

JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象学报 图形学报

2021
12
VOL.26

ISSN1006-8961
CN11-3758/TB





第26卷第12期（总第308期）
2021年12月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

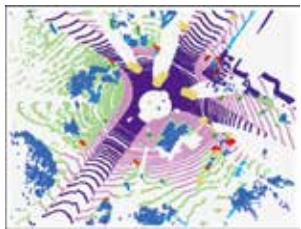
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



环视鱼眼图像处理深度学习
研究进展(第2778页)



结合空间深度卷积和残差
的大尺度点云场景分割
(第2848页)



曼哈顿世界环境下消失点非线
性优化估计算法(第2931页)

综述

深度对抗视觉生成综述

谭明奎, 许守恺, 张书海, 陈奇 2751

人机交互中的智能体情感计算研究

刘婷婷, 刘箴, 柴艳杰, 王瑾, 王媛怡 2767

环视鱼眼图像处理深度学习研究进展

包俊, 刘宏哲, 褚文博 2778

图像处理和编码

图像分解与色彩先验下的多曝光图像融合

李嫛源, 王琴, 朱智勤, 齐观秋 2800

双尺度分解和显著性分析相结合的红外与可见光图像融合

霍星, 邹韵, 陈影, 檀结庆 2813

轻量级注意力特征选择循环网络的超分辨率重建

徐雯捷, 宋慧慧, 袁晓彤, 刘青山 2826

区域级通道注意力融合高频损失的图像超分辨率重建

周波, 李成华, 陈伟 2836

图像分析和识别

结合空间深度卷积和残差的大尺度点云场景分割

刘盛, 黄圣跃, 程豪豪, 沈家瑜, 陈胜勇 2848

细粒度图像分类的互补注意力方法

赵勋, 王家宝, 李阳, 王亚鹏, 苗壮 2860

联合分块谱带优选和深度特征的高光谱人脸识别

谢志华, 李毅, 牛杰一 2870

时空双仿射微分不变量及骨架动作识别

李琪, 墨瀚林, 赵婧涵, 郝宏翔, 李华 2879

多尺度分割先验迭代重加权低秩恢复显著性检测

张荣国, 郑维佳, 赵建, 胡静, 刘小君 2892

融合环境特征与改进YOLOv4的安全帽佩戴检测

葛青青, 张智杰, 袁琰, 李秀梅, 孙军梅 2904

图像理解和计算机视觉

关节轴角先验对3维人体重建结果的影响

姚砾, 张幼安, 张梦雪, 万燕 2918

曼哈顿世界环境下消失点非线性优化估计算法

韩冰心, 卢惠民, 于清华, 张礼廉 2931

双目机器视觉及RetinaNet模型的路侧行人感知定位

连丽容, 罗文婷, 秦勇, 李林 2941

遥感图像处理

遥感图像中油罐目标精确定位与参数提取

江哈, 张月婷, 郭嘉逸, 赵鑫, 李芳芳, 黄丽佳, 胡玉新, 雷斌, 丁赤飏 2953

参数合成空间变换网络的遥感图像一致性配准

陈颖, 张祺, 李文举, 石艳娇, 陈磊 2964

改进Mask R-CNN模型的海洋锋检测

徐慧芳, 黄冬梅, 贺琪, 杜艳玲, 覃学标 2981

噪声鲁棒的轻量级深度遥感场景图像分类检索

王亚鹏, 李阳, 王家宝, 赵勋, 苗壮 2991

改进U-Net型网络的遥感图像道路提取

杨佳林, 郭学俊, 陈泽华 3005

弱监督深度语义分割网络的多源遥感影像水体检测

李鑫伟, 李彦胜, 张永军 3015

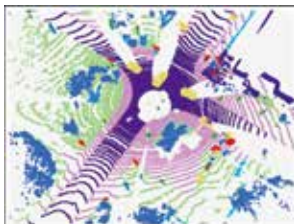
总目次 I

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Research progress of fisheye image processing based on deep learning(P2778)



A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation(P2848)



Vanishing point estimation based on non-linear optimization in manhattan world environments (P2931)

Review

- A review on deep adversarial visual generation
Tan Mingkui, Xu Shoukai, Zhang Shuhai, Chen Qi 2751
- Agent affective computing in human-computer interaction
Liu Tingting, Liu Zhen, Chai Yanjie, Wang Jin, Wang Yuanyi 2767
- Research progress of fisheye image processing based on deep learning
Bao Jun, Liu Hongzhe, Chu Wenbo 2778

Image Processing and Coding

- Multi-exposure image fusion based on image decomposition and color prior
Li Yuanyuan, Wang Qin, Zhu Zhiqin, Qi Guanqiu 2800
- Dual-scale decomposition and saliency analysis based infrared and visible image fusion
Huo Xing, Zou Yun, Chen Ying, Tan Jieqing 2813
- Lightweight attention feature selection recursive network for super-resolution
Xu Wenjie, Song Huihui, Yuan Xiaotong, Liu Qingshan 2826
- Region-level channel attention for single image super-resolution combining high frequency loss
Zhou Bo, Li Chenghua, Chen Wei 2836

Image Analysis and Recognition

- A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation
Liu Sheng, Huang Shengyue, Cheng Haohao, Shen Jiayu, Chen Shengyong 2848
- Complemented attention method for fine-grained image classification
Zhao Xun, Wang Jiabao, Li Yang, Wang Yapeng, Miao Zhuang 2860
- Hyperspectral face recognition based on partition bands optimal selection and deep features
Xie Zhihua, Li Yi, Niu Jieyi 2870
- Spatio-temporal dual affine differential invariants for skeleton-based action recognition
Li Qi, Mo Hanlin, Zhao Jinghan, Hao Hongxiang, Li Hua 2879
- Iterative reweighted low-rank recovery salient object detection with multiscale segmentation prior
Zhang Rongguo, Zheng Weijia, Zhao Jian, Hu Jing, Liu Xiaojun 2892
- Safety helmet wearing detection method of fusing environmental features and improved YOLOv4
Ge Qingqing, Zhang Zhijie, Yuan Long, Li Xiumei, Sun Junmei 2904

Image Understanding and Computer Vision

- The impact of joint axis angle prior on the results of 3D human body reconstruction
Yao Li, Zhang Youan, Zhang Mengxue, Wan Yan 2918
- Vanishing point estimation based on non-linear optimization in manhattan world environments
Han Bingxin, Lu Huimin, Yu Qinghua, Zhang Lilian 2931
- Roadside pedestrian detection and location based on binocular machine vision and RetinaNet
Lian Lirong, Luo Wenting, Qin Yong, Li Lin 2941

Remote Sensing Image Processing

- Accurate localization and parameter extraction of oil tank in remote sensing images
Jiang Han, Zhang Yueting, Guo Jiayi, Zhao Xin, Li Fangfang, Huang Lijia, Hu Yuxin, Lei Bin, Ding Chibiao 2953
- Consistent registration of remote sensing images in parametric synthesized spatial transformation network
Chen Ying, Zhang Qi, Li Wenju, Shi Yanjiao, Chen Lei 2964
- Ocean front detection method based on improved Mask R-CNN
Xu Huifang, Huang Dongmei, He Qi, Du Yanling, Qin Xuebiao 2981
- A robust lightweight deep neural network for remote sensing scene image classification and retrieval under label noise
Wang Yapeng, Li Yang, Wang Jiabao, Zhao Xun, Miao Zhuang 2991
- Road extraction method from remote sensing images based on improved U-Net network
Yang Jialin, Guo Xuejun, Chen Zehua 3005
- Weakly supervised deep semantic segmentation network for water body extraction based on multi-source remote sensing imagery
Li Xinwei, Li Yansheng, Zhang Yongjun 3015

中图法分类号: TP751 文献标识码: A 文章编号: 1006-8961(2021)12-2991-14

论文引用格式: Wang Y P, Li Y, Wang J B, Zhao X and Miao Z. 2021. A robust lightweight deep learning method for remote sensing scene image classification and retrieval under label noise. Journal of Image and Graphics, 26(12): 2991-3004 (王亚鹏, 李阳, 王家宝, 赵勋, 苗壮. 2021. 噪声鲁棒的轻量级深度遥感场景图像分类检索. 中国图象图形学报, 26(12): 2991-3004) [DOI: 10.11834/jig.200538]

噪声鲁棒的轻量级深度遥感场景图像分类检索

王亚鹏, 李阳, 王家宝, 赵勋, 苗壮*

陆军工程大学指挥控制工程学院, 南京 210007

摘要: **目的** 基于深度神经网络的遥感图像处理方法在训练过程中往往需要大量准确标注的数据, 一旦标注数据中存在标签噪声, 将导致深度神经网络性能显著降低。为了解决噪声造成的性能下降问题, 提出了一种噪声鲁棒的轻量级深度遥感场景图像分类检索方法, 能够同时完成分类和哈希检索任务, 有效提高深度神经网络在有标签噪声遥感数据上的分类和哈希检索性能。**方法** 选取轻量级神经网络作为骨干网, 而后设计能够同时完成分类和哈希检索任务的双分支结构, 最后通过设置损失基准的正则化方法, 有效减轻模型对噪声的过拟合, 得到噪声鲁棒的分类检索模型。**结果** 本文在两个公开遥感场景数据集上进行分类测试, 并与 8 种方法进行比较。本文方法在 AID (aerial image datasets) 数据集上, 所有噪声比例下的分类精度比次优方法平均高出 7.8%, 在 NWPU-RESISC45 (benchmark created by Northwestern Polytechnical University for remote sensing image scene classification covering 45 scene classes) 数据集上, 分类精度比次优方法平均高出 8.1%。在效率方面, 本文方法的推理速度比 CLEOT (classification loss with entropic optimal transport) 方法提升了 2.8 倍, 而计算量和参数量均不超过 CLEOT 方法的 5%。在遥感图像哈希检索任务中, 在 AID 数据集上, 本文方法的平均精度均值 (mean average precision, mAP) 在 3 种不同哈希比特下比 MiLaN (metric-learning based deep hashing network) 方法平均提高了 5.9%。**结论** 本文方法可以同时完成遥感图像分类和哈希检索任务, 在保持模型轻量高效的情况下, 有效提升了深度神经网络在有标签噪声遥感数据上的鲁棒性。

关键词: 标签噪声; 鲁棒学习; 图像分类; 图像检索; 哈希学习; 轻量级网络

A robust lightweight deep learning method for remote sensing scene image classification and retrieval under label noise

Wang Yapeng, Li Yang, Wang Jiabao, Zhao Xun, Miao Zhuang*

Command and Control Engineering College, Army Engineering University of PLA, Nanjing 210007, China

Abstract: **Objective** With the development of deep learning technology, deep neural networks have been widely used in various tasks of remote sensing, such as image retrieval, scene classification and change detection. Although these deep learning methods constantly refresh the accuracy of remote sensing applications on specific datasets, they require massive data with millions of reliable annotations, which are impractical or expensive for real-world applications. In contrast, when the accuracy of labels is too low, the performance of these deep learning methods will decline sharply. In order to reduce the labeling cost and improve the labeling speed, researchers have proposed a variety of greedy annotation methods to

收稿日期: 2020-09-14; 修回日期: 2020-12-29; 预印本日期: 2021-01-05

* 通信作者: 苗壮 emiao_beyond@163.com

基金项目: 国家自然科学基金项目 (61806220)

Supported by: National Natural Science Foundation of China (61806220)

improve labeling efficiency via clustering and crowd sourcing information. The performance of deep learning methods will decline dramatically once the label noise is introduced into the dataset. It is necessary to construct a noise robust deep learning method for remote sensing image processing to improve generalization performance. A noise robust and lightweight deep learning method for remote sensing scene classification and retrieval have been proposed to resolve performance degradation, which can effectively improve the classification and hash retrieval performance on remote sensing dataset under label noise. Furthermore, the proposed method can complete classification and hash retrieval tasks at the same time. **Method** First, a lightweight deep neural network named mobile GPU-aware network C (MoGA-C) as the backbone has been used to keep the lightweight of deep learning model, which has been proposed by Xiaomi AI Lab. MoGA-C has been obtained based on mobile GPU-aware (MoGA) neural network structure search algorithm. Various skills of lightweight network design have been integrated to ensure the lightweight of the network in the process of MoGA-C network design. Next, a double-branch structure behind deep neural network has been performed to the tasks of classification and retrieval simultaneously, which can not only avoid the degradation of classification performance caused by the insertion of hash layer, but also effectively increase the classification accuracy under label noise by integrating the results of double-branch. At last, the whole network has been fine-tuned during training process in order to improve the learning ability of deep neural network, which effectively improved the classification performance under low ratio label noise. A loss benchmark in the process of network fine-tuning has been set to reduce over-fitting to label noise in the middle and later stage of training, which limited the lower boundary of training loss and reduced the over-fitting under high ratio noise effectively. **Result** The proposed method has been evaluated via comparing it with other eight state of the art methods on two public remote sensing classification datasets. The research method has performed well under different noise ratios, which is 7.8% higher than sub-optimal method on aerial image datasets (AID) dataset and 8.1% higher on benchmark created by Northwestern Polytechnical University for remote sensing image scene classification covering 45 scene classes (NWPU-RESISC45) dataset in average. The inference speed has reached 2.8 times faster than the classification loss with entropic optimal transport (CLEOT) method. The floating point operations (FLOPs) and parameters are less than 5% of that in CLEOT method. The method has 5.9% average improvement under three different hash bits compared with the metric-learning based deep hashing network (MiLaN) method on AID dataset in the task of remote sensing image retrieval. **Conclusion** A lightweight and noise robust method for remote sensing scene classification and retrieval has been demonstrated to resolve the problem of performance degradation of remote sensing image processing methods under label noise. The proposed method can perform the tasks of classification and hash retrieval at the same time and improve the classification and retrieval performance under label noise effectively. First of all, a lightweight network has been opted as the backbone to ensure the lightweight of the model. Secondly, a parallel double-branch structure has been designed in order to complete the classification and hash retrieval tasks at the same time, the classification performance of the model has been improved further via combining the double-branch prediction results. Finally, the training loss has subjected to a positive value to reduce the over-fitting of label noise effectively via setting a loss benchmark. To compare with other methods, the classification and hash retrieval experiments have been conducted on two public datasets. The experimental results have presented that the proposed method not only has high efficiency, but also has good robustness to different ratios of label noise.

Key words: label noise; robust learning; image classification; image retrieval; hash learning; lightweight network

0 引言

随着深度学习技术的飞速发展,深度神经网络在遥感领域展现出了巨大优越性,并已经广泛应用于场景分类(Liu等,2019;Dong等,2020)、目标检测(Kellenberger等,2018)、像素分割(Huang等,2018)和图像检索(Song等,2019)等遥感图像处理的各项

任务中。但是,使用深度学习技术完成这些任务很大程度上依赖于数据集的大规模标注。如果数据集标注的准确性过低,深度学习方法的性能将受到严重影响。

传统的数据集标注通常由领域专家进行人工标注,标注精度能够得到保证,但非常耗时且成本较高。为了减少标注成本,提高遥感图像的标注速度,研究者提出了多种贪婪标注方法(Xia等,2015;Jin

等,2018),借助聚类和众包信息等手段提高标注效率。这些方法虽然能够节省大量人工标注成本,较快完成大规模数据标注任务,但都不可避免地会引入标签噪声(即错误标记样本类别)。一旦标签噪声引入数据集,深度学习方法性能就会严重下降,原因有二。其一是深度神经网络具有强大的学习能力,网络在训练过程中能够很容易记住训练数据(Zhang等,2017),这种记忆在训练后期往往会导致过拟合,使得网络在测试集上精度下降,泛化能力降低。其二是存在大量标签噪声时,深度神经网络会发生比过拟合更严重的退化现象。因为深度神经网络会记住错误的标签噪声,导致网络过拟合到错误样本,使网络进一步退化(Jiang等,2018)。这种情况下,如果使用常规方法对深度神经网络进行训练,将导致网络泛化能力严重下降,如图1(a)所示。由于存在大量标签噪声,深度神经网络会出现严重的噪声过拟合现象,在训练损失下降到接近零时,测试损失反而上升。因此,构建噪声鲁棒的遥感图像深度学习方法十分必要。

为了减轻标签噪声的影响,计算机视觉领域提出了一系列噪声鲁棒学习方法。这些方法可以简单归纳为3类:第1类方法采取数据清洗的思路,设计了多种清除标签噪声的方案,主要包括有向图模型(Xiao等,2015)、条件随机场(Vahdat,2017)和元学习(Ren等,2018)等。但是,这些方法需要一个额外的干净数据集,以便对数据中的噪声进行建模。第2类方法采取提高损失鲁棒性的思路,设计了多种鲁棒的损失函数,主要包括savage方法(Masnadi-Shirazi和Vasconcelos,2008)、unhinged方法(van Rooyen等,2015)、sigmoid方法和ramp方法(Ghosh等,2015)等。这些方法构建的损失函数在理论上

是有界的,对标签噪声具有固有的鲁棒性,但这些方法大多数都依赖于对称损失函数的假设。第3类方法与第2类方法类似,主要采取损失调整的思路,设计了多种校正方法来调整损失函数,主要包括bootstrapping方法(Reed等,2015)、backward方法和forward方法(Patrini等,2017)等。这些方法可以充分挖掘训练数据中的有效信息,同时调整每个样本的损失。但是,由于无法保证损失的调整始终是正确的,这些方法会导致错误调整累积,影响损失调整的效果,尤其是当标签噪声比例较大时。

遥感领域提出了两种针对遥感图像标签噪声问题的鲁棒学习方法。一是面向遥感场景分类的容错深度学习方法(Li等,2021),该方法参考数据清洗思路,通过集成学习提高标签纠正的准确度。但该方法在训练过程中融合了多个大型网络,并且需要多次迭代,参数量和计算量都大幅增加。二是基于熵最优运输的分类损失(classification loss with entropic optimal transport, CLEOT)(Damodaran等,2020),该方法参考提高损失鲁棒性思路,通过探索图像类间关系设计了一种鲁棒损失,取得了较好效果。该方法在训练时固定了骨干网的参数,虽然能够有效减轻高比例噪声下的过拟合,如图1(b)所示,在固定骨干网参数情况下,测试损失未出现明显上升。但该方法同时限制了深度神经网络的学习能力,导致低比例噪声下的分类性能降低。

为实现高效的噪声鲁棒学习方法,本文提出了一种噪声鲁棒的轻量级深度遥感场景图像分类检索方法,首先以轻量级网络作为骨干网,降低模型复杂度;其次为了同时完成分类和检索任务,不同于以往直接在深度神经网络分类器前插入哈希层的方法,本文方法在网络头部采用双分支结构,不仅避免了

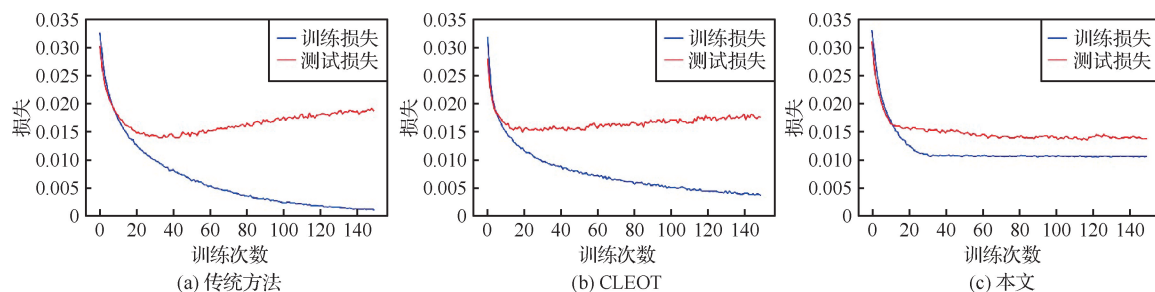


图1 高比例标签噪声情况下遥感图像分类损失曲线

Fig.1 Classification loss curves of remote sensing image under high ratio label noise
((a) traditional method; (b) CLEOT; (c) ours)

插入哈希层导致的分类性能下降,而且通过融合双分支预测结果,有效提升了噪声情况下的分类精度;最后为了充分发挥深度神经网络的学习能力,在训练过程中对整个网络进行微调,有效提高了低比例噪声下的分类精度,并通过设置损失基准的正则化方法有效减轻了高比例噪声情况下的噪声过拟合现象。如图 1(c) 所示,随着训练次数增加,本文方法测试损失不仅没有出现上升,反而继续下降。

本文主要贡献为:1)设计了一种噪声鲁棒的轻量级遥感场景图像分类检索方法,可以同时完成分类和哈希检索任务,并且两个任务可以相互促进,共同提高。2)提出一种轻量级遥感场景图像分类检索模型,引入一种噪声鲁棒的损失正则化方法,有效提高了模型的噪声鲁棒性,在保持标签噪声鲁棒性的同时,提升了数据处理的效率,使推理速度对比方法提升 2.8 倍,参数量和计算量降低 95% 以上。3)在两个遥感场景数据集上不同噪声比例下进行测试,实验结果表明,在遥感场景分类任务中,本文方法在 AID(aerial image datasets) 和 NWPU-RESISC45(benchmark crea-

ted by Northwestern Polytechnical university for remote sensing image scene classification covering 45 scene classes)数据集上平均分类精度均高于其他对比方法。在检索任务中,本文方法在 AID 数据集上的平均检索精度高于 MiLaN(metric-learning based deep hashing network)(Roy 等,2019)方法。

1 轻量级噪声鲁棒分类检索方法

本文提出一种噪声鲁棒的轻量级深度遥感场景图像分类检索方法,主要包括 3 部分。第 1 部分以一个轻量级神经网络为骨干网,用于特征提取;第 2 部分为一个双分支结构,能够同时完成分类和哈希检索任务,同时,在测试过程中通过融合两个分支输出结果可以有效提高分类精度;第 3 部分为噪声鲁棒损失函数,通过在训练过程中设置损失基准,本文方法的损失函数可以减轻深度模型对标签噪声的过拟合,高效实现噪声鲁棒的遥感场景图像分类检索。具体模型如图 2 所示,其中 $\oplus$ 表示相加。

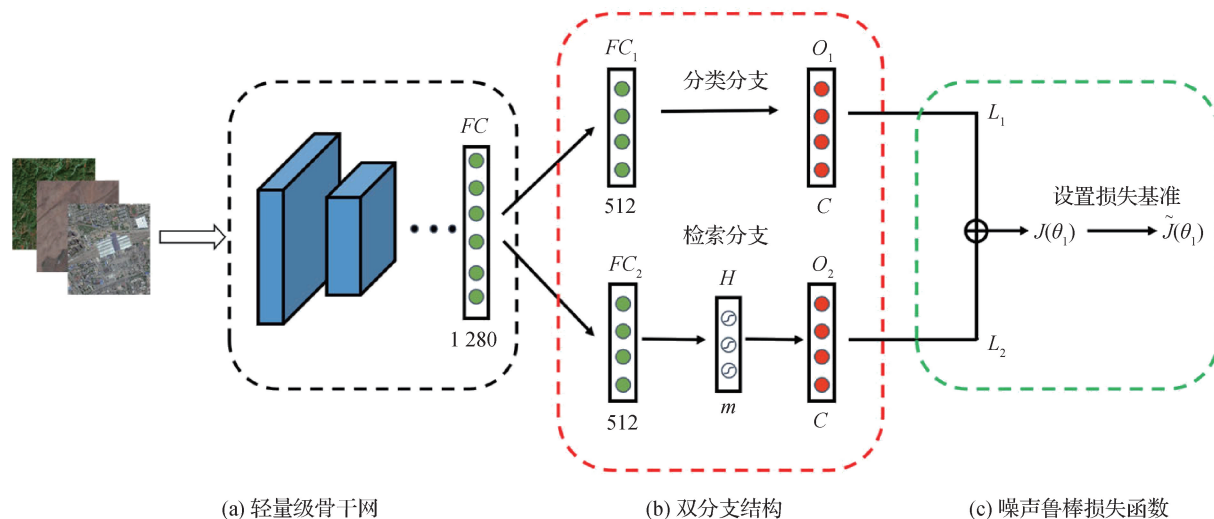


图 2 双分支噪声鲁棒分类检索模型

Fig. 2 Double-branch noise robust classification and retrieval model

((a) lightweight backbone; (b) double-branch structure; (c) noise robust loss function)

1.1 轻量级骨干网

轻量级骨干网采用小米 AI 实验室提出的 MoGA-C(mobile GPU-aware network)网络,由 MoGA(mobile GPU-aware)(Chu 等,2020)神经网络结构搜索算法搜索得到。MoGA-C 网络设计过程中,融入了现有轻量级网络设计的各种技巧。首先以 inverted bottleneck(Sandler 等,2018)卷积模块为基本单元,确保网络的轻量化;其次嵌入 SE(squeeze-

and-excitation)模块(Hu 等,2018),使网络可以关注信息量最大的通道;最后在激活函数选择中,除常用的 ReLU 外,还加入了兼具效率和性能的 Hswish 激活函数,为网络性能带来了进一步提升。在轻量级骨干网中,仅保留 MoGA-C 网络的前 19 个模块为骨干网,结构如表 1 所示。其中,模块 1 的输入大小为 $224^2 \times 3$,b-neck 表示 inverted bottleneck 卷积模块, 5×5 表示卷积核大小为 5×5 ,扩张倍数表示 invert-

ed bottleneck 模块中的扩张倍数, SE 表示是否采用 SE 模块。

表1 MoGA-C 神经网络结构
Table 1 The network structure of MoGA-C

模块	输出大小	操作	扩张倍数	SE	非线性激活
1	$112^2 \times 16$	conv2d, 3×3	—	—	Hswish
2	$112^2 \times 16$	b-neck, 3×3	1	—	ReLU
3	$56^2 \times 24$	b-neck, 5×5	3	—	ReLU
4	$56^2 \times 24$	b-neck, 3×3	3	—	ReLU
5	$28^2 \times 40$	b-neck, 5×5	3	—	ReLU
6	$28^2 \times 40$	b-neck, 3×3	3	—	ReLU
7	$28^2 \times 40$	b-neck, 5×5	3	—	ReLU
8	$14^2 \times 80$	b-neck, 5×5	3	—	Hswish
9	$14^2 \times 80$	b-neck, 5×5	6	✓	Hswish
10	$14^2 \times 80$	b-neck, 5×5	3	—	Hswish
11	$14^2 \times 80$	b-neck, 5×5	3	—	Hswish
12	$14^2 \times 112$	b-neck, 3×3	6	—	Hswish
13	$14^2 \times 112$	b-neck, 3×3	6	✓	Hswish
14	$7^2 \times 160$	b-neck, 3×3	6	✓	Hswish
15	$7^2 \times 160$	b-neck, 3×3	6	✓	Hswish
16	$7^2 \times 160$	b-neck, 3×3	6	✓	Hswish
17	$7^2 \times 960$	conv2d, 1×1	—	—	Hswish
18	$1^2 \times 960$	avgpool, 7×7	—	—	—
19	$1^2 \times 1\ 280$	conv2d, 1×1	—	—	Hswish

注:“—”表示未使用该模块,“✓”表示使用该模块。

假设训练集共有 N 个训练样本,表示为 $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$,则相应的标签集表示为 $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$,其中 $\mathbf{y}_i \in \mathbf{R}^C$ 是样本 $\mathbf{x}_i$ 的真实标签向量,其真实类别对应位置的元素为 1,其他位置元素为 0, C 表示遥感场景的类别数。对于任意遥感图像 $\mathbf{x}_i \in \mathbf{X}$,送入轻量级骨干网 MoGA-C 进行特征提取,在网络最后一个全连接层 FC 后可以得到一个 1 280 维的深度特征向量 $\mathbf{f}_i$,具体为

$$\mathbf{f}_i = \Phi(\mathbf{x}_i; \theta_0) \quad (1)$$

式中, θ_0 表示 MoGA-C 骨干网的所有参数, Φ 表示 MoGA-C 骨干网映射函数,本质上该函数是一种非线性映射函数。经过轻量级骨干网,输出的 1 280 维特征向量 $\mathbf{f}_i$ 将送入双分支结构,进行分类和检索学习任务。

1.2 双分支结构

传统深度哈希检索方法中,通常通过在深度分类网络中插入哈希层,借助于类别标签提供的监督信息实现高性能哈希图像检索。但由于哈希层的存在,导致信息丢失,并进一步影响分类精度。为此,本文提出了一种双分支结构(如图 2(b)),包括分类分支和检索分支,用以同时实现高性能的分类和检索任务。

1.2.1 分类分支

在 MoGA-C 骨干网最后一层 FC 后接 512 维全连接层 FC_1 ,并由概率 $p = 0.5$ 的 dropout 层进行正则化,使 FC_1 层的部分节点失活,以减少网络过拟合,增加泛化能力,而后得到深度特征 $\mathbf{f}_{i1}$,具体为

$$\mathbf{f}_{i1} = \text{dropout}_{p=0.5}(\mathbf{W}_1 \mathbf{f}_i + \mathbf{V}_1) \quad (2)$$

式中, $\mathbf{W}_1 \in \mathbf{R}^{512 \times 1\ 280}$, $\mathbf{V}_1 \in \mathbf{R}^{512}$,分别表示 FC_1 层的权重矩阵和偏置向量。在 FC_1 层后接 C 维全连接层 O_1 并经 softmax 函数激活,得到分类分支的预测结果 $\mathbf{t}_{i1}$,具体为

$$\mathbf{t}_{i1} = \text{softmax}(\mathbf{W}_{o1} \mathbf{f}_{i1} + \mathbf{V}_{o1}) \quad (3)$$

式中, $\mathbf{W}_{o1} \in \mathbf{R}^{C \times 512}$, $\mathbf{V}_{o1} \in \mathbf{R}^C$,分别表示 O_1 层的权重矩阵和偏置向量。

1.2.2 检索分支

在 MoGA-C 骨干网最后一层 FC 后接 512 维全连接层 FC_2 ,并由概率 $p = 0.5$ 的 dropout 层进行正则化,得到深度特征 $\mathbf{f}_{i2}$,具体为

$$\mathbf{f}_{i2} = \text{dropout}_{p=0.5}(\mathbf{W}_2 \mathbf{f}_i + \mathbf{V}_2) \quad (4)$$

式中, $\mathbf{W}_2 \in \mathbf{R}^{512 \times 1\ 280}$, $\mathbf{V}_2 \in \mathbf{R}^{512}$,分别表示 FC_2 层的权重矩阵和偏置向量。为了将深度特征编码为 m 比特的哈希码,在 FC_2 层后接 m 维全连接层 H (本文所称哈希层即为此全连接层),并经 tanh 函数激活,使得哈希层 H 的输出趋近于二值化,以减少量化损失,而后得到哈希层特征 $\mathbf{u}_i$,具体为

$$\mathbf{u}_i = \tanh(\mathbf{W}_H \mathbf{f}_{i2} + \mathbf{V}_H) \quad (5)$$

式中, $\mathbf{W}_H \in \mathbf{R}^{m \times 512}$, $\mathbf{V}_H \in \mathbf{R}^m$,分别表示哈希层 H 的权重矩阵和偏置向量。哈希层 H 后接 C 维全连接层 O_2 并经 softmax 函数激活,得到检索分支的预测结果 $\mathbf{t}_{i2}$,具体为

$$\mathbf{t}_{i2} = \text{softmax}(\mathbf{W}_{o2} \mathbf{u}_i + \mathbf{V}_{o2}) \quad (6)$$

式中, $\mathbf{W}_{o2} \in \mathbf{R}^{C \times m}$, $\mathbf{V}_{o2} \in \mathbf{R}^C$,分别表示 O_2 层的权重矩阵和偏置向量。

当模型训练结束后,即可进行分类和检索任务。

在分类任务中,将图像送入网络后,得到两个分支的预测结果 t_{i1} 和 t_{i2} ,将两者相加取平均,得到最终预测结果 t_i ,具体为

$$t_i = \frac{t_{i1} + t_{i2}}{2} \quad (7)$$

则图像对应类别 c 为

$$c = \operatorname{argmax}_{k=1,2,\dots,C} (t_i^k) \quad (8)$$

值得注意的是,检索分支除了可以完成哈希检索任务外,还为分类分支提供了补充信息。在测试时,通过融合双分支结果,可以有效提高噪声情况下的分类性能。此外,该方法的参数量和计算量几乎没有增加,在保持模型高效的同时提高了分类性能。

在检索任务中,将检索分支得到的哈希层特征 u_i 进行二值化操作 (sign),即可得到图像哈希码 h_i ,具体为

$$h_i = \operatorname{sign}(u_i) \quad (9)$$

在得到哈希码 h_i 后,与数据库图像的哈希码进行对比,计算出汉明距离后按距离由小到大进行排序,即可得到哈希检索结果。

1.3 噪声鲁棒损失函数

深度神经网络具有强大的学习能力,几乎可以拟合所有训练数据。但在有标签噪声情况下,标签噪声会导致模型性能的严重退化(如图 1(a))。为了减轻噪声过拟合问题,CLEOT (Damodaran 等, 2020) 方法在训练模型时固定了骨干网的参数,虽然能够有效减轻高比例噪声下模型的过拟合(如图 1(b)),但却牺牲了骨干网的学习能力,导致低比例噪声下的分类性能降低。

为了充分发挥骨干网的学习能力,采用微调骨干网参数的方式进行训练。微调虽然能够提高深度神经网络的学习能力,但在训练数据存在噪声情况下,微调会导致噪声过拟合。针对噪声过拟合问题,本文引入设置损失基准的正则化方法 (Ishida 等, 2020),通过设置损失基准,限制了训练损失的最小值,使模型有效提高了有噪声情况下的泛化能力。

1.3.1 目标函数

均方误差 (mean square error, MSE) 损失是一种有界损失,是一种比交叉熵损失更加噪声鲁棒的损失函数 (Ghosh 等, 2017)。因此,本文在两个分支中均采用 MSE 作为损失函数。

在分类分支得到 C 维预测结果 t_{i1} 后,首先计算所有训练样本与其真实类别标签 y_i 的均方误差,并对所有样本的均方误差求和,得到分类分支损失 L_1 ,即

$$L_1 = \sum_{i=1}^N \frac{1}{C} \sum_{k=1}^C (t_{i1}^k - y_i^k)^2 \quad (10)$$

同理,在检索分支中,可以得到检索分支损失 L_2 ,具体为

$$L_2 = \sum_{i=1}^N \frac{1}{C} \sum_{k=1}^C (t_{i2}^k - y_i^k)^2 \quad (11)$$

进而得到总损失 L ,具体为

$$L = \lambda L_1 + (1 - \lambda) L_2 \quad (12)$$

式中, $\lambda \in [0, 1]$, 用于调节 L_1 和 L_2 权重。最后得到目标函数,具体为

$$J(\theta_1) = \min(L) \quad (13)$$

式中, θ_1 为使损失函数 L 最小的网络参数,即模型需要学习的参数。

1.3.2 损失基准设置

在深度神经网络训练过程中,随着损失函数的不断优化,训练损失最终将接近于零。但如果训练数据中存在大量噪声,训练损失接近于零往往会导致噪声过拟合。噪声过拟合将导致模型泛化能力下降,在测试数据上识别精度降低。设置损失基准的方法通过限制训练损失的下限,使训练损失始终保持大于一个正值,以减轻模型对噪声数据的过拟合。设置损失基准后的目标函数 $\tilde{J}(\theta_1)$ 可以表示为

$$\tilde{J}(\theta_1) = |J(\theta_1) - b| + b \quad (14)$$

式中, b 为设置的损失基准。当 $J(\theta_1) \geq b$ 时, $\tilde{J}(\theta_1) = J(\theta_1)$, 此时 $\tilde{J}(\theta_1)$ 相对于网络参数 θ_1 的梯度 $\nabla \tilde{J}(\theta_1)$ 与原始目标函数 $J(\theta_1)$ 相对于 θ_1 的梯度 $\nabla J(\theta_1)$ 方向相同,反向传播执行梯度下降;当 $J(\theta_1) < b$ 时, $\tilde{J}(\theta_1) = -J(\theta_1) + 2b$, $\nabla \tilde{J}(\theta_1)$ 与 $\nabla J(\theta_1)$ 方向相反,此时反向传播执行梯度上升。以上方法可以使原目标函数 $J(\theta_1)$ 始终在 b 附近上下浮动,而 $\tilde{J}(\theta_1)$ 则始终保持大于 b , 这可以避免损失进一步降低。图 3 为设置损失基准后的训练损失曲线。蓝色和红色曲线分别为 $J(\theta_1)$ 和 $\tilde{J}(\theta_1)$ 对应的训练损失曲线,绿色直线代表损失基准。可以看出,设置损失基准后,可以避免训练损失过度接近于零,进而防止模型过拟合。

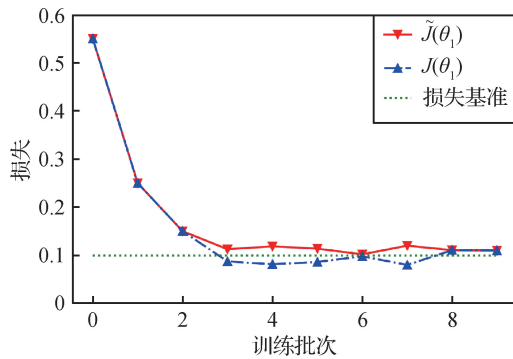


图3 设置损失基准后损失曲线

Fig. 3 The loss curves after setting loss benchmark

2 数据集与实验设置

2.1 数据集

为了对本文方法进行评估,在数据集 AID (aerial image datasets) (Xia 等, 2017) 和 NWPU-RESISC45 (enchmark created by Northwestern Polytechnical University for remote sensing image scene classification covering 45 scene classes) (Cheng 等, 2017) 上进行实验。

AID 数据集由武汉大学于 2017 年发布,共包括 10 000 幅场景图像,涵盖 30 个场景类别,每个场景类别由 220 ~ 420 幅图像组成。图像从谷歌地球中裁剪得到,大小为 600×600 像素,空间分辨率为 $0.5 \sim 8$ m,部分样例见图 4(a)。NWPU-RESISC45 数据集由西北工业大学于 2017 年发布,共 31 500 幅场景图像,涵盖 45 个场景类别,每个场景类别由 700 幅图像组成。图像同样从谷歌地球中裁剪得到,大小为 256×256 像素。除一些空间分辨率较低



图4 遥感场景数据集样例

Fig. 4 Samples of remote sensing scene datasets
((a) AID; (b) NWPU-RESISC45)

的特定类别(如岛屿、湖泊、山脉和冰山)外,大多数场景类别的空间分辨率为 $0.2 \sim 30$ m,部分样例见图 4(b)。

本文在上述两个数据集中加入人工模拟标签噪声,设置方式和具体类别与 CLEOT 方法相同。实验时,根据噪声比例将标签翻转到视觉上相似的类别。两个数据集上的具体标签噪声标注方式如表 2 和表 3 所示。需要说明的是,本文提到的标签翻转比例

表2 AID 数据集上的标签噪声标注对照表

Table 2 Flipped classes according to the label noise on AID dataset

标签噪声
空地↔沙漠
中心→存储罐
教堂→中心、存储罐
密集住宅↔中密度住宅
工业区→中密度住宅
草地→农场
操场→草地
学校、度假村→中密度住宅
学校↔操场
体育场→操场

注:a→b 表示标签 a 翻转成标签 b, a↔b 表示双向翻转标签。

表3 NWPU-RESISC45 数据集上的标签噪声标注对照表

Table 3 Flipped classes according to the label noise on NWPU-RESISC45 dataset

标签噪声
棒球内场→中密度住宅
海滩→河流
密集住宅↔中密度住宅
交叉路口→高速公路
移动家庭公园↔密集住宅
天桥↔交叉路口
网球场→中密度住宅
机场跑道→高速公路
热电站→云
湿地→湖泊
矩形农田→草地
教堂→宫殿
商业区→密集住宅

注:a→b 表示标签 a 翻转成标签 b, a↔b 表示双向翻转标签。

并非整个数据集中所有噪声所占的比例,而是指表 2 和表 3 中翻转的类别按此比例对本类别数据进行标签翻转。

2.2 实验设置

本文实验环境采用 PyTorch 1.3, CUDA 采用 10.0 版本,在两块 NVIDIA 2080Ti 显卡上进行实验。MoGA-C 骨干网先在小型遥感场景数据集 UCM (University of California merced land-use dataset) (Yang 和 Newsam,2010)上进行预训练,以消除领域差异。为了解决类间数据不平衡问题,训练过程中采用加权随机采样方法对训练集进行抽样。抽样时对样本量较少的类别赋予较大的权重,使样本量较少的类别能够有较大的概率被抽中,有效避免了数据类间不平衡问题。

训练过程中,学习率设为 0.1, batch size 设为 128,最大训练次数为 150 轮。采用与 CLEOT 方法相同的方式进行提前停止,保存验证集精度最好的模型,即验证集精度 25 轮内没有提升则停止训练;图像送入网络前统一调整为 224×224 像素;同时,在式(12)中,由于分类分支损失 L_1 和检索分支损失 L_2 均为均方误差损失,在同一数量级上,为平衡两者作用,式中 λ 设为 0.5。

在分类实验中,将检索分支哈希层的神经元个数 m 设置为 16。两个数据集分别按 80%、10% 和 10% 的比例划分训练集、验证集和测试集。其中,训练集和验证集数据为有噪声数据,测试集数据为干

净数据。测试集数据仅用于对各种方法的评价,不参与训练过程。标签翻转比例设置为 {0,0.2,0.4,0.6,0.8},即表 2 和表 3 中对应类别标签按照比例进行翻转,以添加噪声,实际噪声占总体的比例依赖于受噪声影响的类别数。

3 实验结果分析

3.1 分类实验

3.1.1 实验结果

在相同噪声实验条件下,将本文方法与 savage (Masnadi-Shirazi 和 Vasconcelos, 2008)、unhinged (van Rooyen 等,2015)、sigmoid (Ghosh 等,2015)、ramp (Ghosh 等,2015)、bootstrap soft (Reed 等,2015)、backward (Patrini 等,2017)、forward (Patrini 等,2017)和 CLEOT (Damodaran 等,2020)等方法分别在 AID 和 NWPU-RESISC45 数据集上进行分类对比实验,结果如表 4 和表 5 所示,分类精度取 5 次实验结果的平均值。可以看出,本文方法在所有情况下均达到最佳性能。在中低比例(标签翻转比例为 0,0.2,0.4)噪声下,分类精度在 AID 和 NWPU-RESISC45 数据集上平均比次优方法高 8.3% 和 10.9%。在高比例(标签翻转比例为 0.6,0.8)噪声下,比次优方法高 3% 以上。值得注意的是,CLEOT 方法为了避免在高比例噪声下出现严重的噪声过拟合,在训练阶段冻结了骨干网参数,因此在高比例噪

表 4 各种鲁棒学习方法在 AID 数据集上分类精度对比
Table 4 Comparison of classification accuracy among various robust learning methods on AID dataset

方法	标签翻转比例				
	0	0.2	0.4	0.6	0.8
savage (Masnadi-Shirazi 和 Vasconcelos,2008)	83.65 ± 0.10	85.73 ± 0.21	<u>82.28 ± 0.27</u>	62.88 ± 0.50	56.55 ± 0.48
unhinged (van Rooyen 等,2015)	87.64 ± 0.19	<u>86.33 ± 0.19</u>	78.67 ± 0.29	65.93 ± 0.52	57.20 ± 0.16
sigmoid (Ghosh 等,2015)	85.41 ± 0.26	84.71 ± 0.25	82.05 ± 0.17	60.96 ± 0.44	56.18 ± 0.08
ramp (Ghosh 等,2015)	<u>87.74 ± 0.22</u>	86.24 ± 0.23	78.37 ± 0.56	66.04 ± 0.59	57.21 ± 0.27
bootstrap soft (Reed 等,2015)	87.03 ± 0.40	82.54 ± 0.78	73.75 ± 0.82	65.24 ± 1.09	58.00 ± 0.45
backward (Patrini 等,2017)	86.87 ± 0.52	82.63 ± 0.59	74.03 ± 0.56	65.71 ± 1.16	57.90 ± 0.22
forward (Patrini 等,2017)	86.91 ± 0.41	82.30 ± 1.08	73.59 ± 0.76	64.91 ± 0.74	58.43 ± 0.59
CLEOT (Damodaran 等,2020)	87.02 ± 0.63	85.39 ± 1.12	79.19 ± 0.94	<u>71.76 ± 0.66</u>	<u>63.23 ± 0.42</u>
本文	96.42 ± 0.15	94.28 ± 0.24	90.60 ± 1.03	81.20 ± 1.43	68.04 ± 0.93

注:加粗字体为每列最优结果,加下划线字体为每列次优结果。

表5 各种鲁棒学习方法在NWPU-RESISC45数据集上分类精度对比

Table 5 Comparison of classification accuracy among various robust learning methods on NWPU-RESISC45 dataset

方法	标签翻转比例				
	0	0.2	0.4	0.6	0.8
savage(Masnadi-Shirazi 和 Vasconcelos,2008)	76.85 ± 0.15	75.13 ± 0.11	69.96 ± 0.14	59.56 ± 0.03	58.08 ± 0.07
unhinged(van Rooyen 等,2015)	82.81 ± 0.21	82.13 ± 0.14	78.38 ± 0.57	63.07 ± 0.26	61.01 ± 0.01
sigmoid(Ghosh 等,2015)	71.74 ± 0.40	68.08 ± 0.18	65.76 ± 0.50	57.10 ± 0.06	56.61 ± 0.31
ramp(Ghosh 等,2015)	82.99 ± 0.10	<u>82.26 ± 0.20</u>	78.81 ± 0.26	62.97 ± 0.16	60.91 ± 0.32
bootstrap soft(Reed 等,2015)	82.98 ± 0.17	80.65 ± 0.47	75.82 ± 0.88	67.39 ± 0.86	62.22 ± 0.21
backward(Patrini 等,2017)	82.79 ± 0.14	80.65 ± 0.51	75.96 ± 0.72	68.67 ± 0.75	62.45 ± 0.52
forward(Patrini 等,2017)	<u>83.06 ± 0.11</u>	80.87 ± 0.53	74.97 ± 1.02	68.12 ± 1.16	62.56 ± 0.16
CLEOT(Damodaran 等,2020)	82.41 ± 0.27	81.54 ± 0.18	<u>80.84 ± 0.45</u>	<u>76.07 ± 0.35</u>	<u>70.14 ± 0.33</u>
本文	95.34 ± 0.18	93.29 ± 0.26	90.16 ± 0.65	80.89 ± 1.83	73.24 ± 0.7

注:加粗字体为每列最优结果,加下划线字体为每列次优结果。

声下取得了较好的分类性能,但由于冻结参数,限制了深度神经网络的学习能力,使得该方法在低比例噪声情况下的分类性能并不优于其他方法。本文方法通过设置损失基准,有效减轻了高比例噪声情况下微调骨干网时的噪声过拟合问题,同时采用微调骨干网参数的方式,充分发挥了深度神经网络的学习能力,使模型在低比例噪声下也保持了良好性能。

3.1.2 效率分析

为了验证本文轻量级方法的效率,与 CLEOT 方法在模型参数量、模型计算量和模型推理时间等方面进行对比,在推理时间测试中,对 AID 数据集中的测试集数据共 1 000 幅图像进行分类测试,统计 1 000 幅图像推理消耗的总时间。对比结果如图 5 所示。可以看出,本文方法和 CLEOT 方法的模型参数量分别为 5.6 M 和 136.4 M,模型计算量分别为 0.22 G FLOPs (floating point operations) 和 15.5 G FLOPs,1 000 幅图像推理时间分别为 45.8 ms 和 175.1 ms。与 CLEOT 方法相比,本文方法的模型参数减少至 4.1%,计算量减少至 1.4%,推理速度提高 2.8 倍。

3.1.3 消融实验

为了验证本文方法的有效性,在 AID 数据集上以 MoGA-C 网络为骨干网进行 4 组实验,分别对比了骨干网参数调节、双分支结构、损失基准等 3 个条件对深度网络鲁棒性的影响,实验结果如表 6 所示。对比实验 1 和实验 2 可知,在高比例噪声情况下微调参数比固定参数更容易发生过拟合,导致精度降

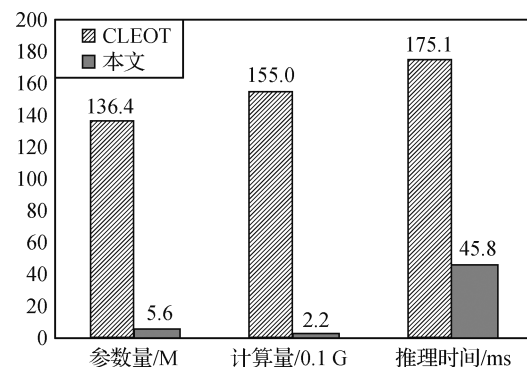


图5 本文方法与 CLEOT 方法效率对比

Fig. 5 Comparison of efficiency between CLEOT method and ours

低。而在低比例噪声情况下,微调参数可以充分发挥深度神经网络的学习能力。对比实验 2 和实验 3 可知,双分支结构在高比例噪声情况下具有较好的鲁棒性。在检索分支中,由于哈希层的存在,tanh 激活函数使哈希层输出趋近于二值化,丢弃了部分信息,客观上为模型提供了正则化,使得高比例噪声情况下有效减轻了模型对噪声的过拟合。对比实验 3 和实验 4 可知,通过设置损失基准能够有效减轻噪声过拟合带来的影响,分类精度显著提高。这种现象充分说明噪声过拟合问题是导致噪声情况下遥感图像分类精度过低的主要原因。由此可见,本文方法通过有效整合轻量级网络、双分支结构和损失基准 3 种策略,较好地解决了遥感图像分类中的噪声过拟合问题。轻量级网络利用模型参数少的特性,

避免参数过多导致过拟合;双分支结构通过多任务学习,避免单一任务过拟合;损失函数与损失基准设置,通过限制训练损失最小值,避免优化过程中产生过拟合。

表 6 在 AID 数据集上消融实验的分类精度
Table 6 Classification accuracy in ablation experiments on AID dataset

实验序号	方法	标签翻转比例				
		0	0.2	0.4	0.6	0.8
1	参数固定 + MSE	91.96 ± 0.26	91.80 ± 0.48	86.64 ± 0.7	74.60 ± 0.71	65.74 ± 0.73
2	参数微调 + MSE	96.60 ± 0.23	93.50 ± 0.34	90.14 ± 0.62	73.32 ± 1.08	65.32 ± 0.39
3	参数微调 + MSE + 双分支	96.42 ± 0.15	93.70 ± 0.24	90.15 ± 0.66	74.58 ± 1.30	65.80 ± 0.52
4	参数微调 + MSE + 双分支 + 损失基准	96.42 ± 0.15	94.28 ± 0.24	90.60 ± 1.03	81.20 ± 1.43	68.04 ± 0.93

注:加粗字体为每列最优结果。在无噪声情况下,不设置损失基准。

3.1.4 损失基准的选择

在本文中,损失基准 b 的选择对于抑制噪声过拟合问题十分重要,该参数的选择依赖于训练损失和验证损失之间的关系确定。图 6 显示了 AID 数据集中标签翻转比例设置为 0.8 情况下未设置损失基准时的验证损失和训练损失曲线。可以看出,在训练和验证过程中,可以利用训练和验证损失曲线将该过程大致划分为 A、B、C 这 3 个阶段。在阶段 A,两种损失均呈快速下降趋势,未发生过拟合,网络参数处于快速更新阶段,训练损失下降到 0.013 附近时,阶段 A 结束;在阶段 B,验证损失和训练损失缓慢下降,验证损失和训练损失差值增大,此时已经开始发生过拟合,训练损失下降到 0.007 附近时,阶段 B 结束;在阶段 C,验证损失趋于平稳,训练损失继续缓慢减小,逐渐逼近于 0,网络参数处于严重过拟合状态,此阶段训练损失继续下降将导致模型对噪声的严重过拟合。

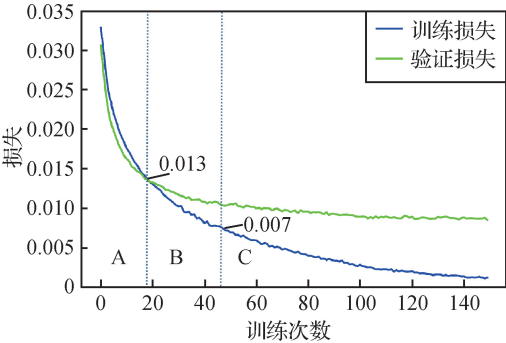


图 6 训练阶段划分

Fig. 6 Division of training stages

因此,本文将损失基准 b 限制在阶段 B,针对

AID 数据集标签翻转比例设置为 0.8 情况下,损失基准 b 取值范围为 $[0.007, 0.013]$ 。由于不同数据集的类别不同,数据量不同,拟合的难易程度也有所不同,所以损失曲线会有所差异。根据前文所述阶段划分原则,本文在 NWPU-RESISC45 数据集上,标签翻转比例设置为 0.8 情况下,测得的损失基准 b 的取值范围为 $[0.004, 0.01]$ 。表 7 和表 8 为两个数据集上标签翻转比例设置为 0.8 情况下 b 取不同值测得的分类精度。可以看出,当 b 取范围中值(即在 AID 数据集上 b 取 0.010,在 NWPU-RESISC45 数据集上 b 取 0.007)时,模型分类精度最高。

表 7 在 AID 数据集上不同 b 值的分类精度

Table 7 Classification accuracy under different b on AID dataset

b 值	分类精度/%
0.007	66.40 ± 1.66
0.010	68.04 ± 0.93
0.013	67.82 ± 1.19

注:加粗字体为最优结果。

表 8 在 NWPU-RESISC45 数据集上不同 b 值的分类精度

Table 8 Classification accuracy under different b on NWPU-RESISC45 dataset

b 值	分类精度/%
0.004	71.67 ± 0.69
0.007	73.24 ± 0.70
0.010	72.71 ± 0.56

注:加粗字体为最优结果。

3.2 哈希检索实验

3.2.1 评价指标

在查询阶段,首先通过计算查询样本与数据库样本之间的汉明距离对所有样本排序。获得数据库样本排序列表后,即可得到每个查询图像的平均精度(average precision, AP),具体为

$$AP = \frac{1}{n_i} \sum_{j=1}^{n_i} P(i, j) \quad (15)$$

式中, n_i 是检索结果中所有与第 i 个查询图像相似的图像数量, $P(i, j)$ 是第 i 个查询图像的检索结果中第 j 幅与图像 i 相似图像的精度。例如,检索结果中第 3 幅与查询图像相似的图像排在第 5 个,则其精度 $P = 3/5 = 0.6$ 。而后,对所有查询图像的 AP 值取平均计算平均精度均值(mean average precision, MAP),具体为

$$f_{\text{MAP}} = \frac{1}{Q} \sum_{i=1}^Q \frac{1}{n_i} \sum_{j=1}^{n_i} P(i, j) \quad (16)$$

式中, Q 表示查询图像的数量。MAP 反映了检索结果的准确度, MAP 值越高,检索效果越好。

3.2.2 实验结果

为了验证本文哈希检索方法的有效性,本文与 MiLaN (metric-learning based deep hashing network) (Roy 等, 2019) 和 KSLSH (kernel-based supervised locality-sensitive hashing) (Demir 和 Bruzzone, 2016) 方法进行了对比。KSLSH 方法是一种非深度学习的哈希方法,通过核函数将图像映射到汉明空间,并将图像的标签作为监督信息,有效提高了核函数的编码性能,是一种典型的传统哈希检索方法。MiLaN 方法是一种典型的深度哈希检索方法,利用深度网络对图像进行特征提取,并通过三元组损失 (Schroff 等, 2015) 拉近相似图像距离,拉远不相似

图像距离,实现了高效的深度哈希检索。本文方法通过将检索和分类任务统一在同一个框架内,有效利用图像类别信息,使相同类别图像生成相似的哈希码,实现了更高性能的哈希检索。检索实验条件设置与 MiLaN 方法相同,按照 6:4 划分训练集和测试集,实验结果如图 7 所示。其中 MAP@20 表示计算检索结果中的前 20 个图像的 MAP 值。通过对比可以发现,本文方法在 16、24、32 bits 下均达到了最高精度。与 MiLaN 方法相比,本文方法检索精度平均提高了 5.9%。

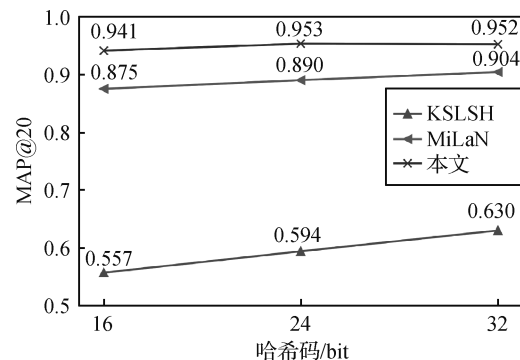


图7 检索精度对比

Fig. 7 The comparison of retrieval accuracy

图 8 给出了在 AID 数据集上标签翻转比例设置为 0.8 情况下,3 幅遥感场景图像 32 bits 哈希码的检索结果。图中第 1 列为查询图像,后 10 列为前 10 个检索结果,绿色框为正确类别,红色框为错误类别。从检索结果可以看出,在高比例噪声条件下,本文方法仍能达到较好的检索效果。

同时,为了扩展遥感场景图像哈希检索的研究内容,本文还针对标签噪声情况下的遥感场景图像哈希检索在两个数据集上进行测试,计算检索结果中的前 20 个图像的 MAP 值,如表 9 和表 10 所示。

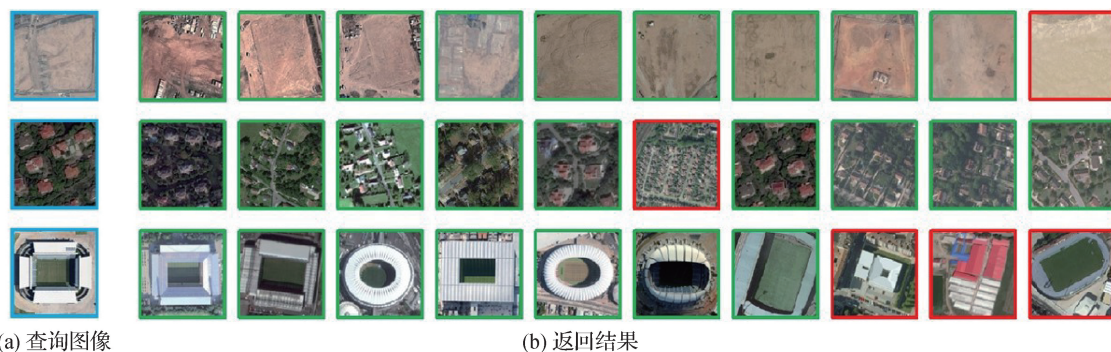


图8 检索结果样例

Fig. 8 Samples of retrieval results ((a) query images; (b) returned images)

表 9 在 AID 数据集上不同标签噪声比例下的检索精度
Table 9 Retrieval accuracy under different label noise ratios on AID dataset

哈希码/bit	标签翻转比例				
	0	0.2	0.4	0.6	0.8
16	0.941	0.904	0.864	0.820	0.774
24	0.953	0.908	0.873	0.840	0.849
32	0.952	0.916	0.882	0.860	0.854

表 10 在 NWPU-RESISC45 数据集上不同标签噪声比例下的检索精度
Table 10 Retrieval accuracy under different label noise ratios on NWPU-RESISC45 dataset

哈希码/bit	标签翻转比例				
	0	0.2	0.4	0.6	0.8
16	0.942	0.838	0.785	0.769	0.714
24	0.945	0.888	0.852	0.811	0.760
32	0.943	0.901	0.870	0.834	0.774

可以看出,遥感图像哈希检索的精度会随噪声数据比例的增加而降低。此外,随着哈希比特编码长度的增加,遥感图像哈希检索的精度会随着哈希编码长度增加而增加。该项实验是遥感场景图像哈希检索第 1 次在考虑噪声情况下进行的测试,因此该结果可以作为今后研究的测试基准。

4 结 论

为了解决在有标签噪声情况下深度遥感场景图像处理方法性能下降问题,本文提出了一种噪声鲁棒的轻量级深度遥感场景图像分类检索方法,能够同时完成分类和哈希检索任务,有效提高深度神经网络在有标签噪声遥感数据上的分类和检索性能。选取轻量级网络作为骨干网,确保模型轻量化。而后为了同时完成分类和哈希检索任务,设计了平行的双分支结构,并通过融合双分支预测结果,使模型的性能得到进一步提高。通过设置损失基准的方法,避免了训练损失过度趋近于零,有效减轻了模型对噪声的过拟合。

本文在两个公开数据集上进行了分类和哈希检索实验,并与其他方法进行对比。在分类实验中,本文方法在所有情况下均达到最佳性能。在中低比例

噪声下,本文方法分类精度在 AID 数据集上平均比次优方法高 8.3%;在 NWPU-RESISC45 数据集上,平均比次优方法高 10.9%。在高比例噪声下,比次优方法高 3% 以上。在检索实验中,本文方法检索精度平均比 MiLaN 方法高 5.9%。在效率方面,本文轻量级模型使模型参数减少至 4.1%,计算量减少至 1.4%,推理速度提高 2.8 倍。实验结果表明,本文方法不仅具有较高效率,而且对各种噪声比例均有良好的鲁棒性。但是,由于本文方法涉及到损失基准 b 值的选择,未能实现端到端的训练。为此,本文给出了损失基准 b 值的选择原则,实验过程中需要根据具体的损失曲线确定具体的损失基准 b 。

在将来的研究工作中,将致力于遥感场景图像无监督哈希检索的研究。通过无监督学习方式对无标签遥感数据进行标注,得到带标签噪声的伪标签,而后利用得到的伪标签训练哈希检索模型,以减少深度模型对标注数据的依赖,充分挖掘无标签遥感数据的价值。

参考文献 (References)

Cheng G, Han J W and Lu X Q. 2017. Remote sensing image scene classification: benchmark and state of the art. Proceedings of the IEEE, 105 (10): 1865-1883 [DOI: 10.1109/JPROC.2017.2675998]

Chu X X, Zhang B and Xu R J. 2020. MoGA: searching beyond Mobile-netv3//Proceedings of ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE: 4042-4046 [DOI: 10.1109/ICASSP40776.2020.9054428]

Damodaran B B, Flamary R, Seguy V and Courty N. 2020. An Entropic Optimal transport loss for learning deep neural networks under label noise in remote sensing images. Computer Vision and Image Understanding, 191: #102863 [DOI: 10.1016/j.cviu.2019.102863]

Demir B and Bruzzone L. 2016. Hashing-based scalable remote sensing image search and retrieval in large archives. IEEE Transactions on Geoscience and Remote Sensing, 54 (2): 892-904 [DOI: 10.1109/TGRS.2015.2469138]

Dong R C, Xu D Z, Jiao L C, Zhao J and An J G. 2020. A fast deep perception network for remote sensing scene classification. Remote Sensing, 12(4): #729 [DOI: 10.3390/rs12040729]

Ghosh A, Kumar H and Sastry P S. 2017. Robust loss functions under label noise for deep neural networks//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI: 1919-1925

- Ghosh A, Manwani N and Sastry P S. 2015. Making risk minimization tolerant to label noise. *Neurocomputing*, 160: 93-107 [DOI: 10.1016/j.neucom.2014.09.081]
- Hu J, Shen L and Sun G. 2018. Squeeze-and-Excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Huang B H, Lu K K, Audeberr N, Khalel A, Tarabalka Y, Malof J, Boulch A, Le Saux B, Collins L, Bradbury K, Lefèvre S and ElSaban M. 2018. Large-scale semantic classification: outcome of the first year of Inria aerial image labeling benchmark//Proceedings of IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia, Spain; IEEE: 6947-6950 [DOI: 10.1109/IGARSS.2018.8518525]
- Ishida T, Yamane I, Sakai T, Niu G and Sugiyama M. 2020. Do we need zero training loss after achieving zero training error?//Proceedings of the 37th International Conference on Machine Learning. [s. l.]: [s. n.].
- Jiang L, Zhou Z Y, Leung T, Li L J and Li F F. 2018. MentorNet: learning data-driven curriculum for very deep neural networks on corrupted labels [EB/OL]. [2020-08-03]. <https://export.arxiv.org/pdf/1712.05055.pdf>
- Jin P, Xia G S, Hu F, Lu Q K and Zhang L P. 2018. AID ++: an updated version of aid on scene classification//Proceedings of IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia, Spain; IEEE: 4721-4724 [DOI: 10.1109/IGARSS.2018.8518882]
- Kellenberger B, Marcos D and Tuia D. 2018. Detecting mammals in UAV images: best practices to address a substantially imbalanced dataset with deep learning [EB/OL]. [2020-08-03]. <https://arxiv.org/pdf/1806.11368.pdf>
- Li Y S, Zhang Y J and Zhu Z H. 2021. Error-tolerant deep learning for remote sensing image scene classification. *IEEE Transactions on Cybernetics*, 51(4): 1756-1768 [DOI: 10.1109/TCYB.2020.2989241]
- Liu Y S, Liu Y B and Ding L W. 2019. Scene classification by coupling convolutional neural networks with wasserstein distance. *IEEE Geoscience and Remote Sensing Letters*, 16(5): 722-726 [DOI: 10.1109/LGRS.2018.2883310]
- Masnadi-Shirazi H and Vasconcelos N. 2008. On the design of loss functions for classification: theory, robustness to outliers, and Savage-Boost//Advances in Neural Information Processing Systems 21 (NIPS 2008). Vancouver, Canada; NIPS: 1049-1056
- Patrini G, Rozza A, Menon A K, Nock R and Qu L Z. 2017. Making deep neural networks robust to label noise: a loss correction approach//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA; IEEE: 2233-2241 [DOI: 10.1109/CVPR.2017.240]
- Reed S, Lee H, Anguelov D, Szegedy C, Erhan D and Rabinovich A. 2015. Training deep neural networks on noisy labels with bootstrapping//Accepted as a workshop contribution at ICLR 2015. San Diego, USA; ICLR
- Ren M Y, Zeng W Y, Yang B and Urtasun R. 2018. Learning to reweight examples for robust deep learning//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden; PMLR: 4331-4340
- Roy S, Sangineto E, Demir B and Sebe N. 2019. Metric-learning based deep hashing network for content based retrieval of remote sensing images [EB/OL]. [2020-08-03]. <https://arxiv.org/pdf/1904.01258.pdf>
- Sandler M, Howard A, Zhu M L, Zhmoginov A and Chen L C. 2018. MobileNetV2: inverted residuals and linear bottlenecks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 4510-4520 [DOI: 10.1109/CVPR.2018.00474]
- Schroff F, Kalenichenko D and Philbin J. 2015. FaceNet: a unified embedding for face recognition and clustering//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA; IEEE: 815-823 [DOI: 10.1109/CVPR.2015.7298682]
- Song W W, Li S T and Benediktsson J A. 2019. Deep hashing learning for visual and semantic retrieval of remote sensing images [EB/OL]. [2020-08-03]. <https://arxiv.org/pdf/1909.04614.pdf>
- Vahdat A. 2017. Toward robustness against label noise in training deep discriminative neural networks//Advances in Neural Information Processing Systems 30 (NIPS 2017). Long Beach, USA; NIPS: 5596-5605
- van Rooyen B, Menon A and Williamson R C. 2015. Learning with symmetric label noise: the importance of being unhinged//Advances in Neural Information Processing Systems 28 (NIPS 2015). Montreal, Canada; NIPS: 10-18
- Xia G S, Hu J W, Hu F, Shi B G, Bai X, Zhong Y F, Zhang L P and Lu X P. 2017. AID: a benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7): 3965-3981 [DOI: 10.1109/TGRS.2017.2685945]
- Xia G S, Wang Z F, Xiong C M and Zhang L P. 2015. Accurate annotation of remote sensing images via active spectral clustering with little expert knowledge. *Remote Sensing*, 7(11): 15014-15045 [DOI: 10.3390/rs71115014]
- Xiao T, Xia T, Yang Y, Huang C and Wang X G. 2015. Learning from massive noisy labeled data for image classification//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA; IEEE: 2691-2699 [DOI: 10.1109/CVPR.2015.7298885]
- Yang Y and Newsam S. 2010. Bag-of-visual-words and spatial extensions for land-use classification//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Sys-

tems. San Jose, USA: ACM: 270-279 [DOI: 10.1145/1869790.1869829]

Zhang C Y, Bengio S, Hardt M, Recht B and Vinyals O. 2017. Understanding deep learning requires rethinking generalization//Proceedings of the 5th International Conference on Learning Representations. Toulon, France: ICLR: 1-15

作者简介



王亚鹏, 1995 年生, 男, 硕士研究生, 主要研究方向为遥感图像分类与检索。

E-mail: 18921555900@163.com



苗壮, 通信作者, 男, 副教授, 主要研究方向为图像视频处理。

E-mail: emiao_beyond@163.com

李阳, 男, 副教授, 主要研究方向为机器视觉与图像检索。

E-mail: solarleeon@outlook.com

王家宝, 男, 讲师, 主要研究方向为计算机视觉与机器学习。

E-mail: jiabao_1108@163.com

赵勋, 男, 硕士研究生, 主要研究方向为细粒度图像分类。

E-mail: zhaoxun1125@163.com