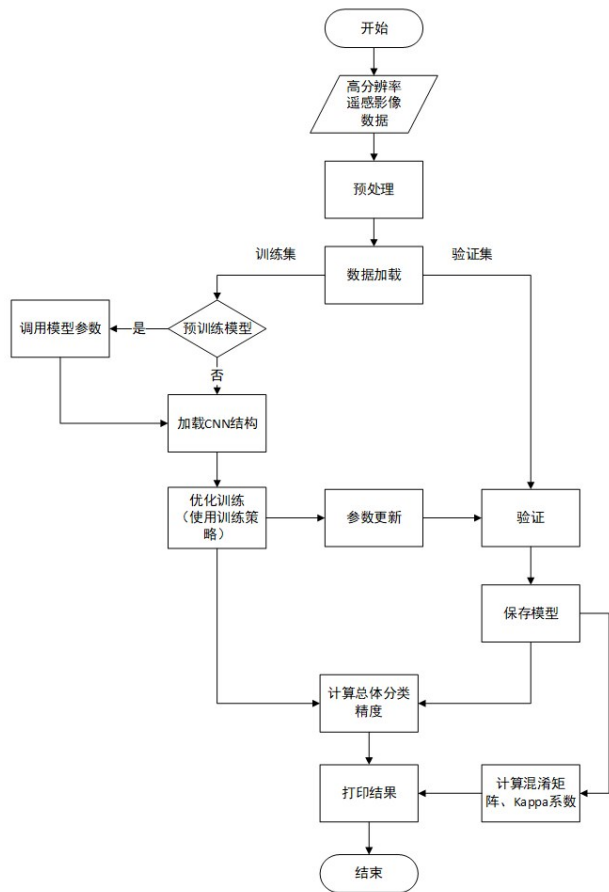


图2 实验流程

2.1 学习率热身训练

学习率热身训练(Learning Rate Warmup Training)是一种从较小学习率(Learning Rate, LR)缓慢上升到初始学习率的方式,可以减缓模型在训练初期由小批次(Mini Batch)内数据分布方差较大引起的过拟合现象,保持数据分布的平稳,同时保持模型深层的稳定性^[50,56-58]. 一种渐进式预热策略,是将学习率由0线性增加到初始学习率^[50],即

$$LR = \frac{i\eta}{m} \quad (1)$$

其中, m 为预热批次; η 为初始化学习率; $1 \leq i \leq m$.

2.2 线性比例学习率

线性比例学习率(Linear Scaling Learning Rate)常应用于大批次训练的分类任务. 实验中会通过提高学习率的方式加快实验进程^[48,50,59]. 假设实验中 batch size 设置为 256,初始化学习率设置为 0.1,若更改 batch size 为 b ,初始化学习率设置原则为

$$LR = 0.1 \times \frac{b}{256} \quad (2)$$

2.3 无偏置衰减

无偏置衰减(No Bias Decay)作为一种正则化手段,可用于权重和偏置的衰减任务,相当于对参数应用L2正则化,使值接近于0. 但Jia等人^[51]在ResNet-50和AlexNet上的试验结果表明:仅全连接层和卷积层的权重适合衰减操作,可达到约束训练权重,避免过度拟合的目的.

2.4 余弦学习率衰减

根据训练过程调整学习率是提高训练质量和效率的关键,He等人^[59]在实验中使用每30个周期降低0.1,Szegedy等人^[52]在实验中使用每2个周期降低0.94.Loshchilov等人^[55]提出的余弦学习率衰减(Cosine Learning Rate Decay)策略可以有效改善训练进度. 假设 batch size 总数为 T (不包含预热阶段的 batch size),则在某一批次 t ,学习率 η_t 为

$$\eta_t = \frac{1}{2} \left(1 + \cos \left(\frac{t\pi}{T} \right) \right) \eta \quad (3)$$

其中, η 为初始学习率.

2.5 标签平滑

标签平滑(Label Smoothing)本质上是一种正则化方法,可以有效增加类间距离,减少类内距离.Szegedy等人^[52]、He等人^[48]和Müller等人^[60]都已经在自然图像中验证了其有效性.

在多分类训练任务中,假设对于网络的输入图像, K 表示标签数(类别数), z_i 表示第 i 类的置信度分数, $q = \text{softmax}(z)$ 表示 Softmax 归一化输出,所以第 i 类预测概率 q_i 可表示为

$$q_i = \frac{\exp(z_i)}{\sum_{j=1}^K \exp(z_j)} \quad (4)$$

若输入图像真实标签为 y ,真实概率分布设置为

$$p_i = \begin{cases} 1, & \text{if } (i=y) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

使用交叉熵函数计算损失:

$$\text{loss}(p, q) = - \sum_{i=1}^K p_i \log q_i \quad (6)$$

通过最小化预测概率和标签真实概率的交叉熵,

从而得到最优的预测概率分布:

$$z_i^* = \begin{cases} +\infty, & \text{if } (i=y) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

标签平滑方法首先被提出应用于训练 Inception-v2 模型,它将真实概率分布修改为

$$p_i' = \begin{cases} 1-\varepsilon, & \text{if } (i=y) \\ \frac{\varepsilon}{K-1}, & \text{otherwise} \end{cases} \quad (8)$$

其中, ε 为值很小的常量.

交叉熵损失函数计算损失修改为

$$\text{loss}(p, q)' = \begin{cases} (1-\varepsilon) \cdot \text{loss}, & \text{if } (i=y) \\ \varepsilon \cdot \text{loss}, & \text{otherwise} \end{cases} \quad (9)$$

此时,最优解变为

$$z_i^{*'} = \begin{cases} \log \frac{(K-1)(1-\varepsilon)}{\varepsilon} + \alpha, & \text{if } (i=y) \\ \alpha, & \text{otherwise} \end{cases} \quad (10)$$

其中, α 为任意实数.

最终模型通过抑制正负样本输出差值,使得网络有更强的泛化能力.

2.6 混合训练

混合训练 (Mixup Training) 相比于传统图像增强操作 (旋转、裁剪、缩放等) 是一种混合增强训练方法. 与传统的经验风险最小化方法 (ERM) 相比, 混合训练方法具有更好的泛化能力^[61,62]. 在混合过程中, 通过对多个样本及其标签进行线性插值来生成新样本 (\hat{x}, \hat{y}) , 在混合训练中只使用新数据:

$$\hat{x} = \lambda x_i + (1-\lambda) x_j \quad (11)$$

$$\hat{y} = \lambda y_i + (1-\lambda) y_j \quad (12)$$

其中, $\lambda \in [0, 1]$, 服从 Beta 分布; (x_i, y_i) 和 (x_j, y_j) 表示两个随机样本.

2.7 混合精度训练

混合精度训练 (Mixed Precision Training) 是在尽可能减少精度损失的情况下, 利用半精度浮点数 (FP16) 加速训练. 精度损失是 FP16 代替 FP32 计算出现的主要问题. 针对上述问题, Micikevicius 等人^[63] 提出混合精度训练的概念, 文中提出三种方法: 保留 FP32 副本、损失放大 (Loss Scaling) 和混合计算 $\text{FP16} \times \text{FP16} + \text{FP32}$. 根据文中实验结果可以发现, 混合精度的训练结果基本没有精度损失.

3 实验结果与分析

3.1 实验环境

实验均在 Pytorch 框架下进行, 集成开发环境 (IDE) 为 Pycharm, 编译语言为 Python, CPU 为英特尔 E5-

2620V4, GPU 为 TIAN XP. 本文实验为提高训练速度采用单机多显卡分布式训练操作.

3.2 训练策略

为广泛评估神经网络训练策略对遥感图像场景分类性能的影响, 本文使用第 2 节中的神经网络训练策略参与实验对比. 以上 7 种训练策略基本涵盖图像分类中主流神经网络训练策略, 也是在高精度遥感图像场景分类任务中可能用到的训练策略.

3.3 实验数据

为了研究训练策略在遥感图像场景分类任务中的有效性和适用性, 并保证实验结果的可靠性. 本文采用 UC Merced Land-Use Dataset^[45]、AID Dataset^[46] 和 NWPU-RESISC45 Dataset^[47] 进行实验, 分析不同神经网络训练策略对高分辨率遥感图像场景分类性能的影响, 示例如图 3 所示.

UC Merced 遥感影像取自 USGS National Map Urban Area Imagery 系列, 是土地利用图像遥感数据集, 包含 21 个类别场景, 每个类别包含 100 幅图像, 共计 2100 幅遥感图像, 其空间分辨率为 1 英尺 (约 0.3m), 图像大小为 256×256 , 是遥感图像场景分类任务中常用的数据集.

AID 遥感影像由 Google Earth 不同远程成像传感器提取自不同地区、时间和成像条件, 包含有 30 个类别的场景, 每个类别包含 220~420 幅图像, 共计 10000 幅遥感图像, 其空间分辨率为 8~0.5m, 图像大小为 600×600 , 与 UC-Merced 数据集等单一来源的遥感图像数据集相比, 具有更高的类内多样性.

NWPU-RESISC45 遥感影像取自 Google Earth, 包含 45 个类别的场景, 每个类别包含 700 张图像, 共计 31500 幅遥感图像, 其空间分辨率为 30~0.2 m (岛、湖、山、雪山类图像空间分辨率较低), 图像大小为 256×256 , 场景类别丰富, 类内多样性和类间相似性较高, 对遥感图像场景分类具有更高的挑战性.

以上 3 个数据集比例划分与现有文献^[8,45~47] 保持一致, 均随机选择样本, 如表 1 所示.

3.4 评价指标

为评价神经网络训练策略对分类性能的影响, 本文采用总体分类精度、混淆矩阵、Kappa 系数和各类别分类精度标准差作为评估指标对实验结果进行分析. 此外, 为避免 CNN 模型中权值参数随机初始化和算法结果随机性, 保证实验结果稳定, 每组实验均设置相同随机种子 (Random Seed).

总体分类精度 (Overall Accuracy) 表示被正确分类的样本数占分类样本总数的比例. 计算公式为

$$P_{\text{OA}} = \frac{P_{\text{true}}}{P_{\text{all}}} \times 100\% \quad (13)$$

其中, P_{true} 表示被正确分类的样本数, P_{all} 分类样本

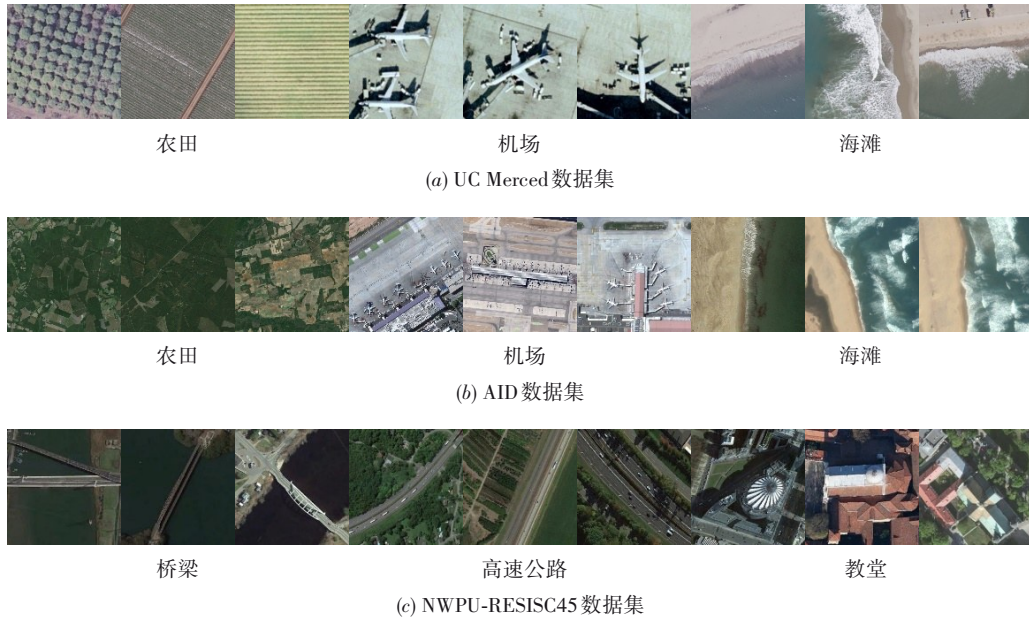


图3 3个高分辨率遥感图像场景分类数据集中的图像示例

表1 数据集训练和测试比例

Dataset	Train	Test
UC Merced	80%	20%
AID	20%/50%	80%/50%
NWPU-RESISC45	10%/20%	90%/80%

注:20%/50%、10%/20%等表示不同训练集比例,即训练样本占数据集总样本数的比例,如AID(20%)表示AID数据集中训练集比例为20%。

总数。

混淆矩阵(Confusion Matrix)在机器学习问题中常被用作多分类结果的可视化工具。矩阵的行和列分别代表分类样本的真实值和预测值,如表2所示。其中,矩阵中元素 x_{ij} 表示将第*i*种场景类别预测为第*j*种场景类别的样本数占该类别样本总数的比例。

Kappa系数基于混淆矩阵进行计算,是一致性检验的指标,通常范围为[0,1],常用于衡量分类的效果。

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (14)$$

即

$$Kappa = \frac{N \sum_{i=1}^K x_{ii} - \sum_{i=1}^K (a_i \cdot b_i)}{N^2 - \sum_{i=1}^K (a_i \cdot b_i)} \quad (15)$$

其中, N 代表样本总数, K 代表样本类别数, x_{ii} 是混淆矩阵对角元素, a_i 是混淆矩阵第*i*行元素总和, b_i 是混淆矩阵第*i*列元素总和。

3.5 基准训练程序

为减少实验中算力压力和节约CNN训练时间,各

表2 混淆矩阵

Actual	E_1	x_{11}	x_{12}	x_{1n}
	E_2	x_{21}	\ddots			
	\vdots	\vdots		x_{ij}		
	\vdots	\vdots			\ddots	
	\vdots	\vdots				x_{m1}
	E_n	x_{n1}				
		E_1	E_2	E_n
		Predicted				

注:混淆矩阵以二维表形式表示,行和列分别代表:“Actual”为真实类别标签,“Predicted”为预测类别标签。其中, $E_1 \sim E_n$ 为类别标签, $x_{11} \sim x_{nn}$ 表示二维表中元素。

项实验均基于ImageNet预训练模型。CNN训练阶段基准训练采用初始训练设置,过程如下。

- (1)输入数据使用随机采样,图像解码为32-Float。
- (2)基准训练程序不进行图像增强处理(旋转、反转、平移、裁剪等)操作。
- (3)重置图像分辨率为224×224,打乱数据读取次序。
- (4)归一化RGB通道,均值为[0.5,0.5,0.5],方差为[0.5,0.5,0.5]。
- (5)基准训练程序下使用交叉熵损失函数(CrossEntropy Loss)和随机梯度下降(SGD),学习率衰减设置为周期学习率衰减(StepLR)。
- (6)训练批次、循环训练次数和权重衰减等参数设置如表3所示。

本节对 AlexNet、VGG-16、ResNet-50 和 GoogLeNet

模型的实验采用相同数据集,对应的 Baseline 训练程序具体设置如表 3 所示,总体分类精度如表 4 所示,混淆矩阵如图 4 所示, Kappa 系数如表 5 所示. 由于篇幅限

制,正文只展示基于 GoogLeNet、ResNet-50 模型,在 UC Merced(80%)、AID(20%)和 NWPU-RESISC45(10%)上的 Baseline 训练程序混淆矩阵,如图 4 所示.

表 3 初始训练程序参数设置

参数名称	参数值	参数名称	参数值
epochs	100	Weight decay	0.0001
batch size	128	LR step size	30
learning rate	0.001	seed	888

表 4 Baseline 训练程序总体分类精度/%

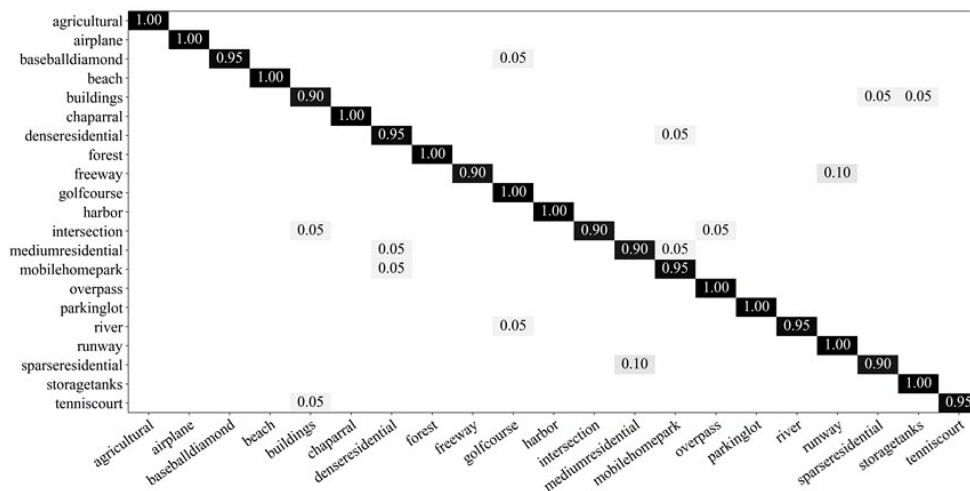
预训练模型	UC Merced	AID		NWPU-RESISC45	
	80%	20%	50%	10%	20%
AlexNet	92.86±0.09	82.65±0.12	88.96±0.08	74.56±0.09	80.91±0.07
VGG-16	95.00±0.08	84.69±0.12	90.50±0.08	76.01±0.13	84.08±0.09
ResNet-50	99.52±0.01	91.89±0.09	94.90±0.06	87.75±0.08	90.858±0.07
GoogLeNet	96.43±0.04	87.95±0.12	92.84±0.08	81.44±0.1	87.22±0.08

注:10%、20%、50%、80%代表训练集样本占数据集总样本数的比例,±为各分类精度的标准差.

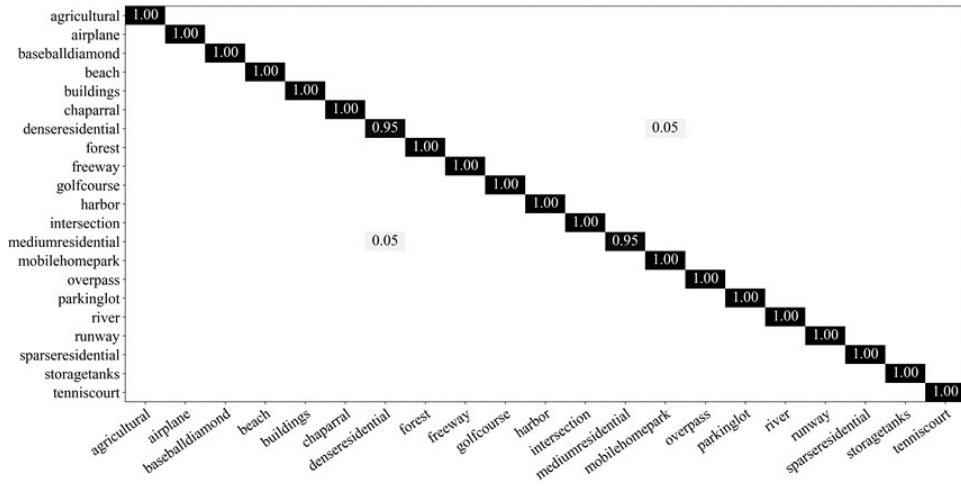
表 5 Baseline 训练程序 Kappa 系数/%

预训练模型	UC Merced	AID		NWPU-RESISC45	
	80%	20%	50%	10%	20%
AlexNet	0.925	0.820	0.886	0.744	0.809
VGG-16	0.948	0.841	0.902	0.755	0.837
ResNet-50	0.995	0.916	0.948	0.875	0.906
GoogLeNet	0.963	0.875	0.926	0.810	0.869

注:10%、20%、50%、80%代表训练集样本占数据集总样本数的比例.

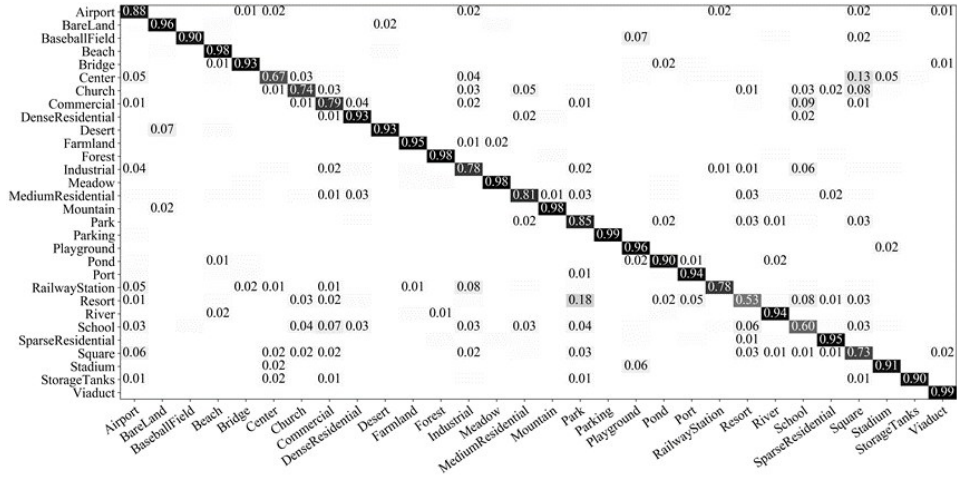


(a1) 基于 GoogLeNet 的混淆矩阵

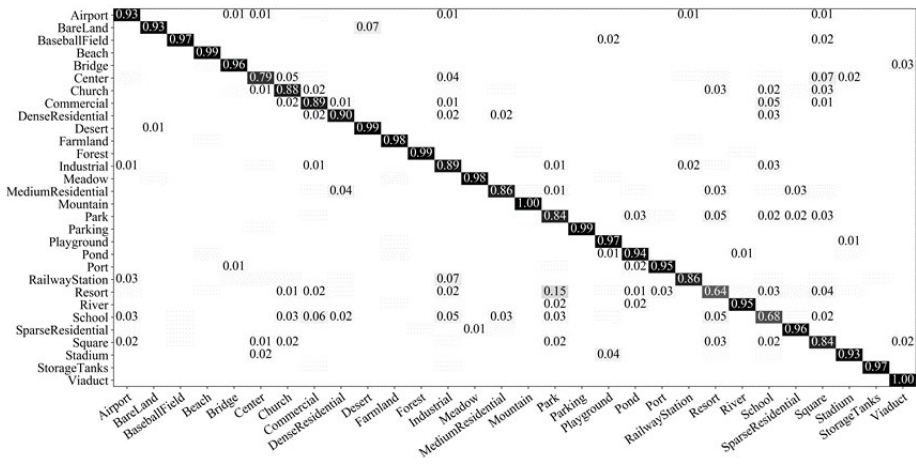


(a2) 基于 ResNet-50 的混淆矩阵

(a) GoogLeNet 和 ResNet-50 在 UC Merced(80%)数据集上的混淆矩阵

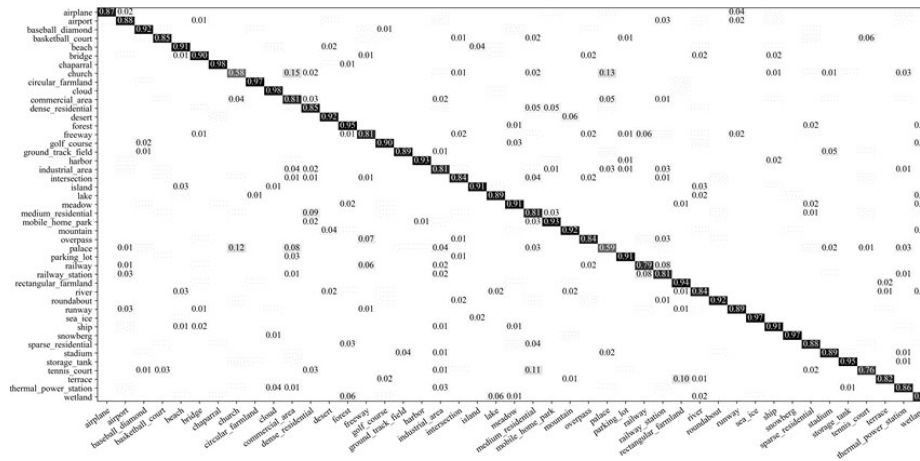


(b1) 基于 GoogLeNet 的混淆矩阵

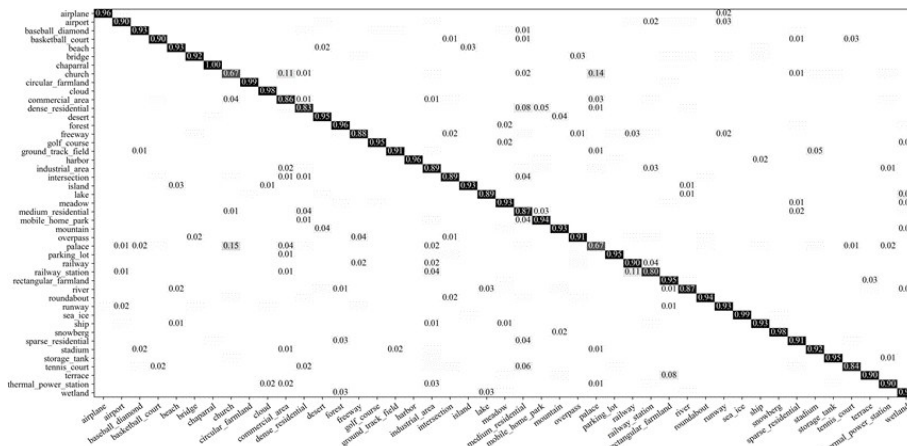


(b2) 基于 ResNet-50 的混淆矩阵

(b) GoogLeNet 和 ResNet-50 在 AID(20%)数据集上的混淆矩阵



(c1) 基于 GoogLeNet 的混淆矩阵



(c2) 基于 ResNet-50 的混淆矩阵

(c) GoogLeNet 和 ResNet-50 在 NWPU-RESISC45(10%)数据集上的混淆矩阵

图4 Baseline 训练程序混淆矩阵

3.6 实验结果

为对比训练策略在不同数据集和网络模型上的实验结果,保证实验客观有效,本节中所有实验均采用与 3.5 节相同的实验环境和设置. 以 3.6.1 节为例,单一训练策略在 AlexNet、VGG-16、ResNet-50、GoogLeNet 上的实验表现如表 6~9 所示,对于不同的深度学习模型均采用相同的数据集进行训练和测试,数据集中的训练集和测试集均按照表 1 进行划分. 表 10 中,对不同模型的训练策略叠加实验同样采用相同的数据集进行训练和测试,划分标准如表 1 所示. 3.6.2 节和 3.6.3 节中的实验设置与 3.6.1 节保持一致.

3.6.1 基于总体分类精度的定量评估

(1) 基于消融研究的实验结果. 基于 AlexNet 预训练模型,7 种神经网络训练策略在数据集上的总体分类精度如表 6 所示. 使用相同实验设置,分别对 VGG-16、ResNet-50 和 GoogLeNet 进行消融实验,表 7~9 分别表示

VGG-16、ResNet-50 和 GoogLeNet 的总体分类精度. 对比表 6~9 的实验结果,分析如下.

① 在基于 AlexNet 的实验中,大部分训练策略对于总体分类精度提升了 0.26%~2.04%,全部训练策略对 UC Merced(80%)、AID(50%)数据集分类效果均有提高.

② 在基于 VGG-16 的实验中,大部分训练策略对于总体分类精度提升了 0.08%~3.01%,全部训练策略对 AID(50%)数据集分类效果均有提高.

③ 在基于 ResNet-50 和 GoogLeNet 的实验中,大部分训练策略对各数据集分类效果提升不明显.

④ 训练策略对基于 AlexNet 和 VGG-16 模型分类精度有较普遍提升,对基于 ResNet-50 和 GoogLeNet 模型分类精度提升不明显.

⑤ 训练策略对 AlexNet 和 UC Merced 的平均提升精度最高,提升了约 1.24%.

(2) 基于叠加实验研究. 将上述对训练策略进行叠加实验,实验对象和相关设置与消融实验保持一致. 实

表 6 基于 7 种训练策略 AlexNet 的总体分类精度/%

训练策略	UC Merced		AID		NWPU-RE-SISC45	
	80%	20%	50%	10%	20%	
	Baseline	92.86	82.65	88.96	74.56	80.91
Warmup training	95.00	83.09	90.10	75.75	81.68	
No bias decay	94.52	82.95	89.44	75.56	81.39	
Linear scaling learning rate	93.33	81.08	88.98	73.80	80.42	
Low-precision training	93.33	81.95	89.24	75.01	81.04	
Cosine Learning Rate Decay	94.05	84.13	90.50	76.65	82.14	
Label Smoothing	93.10	81.28	89.12	73.56	79.87	
Mixup training	94.76	82.45	89.76	75.26	80.88	

注:10%、20%、50%、80%代表训练集样本占数据集总样本数的比例。加粗文字表示总体分类精度有提高。

表 7 基于 7 种训练策略 VGG-16 的总体分类精度/%

训练策略	UC Merced		AID		NWPU-RE-SISC45	
	80%	20%	50%	10%	20%	
	Baseline	95.00	84.69	90.50	76.01	84.08
Warmup training	95.48	84.75	91.48	78.26	85.21	
No bias decay	95.24	83.70	91.00	76.81	84.15	
Linear scaling learning rate	95.24	84.18	90.62	76.00	83.57	
Low-precision training	95.24	84.49	90.90	76.75	84.06	
Cosine Learning Rate Decay	95.00	85.24	91.32	78.31	85.04	
Label Smoothing	95.00	84.43	90.82	75.76	83.86	
Mixup training	97.86	85.85	92.04	77.29	85.03	

注:10%、20%、50%、80%代表训练集样本占数据集总样本数的比例。加粗文字表示总体分类精度有提高。

表 8 基于 7 种训练策略 ResNet-50 的总体分类精度/%

训练策略	UC Merced		AID		NWPU-RE-SISC45	
	80%	20%	50%	10%	20%	
	Baseline	99.52	91.89	94.90	87.75	90.85
Warmup training	99.52	91.48	94.72	87.34	90.73	
No bias decay	99.52	91.84	94.86	87.69	90.85	
Linear scaling learning rate	98.57	91.35	95.02	87.08	90.60	
Low-precision training	99.52	91.83	94.90	87.68	90.78	
Cosine Learning Rate Decay	99.52	92.01	94.98	88.05	91.04	
Label Smoothing	98.81	91.51	94.46	87.40	90.38	
Mixup training	98.81	91.41	95.06	87.63	91.03	

注:10%、20%、50%、80%代表训练集样本占数据集总样本数的比例。加粗文字表示总体分类精度有提高。

实验结果如表 10 所示,对比表 4 的 Baseline 训练程序总体分类精度,实验结果如下。

①叠加实验在各实验中的分类精度普遍提高,但是在 AlexNet、AID (20%) 和 ResNet-50、UC Merced (80%) 中的效果不佳。

表 9 基于 7 种训练策略在 GoogLeNet 预训练模型的总体分类精度/%

训练策略	UC Merced		AID		NWPU-RE-SISC45	
	80%	20%	50%	10%	20%	
	Baseline	96.43	87.95	92.84	81.44	87.22
Warmup training	95.71	86.40	92.14	79.52	86.12	
No bias decay	96.43	87.76	92.70	81.16	87.08	
Linear scaling learning rate	95.95	87.56	92.94	81.48	87.12	
Low-precision training	96.67	87.78	92.76	81.12	87.10	
Cosine Learning Rate Decay	96.43	88.78	93.44	83.03	87.80	
Label Smoothing	96.19	87.088	92.5	80.462	86.603	
Mixup training	95.48	85.73	92.36	79.50	86.72	

注:10%、20%、50%、80%代表训练集样本占数据集总样本数的比例。加粗文字表示总体分类精度有提高。

表 10 训练策略叠加实验总体分类精度/%

训练策略	UC Merced		AID		NWPU-RESISC45	
	80%	20%	50%	10%	20%	
	AlexNet	95.95	82.76	90.44	76.28	82.04
VGG-16	96.67	86.1	93.56	79.86	87.48	
ResNet-50	99.65	92.59	95.44	89.49	92.04	
GoogLeNet	98.57	91.09	94.56	86.52	90.03	

注:10%、20%、50%、80%代表训练集样本占数据集总样本数的比例。加粗文字表示总体分类精度有提高。

②叠加实验中各实验对象分类精度提升幅度普遍较高。

③叠加实验中,GoogLeNet、NWPU-RESISC45 (10%) 分类精度提高了约 6.2%,提升最为明显。

3.6.2 基于混淆矩阵的定量评估

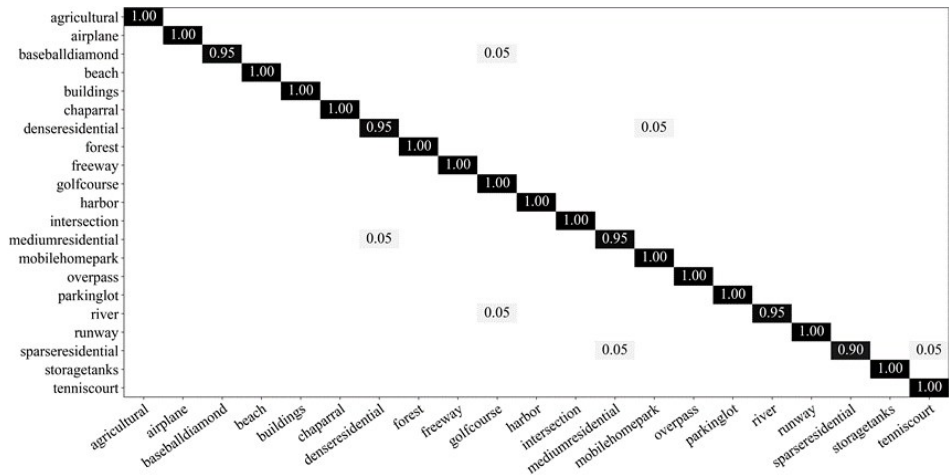
由于篇幅所限,正文只展示叠加训练策略基于 GoogLeNet、ResNet-50 模型,在 UC Merced (80%)、AID (20%) 和 NWPU-RESISC45 (10%) 上的混淆矩阵,如图 5 所示。为便于观察实验结果,根据矩阵中数值大小进行标黑处理,背景颜色越深,表示混淆程度越低,分类性能越好。其中,为更加清晰展示实验结果,矩阵中数值小于 0.01 的部分未显示。对比混淆矩阵,实验结果如下。

(1) 同一数据集下,基于各模型的混淆程度: AlexNet>VGG-16>GoogLeNet>ResNet-50。

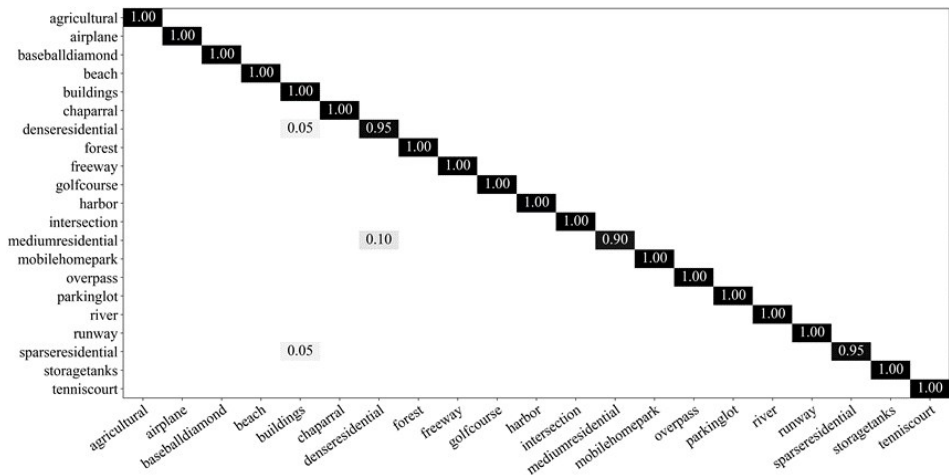
(2) 同一模型下,各数据集混淆程度: NWPU-RESISC45 (10%)>NWPU-RESISC45 (20%)>AID (20%)>AID (50%)>UC Merced (80%)。

(3) 根据消融实验结果:在基于 AlexNet 和 VGG-16 模型的实验中,大部分训练策略对降低混淆程度有效,且表现较明显;在基于 ResNet-50 和 GoogLeNet 模型的实验中,小部分训练策略对降低混淆程度有效,且表现不明显。

(4) 根据叠加实验结果:在基于不同模型和数据集的实验中,全部训练策略对于降低混淆程度有效,且表现较明显。

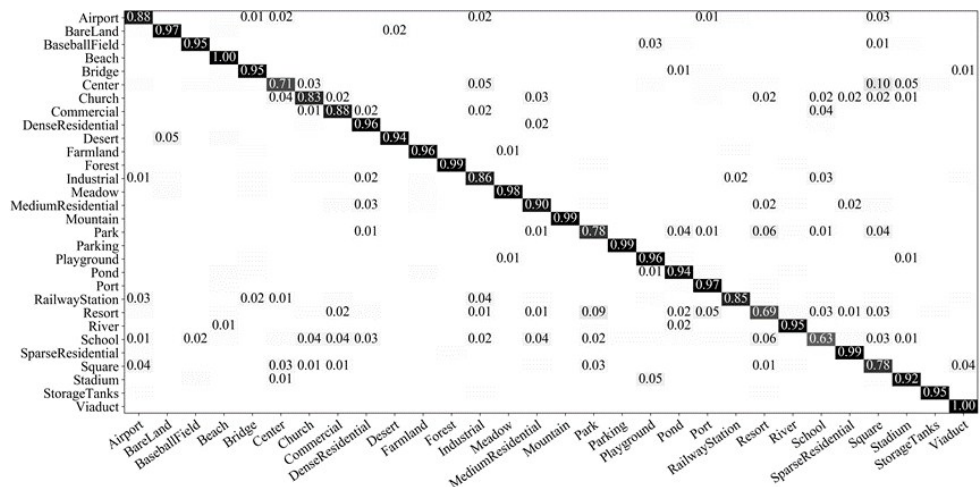


(a1) 基于 GoogLeNet 的混淆矩阵

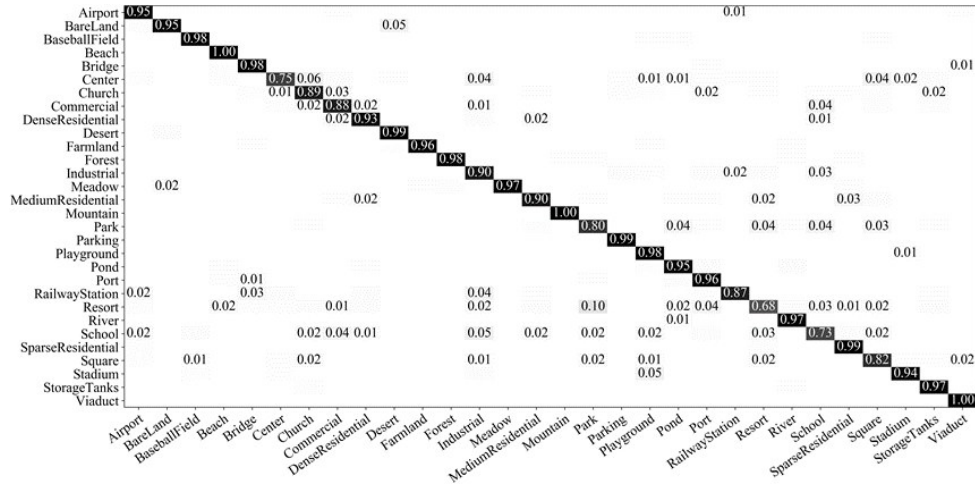


(a2) 基于 ResNet-50 的混淆矩阵

(a) GoogLeNet 和 ResNet-50 在 UC Merced(80%)数据集上的混淆矩阵

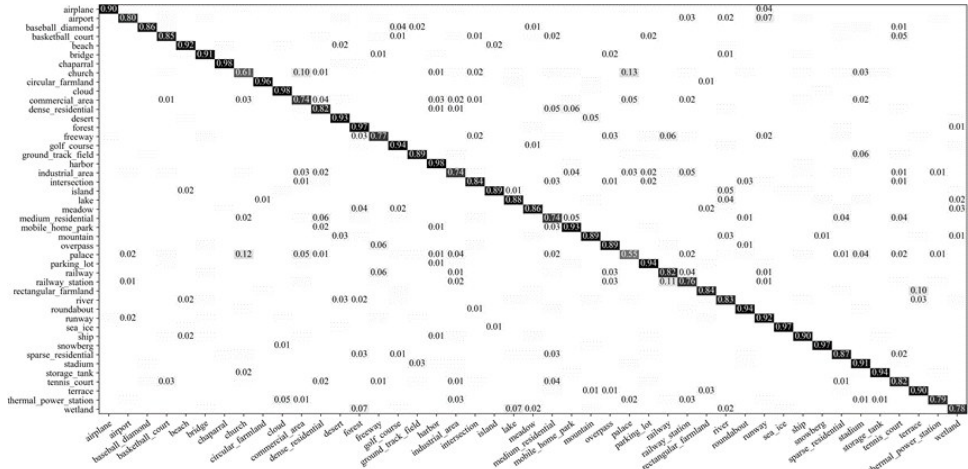


(b1) 基于 GoogLeNet 的混淆矩阵

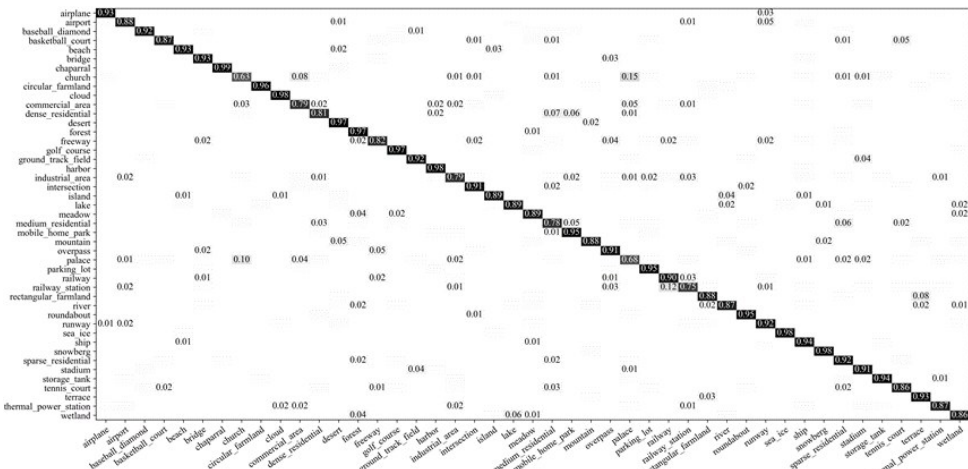


(b2) 基于 ResNet-50 的混淆矩阵

(b) GoogLeNet 和 ResNet-50 在 AID(20%)数据集上的混淆矩阵



(c1) 基于 GoogLeNet 的混淆矩阵



(c2) 基于 ResNet-50 的混淆矩阵

(c) GoogLeNet 和 ResNet-50 在 NWPU-RESISC45(10%)数据集上的混淆矩阵

图5 叠加训练策略基于 ResNet-50 和 GoogLeNet 模型在 UC Merced(80%)、AID(20%)和 NWPU-RESISC45(10%)数据集上的混淆矩阵

3.6.3 基于 Kappa 系数的定量评估

基于 AlexNet、VGG-16、GoogLeNet 和 ResNet-50 模型,在 UC Merced、AID 和 NWPU-RESISC45 数据集上,使用训练策略整体叠加后 Kappa 系数如表 11 所示,对比表 5 的 Baseline 训练程序的 Kappa 系数,实验结果如下。

① 在 UC Merced、AID 和 NWPU-RESISC45 数据集下,Kappa 系数在大部分实验中有提高。

② 基于 AlexNet、VGG-16、GoogLeNet 和 ResNet-50 模型的 Kappa 系数均有提高。

③ 同一数据集下,Kappa 系数大小为:ResNet-50 > GoogLeNet > VGG-16 > AlexNet。

④ 基于同一模型,Kappa 系数大小为:UC Merced (80%) > AID (50%) > AID (20%) > NWPU-RESISC45 (20%) > NWPU-RESISC45 (10%)

⑤ 在 UC Merced 数据集上,基于 AlexNet 的 Kappa 系数提高了约 3.6%,其提升最为明显。

表 11 基于叠加训练策略的 Kappa 系数/%

预训练模型	UC Merced	AID		NWPU-RESISC45	
	80%	20%	50%	10%	20%
AlexNet	0.958	0.815	0.901	0.757	0.816
VGG-16	0.965	0.856	0.933	0.794	0.872
GoogLeNet	0.985	0.908	0.944	0.862	0.898
ResNet-50	0.990	0.923	0.953	0.893	0.919

注:10%、20%、50%、80%代表训练集样本占数据集总样本数的比例。加粗文字表示 Kappa 系数有提高。

3.7 定量评估结果分析

3.7.1 综合分析

如表 6~9 所示,本研究通过大量实验,基于 AlexNet、VGG-16、GoogLeNet 和 ResNet-50 模型,在 UC Merced、AID、NWPU-RESISC45 数据集下,获得了 7 种神经网络训练策略总体分类精度的详细实验结果。结合表 10 和表 11,对比 3.5 节基准训练程序结果,可以得到以下结论。

综合总体分类精度、混淆矩阵、Kappa 系数来看,训练策略在不同模型和数据集上的表现各有特点。AlexNet、VGG-16 模型网络层数较少,结构简单,受训练策略影响较为明显。根据对不同复杂度和训练难度数据集的集中实验,实验结果表明:数据集复杂度和训练难度对实验结果影响较弱。其中,在模型结构较为简单的 AlexNet、VGG-16 模型的实验中,分类难度最低的 UC Merced 和分类难度最大的 NWPU-RESISC45,训练策略都表现出有效的积极影响;同时,在相同数据集,不同训练比例下,训练策略都表现出有效的积极影响,包括:AID (20%)、AID (50%) 和 NWPU-RESISC45 (10%)、NWPU-RESISC45 (20%)。

单一训练策略对模型总体分类精度的提高有限,随着网络层数增加和模型结构复杂,单一训练策略对精度提高的有效性降低,7 种训练策略对 AlexNet 模型产生有效的积极影响较为明显,VGG-16 次之,对 GoogLeNet 和 ResNet-50 模型的影响不明显;整体叠加的训练策略对分类精度的有效性影响较为明显,且提升分类精度的条件不限制于网络结构和数据集的复杂度,对不同网络模型的适用性较好。

3.7.2 逐类分析

使用相同训练策略和相同模型的情况下,各分类对象之间的混淆程度不同,这与数据集中各类别土地利用场景的地貌特征存在密切联系。因此,本研究从地物特点的角度出发,结合叠加训练策略在不同场景类别混淆矩阵的实验表现,分析训练策略对不同场景类别分类效果的影响。本研究主要分析基于 ResNet-50 模型的混淆矩阵,该模型在高分辨率遥感图像场景分类任务中应用频率较高,分类效果较好;数据集使用分类和训练难度最大的 NWPU-RESISC45 (10%),混淆矩阵如图 4 (c2)、图 5 (c2) 所示,图中对角线数据表示各类别分类正确的比例,对角线以外的数据表示各类别发生混淆的比例。依据混淆矩阵实验结果进行分析,结论如下。

(1) 根据图 4 (c2) ResNet-50 对角线数据,教堂、商业区、高速公路、工业区、住宅、火车站和网球场等场景类别的分类精度(分类精度 ≤ 0.8)远低于总体分类精度。篮球场、十字路口和湖泊等场景类别分类精度(0.8 \leq 分类精度 ≤ 0.87)略低于总体分类精度。

(2) 根据图 4 (c2) ResNet-50 对角线以外数据,混淆主要发生在教堂与商业区、教堂与住宅区、火车站与铁路等几个场景类别之间(混淆比例 ≥ 0.1)。这几类场景类别中包含相似的建筑物结构,特征区分不明显,可能导致较高的混淆比例。此外,矩形农田与梯田、湖泊与湿地、岛屿与河流等场景类别也容易发生混淆(0.05 \leq 混淆比例 < 0.1),这类场景本身地貌特征具有一定的相似性,同样可能是导致混淆比例偏高的原因。

(3) 根据图 5 (c2) 中的实验数据,叠加训练策略可以使混淆程度较高的几类场景(教堂与商业区、教堂与住宅区、火车站与铁路等)的混淆比例降低 0.02~0.03。同样,也可以使矩形农田与梯田、湖泊与湿地、岛屿与河流等场景类别的混淆比例降低 0.01~0.02。这是通过训练策略优化 CNN 训练过程,卷积核更有效提取图像特征,从而提升分类精度。

4 结论

本研究将神经网络训练策略引入高分辨率遥感图像场景分类的研究中,对现有神经网络训练策略进行

总结,实验采用3种遥感影像数据集(按训练集比例分为5组)和4种CNN模型,从理论分析和实验对比两个层面将训练策略对场景分类性能的影响进行评估,并对组合策略进行实验评估,得到以下结论.

(1)通过实验筛选,在AlexNet、VGG-16模型的性能表现中,Warmup training、No bias decay、Cosine Learning Rate Decay 3种训练策略表现最优,其他训练策略次之;在GoogLeNet和ResNet-50模型的性能表现中,Cosine Learning Rate Decay表现良好,其他策略表现不明显.因此,单一训练策略对网络层数较少、结构较简单的AlexNet、VGG-16产生有效的积极影响,对网络层数较多、结构复杂的GoogLeNet、ResNet-50的影响不明显.

(2)通过实验筛选,单一训练策略在不同分类难度和训练难度的数据集(UC Merced(80%)、AID(20%)、AID(50%)、NWPU-RESISC45(10%)、NWPU-RESISC45(20%))上的实验结果受模型影响较大,若是AlexNet和VGG-16,则性能表现较好;若是GoogLeNet和ResNet-50,则性能表现不明显.因此,单一训练策略对分类难度不同的数据集影响不明显,对不同训练难度的数据集影响不明显.

(3)通过叠加实验,7种训练策略的组合策略对AlexNet、VGG-16、GoogLeNet和ResNet-50模型的性能都具有较明显的提升,在不同分类难度和训练难度的数据集上,都具有较好的性能表现.因此,组合训练策略对遥感图像场景分类任务具有良好的适用性.

(4)通过对混淆矩阵的分析,发现建筑物场景相似的土地利用类型容易产生较大混淆,对高分辨率遥感图像场景分类任务的总体分类精度影响较大,土地覆盖类型相似的场景类别因地貌特征相似也易发生混淆.

(5)组合训练策略对降低易混淆的场景类别间的混淆程度有效,对降低建筑物场景类别混淆程度最为明显,对土地覆盖类型相似的场景类别也具有较为明显的作用.

目前,基于CNN模型的图像分类方法性能表现优异,是本领域的研究重点.相比于优化网络结构、改进损失函数、设计特征提取方法等手段,应用神经网络训练策略提高分类精度的方法,研究成本更低,应用范围更广.因此,在未来进行遥感图像场景分类的相关研究中可以引入神经网络训练策略方法,进一步提升研究方案的实验效果.另一方面,神经网络训练策略只是针对训练过程的优化,对CNN模型特征提取能力和分类性能的提升有限.将度量学习和生成对抗网络引入遥感场景分类任务中可能是下一步的研究热点之一,度量学习(Deep Metric Learning)是通过机器学习的形式,

依据训练数据构造一种基于特定任务的度量函数,是解决遥感图像中类间相似性高、类内差异性大问题的研究方法;生成对抗网络(Generative Adversarial Network, GAN)是在训练过程中,让两个神经网络通过相互博弈的方式进行学习,可以对遥感场景中易混淆类别进行重点学习,解决某些场景类别混淆程度较大的问题.

此外,本研究主要在公开数据集上进行实验研究,此类数据集只包含R、G、B三个波段的伪彩色图像,因此实验中无法利用遥感影像中更多的光谱信息进行分析.因此,将NDVI、NDWI等遥感参量和多波段信息应用到CNN模型训练学习中也是下一步研究的工作重点之一.

参考文献

- [1] 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.
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. (in Chinese)
- [2] 许凤晖,慕晓冬,赵鹏,等.利用多尺度特征与深度网络对遥感影像进行场景分类[J].测绘学报,2016,45(7):834-840.
Xu S H, Mu X D, Zhao P, et al. Scene classification of remote sensing image based on multi-scale feature and deep neural network[J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(7): 834-840. (in Chinese)
- [3] 钱晓亮,李佳,程臻,等.特征提取策略对高分辨率遥感图像场景分类性能影响的评估[J].遥感学报,2018,22(5):758-776.
Qian X L, Li J, Cheng G, et al. Evaluation of the effect of feature extraction strategy on the performance of high-resolution remote sensing image scene classification[J]. Journal of Remote Sensing, 2018, 22(5): 758-776. (in Chinese)
- [4] Krizhevsky A, Sutskever I, Hinton G. Image net classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [5] 罗会兰,陈鸿坤.基于深度学习的目标检测研究综述[J].电子学报,2020,48(6):1230-1239.
Luo H L, Chen H K. Survey of object detection based on deep learning[J]. Acta Electronica Sinica, 2020, 48(6): 1230-1239. (in Chinese)
- [6] 刘颖,刘红燕,范九伦,等.基于深度学习的小目标检测研究与应用综述[J].电子学报,2020,48(3):590-601.
Liu Y, Liu H Y, Fan J L, et al. A survey of research and ap-

- plication of small object detection based on deep learning [J]. *Acta Electronica Sinica*, 2020, 48(3): 590 – 601. (in Chinese)
- [7] Zhang W, Tang P, Zhao L. Remote sensing image scene classification using CNN-CapsNet[J]. *Remote Sensing*, 2019, 11(5): 494.
- [8] Cheng G, Yang C, Yao X, 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.
- [9] Cheng G, Li Z, Yao X, et al. Remote sensing image scene classification using bag of convolutional features[J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(10): 1735 – 1739.
- [10] 陈雅琼, 强振平, 陈旭, 等. 利用微调卷积神经网络的土地利用场景分类[J]. *遥感信息*, 2019, 34(3): 70 – 77.
Chen Y Q, Qiang Z P, Chen X, et al. Classification of land use scenarios based on fine-tuning convolution neural network[J]. *Remote Sensing Information*, 2019, 34(3): 70 – 77. (in Chinese)
- [11] 柳潜. 基于深度学习的遥感图像场景分类研究 [D]. 北京: 北京邮电大学, 2019.
Liu Q. Research of remote sensing image scene classification based on deep learning[D]. Beijing, China: Beijing University of Posts and Telecommunications, 2019. (in Chinese)
- [12] 李金玲. 基于预训练神经网络的遥感图像场景分类方法研究[D]. 武汉: 华中科技大学, 2019.
Li J L. Remote sensing image scene classification based on pre-trained CNN[D]. Wuhan, China: Huazhong University of Science and Technology, 2019. (in Chinese)
- [13] 黄鸿, 徐科杰, 石光耀. 联合多尺度多特征的高分遥感图像场景分类[J]. *电子学报*, 2020, 48(9): 1824 – 1833.
Huang H, Xu K J, Shi G Y. Scene classification of high-resolution remote sensing image by multi-scale and multi-feature fusion[J]. *Acta Electronica Sinica*, 2020, 48(9): 1824 – 1833. (in Chinese)
- [14] Goldblatt R, Stuhlmacher M, Tellman B, et al. Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover[J]. *Remote Sensing of Environment*, 2018, 205: 253 – 275.
- [15] Du S, Zhang F, Zhang X. Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach[J]. *Isprs Journal of Photogrammetry and Remote Sensing*, 2015, 105: 107 – 119.
- [16] Yuan Y, Fang J, Lu X, et al. Remote sensing image scene classification using rearranged local features[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(3): 1779 – 1792.
- [17] Liu Q, Hang R, Song H, et al. Learning multiscale deep features for high-resolution satellite image scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(1): 117 – 126.
- [18] Zhu Q, Zhong Y, Zhang L, et al. Scene classification based on the fully sparse semantic topic model[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(10): 5525 – 5538.
- [19] Yan L, Zhu R, Mo N, et al. Cross-domain distance metric learning framework with limited target samples for scene classification of aerial images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(6): 3840 – 3857.
- [20] Zhu Q, Zhong Y, Wu S, et al. Scene classification based on the sparse homogeneous – heterogeneous topic feature model[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(5): 2689 – 2703.
- [21] Zhu Q, Zhong Y, Zhang L, et al. Adaptive deep sparse semantic modeling framework for high spatial resolution image scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(10): 6180 – 6195.
- [22] Wang Y, Zhang L, Tong X, et al. LRAGE: Learning latent relationships with adaptive graph embedding for aerial scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(2): 621 – 634.
- [23] Zhong Z, Fan B, Ding K, et al. Efficient multiple feature fusion with hashing for hyperspectral imagery classification: A comparative study[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(8): 4461 – 4478.
- [24] Xu X, Li J, Huang X, et al. Multiple morphological component analysis based decomposition for remote sensing image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(5): 3083 – 3102.
- [25] Pham M T, Lefèvre S, Aptoula E. Local feature-based attribute profiles for optical remote sensing image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 56(2): 1199 – 1212.
- [26] Interdonato R, Ienco D, Gaetano R, et al. DuPLO: A Dual view point deep learning architecture for time series classification[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, 149: 91 – 104.
- [27] Gavish, Yoni, O'connell, et al. Comparing the perfor-

- mance of flat and hierarchical Habitat/Land-Cover classification models in a NATURA 2000 site[J]. *ISPRS Journal of Photogrammetry & Remote Sensing*, 2018, 136: 1 – 12.
- [28] Liu Y, Zhong Y, Qin Q. Scene classification based on multiscale convolutional neural network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(12): 7109 – 7121.
- [29] Kemker R, Kanan C. Self-taught feature learning for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(5): 2693 – 2705.
- [30] Shu Y, Tang H, Li J, et al. Object-based unsupervised classification of VHR panchromatic satellite images by combining the HDP and IBP on multiple scenes[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(11): 6148 – 6162.
- [31] Zheng X, Yuan Y, Lu X. A deep scene representation for aerial scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(7): 4799 – 4809.
- [32] Romero A, Gatta C, Campsvalls G. Unsupervised deep feature extraction for remote sensing image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(3): 1349 – 1362.
- [33] Zhong P, Gong Z, Li S, et al. Learning to diversify deep belief networks for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(6): 3516 – 3530.
- [34] Zhong Z, Li J, Luo Z, et al. Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2017, 56(2): 847 – 858.
- [35] Zhao W, Jiao L, Ma W, et al. Superpixel-based multiple local CNN for panchromatic and multispectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(7): 4141 – 4156.
- [36] Wang Q, Liu S, Chanussot J, et al. Scene classification with recurrent attention of VHR remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 57(2): 1155 – 1167.
- [37] Gong Z, Zhong P, Yu Y, et al. Diversity-promoting deep structural metric learning for remote sensing scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(1): 371 – 390.
- [38] Cheng G, Han J, Zhou P, 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(98): 119 – 132.
- [39] Wang W, Du L, Gao Y, et al. A discriminative learned CNN embedding for remote sensing image scene classification[J]. *arXiv: Computer Vision and Pattern Recognition*, 2019: arXiv:1911.12517.
- [40] 刘冰, 李瑞麟, 封举富. 深度度量学习综述[J]. *智能系统学报*, 2019, 14(6): 1064 – 1072.
- Liu B, Li R L, Feng J F. A brief introduction to deep metric learning[J]. *CAAI Transactions on Intelligent Systems*, 2019, 14(6): 1064 – 1072.(in Chinese)
- [41] 刘瑄, 池明旻. 基于多粒度特征蒸馏的遥感图像场景分类研究 [J/OL]. <https://doi.org/10.19678/j.issn.1000-3428.0056798>. [2020-12-07].
- Liu X, Chi M M. Remote sensing image scene classification based on multi-granular feature distillation deep convolution neural networks[J/OL]. <https://doi.org/10.19678/j.issn.1000-3428.0056798>. [2020-12-07]. (in Chinese)
- [42] 祁昆仑. 基于视觉特征的高分辨率光学遥感影像多任务分类研究[J]. *测绘学报*, 2017, 46(6): 802.
- Qi K L. Multi-task classification of high resolution optic remote sensing images based on visual features[J]. *Acta Geodaetica et Cartographica Sinica*, 2017, 46(6): 802. (in Chinese)
- [43] 张馨月. 基于 DCNN 的高分辨率遥感图像场景分类 [D]. 长春: 吉林大学, 2019.
- Zhang X Y. Scene classification of high-resolution remotely sensed images based on DCNN[D]. Changchun, China: Jilin University, 2019. (in Chinese)
- [44] 崔先亮, 陈立福, 邢学敏, 等. 基于频带特征融合的 GL-CNN 遥感图像场景分类[J]. *遥感技术与应用*, 2019, 34(4): 712 – 719.
- Cui X L, Chen L F, Xing X M, et al. Remote sensing image scene classification based on frequency band feature fusion and GL-CNN[J]. *Remote Sensing Technology and Application*, 2019, 34(4): 712 – 719.(in Chinese)
- [45] Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification[A]. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*[C]. New York, USA: ACM, 2010. 270 – 279.
- [46] Xia G S, Hu J, 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.
- [47] Cheng G, Han J, Lu X. Remote sensing image scene clas-

- sification: Benchmark and state of the art[J]. Proceedings of the IEEE, 2017, 105(10): 1865 – 1883.
- [48] He T, Zhang Z, Zhang H, et al. Bag of tricks for image classification with convolutional neural networks[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Long Beach, USA: IEEE, 2019. 558 – 567.
- [49] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[A]. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics[C]. Sardinia, Italy: PMLR, 2010. 249 – 256.
- [50] Goyal P, Dollár P, Girshick R, et al. Accurate, large mini-batch sgd: Training imagenet in 1 hour[J]. arXiv Preprint, 2017, arXiv: 1706.02677.
- [51] Jia X, Song S, He W, et al. Highly scalable deep learning training system with mixed-precision: Training imagenet in four minutes[J]. arXiv Preprint, 2018, arXiv: 1807.11205.
- [52] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Las Vegas, USA: IEEE, 2016. 2818 – 2826.
- [53] Zhong Z, Zheng L, Kang G, et al. Random Erasing Data Augmentation[J]. Computer Vision and Pattern Recognition, 2020, arXiv:1708.04896.
- [54] DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[J]. arXiv Preprint, 2017, arXiv:1708.04552.
- [55] Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts[J]. arXiv Preprint, 2016, arXiv: 1608.03983.
- [56] Xiong R, Yang Y, He D, et al. On layer normalization in the transformer architecture[J]. arXiv Preprint, 2020, arXiv:2002.04745.
- [57] Gotmare A, Keskar N S, Xiong C, et al. A closer look at deep learning heuristics: Learning rate restarts, warmup and distillation[J]. arXiv Preprint, 2018, arXiv:1810.13243.
- [58] You Y, Gitman I, Ginsburg B. Large batch training of convolutional networks[J]. arXiv Preprint, 2017, arXiv: 1708.03888.
- [59] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. as Vegas, USA: IEEE, 2016. 770 – 778.
- [60] Müller R, Kornblith S, Hinton G E. When does label smoothing help?[A]. Advances in Neural Information Processing Systems[C]. Vancouver, Canada: DBLP, 2019. 4694 – 4703.
- [61] Liang D, Yang F, Zhang T, et al. Understanding mixup training methods[J]. IEEE Access, 2018, 6: 58774 – 58783.
- [62] Zhang H, Cisse M, Dauphin Y N, et al. Mixup: Beyond empirical risk minimization[J]. arXiv Preprint, 2017, arXiv:1710.09412.
- [63] Micikevicius P, Narang S, Alben J, et al. Mixed precision training[J]. arXiv Preprint, 2017, arXiv:1710.03740.

作者简介



郑海颖 男,1994年8月出生,山东潍坊人.华北水利水电大学信息工程学院硕士研究生,主要研究方向为遥感图像场景分类、目标检测.

E-mail:zhy20130901@163.com



王峰 男,1970年11月生,河南汤阴人.华北水利水电大学信息工程学院副教授,硕士生导师,主要研究方向为软件工程技术、数据库技术、图形图像处理、机器学习等.

E-mail:wangfeng@newu.edu.cn



姜维(通讯作者) 男,1981年12月生,河南郑州人.现为华北水利水电大学副教授,主要研究方向为场景文字检测与识别、遥感目标的检测与识别.

E-mail:jiangwei@newu.edu.cn