



计算机工程
Computer Engineering
ISSN 1000-3428, CN 31-1289/TP

《计算机工程》网络首发论文

题目: 联合集成学习与 EfficientNet 的光学遥感图像场景分类
作者: 宝音图, 刘伟, 牛朝阳, 李润生, 张浩波
DOI: 10.19678/j.issn.1000-3428.0059128
网络首发日期: 2020-10-23
引用格式: 宝音图, 刘伟, 牛朝阳, 李润生, 张浩波. 联合集成学习与 EfficientNet 的光学遥感图像场景分类. 计算机工程.
<https://doi.org/10.19678/j.issn.1000-3428.0059128>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。



联合集成学习与 EfficientNet 的光学遥感图像场景分类

宝音图, 刘伟, 牛朝阳, 李润生, 张浩波

(信息工程大学数据与目标工程学院, 河南 郑州 450001)

摘 要：深度学习提高了光学遥感图像场景分类的准确度和效率, 但光学遥感图像语义丰富, 部分场景仍存在易误分类的情况。同时由网络模型规模扩大带来的硬件、时间成本消耗过大问题, 制约着研究成果的普及应用。针对这些问题, 提出基于轻量化网络模型的光学遥感图像场景分类方法, 通过 EfficientNet 网络提取图像特征, 然后对图像特征复合提取生成语义信息更丰富的新特征, 再利用多个子分类器构建集成学习模块解析新特征得到预分类结果, 最后集成加权预分类结果以获得最终的分类结果。通过在 AID 和 NWPU-RESISC45 数据集上进行实验, 即使只训练 20% 的数据样本, 该方法达到了 94.32% 和 93.36% 的准确率, 对易误分类场景有更好的分类效果, 且参数量和浮点运算量大幅减少。

关键词：光学遥感图像; 场景分类; 深度学习; 集成学习; EfficientNet



开放科学(资源服务)标志码(OSID)

Scene Classification of Optical Remote Sensing Image Based on Ensemble Learning and EfficientNet

Bao Yintu, Liu Wei, Niu Chaoyang, Li Runsheng, Zhang Haobo

(School of Data and Target Engineering, Information Engineering University, Zhengzhou, Henan 450001, China)

Abstract: Deep learning improved the efficiency of scene classification of optical remote sensing images, but some scenes are easily misclassified for the rich semantic. At the same time, due to the expansion of the network model, the cost of hardware and time is too large, which restricted the application of research results. In this paper, the application of lightweight network in scene classification of optical remote sensing image is investigated. Firstly, image features are extracted based on EfficientNet. Secondly, new features with richer information are generated from image features. Then, multiple sub classifiers are used to construct an ensemble learning module to analyze the new features and get the pre-classification results. Finally, the pre-classification results are weighted to obtain the final classification results. The method achieves higher accuracy on AID and NWPU-RESISC45 datasets, and has a better effect for misclassification scenarios, and greatly reduces the number of parameters and FLOPS.

Key words: optical remote sensing image; scene classification; deep learning; ensemble learning; EfficientNet

DOI:10.19678/j.issn.1000-3428.0059128

基金项目：国家自然科学基金 (41901378)

作者简介：宝音图, 1988 年, 男, 硕士研究生, 主研方向: 智能信息处理; 刘伟, 博士/副教授; 牛朝阳, 博士/副教授; 李润生, 博士/讲师; 张浩波, 硕士研究生。E-mail: bao258456@163.com



0 概述

随着对地观测技术的进步,光学遥感图像的数量急剧增加,如何充分利用不断增多的光学遥感图像进行智能地球观测成为研究的重点,科学高效地解析海量光学遥感图像变得非常重要。场景分类将不同的场景图像通过给定的语义类别进行分类标识,是一种有效解析光学遥感图像的技术。光学遥感图像场景分类技术在城市规划、灾害检测、环境监测等领域有着广泛的应用^[1]。

相较于光学自然图像分类任务,光学遥感图像场景分类任务由于图像特征类型复杂、分布范围广,有效地提取语义特征、理解语义内容难度更大^[1]。根据不同的应用需求,已经出现了很多提取光学遥感图像特征的方法,主要分为三类:一是初级特征提取方法,其特点是针对光学遥感图像特点在局部或全局设计人类工程特征^{[1][2]}。初级特征提取的方法主要包括尺度不变特征变换(Scale-invariant Feature Transform, SIFT)^[3]、局部二进制模式(Local Binary Patterns, LBP)^[4]和定向梯度直方图(Histogram of Oriented Gradient, HOG)^[5]等,这些方法只能针对特定的目标,缺乏灵活性,泛化能力较弱,且只有局部特性,不能直接表示整幅图像^[6];二是中级特征提取方法,其特点是先提取图像的局部特征,然后对多个局部特征进行编码以获得遥感图像的中级特征^[2]。中级特征提取的方法主要包括视觉词袋模型(Bag of Visual Word, BoVW)^[7]和稀疏编码 Sparse Coding^[8]等,由于其简单有效得到广泛应用。但是因为表达能力的限制,其分类精度不高;三是高级特征提取方法,基于深度学习提取光学遥感图像特征。与初级和中级方法不同,深度学习模型通过架构神经网络来学习更强大、抽象和深层次的特征,这极大地推动了深度学习在光学遥感图像中的应用。大多数深度学习模型,尤其是卷积神经网络(Convolution Neural Network, CNN),包括 AlexNet^[9]、VGGNet^[10]、GoogLeNet^[11]、ResNet^[12]等,都能适用于光学遥感图像场景分类,并取得了较好分类效果。

虽然基于深度学习的方法已大大提高了分类准确性,但是光学遥感图像除了有颜色、纹理、空间信息等浅层次的信息^[13],还有很多深层次的信息,具体包括:遥感图像同类场景的多样性;遥感图像不同类场景的相似性;遥感图像场景的尺寸、方向变化较多。这些深层次的信息无疑增加了场景分类的难度^[14],更难以实现精准分类^[2]。

深度学习网络模型的深度和宽度对光学遥感图像分类精度有至关重要的影响^{[14][15][16]}。然而随着网络规模的扩大,容易发生梯度消失或梯度爆炸,还会引起网络退化问题,致使训练误差增大^[17]。同时,网络规模的扩大也会带来权重参数的大量冗余,导致训练周期过长、硬件要求过高,不适合末端应用部署,难以大范围普及^[18]。随着深度学习技术进一步发展,MobileNets^[19]等注重于易训练、易部署的轻量化网络模型逐渐成为研究重点,但是这些网络模型通常是针对自然图像提出的,在光学遥感图像领域应用还需有针对性地调整网络结构。

本文提出一种基于轻量化网络模型的光学遥感图像场景分类方法。该方法通过 EfficientNet^[18]网络模型提取图像特征,针对光学遥感图像场景易误分类的情况,构建集成学习模块来提升网络模型的分类性能。

1 原理与方法

1.1 EfficientNet 网络模型

为使网络模型获得更好的精度,通常采取优化网络深度、网络宽度和图像分辨率的方式,如 ResNet、DenseNet 等,但是这些网络模型往往只改变三个维度之一(网络深度、网络宽度和图像分辨率),且需要繁琐的手动调整参数,并且仍然经常会产生次优的精度和效率。Google Brain Team 通过复合比较网络深度、网络宽度、图像分辨率对精度、参数量和浮点运算量的影响,提出了 EfficientNet^[18]网络模型。

EfficientNet 网络模型将网络深度、网络宽度、图像分辨率的不同组合变成优化问题,可由式表示,



$$\begin{aligned} & \max_{d,w,r} \text{Accuracy}(N(d,w,r)) \\ & s.t. \quad N(d,w,r) = \bigodot_{i=1 \sim s} \hat{F}_i^{d \cdot \hat{L}_i} (X_{\langle r \cdot \hat{H}_i, r \cdot \hat{W}_i, w \cdot \hat{C}_i \rangle}) \quad \backslash * \text{MERGEFORMAT (1)} \\ & \quad \text{Memory}(N) \leq \text{target_memory} \\ & \quad \text{FLOPS}(N) \leq \text{target_flops} \end{aligned}$$

式中, N 为网络模型, d, w, r 是缩放网络的深度、宽度和图像分辨率的系数, i 是网络模型的组件, s 是组件的总个数, \hat{F}_i 是预设的网络层结构, \hat{L}_i 是预设的层数, \hat{H}_i 和 \hat{W}_i 是预设的分辨率, \hat{C}_i 是预设的通道数, X 为将参数按缩放系数调整, $\text{Memory}(N)$ 为网络的参数量, $\text{FLOPS}(N)$ 为网络的浮点运算量, \bigodot 为将每个组件设置为调整后的参数并搭建网络模型, target_memory 为参数量的阈值, target_flops 为浮点运算量的阈值, $\max \text{Accuracy}$ 为求网络模型的最大精度。

平衡网络的深度、宽度和分辨率三个维度之间的联系可以获得更好的准确率和效率, EfficientNet 使用复合系数 ϕ 对三个维度进行缩放, 可由式表示,

$$\begin{aligned} & d = \alpha^\phi, w = \beta^\phi, r = \gamma^\phi \\ & s.t. \quad \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad \backslash * \text{MERGEFORMAT (2)} \\ & \quad \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{aligned}$$

式中, d, w, r 是三个维度的系数, α, β, γ 是对应维度的资源分配参数, ϕ 是资源控制系数, 根据模型可利用资源进行调整。在满足约束条件的情况下, 通过神经架构搜索 (Neural Architecture Search, NAS) 对各参数进行优化调整。在参数量和运算量不增加的基础上, EfficientNet 模型通过不断调整网络深度、网络宽度和图像分辨率的系数达到最优的精度。

根据不同的图像分辨率, EfficientNet 网络构建了 B0-B7 共 8 个模型。其中 B7 模型在自然图像 ImageNet 数据集上 Top-1 和 Top-5 的精度分别为 84.4% 的和 97.1%, 已达到当时最优精度。并且在达到同等精度的网络中, B0-B7 模型的参数量和运算量都是最少的, 具体性能对比如表 1 所示

表 1 在 ImageNet 数据集上各模型性能对比

模型(图像分辨率)	Top-1 精度	Top-5 精度	参数量	浮点运算量
EfficientNet-B0(224*224pi)	77.30%	93.50%	5.3M	0.4B
ResNet-50	76.00%	93.00%	26.1M	4.1B
DenseNet-169	76.20%	93.20%	14.2M	3.5B
EfficientNet-B1(240*240pi)	79.20%	94.50%	7.8M	0.7B
ResNet-152	77.80%	93.80%	60.1M	11.0B
DenseNet-264	77.90%	93.90%	34.1M	6.0B
Inception-v3	78.80%	94.40%	24.2M	5.7B
EfficientNet-B2(260*260pi)	80.30%	95.00%	9.2M	1.0B
Inception-v4	80.00%	95.00%	48.1M	13.1B
Inception-resnet-v2	80.10%	95.10%	54.9M	13.1B
EfficientNet-B3(300*300pi)	81.70%	95.60%	12.0M	1.8B
ResNeXt-101	80.90%	95.60%	84.2M	32.0B
EfficientNet-B4(380*380pi)	83.00%	96.30%	19.0M	4.2B
SENet	82.70%	96.20%	146.1M	42.0B
EfficientNet-B5(456*456pi)	83.70%	96.70%	30.0M	9.9B
AmoebaNet-C	83.50%	96.50%	155.1M	41.1B
EfficientNet-B6(528*528pi)	84.20%	96.80%	43.0M	19.1B
EfficientNet-B7(600*600pi)	84.40%	97.10%	66.2M	37.1B

1.2 联合集成学习的 EfficientNet 网络

EfficientNet 网络模型在自然场景图像上达到了较高精度, 但光学遥感图像场景包含着更丰富的语义信息, 场景的类内相异性、类间相似性以及场景的尺寸方向变换等情况影响了 EfficientNet 网络模型对光学遥感图像场景的分类精度。为解决上述问题, 本文通过 EfficientNet 网络模型提取图像特征, 利用多个子分类器构建集成学习模块提升分类效果, 使用 Dropout 算法防止发生过拟合, 采用交叉熵损失函数辅助网络训练, 提高了网络模型对光学遥感图像场景的分类精度。



1.2.1 网络模型结构

集成学习通过构建多个子分类器可以提升分类效果^[6]。本文方法基于 EfficientNet 网络提取图像特征（一级特征），将所有的一级特征按照设定比率随机选取，而后进行融合产生新的特征（二级特征），特征提取比率过低会因为选取特征不够充分影响精度，特征提取比率过高会极大增加模型的参数量，为在提升精度的同时不过多地增加模型的参数量，本文将特征提取比率设定为 60%，给每个二级特征分配一个子分类器进行预分类，集成加权所有子分类器的预分类结果，产生最终的分类结果。因为每个子分类器选取的一级特征不完全一样，能够更充分地解析图像的语义信息，实现子分类器之间的优势互补，从而获得比单一分类器更好的分类效果。网络模型结构如图 1 所示。

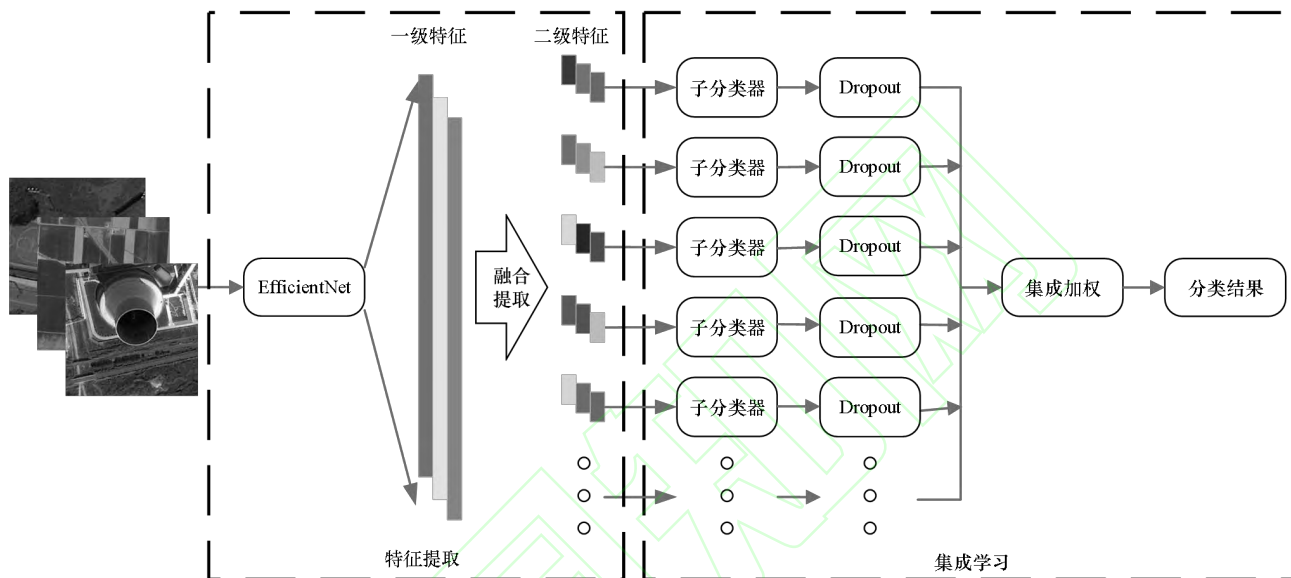


图 1 本文网络模型结构图

Fig.1 The structure of network model in this paper

1.2.2 Dropout 算法

在深度学习中，当参数多而训练样本又比较少时，网络模型容易产生过拟合现象。Dropout 算法能够缓解过拟合，起到正则化效果^[20]。Dropout 算法的表达式如式所示，

$$\begin{aligned}\hat{x}^{(l)} &= r^{(l)} * x^{(l)} \\ y^{(l+1)} &= w^{(l+1)} * \hat{x}^{(l)} + b^{(l)} \\ &\backslash * \text{MERGEFORMAT (3)}\end{aligned}$$

式中， r 为以给定概率产生的 0、1 序列， x 为对应层的神经元， \hat{x} 为经过 Dropout 算法保留下的神经元， w 为权重， b 为偏置， y 为输出的神经元， l 为神经元所在的层数。

在训练过程中，每个神经元以一定概率保留，每次向前传播保留下来的神经元都不同，这样可以使得模型不依赖于某些局部特征，泛化能力更强^[20]，Dropout 算法基本思想如图 2 所示。因光学遥感图像数据集的训练样本偏少，为了防止过拟合发生，本文在每个子分类器中都添加了 Dropout 算法。

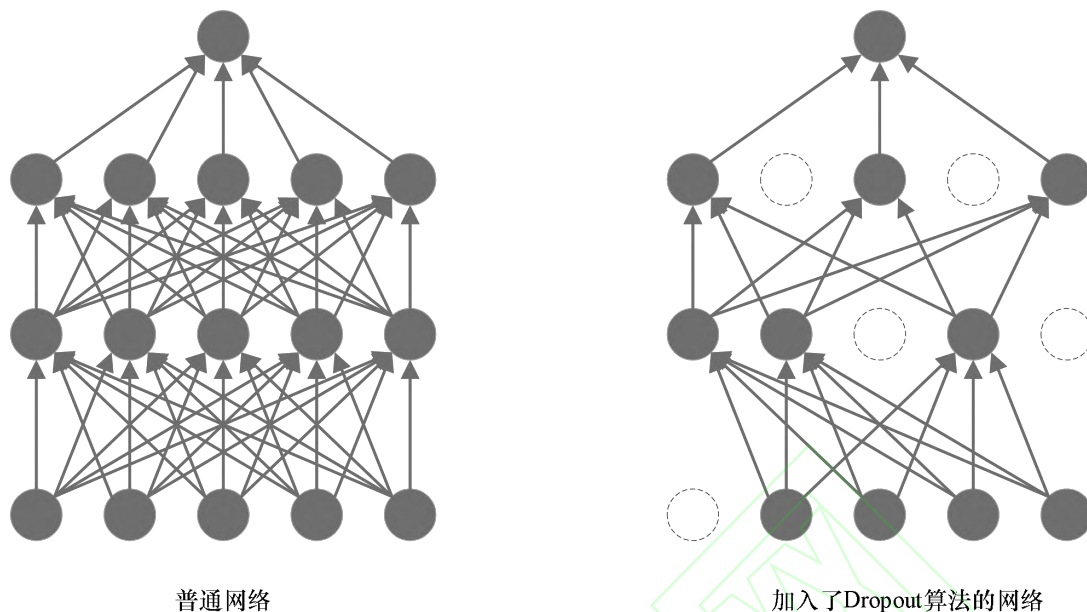


图 2 Dropout 算法
Fig.2 Dropout algorithm

1.2.3 交叉熵损失函数

深度学习中用损失函数来度量网络模型得到的预测标签和数据真实标签之间的差距，可以衡量网络模型的泛化能力好坏^[21]。根据不同的应用需求，选择合适的损失函数可以使模型得到更好的训练效果。本文属于多场景分类问题，采用更适合多分类的交叉熵损失函数^[22]。交叉熵损失函数可由式表示，

$$L(y, \hat{y}) = - \sum_{i=1}^N y^{(i)} \log_{10} \hat{y}^{(i)} + (1 - y^{(i)}) \log_{10} (1 - \hat{y}^{(i)})$$

* MERGEFORMAT (4)

式中， y 为样本真实标签， \hat{y} 为样本通过网络模型得到的预测标签， N 为样本数。通过交叉熵损失函数，可以得到样本预测标签和真实标签之间的损失值，损失值越小预测标签越接近真实标签，损失值越大则预测标签和真实标签差距越大。这样通过损失值就可以判断预测结果的好坏，从而更有效地训练网络模型。

2 实验及结果分析

2.1 实验数据集

为验证本文网络模型的有效性，实验选用两个公开的大规模遥感图像数据集 AID 数据集^[23]和 NWPU-RESISC45 数据集^[24]。

AID 数据集是由华中科技大学和武汉大学联合创建的大规模遥感图像数据集，该数据集总共有 10,000 张图像，分为机场、河流、草地等 30 个场景类别，不同场景类型的图像数量从 220~420 张不等，每张图片尺寸为 600×600 像素，图像空间分辨率为 0.5~8 米，该数据集从 Google Earth 影像中收集，图像选取自不同国家、地区、时间和季节，以及不同的成像条件^[23]，如图 3 所示。

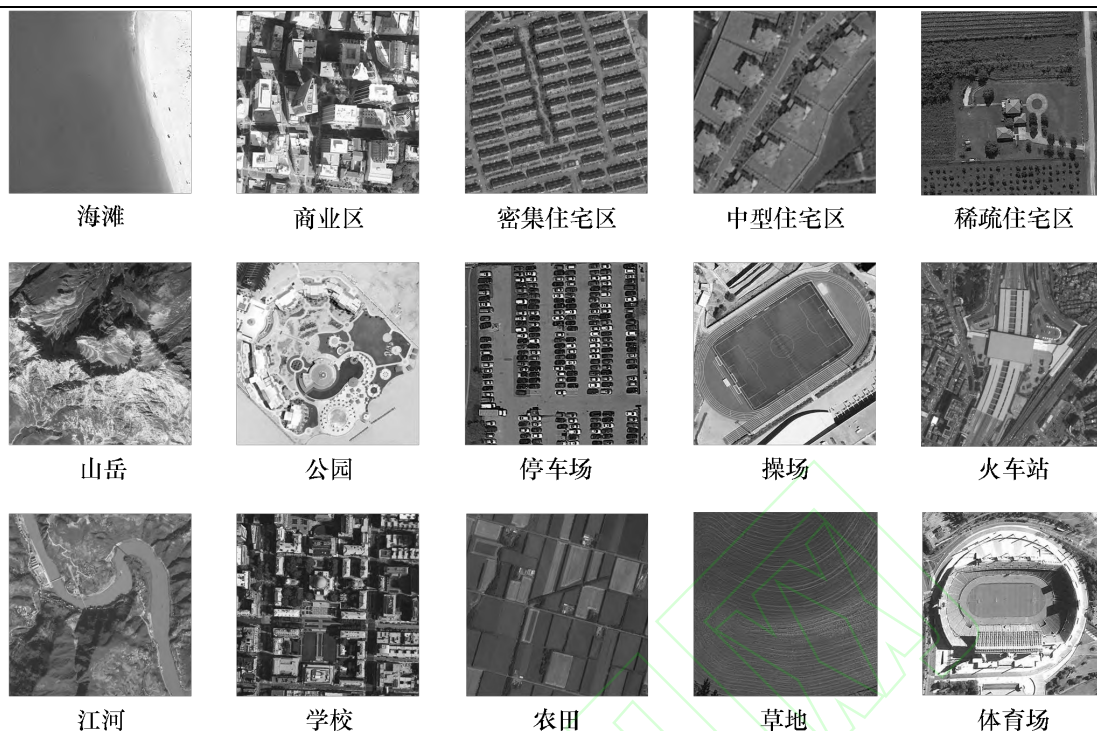


图 3 AID 数据集部分场景示例

Fig.3 Example images of AID dataset

NWPU-RESISC45 数据集是由西北工业大学创建的大规模遥感图像数据集, 该数据集比 AID 数据集更为复杂, 共有 31,500 张遥感图像, 区分飞机、湖、云等 45 个场景类别, 每个类别有 700 张图像, 每张图像尺寸为 256×256 像素, 图像的空间分辨率为 0.2~30 米, 该数据集从 Google Earth 中提取, 选自全球 100 多个国家和地区, 包含不同天气、季节的图像, 为了增加难度, 选取了含有遮挡、重叠等复杂因素的遥感图像, 场景更加复杂, 分类的难度更大^[24], 如图 4 所示。

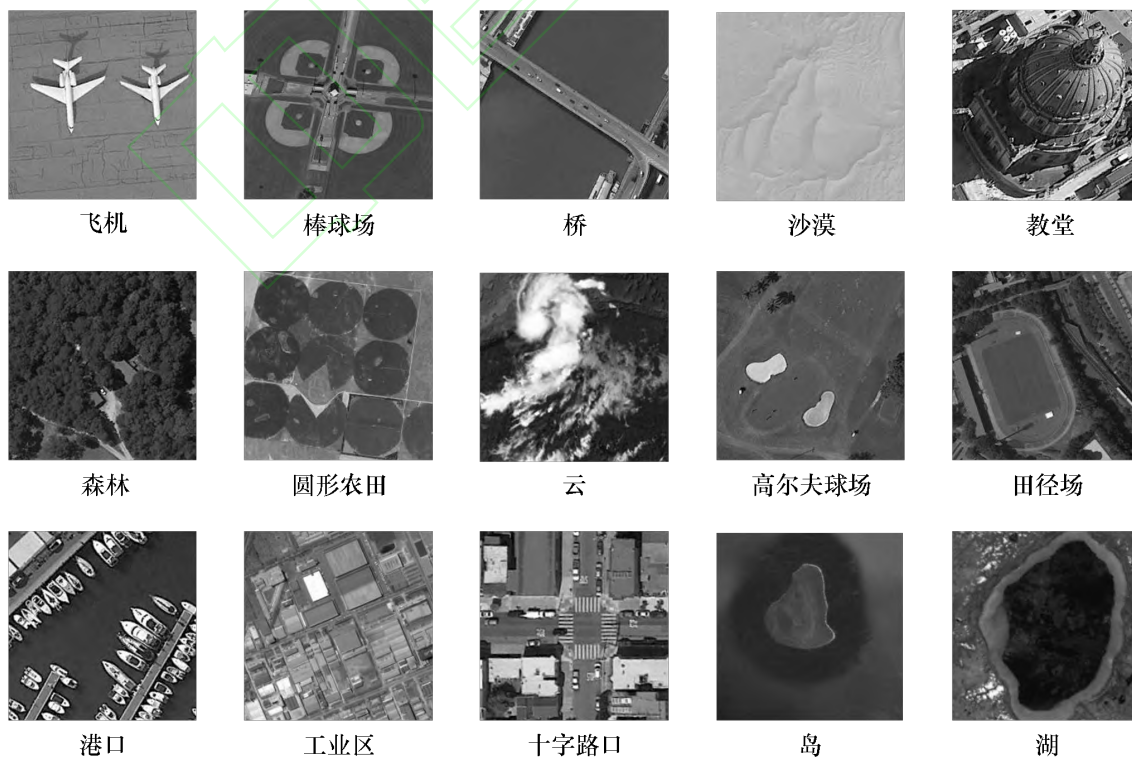


图 4 NWPU-RESISC45 数据集部分场景示例

Fig.4 Example images of NWPU-RESISC45 dataset



2.2 实验参数设置

实验环境为 PyTorch 1.5 框架, 操作系统为 Windows10 系统, CPU 为 12 核的 Intel Xeon Gold 6226, 主频为 2.70 GHz, 内存为 64GB DDR4, GPU 为 NVIDIA 公司的 TITAN XP, 显存为 12GB, CUDA 版本为 10.2, cudnn 版本为 7.6.5。本文实验选择 EfficientNet-B3 网络模型。

为了保证实验的公平性、合理性, 数据集的训练比率与文献[1][2]一致。对于 AID 数据集, 随机选取每类场景数据总量的 20%、50%作为训练数据, 余下的 80%、50%作为测试数据。对于 NWPU-RESISC45 数据集, 随机选取每类场景数据总量的 10%、20%作为训练数据, 余下的 90%、80%作为测试数据。

迁移学习不仅能解决过拟合问题, 同时也能缩减模型训练的时间^[15], 本文实验使用了在 ImageNet 数据集上训练好的权重进行迁移训练。为防止训练期间发生梯度消失或梯度爆炸, 本次实验分两个阶段进行训练, 第一个阶段为预训练阶段, 冻结除分类层以外的网络层的权重, 损失函数使用交叉熵损失函数, 优化器使用随机梯度下降法 (Stochastic Gradient Descent, SGD), 初始学习率为 $1e-1$, 权重衰减为 $5e-3$, 动量为 0.9, 每 2 个 epoch 学习率下降一次, 下降系数为 0.9, 共训练 10 个 epoch。第二个阶段在预训练的基础上继续训练, 解冻所有网络层的权重, 损失函数使用交叉熵损失函数, 优化器使用 SGD, 初始学习率为 $1e-3$, 权重衰减为 $5e-3$, 动量为 0.9, 每 5 个 epoch 学习率下降一次, 下降系数为 0.95, 共训练 200 个 epoch。为了增强网络模型的训练效果, 实验中使用了 PyTorch 框架下的 transforms 库中的水平翻转、垂直翻转、颜色亮度变换、角度旋转等方式进行数据增强, 如图 5 所示。为了确保实验的准确性, 每种测试都重复了 10 次。



图 5 数据增强示例

Fig.5 Examples of data enhancement

图 6 为本文模型训练过程中随循环次数变化的损失值和准确率曲线, 从图 6 可以看出, 经过 200 次循环, 损失值和准确率逐渐收敛, 说明网络模型较为稳定。图 6 (a) 中 180 次循环后损失值不再下降, 保持在 0.1 以下, 图 6 (b) 中准确率在 180 次循环后不再增加, 最终达到 97.52%。

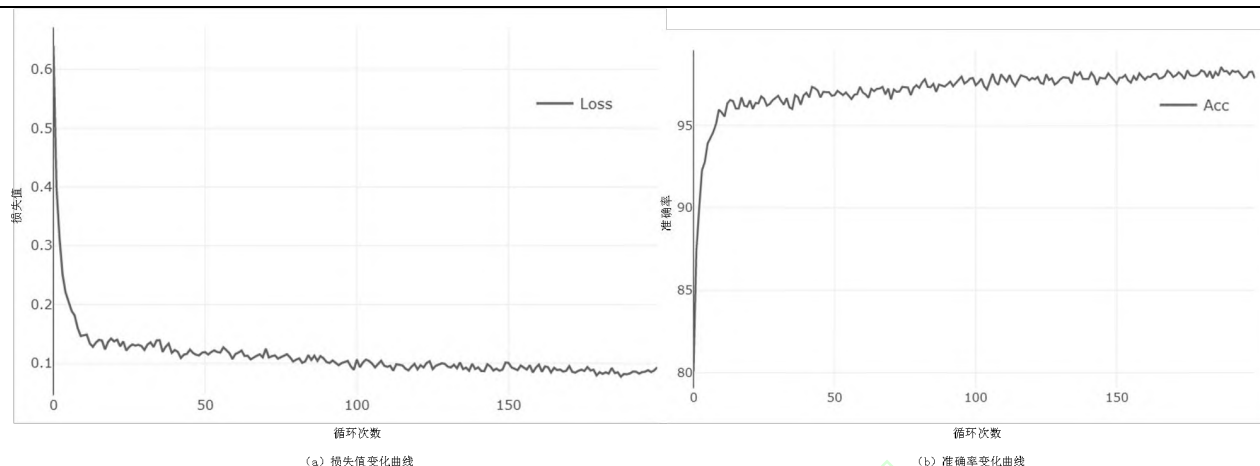


图 6 模型训练过程

Fig.6 Training process of model in this paper

2.3 AID 数据集场景分类

将 AID 数据集分为训练集和测试集,用本文网络模型训练测试,统计测试结果得到混淆矩阵,其中 20% 训练比率的混淆矩阵如图 7 所示,对角线元素为各场景分类的准确率。可以看到 30 类场景中有 25 类场景的准确率达到 90% 以上,其中海滩、森林和高架桥等场景的准确率已接近 100%,但是度假胜地和中心区两个场景误分类较多,准确率不足 80%,度假胜地场景的准确率最低为 75%。

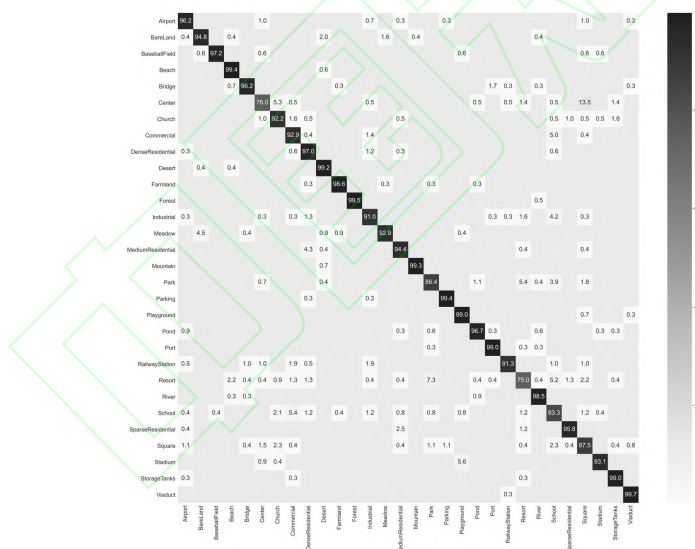


图 7 AID 数据集训练比率为 20% 时的混淆矩阵

Fig.7 Confusion matrices of the AID dataset under the training ratio of 20%

为了比较模型性能,选取了现阶段光学遥感图像场景分类中几种典型网络模型进行比对。GoogLeNet 是 2014 年提出的一种 22 层的网络模型;D-CNNs^[25]是 2018 年提出的网络模型,通过引入度量学习正则化来提升分类能力;CNN-CapsNet^[2]是 2019 年提出的网络模型,通过胶囊网络捕获空间信息来增强特征提取能力。表 2 为各网络模型在 AID 数据集上的性能对比,可以看到本文方法取得了现阶段较好的精度。与 D-CNNs 和 CNN-CapsNet 网络模型相比,在达到近似精度的情况下,本文方法所需的参数量和浮点运算量成倍数的大幅减少。与 EfficientNet 网络模型相比,分类准确率有 3~4% 的提升,参数量仅增加了 2M,浮点运算量没变,说明引入集成学习模块后的网络模型仅需较少的硬件消耗,在不影响运算时间的情况下带来了明显的分类性能提升。



表 2 各模型在 AID 数据集上的性能对比

Tab.2 Comparison of performance for different models on AID dataset				
模型	20%训练比率准确度	50%训练比率准确度	参数量	浮点运算量
GoogLeNet ^[2]	83.44±0.40	86.39±0.55	7M	3.2B
D-CNN ^s ^[25]	90.82±0.16	96.89±0.10	138M	31.1B
CNN-CapsNet ^[2]	93.79±0.13	96.32±0.12	75M	6.3B
EfficientNet	89.34±0.15	91.52±0.13	11M	2.1B
本文方法	94.32±0.12	96.46±0.07	13M	2.1B

2.4 NWPU-RESISC45 数据集场景分类

将 NWPU-RESISC45 数据集分为训练集和测试集,用本文网络模型训练测试,统计测试结果得到混淆矩阵,其中 20%训练比率的分类结果如图 8 所示,对角线元素为各场景分类的准确率。通过混淆矩阵可以看到 45 种场景中有 39 种场景的准确率在 90%以上,其中灌木丛场景的准确率已接近 100%,不过教堂、中型住宅和宫殿三个场景的准确率不足 80%,宫殿场景的准确率最低为 75.2%。

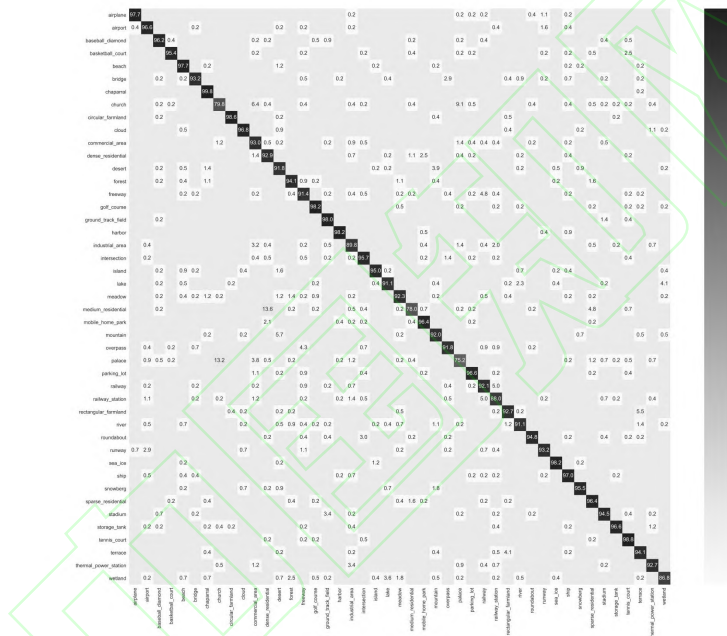


图 8 NWPU-RESISC45 数据集训练比率为 20%时的混淆矩阵

Fig.8 Confusion matrices of the NWPU-RESISC45 dataset under the training ratio of 20%

表 3 为本文方法与其他网络模型在 NWPU-RESISC45 数据集上的性能对比,可以看到本文方法在 NWPU-RESISC45 数据集上同样取得了现阶段较好的精度,与在 AID 数据集上的结果类似,本文方法不仅精度更高,参数量和浮点运算量都是最少的。在两个数据集上都取得较好结果,说明本文方法较为稳定,鲁棒性较强。

表 3 各网络模型在 NWPU-RESISC45 数据集上的性能对比

Tab.3 Comparison of performance for different models on NWPU-RESISC45 dataset				
模型	10%训练比率准确度	20%训练比率准确度	参数量	浮点运算量
GoogLeNet	76.19±0.38	86.39±0.55	7M	3.2B
D-CNNs	89.22±0.50	91.89±0.22	138M	31.1B
CNN-CapsNet	89.03±0.21	92.60±0.11	75M	6.3B
EfficientNet	88.42±0.11	90.02±0.12	11M	2.1B
本文方法	91.56±0.13	93.36±0.06	13M	2.1B

为了更直观地比较网络模型对易误分类场景的辨识性能,选取准确率达不到 90%的场景,绘制了各场



景在不同网络中准确率的对比图,如图9所示。可以看到本文方法在大多数易误分类场景中表现了最好的分类效果,说明本文方法在易误分类场景中也有优势。

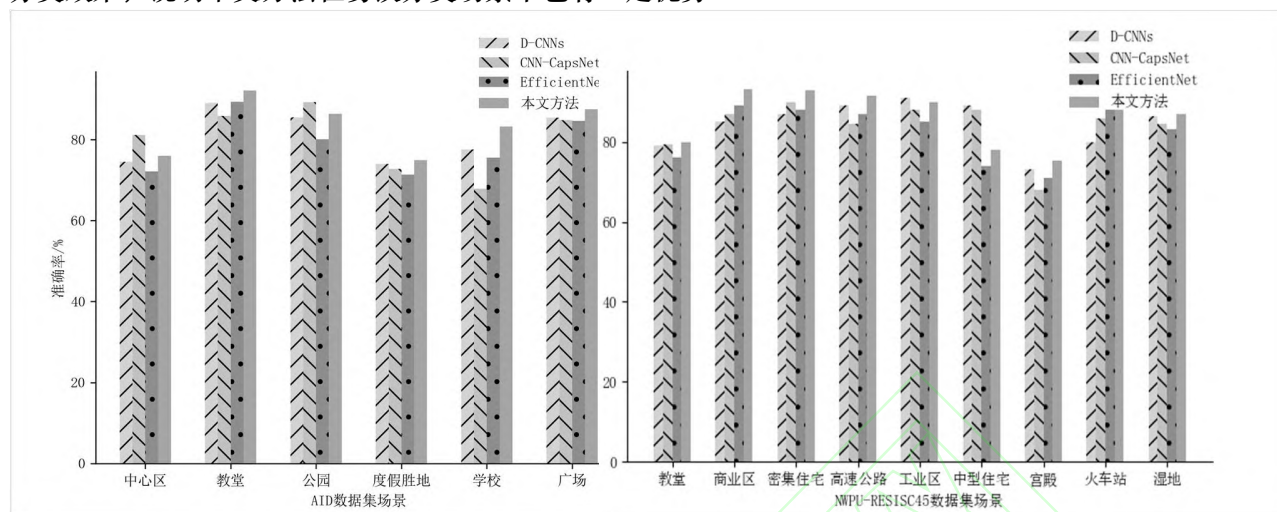


图9 各网络模型的易误分类场景性能对比

Fig.9 Comparison of performance for different models on easily misclassified scene images

3 结束语

本文研究了深度学习在光学遥感图像场景分类中的应用,联合集成学习和轻量化网络模型的优势,提出一种对硬件要求低、运算时间短的光学遥感图像场景分类方法。本文采用 EfficientNet 网络模型来减少参数量和浮点运算量,通过集成学习模块充分解析图像特征的语义信息提升类性能。经在两个光学遥感图像数据集上进行实验比对,本文提出的网络模型获得了较高的准确率,尤其是在大多数易误分类场景中达到了最好分类效果。对比其他网络模型而言,本文方法所需的参数量和浮点运算量大幅减少,对硬件要求低、运算时间短,更适合末端部署,便于普及推广。然而,本文方法在个别场景上的分类精度仍有待进一步提高,以及如何在末端部署实现本文方法是今后研究重点。

参考文献:

- [1] Cheng G, Xie X, Han J, et al. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, PP(99): 1-1.
- [2] Zhang W, Tang P, Zhao L. Remote Sensing Image Scene Classification Using CNN-CapsNet[J]. Remote Sensing, 2019, 11(5): 494.
- [3] D. G. Lowe. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, Vol.60 (2), pp.91-110.
- [4] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(7): 971-987.
- [5] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, CA, USA: IEEE, 2005 vol.1: 886-893.
- [6] YU D H, ZHANG B M, ZHAO C, et al. Scene classification of remote sensing image using ensemble convolutional neural network. 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(06): 717-727.
- [7] Li F. A Bayesian hierarchical model for learning natural scene categories[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, CA, USA: IEEE, 2005 vol.2: 523-531.
- [8] 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.

- [9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [10] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]// 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'14). Columbus, Ohio, USA: IEEE, 2014: 1-1.
- [11] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[C]// 2015 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'15). Boston, USA: IEEE, 2015: 1-9.
- [12] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. New York, USA: IEEE Computer Society, 2016: 770-778.
- [13] ZHANG Yu-Jin. Image Engineering(II) Image Analysis(Fourth Edition)[M]. Beijing: Tsinghua University Press, 2018. 章毓晋. 图像工程(中册)图像分析(第四版)[M]. 北京: 清华大学出版社, 2018.
- [14] LIANG Wentao, KANG Yan, LI Hao, Adaptive Remote Sensing Scene Classification Based on Complexity Clustering[J/OL]. Computer Engineering: 1-12[2020-07-16]. 梁文韬, 康雁, 李浩, 等. 基于复杂度聚类的自适应遥感场景分类[J/OL]. 计算机工程: 1-12[2020-07-16]. <https://doi.org/10.19678/j.issn.1000-3428.0056209>.
- [15] QIAO Tingting, LI Luqun. Scene classification of high — resolution remote sensing image combining data augmentation and transfer learning[J]. Bulletin of Surveying and Mapping, 2020(02):37-42. 乔婷婷, 李鲁群. 结合数据增广和迁移学习的高分辨率遥感影像场景分类[J]. 测绘通报, 2020(02): 37-42.
- [16] ZHANG Chen, XIA Kai, YANG Yinhui, et al. Remote sensing scene classification algorithm based on improved residual network[J/OL]. Science of Surveying and Mapping: 1-8[2020-07-16]. <http://kns.cnki.net/kcms/detail/11.4415.P.20200616.0843.004.html>. 章晨, 夏凯, 杨垠晖, 等. 改进残差网络的遥感图像场景分类[J/OL]. 测绘科学: 1-8[2020-07-16]. <http://kns.cnki.net/kcms/detail/11.4415.P.20200616.0843.004.html>.
- [17] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//2016 Computer Vision and Pattern Recognition. Las Vegas, Nevada, USA: IEEE, 2016: 779-788.
- [18] Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[C]// International Conference on Machine Learning. Long Beach, USA: ICML, 2019: 6105-6114.
- [19] Howard, A G, Zhu, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[J/OL]. <https://arxiv.org/abs/1704.04861>, 2017.
- [20] DONG Hongyi. Object Detection by Deep Learning: Core Technologies and Practices[M]. Beijing: China Machine Press, 2020. 董洪义. 深度学习之 PyTorch 物体检测实战[M]. 北京: 机械工业出版社, 2020.
- [21] TANG Jinmin. Practical Computer Vision of Deep Learning by PyTorch[M]. Beijing: Publishing House of Electronics Industry, 2018. 唐进民. 深度学习之 PyTorch 实战计算机视觉[M]. 北京: 电子工业出版社, 2018.
- [22] Berman M, Triki A R, Blaschko M B. The Lov'asz-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks[C]// 2018 Computer Vision and Pattern Recognition. 2018. Salt Lake City, USA: IEEE, 2018: 4413-4421.
- [23] Xia G, Hu J, Hu F, et al. AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification[J]. IEEE Transactions on Geoenvironment and Remote Sensing, 2017, 55(7):3965-3981.
- [24] 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.
- [25] Cheng G, Yang C, Yao X, et al. When Deep Learning Meets Metric Learning: Remote Sensing Image Scene Classification via Learning Discriminative CNNs. IEEE Transactions on Geoenvironment and Remote Sensing, 2018, 56:2811-2821.