



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

中国图象图形学报

2021
11
VOL.26

ISSN1006-8961
CN11-3758/TB



电力视觉前沿技术



第26卷第11期（总第307期）
2021年11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

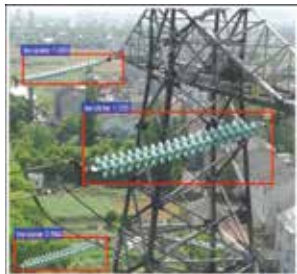
Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

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

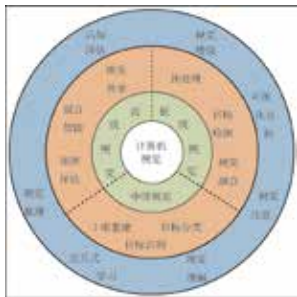
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

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



面向改进尺度缩放网络的绝缘子识别(第2561页)



混合增强视觉认知架构及其关键技术进展(第2619页)



激光点云的稀疏体素金字塔邻域构建与分类(第2703页)

编者按 I

电力视觉前沿技术

输电线路部件视觉缺陷检测综述

赵振兵, 蒋志钢, 李延旭, 戚银城, 翟永杰, 赵文清, 张珂 2545

面向改进尺度缩放网络的绝缘子识别

赵文清, 张海明, 徐敏夫 2561

最优知识传递宽残差网络输电线路螺栓缺陷图像分类

戚银城, 金超熊, 赵振兵, 丁洁涛, 吕斌 2571

可变形NTS-Net的螺栓属性多标签分类

张珂, 何颖宣, 赵凯, 冯晓晗, 赵振兵, 马占宇 2582

嵌入双注意力机制的Faster R-CNN航拍输电线路螺栓缺陷检测

戚银城, 武学良, 赵振兵, 史博强, 聂礼强 2594

轻量化航拍图像电力线语义分割

许刚, 李果 2605

综述

混合增强视觉认知架构及其关键技术进展

王培元, 关欣 2619

深度学习人脸特征点自动定位综述

徐亚丽, 赵俊莉, 吕智涵, 张志梅, 李劲华, 潘振宽 2630

图像处理和编码

多尺度特征复用混合注意力网络的图像重建

卢正浩, 刘丛 2645

图像分析和识别

自监督深度离散哈希图像检索

万方, 强浩鹏, 雷光波 2659

L-UNet: 轻量化云遮挡道路提取网络

许苗, 李元祥, 钟娟娟, 左宗成, 熊伟 2670

基于多尺度特征多对抗网络的雾天图像识别

陈硕, 钟汇才, 李勇周, 王师峰, 杨建刚 2680

图像理解和计算机视觉

结合动态图卷积和空间注意力的点云分类与分割

宋巍, 蔡万源, 何盛琪, 李文俊 2691

激光点云的稀疏体素金字塔邻域构建与分类

陶帅兵, 梁冲, 蒋腾平, 杨玉娇, 王永君 2703

计算机图形学

单幅植物叶片图像的3维重建

任非儿, 刘通, 杨龙 2713

医学图像处理

深度学习多部位病灶检测与分割

黎生丹, 柏正尧 2723

遥感图像处理

对抗学习遥感图像场景识别

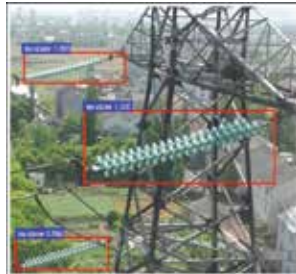
李彤, 张钧萍 2732

融合特征的卫星视频车辆单目标跟踪

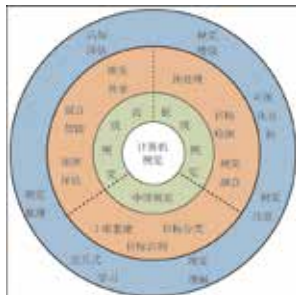
韩鸣飞, 李盛阳, 万雪, 轩诗宇, 赵子飞, 谭洪, 张万峰 2741

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Insulator recognition based on an improved scale-transferrable network(P2561)



Hybrid enhanced visual cognition framework and its key technologies(P2619)



Sparse voxel pyramid neighborhood construction and classification of LiDAR point cloud(P2703)

Frontier Technology of Power Computer Vision

Overview of visual defect detection of transmission line components

Zhao Zhenbing, Jiang Zhigang, Li Yanxu, Qi Yincheng, Zhai Yongjie, Zhao Wenqing, Zhang Ke 2545

Insulator recognition based on an improved scale-transferrable network

Zhao Wenqing, Zhang Haiming, Xu Minfu 2561

Image classification method of transmission line bolt defects using the optimal knowledge transfer wide residual network

Qi Yincheng, Jin Chaoxiong, Zhao Zhenbing, Ding Jietao, Lyu Bin 2571

Multi-label classification method of bolt attributes based on deformable NTS-Net

Zhang Ke, He Yingxuan, Zhao Kai, Feng Xiaohan, Zhao Zhenbing, Ma Zhanyu 2582

Bolt defect detection for aerial transmission lines using Faster R-CNN with an embedded dual attention mechanism

Qi Yincheng, Wu Xueliang, Zhao Zhenbing, Shi Boqiang, Nie Liqiang 2594

Research on lightweight neural network of aerial powerline image segmentation

Xu Gang, Li Guo 2605

Review

Hybrid enhanced visual cognition framework and its key technologies

Wang Peiyuan, Guan Xin 2619

Automatic facial feature points location based on deep learning: a review

Xu Yali, Zhao Junli, Lyu Zhihan, Zhang Zhimei, Li Jinhua, Pan Zhenkuan 2630

Image Processing and Coding

Multiscale feature reuse mixed attention network for image reconstruction

Lu Zhenghao, Liu Cong 2645

Image Analysis and Recognition

Self-supervised deep discrete hashing for image retrieval

Wan Fang, Qiang Haopeng, Lei Guangbo 2659

L-UNet: lightweight network for road extraction in cloud occlusion scene

Xu Miao, Li Yuanxiang, Zhong Juanjuan, Zuo Zongcheng, Xiong Wei 2670

Haze image recognition based on multi-scale feature and multi-adversarial networks

Chen Shuo, Zhong Huicai, Li Yongzhou, Wang Shizheng, Yang Jiangang 2680

Image Understanding and Computer Vision

Dynamic graph convolution with spatial attention for point cloud classification and segmentation

Song Wei, Cai Wanyuan, He Shengqi, Li Wenjun 2691

Sparse voxel pyramid neighborhood construction and classification of LiDAR point cloud

Tao Shuaibing, Liang Chong, Jiang Tengping, Yang Yujiao, Wang Yongjun 2703

Computer Graphics

3D reconstruction of a single plant leaf image

Ren Feier, Liu Tong, Yang Long 2713

Medical Image Processing

Multiorgan lesion detection and segmentation based on deep learning

Li Shengdan, Bai Zhengyao 2723

Remote Sensing Image Processing

Remote sensing image scene recognition based on adversarial learning

Li Tong, Zhang Junping 2732

Integrating multiple features for tracking vehicles in satellite videos

Han Mingfei, Li Shengyang, Wan Xue, Xuan Shiyu, Zhao Zifei, Tan Hong, Zhang Wanfeng 2741

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2021)11-2680-11

论文引用格式: Chen S, Zhong H C, Li Y Z, Wang S Z and Yang J G. 2021. Haze image recognition based on multi-scale feature and multi-adversarial networks. *Journal of Image and Graphics*, 26(11): 2680-2690 (陈硕, 钟汇才, 李勇周, 王师峥, 杨建刚. 2021. 基于多尺度特征多对抗网络的雾天图像识别. *中国图象图形学报*, 26(11): 2680-2690) [DOI:10.11834/jig.200491]

基于多尺度特征多对抗网络的雾天图像识别

陈硕^{1,2}, 钟汇才^{1*}, 李勇周¹, 王师峥¹, 杨建刚¹

1. 中国科学院微电子研究所, 北京 100029; 2. 中国科学院大学, 北京 100049

摘要: 目的 当前的大型数据集, 例如 ImageNet, 以及一些主流的网络模型, 如 ResNet 等能直接高效地应用于正常场景的分类, 但在雾天场景下则会出现较大的精度损失。雾天场景复杂多样, 大量标注雾天数据成本过高, 在现有条件下, 高效地利用大量已有场景的标注数据和网络模型完成雾天场景下的分类识别任务至关重要。方法 本文使用了一种低成本的数据增强方法, 有效减小图像在像素域上的差异。基于特征多样性和特征对抗的思想, 提出多尺度特征多对抗网络, 通过提取数据的多尺度特征, 增强特征在特征域分布的代表性, 利用对抗机制, 在多个特征上减少特征域上的分布差异。通过缩小像素域和特征域分布差异, 进一步减小领域偏移, 提升雾天场景的分类识别精度。结果 在真实的多样性雾天场景数据上, 通过消融实验, 使用像素域数据增强方法后, 带有标签的清晰图像数据在风格上更趋向于带雾图像, 总的分类精度提升了 8.2%, 相比其他的数据增强方法, 至少提升了 6.3%, 同时在特征域上使用多尺度特征多对抗网络, 相比其他的网络, 准确率至少提升了 8.0%。结论 像素域数据增强以及多尺度特征多对抗网络结合的雾天图像识别方法, 综合考虑了像素域和特征域的领域分布差异, 结合了多尺度的丰富特征信息, 同时使用多对抗来缩小雾天数据的领域偏移, 在真实多样性雾天数据集上获得了更好的图像分类识别效果。

关键词: 雾天场景; 图像识别; 多尺度特征; 多对抗; 领域偏移

Haze image recognition based on multi-scale feature and multi-adversarial networks

Chen Shuo^{1,2}, Zhong Huicai^{1*}, Li Yongzhou¹, Wang Shizheng¹, Yang Jiangang¹

1. Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Objective While dealing with high-resolution remote sensing image scene recognition, classical supervised machine learning algorithms are considered effective on two conditions, namely, 1) test samples should be in the same feature space with training samples, and 2) adequate labeled samples should be provided to train the model fully. Deep learning algorithms, which achieve remarkable results in image classification and object detection for the past few years, generally require a large number of labeled samples to learn the accurate parameters. The main image classification methods select training and test samples randomly from the same dataset, and adopt cross validation to testify the effectiveness of the model. However, obtaining scene labels is time consuming and expensive for remote sensing images. To deal with the insufficiency of labeled samples in remote sensing image scene recognition and the problem that labeled samples cannot be shared between different datasets due to different sensors and complex light conditions, deep learning architecture and adversarial

收稿日期: 2020-08-19; 修回日期: 2020-12-08; 预印本日期: 2020-12-15

* 通信作者: 钟汇才 zhonghuicai@ime.ac.cn

learning are investigated. A feature transfer method based on adversarial variational autoencoder is proposed. **Method** Feature transfer architecture can be divided into three parts. The first part is the pretrain model. Given the limited samples with scene labels, the unsupervised learning model, variational autoencoder (VAE), is adopted. The VAE is unsupervised trained on the source dataset, and the encoder part in the VAE is finetuned together with classifier network using labeled samples in the source dataset. The second part is adversarial learning module. In most of the research, adversarial learning is adopted to generate new samples, while the idea is used to transfer the features from source domain to target domain in this paper. Parameters of the finetuned encoder network for the source dataset are then used to initialize the target encoder. Using the idea of adversarial training in generative adversarial networks (GAN), a discrimination network is introduced into the training of the target encoder. The goal of the target encoder is to extract features in the target domain to have as much affinity to those of the source domain as possible, such that the discrimination network cannot distinguish the features are from either the source domain or target domain. The goal of the discrimination network is to optimize the parameters for better distinction. It is called adversarial learning because of the contradiction between the purpose of encoder and discrimination network. The features extracted by the target encoder increasingly resemble those by the source encoder by training and updating the parameters of the target encoder and the discrimination network alternately. In this manner, by the time the discrimination network can no longer differentiate between source features and target features, we can assume that the target encoder can extract similar features to the source samples, and remote sensing feature transfer between the source domain and target domain is accomplished. The third part is target finetuning and test module. A small number of labeled samples in target domain is employed to finetune the target encoder and source classifier, and the other samples are used for evaluation. **Result** Two remote sensing scene recognition datasets, UCMerced-21 and NWPU-RESISC45, are adopted to prove the effectiveness of the proposed feature transfer method. SUN397, a natural scene recognition dataset is employed as an attempt for the cross-view feature transfer. Eight common scene types between the three datasets, namely, baseball field, beach, farmland, forest, harbor, industrial area, overpass, and river/lake, are selected for the feature transfer task. Correlation alignment (CORAL) and balanced distribution adaptation (BDA) are used as comparisons. In the experiments of adversarial learning between two remote sensing scene recognition datasets, the proposed method boosts the recognition accuracy by about 10% compared with the network trained only by the samples in the source domain. Results improve more substantially when few samples in the target domain are involved. Compared with CORAL and BDA, the proposed method improves scene recognition accuracy by more than 3% when using a few samples in the target domain and between 10%~40% without samples in the target domain. When using the information of a natural scene image, the improvement is not as much as that of a remote sensing image, but the scene recognition accuracy using the proposed feature transfer method is still increased by approximately 6% after unsupervised feature transfer and 36% after a small number of samples in the target domain are involved in finetuning. **Conclusion** In this paper, an adversarial transfer learning network is proposed. The experimental results show that the proposed adversarial learning method can make the most of sample information of other dataset when the labeled samples are insufficient in the target domain. The proposed method can achieve the feature transfer between different datasets and scene recognition effectively, and remarkably improve the scene recognition accuracy.

Key words: haze scene; image recognition; multiple features; multiple adversarial; domain shift

0 引言

雾天作为一个重要的场景广泛存在于日常生活中,而针对雾天场景的各种任务需求也变得十分迫切,逐渐成为计算机视觉的热点之一。为了成功地完成雾天图像分类识别任务,需要事先为每个特定雾天场景提供足够数量的手动标注图像。然而在实

际情况下,获得足够的多样性雾天标记数据样本的成本较高,而随着大量标记训练数据和大量计算资源的出现,深度神经网络可以达到令人满意的性能(He等,2016),这些数据和深度神经网络模型可以直接广泛地应用于多数场景下的图像分类。然而,在处理雾天场景下的分类时会产生极大的分类误差,要想利用现有条件进行雾天场景的图像识别,就要解决数据分布差异的问题(Pan和Yang,2010)。

而雾天场景数据的分布较其他常规的情况有着更大的差异,同时,具有这样分布差异的数据也存在更大领域偏移(domain shift)。卷积神经网络难以学习到数据的不变性特征(Pan 和 Yang,2010)。

针对上述问题,现有研究方法提供了一些解决手段,主要分为两类。一方面在现有的卷积神经网络的基础上,结合数据增强的方法,例如使用 PMS-Net(robust haze removal based on patch map for single images)(Chen 等 2019)、DCPDN(densely connected pyramid dehazing network)(Zhang 和 Patel,2018)等一些去雾算法对雾天数据去雾,以缩小与正常场景数据的分布差异。然而大多数去雾算法需要大量成对的图像进行训练,由于雾天环境的特殊性,训练数据中雾天场景图像多是合成的,很难在像素域(Bousmalis 等,2017)上减小分布差异,根据 Pei 等人(2018a)的实验,利用去雾算法进行数据增强在真实场景的雾天分类识别任务上没有显著的增益,同时收集和生成更多的数据还会加大训练时间和成本,此外还有利用大气退化模型(He 等,2011)对训练数据进行加雾处理。然而,这种方法难以适用于真实多样性的雾天场景。

另一方面,由于不同数据之间会存在不同的分布差异,仅使用数据增强深度神经网络难以在特征域(Kouw 等,2016)上提取到不变性特征,而领域自适应算法(Pan 和 Yang,2010)可以缩小领域偏移,学习不同数据的不变性特征。现有的领域自适应算法通过不同特征对齐方法对卷积神经网络提取的特征进行对齐,其中一些领域自适应算法是基于最大均值差异来减小领域偏移,如 DANN(domain adaptive neural network)(Ghifary 等,2014)使用最大均值差异度量减少由从不同域抽取的样本在两个隐藏层之间的特征分布;DDC(deep domain confusion)(Tzeng 等,2014)利用适应层将两个并行的 CNN(convolutional neural network)网络连接到一起,同样通过 MMD(maximum mean discrepancy)定义损失函数,缩小源域和目标域的领域偏移;DeepCORAL(correlation alignment for deep domain adaptation)(Sun 和 Saenko,2016);基于 CORAL 方法用非线性变换方法将源域和目标域分布的二阶统计特征进行对齐。另一种基于对抗的思想,如 DANN-A(domain-adversarial training of neural networks)(区别于 DANN,该方法使用了对抗)(Ganin 等,2017),通过

对抗网络来获取具有领域不变性的特征;MCD(maximum classifier discrepancy)(Saito 等,2018)利用两个不同分类器在目标样本预测上的差异来检测这些远离支持点的目标样本,找到两个分类器的预测差异,固定分类器,训练特征生成器,让生成的特征更靠近源域。现有领域自适应算法大都关注于对齐单一特征的边缘分布,而单一特征仅仅代表少部分数据的特征信息,其中部分领域自适应算法在出现较大的领域偏移时还会出现训练不稳定,鲁棒性欠佳等情况。

针对以上问题提出一个框架,首先使用 Cycle-GAN(unpaired image-to-image translation using cycle-consistent adversarial networks)(Zhu 等,2017)进行数据增强,一方面在正常场景下(源域)进行数据增强,可以生成带标签并且风格与雾天场景相近的数据;另一方面 Cycle-GAN 使用了常规数据和雾天场景数据(目标域)进行对抗训练,能在像素域上减小二者的数据分布差异,由于 Cycle-GAN 不需要使用成对图像,也就不需要额外的数据,极大地节省了训练时间和训练成本。其次提出一个有效的自适应网络,即多尺度特征多对抗网络(multi-scale features and multi-adversarial network, MFMAN),为了在特征域上充分对齐数据分布,减小领域偏移,提出了 IAM(inception adversarial module)感知对抗模块,在通过卷积神经网络进行特征提取后,使用 IAM 提取多个不同尺度的特征来替代单一特征,充分利用卷积神经网络提取的特征信息,多个域判别器形成对抗减少特征分布差异。其次为了利用复杂的多模结构,在类别上使用多个判别器来学习数据分布的多模结构(Pei 等,2018b),增加了对类别的判别能力,从而增加了正迁移,减少了负迁移,特征得以在各个类别上减小域差异,学习到领域不变性特征。

本文贡献主要在于:1)提出真实雾天场景数据,包含众多类似场景,如霾、烟、沙尘等,其中雾的浓度也具有多样性。2)提出一个雾天图像识别框架。其中使用 Cycle-GAN 作为数据增强方法,有效地减少雾天场景在像素域上与正常场景的分布差异,生成有标注的类雾天风格数据集,降低标注成本。3)提出多尺度特征多对抗网络提取更多样的数据特征,在多尺度特征和类别上使用多个判别器形成多对抗,减小数据分布,缩小领域差异,进一步提高雾天图像分类识别精度。这里提出的雾天图像

识别框架在提出的数据集上取得一个较好的表现。

1 研究方法

1.1 数据集

根据实验需求,需要真实的雾天场景分类数据集和正常场景分类数据集用来进行训练和测试,当前一些较大的公开数据集,例如 Caltech-256, PAS-CALVOC (pattern analysis, statistical modeling and computational learning visual object classes) 和 ImageNet 主要是正常场景下的图像,这里使用的正常场

景下的图像数据主要来源于这些大型开源数据集,提取其中 12 个类别,组成 RealClear-12,这些数据通常便于获取,无需标注。雾天图像主要在网络上的图像、视频中获取,逐幅进行筛选,手动分类进行标注,将其命名为 RealHaze-12,雾天的图像数据覆盖了日常生活中的多数场景,雾天场景雾的浓度和雾的种类也十分丰富,包含雾、霾、烟和沙尘等。RealClear-12 共有 4 797 幅,其中源域数据 3 602 幅,源域测试集 1 195 幅;RealHaze-12 共有 2 306 幅作为目标域数据,图像的详细信息见表 1,数据集中部分图像展示见图 1。

表 1 RealClear-12 和 RealHaze-12 的详细信息

Table 1 RealClear-12 and RealHaze-12 datasets specific information

名称	truck	tree	train	streetlamp	plane	people	car	bus	building	bridge	boat	bird	total
RealClear-12 (train)	272	304	303	282	336	297	304	282	306	314	300	302	3 602
RealClear-12 (test)	92	101	100	94	112	98	101	93	102	104	100	100	1 195
RealHaze-12	171	202	206	171	231	194	205	185	202	213	200	182	2 362

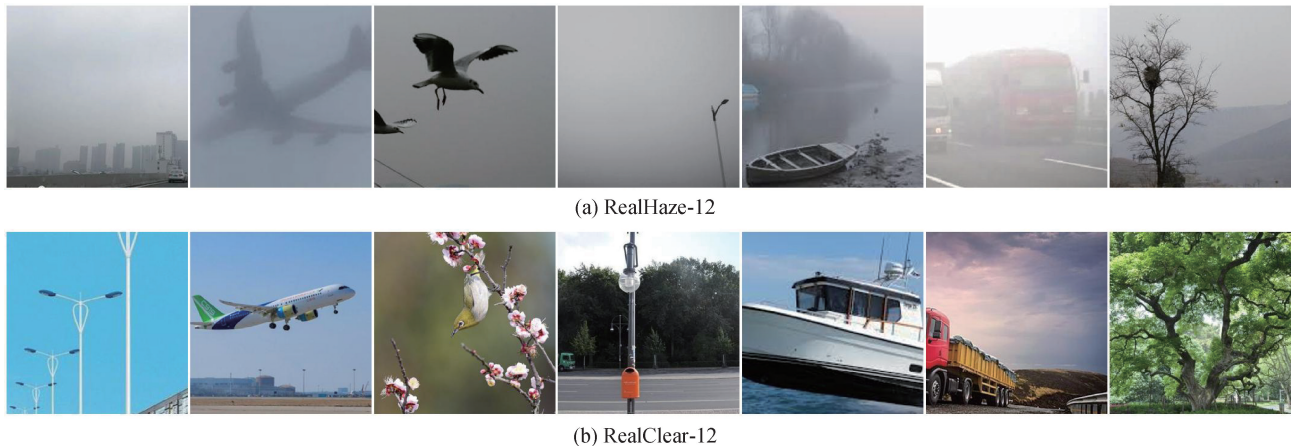


图 1 部分 RealHaze-12 图像和 RealClear-12 图像

Fig. 1 Partial RealHaze-12 images and RealClear-12 images ((a) RealHaze-12; (b) RealClear-12)

1.2 框架结构

1.2.1 数据增强

使用 Cycle-GAN 算法 (Zhu 等, 2017) 如图 2 所示进行数据增强, 其结构包括一对生成网络和一对判别网络, 输入一组非对齐的图像进行训练来学习输入图像和输出图像之间的像素级映射, 将图像从源域 A 转换为目标域 B 。目标是学习一个映射 $G_{AB}: A \rightarrow B$, 使用判别器 D_B 拉近 $Fake_B$ 与 B 的分布, 同理在 $G_{BA}: B \rightarrow A$, 使用判别器 D_A 拉近 $Fake_A$ 与 A 的分布,

并引入循环一致性损失以强制执行 $F(G(A) \approx B$ (反之亦然), 在使用 Cycle-GAN 进行处理时, 其中 Real-

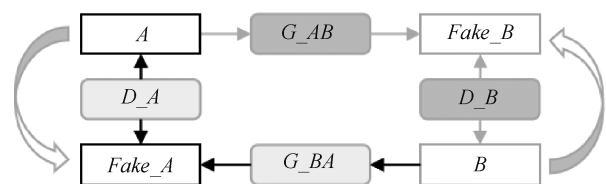


图 2 循环生成对抗网络结构

Fig. 2 Cycle-consistent adversarial network structure

Clear-12 作为源域 A , RealHazer-12 作为目标域 B , 通过这样的循环对抗网络结构既可以得到与雾天场景风格类似的有标签数据, 也就是 $Fake_B$, 同时在像素域上初步缩小两个场景数据的分布差异。

1. 2. 2 多尺度特征多对抗网络

在利用 Cycle-GAN 进行像素域的处理后, Real-Clear-12 的有标注数据生成了与 RealHaze-12 风格接近的数据, 在特征域上使用此数据作为源域数据进一步在特征域上利用多尺度特征多对抗网络处理。网络结构如图 3 所示, 由深度特征提取器 (feature extractor)、全局域判别器 (global domain discriminator)、局部域判别器 (local domain discriminator)、标签分类器 (label classier) 组成。首先第 1 部分的深度特征提取器分别由图 3 中的卷积神经网络 CNN 以及 IAM 模块构成, 特征提取器去掉卷积神经网络的全连接层, 连接 IAM 模块, 该部分的主要功能是在特定的空间中将数据进行映射, 该部件的主要功能是提取深度特征, 而加入 IAM 模块则会对数据包含的不同尺度大

小的特征进行提取。第 2 部分标签分类器对来自源域的数据进行分类, 尽可能分出正确的标签。第 3 部分域判别器, 是为了对特征空间的数据进行分类, 尽可能分出数据来自哪个域, 由全局域判别器和局部域判别器组成, 其中局部域判别器使用 IAM 中提取的多尺度特征融合后 (图 3 中 Feature f) 与每一个类别预测结果 $\hat{y}$ 相乘, 依次连接与类别数量相同的判别器 (所用数据具有几个类别, 则会针对每一个类别进行域判别) 能够在一定程度上将来源于不同数据域的数据进行区分, 全局域判别器则由 IAM 和感知对抗模块组成, 通过感知对抗模块提取多尺度的特征, 每一个特征都进行域判别, 形成全局域判别。特征提取器和标签分类器这两个部件组成了一个可以前向传播的神经网络, 特征提取器后面接上判别器, 对数据域的来源进行判断, 原理与领域自适应对抗网络相同, 中间通过一个梯度反转层连接, 随后可以通过这个反转层, 与标签分类器损失通过极大极小的策略实现对抗, 进行领域自适应, 最终达到缩小不同数据的域偏移。

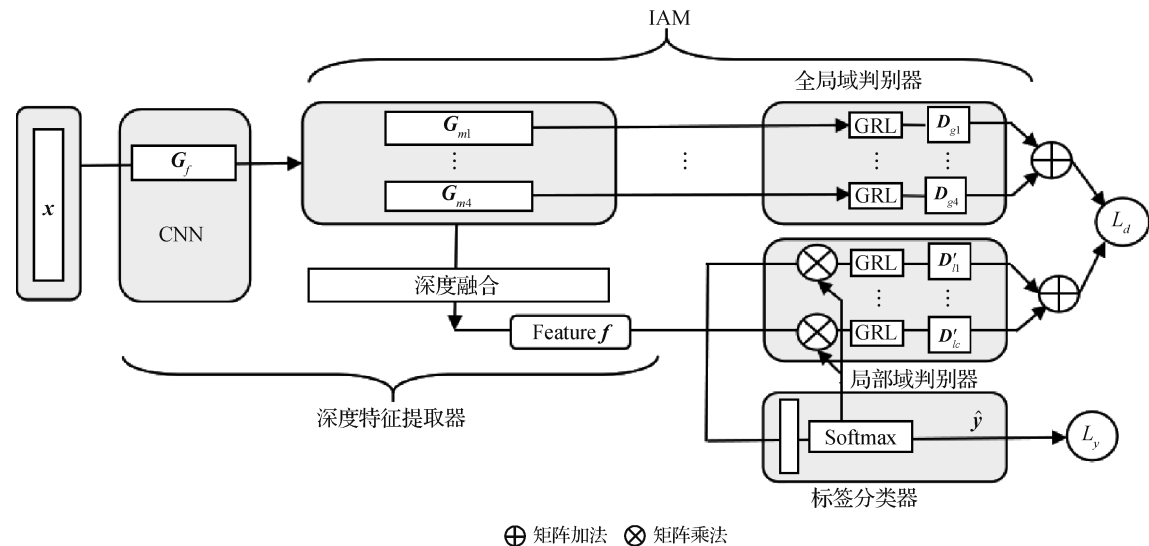


图 3 多尺度特征多对抗网络结构

Fig. 3 Multi-scale features and multi-adversarial network structure

1) 多尺度感知对抗模块 (inception adversarial module, IAM)。针对特征提取器只能提取单一特征进行改进, 由于不同的网络结构会提取出不同的特征, 因此采用一种特殊的结构对特征进行多尺度的提取, 在该结构中采用的构造参照了 GoogLeNet (Szegedy 等, 2015) 中的 Inception 模块。如图 4 所示, IAM 连接去掉全连接层的卷积神经网络 (图中 previous layer), 使用感知模块中的一种提取多尺度

特征的结构, 此结构利用卷积神经网络中常用的结构, 例如 $(1 \times 1, 3 \times 3, 5 \times 5)$ 等卷积、 (3×3) 池化操作组合起来, 其中的池化以及卷积后的尺寸一样, 通道相加, 极大地扩展了神经网络的横向宽度, 与此同时也有效地提升了网络针对不同尺度的适应性。使用的卷积层会取得输入数据特征的所有详细信息, 与此同时, 5×5 的滤波器也能够覆盖其中大部分接收层的特征。特征的空间大小经过处理, 可以降低

过拟合。在这些层之上,使用线性激活(ReLU)来增强网络的非线性特征。在IAM模块中, 1×1 卷积的功能是减少在之前的特征融合处理中产生的较大的维度。同时也能够对线性激活函数进行修正。如图4所示,在不同尺度特征后均连接一个域判别器,用于对齐多个不同尺度特征的分布,可以充分利用多个不同特征分布代替单一特征分布。

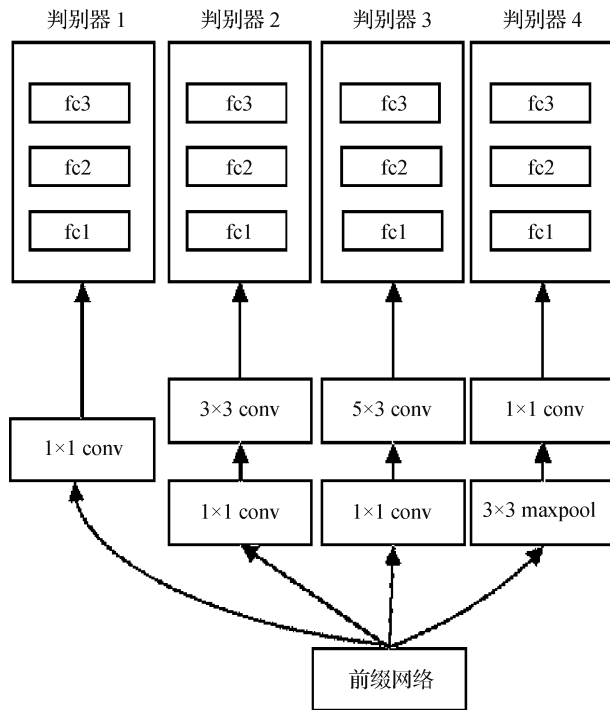


图4 IAM 的结构

Fig.4 The structure of IAM

2)全局域判别器(如图3)。用来对齐提取的多尺度特征,每一个域判别器类似 DANN(Tzeng 等, 2014)中的通用的域判别器。特征提取器提取出4个尺度不同的特征,这里以其中的一个特征为例分析全局域判别器,其中卷积神经网络特征提取器为 G_f ,图中 G_{m1} 到 G_{m4} 为IAM中的多尺度特征提取单元,全局域判别器为 D_g ,输入为 x ,其中一个特征的域判别器损失函数定义为

$$L_{g1} = \frac{1}{n_s + n_t} \sum_{x_i \in D_s \cup D_t} L_{d1}(D_g(G_{m1}(G_f(x_i))), d_i) \quad (1)$$

式中, n_s , n_t 分别为源域和目标域的数据数量, D_s 和 D_t 分指代源域和目标域, x_i 为输入样本, d_i 是领域标签, L_{d1} 为每一个样本的域判别器损失(cross-entropy),则全局域判别损失函数定义为

$$L_g = \sum_{i=1}^4 L_{gi} \quad (2)$$

最后的全局域判别器损失函数为 L_g 。

3)局部域判别器。从图3可以看到,在局部域判别器需要使用标签分类器的预测结果与特征提取器(feature extractor)提取的特征 f 相乘,其中IAM中提取的多尺度特征融合为

$$G_m = \sum_{j=1}^{n_f} G_{mj} \quad (3)$$

所以,局部域判别器的损失函数定义为

$$L_l = \frac{1}{n_s + n_t} \sum_{c=1}^C \sum_{x_i \in D_s \cup D_t} L_d^c(D_l(\hat{y}_i^c G_m(G_f(x_i))), d_i) \quad (4)$$

式中, C 为数据包含的类别数, D_l 为局部域判别器, L_d^c 为每一个类别的局部判别损失, $\hat{y}_i^c$ 是标签分类器预测结果,标签预测的结果为 C 类,同样设置 C 个域判别器,用相乘的结果来表示数据在 C 个判别领域的参与程度,根据 Pei 等人(2018b)使用多个域判别器进行对抗训练会加强正迁移,最大化对齐不同域的多模结构,同时会减弱负迁移,避免不同域分布模式的错误对齐。

4)标签分类器(label classifier)。使用全连接层连接,最后利用 cross-entropy 损失进行训练,标签分类器的类别损失定义为

$$L_y = \frac{1}{n_s} \sum_{x_i \in D_s} \sum_{c=1}^C P_{x_i \rightarrow c} \log G_y(G_m(G_f(x_i))) \quad (5)$$

式中, x_i 是有标注的源域数据, $P_{x_i \rightarrow c}$ 是 x_i 属于 C 个类别中的预测, G_y 是标签类别分类器。

1.3 损失函数及训练过程

1)多元特征多对抗网络的损失函数为全局域判别器的损失函数、局部域判别器的损失函数以及标签预测器的损失函数混合而成,定义为

$$L(\theta_f, \theta_m, \theta_y, \theta_d, \theta_d^c |_{c=1}^C) = L_y + \lambda(L_g + L_l) \quad (6)$$

式中, θ_f 为骨干网络参数, θ_m 为多尺度特征参数, θ_y 为标签分类器参数, θ_d 为全局域判别器损失, θ_d^c 为局部域判别器损失, λ 为超参数,用来决定局部损失和全局损失在总体损失中所占的比重。

$$\theta_f \leftarrow \theta_f - \mu \left(\frac{\partial L_y}{\partial \theta_f} - \lambda \left(\frac{\partial L_g}{\partial \theta_f} - \frac{\partial L_l}{\partial \theta_f} \right) \right) \quad (7)$$

$$\theta_m \leftarrow \theta_m - \mu \left(\frac{\partial L_y}{\partial \theta_m} - \lambda \left(\frac{\partial L_g}{\partial \theta_m} - \frac{\partial L_l}{\partial \theta_m} \right) \right) \quad (8)$$

$$\theta_y \leftarrow \theta_y - \mu \frac{\partial L_y}{\partial \theta_y} \quad (9)$$

$$\theta_m \leftarrow \theta_m - \mu \frac{\partial L_m}{\partial \theta_m} \quad (10)$$

$$\theta_d \leftarrow \theta_d - \mu \frac{\partial L_g}{\partial \theta_d} \quad (11)$$

$$\theta_d^c \leftarrow \theta_d^c - \mu \frac{\partial L_l}{\partial \theta_d^c} \quad (12)$$

式中, μ 为学习率。

2) 优化过程, 其中使用随机梯度下降 (stochastic gradient descent, SGD) 作为优化器优化上面的参数, 但是在式 (7) 中, 域判别器损失和标签分类器损失符号相反, 由于域判别损失最小化会使两个数据的领域偏移越来越大, 达不到需要对抗的目的, 类似的优化器, 如 Adagrad (Duchi 等, 2011) 或 Adaelta (Zeiler, 2012) 都无法进行优化, 在这里加入 GRL (gradient reversal layer) (Ganin 等, 2017), 该网络层中没有参数与之相关, 在前向传播时是一个正常的连接层, 不会对网络传播产生影响, 在反向传播的时候会改变梯度的符号, 将其乘以 -1 , 之后继续反向传播, 通过这个网络层就会达到反转梯度的目的, 从而能够使用优化器进行优化。

2 实验与分析

2.1 数据集和实验设置

在此使用提出的数据集 RealHaze-12 和 RealClear-12, 如图 1 所示的部分数据集样例, 其中包含大量浓度种类不同的雾天场景, 数据集具有多样性, 雾天图像多取自日常生活中的常见场景, 选取 RealClear-12 中的 3 602 幅图像作为源域数据, 选取其中的 1 195 幅图像作为源域数据的测试集, 用于训练基础的分类网络, 选取 RealHaze-12 中的 2 306 幅图像作为雾天分类识别的目标域数据集。

实验平台硬件配置为 Intel (R) Xeon (R) Gold 5118 CPU @ 2.30 GHz、内存 256 GB、4 张 NVIDIA Tesla (16 GB) 显卡, 软件配置为 Centos7.5 系统、使用 pytorch 深度学习框架。将所有图像裁剪到 256×256 像素, 再使用 random crop 随机裁剪成 224×224 像素, 基于 GPU 显存利用率, 以及训练的稳定性, batchsize 设置为 128, 使用预训练的骨干网络初始学习率为 0.001, 其他网络层结果的初始学习率为 0.01, SGD 作为优化方案, 学习率变化策略遵循已有工作, 权重衰减为 0.000 5, 动量设置为

0.9, 共训练 50 个轮次。根据已有经验 (Wang 等, 2018), 进行网格化搜索实验, 将参数由 0 到 1, 以 0.1 为单位进行实验, 在本数据集上 λ 取 0.3 时, 收敛更快、精度更高。实验主要测试各个类别的准确预测数量与目标域中类别数量的比值作为类别精度, 全部类别准确预测数量与雾天场景数据 (目标域) 数量的比值作为总的精度 (Tot)。

2.2 实验流程

为了有效评估算法框架在雾天无监督分类的有效性, 使用经典网络: ResNet50 (He 等, 2016) 作为基础分类网络以及骨干网络, 选取的基础分类网络以及领域自适应算法有: ResNet50、DANN、DDC (deep domain confusion)、D-CORAL (correlation alignment for deep domain adaptation)、DANN-A、MCD。为了验证多尺度特征的有效性和数据增强的有效性, 进行了消融实验, 针对不同数据增强方法在基础分类网络上进行实验。1) 为了验证使用 Cycle-GAN 在像素域进行数据增强相比于其他去雾算法更有效, 在此处以 RealClear-12 作为训练集, 在基础的分类网络 (ResNet50) 上进行训练, RealHaze-12 作为测试集, 首先利用去雾算法对雾天数据进行数据增强, 也就是对测试集 RealHaze-12 去雾, 其中分别使用了经典的以及当前主流的去雾算法, 如 DCP (single image haze removal using dark channel prior) (He 等, 2011)、DCPDN (densely connected pyramid dehazing network) (Zhang 和 Patel, 2018)、PMS-Net (Chen 等, 2019)、PMHLD (patch map-based hybrid learning DehazeNet for single image haze removal) (Chen 等, 2020)。与之相比, Cycle-GAN 是对训练集 RealClear-12 进行数据增强, 生成类似雾天场景的数据, 在测试集 RealHaze-12 上进行测试。2) 对比多尺度特征多对抗网络和不添加感知对抗 (IAM) 的网络, 为了验证感知对抗模块对分类准确度的影响, 均使用 Cycle-GAN 对 RealClear-12 进行数据增强后生成的数据作为源域数据, 使用 RealHaze-12 作为目标域数据, 未加 IAM 的算法为 MFMAN (N-IAM)。3) 最终同样使用 Cycle-GAN 对 RealClear-12 进行数据增强后生成的数据作为源域数据, 使用 RealHaze-12 作为目标域数据, 验证以上提出特征域领域自适应网络 (MFMAN) 相比于其他领域自适应算法的有效性。

2.3 结果分析

1) 表 2 中的 ResNet50 是使用 RealClear-12 作为

训练集, RealHaze-12 作为测试集的结果, 通过表 2 可以看到利用去雾算法对 RealHaze-12 数据进行去雾的数据增强, 发现其中 PMS-Net、DCPDN、PMHLD 算法相比于表 2 中的 ResNet50 在分类精度上分别有 0.8%、1.5%、1.9% 的增长, 而 DCP 算法则出现了 1.6% 的下降, 根据实验可以看出使用去雾的算法进行数据增强在真实多样性雾天场景下难以取得较好的效果, 多数去雾算法使用的合成的成对雾天数据集训练出的去雾模型会在合成数据集上产生过拟合, 并且不能很好地拉近像素域的分布, 去雾算法会对图像造成某些信息丢失 (Pei 等, 2018a)。使用 Cycle-GAN

处理的图像如图 5 所示。其中由表 2 中 Cycle-GAN 的结果, 本文提出使用 Cycle-GAN 进行类似加雾的数据增强, 在 3 个基础分类识别网络中均能取得最好的效果, 准确率相比直接识别雾天图像有了明显的提升。2) 由表 3 中 MFMAN 和 MFMAN(N-IAM) 的实验结果看, 使用感知对抗模块分类准确度提升了 3.3%, 多尺度特征能够比较充分地对齐特征域上的特征分布, 有效减小领域偏移。3) 最后以 MFMAN(IAM) 对比表 3 中的其他领域自适应算法, 提出的多尺度特征多对抗网络相比于其他的领域自适应算法在目标域 (RealHaze-12) 上的识别精度都有较大的提升。

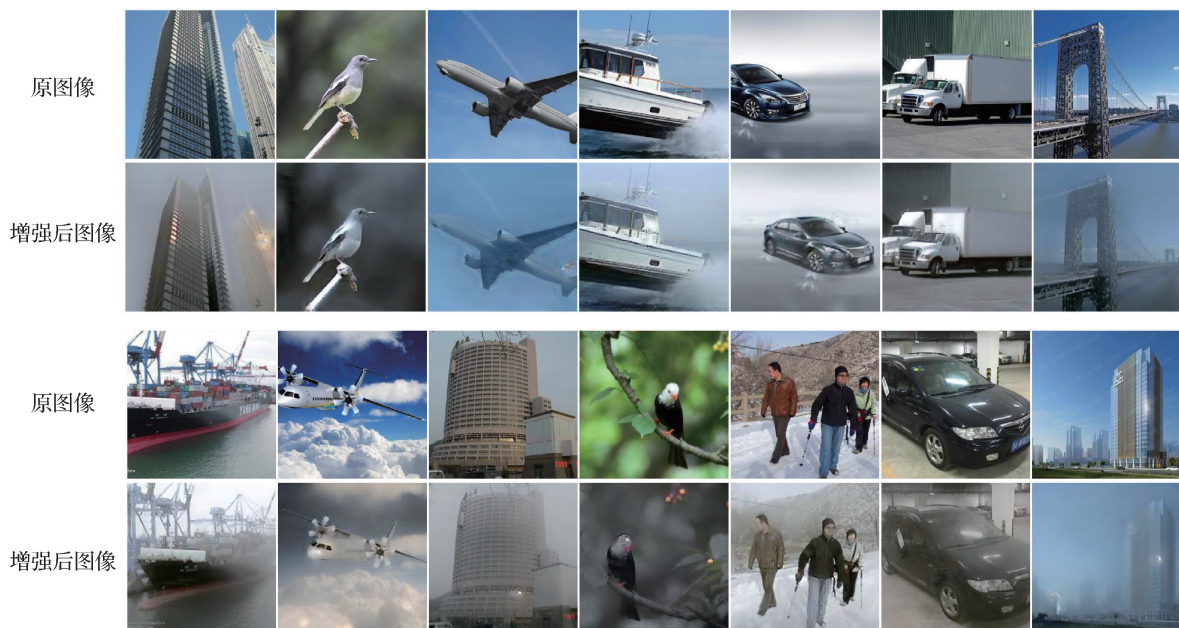


图5 数据增强后的图像与原图像

Fig. 5 After data argument images and original images

表2 雾天数据使用去雾算法在 ResNet50 上的分类精度结果

Table 2 The results of classification accuracy of haze data on ResNet50 using haze remove algorithms

方法	truck	tree	train	streetlamp	plane	people	car	bus	building	bridge	boat	bird	Tot
ResNet50	58.4	92.0	45.6	94.7	81.3	84.0	51.7	32.9	76.7	97.1	84.0	70.3	72.7
DCP	53.8	98.0	56.3	87.1	79.6	79.3	29.7	32.9	90.5	93.8	80.0	67.5	71.1
DCPDN	53.2	95.5	66.0	90.6	88.7	69.5	54.6	31.8	88.1	92.4	78.5	74.1	74.2
PMS-Net	54.9	94.5	68.9	93.5	81.8	77.8	40.9	33.5	78.7	93.4	85.0	79.1	73.5
PMHLD	54.9	97.5	71.3	91.2	80.9	71.1	49.7	46.4	90.5	92.0	84.5	78.0	74.6
Cycle-GAN	65.4	94.0	72.3	94.1	91.7	85.0	62.4	49.7	91.5	98.1	90.5	71.4	80.9

注: 加粗字体为每列最优值, Tot 表示总精度。

表 3 使用 Cycle-GAN 算法处理后的分类精度
Table 3 Classification accuracy after processing with Cycle-GAN algorithms

方法	数据集类别													/%
	truck	tree	train	streetlamp	plane	people	car	bus	building	bridge	boat	bird	Tot	
ResNet50	65.4	94.0	72.3	94.1	91.7	85.0	62.4	49.7	91.5	98.1	90.5	71.4	80.9	
DANN-A	64.9	93.5	78.1	88.3	95.2	89.1	69.2	54.5	90.5	99.0	89.0	72.5	82.6	
DDC	70.7	92.0	81.5	89.4	94.8	86.0	73.6	52.9	91.5	99.5	88.0	72.5	83.3	
D-CORAL	69.0	92.5	81.5	89.4	94.8	86.5	68.2	54.5	91.5	99.0	90.0	69.7	82.9	
DANN	42.6	87.6	52.4	88.8	86.6	81.9	72.1	46.4	87.1	81.2	73.5	76.9	73.1	
MCD	73.0	95.5	81.5	89.4	90.9	85.5	71.2	51.3	94.0	99.0	90.0	70.3	83.9	
MFMAN(N-IAM)(本文)	65.4	94.5	79.1	91.2	93.9	89.6	68.7	52.9	92.0	98.1	89.0	75.2	83.1	
MFMAN(IAM)(本文)	70.7	96.0	84.4	86.5	93.9	89.6	85.8	61.6	92.0	97.1	90.5	81.3	86.4	

注:加粗字体为每列最优值。

对比表 2 中 ResNet50 的结果,这里提出的使用 Cycle-GAN 先进行数据增强,再通过 MFMAN 进行减小数据分布、缩小领域偏移的框架总体精度提升了 13.7%,有效地改善了无标签雾天数据集在现有标注数据和基于基础网络条件下表现不佳的情况。在训练过程中的在雾天数据集上的精度变化曲线如图 6 所示,可以看出 MFMAF(IAM)在训练时精度初始值相较于大多数算法来说都比较高,准确率增长更快,在图中可以看到当前提出的算法精度 5 个 epoch 后就基本保持稳定,而且不会出现类似于 DANN 那种很大的波动,说明算法具有稳定性和良好的收敛性。

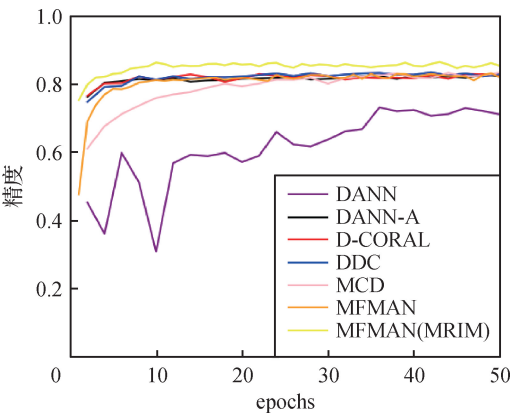


图 6 算法精度曲线

Fig. 6 Algorithms accuracy curve

图 7 展示了 MFMAN 在雾天数据集上 loss 的变化曲线。从图 7 中可以看到训练集上的 loss 变化曲线比较理想,网络拟合速度较快,在 10 个 epoch 就

可以降到稳定值,训练比较稳定,网络收敛速度较快,未产生过拟合现象。

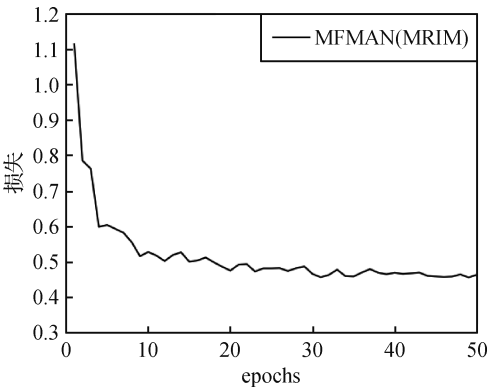


图 7 训练损失变化曲线

Fig. 7 Training loss curve

3 结 论

针对无标注雾天场景图像难以在现有标注数据、基于现有的基础分类网络下取得很好的分类识别效果的问题,提出了一个框架,使用循环生成对抗网络(Cycle-GAN)对已有数据进行数据增强,简单有效地使其在像素域的风格和数据分布上接近雾天场景。其次利用多尺度特征多对抗网络,通过多尺度特征多对抗模块充分利用数据特征信息,有效地在特征域上进一步对齐数据分布,减小领域偏移,使得雾天分类识别的准确性大大提高。通过实验以上提出的框架相比使用现有标注数据训练的基础分类网络在雾天场景下的分类识别精度有显著提升,而

提出的域多尺度特征多对抗网络优于目前其他有竞争力的无监督领域自适应方法。

当前方法在分类精度和结构优化方面仍有待进一步提高,后续研究将探索在像素域和特征域上算法的融合,改善整体的框架结构,进一步丰富数据集,使得整体性能朝着更加高效的方向发展。

参考文献 (References)

- Bousmalis K, Silberman N, Dohan D, Erhan D and Krishnan D. 2017. Unsupervised pixel-level domain adaptation with generative adversarial networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 95-104 [DOI: 10.1109/CVPR.2017.18]
- Chen W T, Ding J J and Kuo S Y. 2019. PMS-Net: robust haze removal based on patch map for single images//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 11673-11681 [DOI: 10.1109/CVPR.2019.01195]
- Chen W T, Fang H Y, Ding J J and Kuo S Y. 2020. PMHLD: patch map-based hybrid learning DehazeNet for single image haze removal. IEEE Transactions on Image Processing, 29: 6773-6788 [DOI: 10.1109/TIP.2020.2993407]
- Duchi J, Hazan E and Singer Y. 2011. Adaptive subgradient methods for online learning and stochastic optimization. The Journal of Machine Learning Research, 12: 2121-2159
- Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F and Marchand M. 2017. Domain-adversarial training of neural networks//Csurka G, ed. Domain Adaptation in Computer Vision Applications. Cham: Springer: 2096-2030 [10.1007/978-3-319-58347-1_10]
- Ghifary M, Kleijn W B and Zhang M J. 2014. Domain adaptive neural networks for object recognition//Proceedings of the 13th Pacific Rim International Conference on Artificial Intelligence. Gold Coast, Australia: Springer: 898-904 [DOI: 10.1007/978-3-319-13560-1_76]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- He K M, Sun J and Tang X O. 2011. Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence. 33(12):2341-2353
- Kouw W M, van der Maaten L J P, Krijthe J H and Loog M. 2016. Feature-level domain adaptation. Journal of Machine Learning Research, 17: 1-32
- Pan S J and Yang Q. 2010. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10): 1345-1359 [DOI: 10.1109/TKDE.2009.191]
- Pei Y T, Huang Y P, Zou Q, Lu Y H and Wang S. 2018a. Does haze removal help CNN-based image classification?//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 697-712 [DOI: 10.1007/978-3-030-01249-6_42]
- Pei Z Y, Cao Z J, Long M S and Wang J M. 2018b. Multi-adversarial domain adaptation//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Louisiana, USA: AAAI: 3934-3941
- Saito K, Watanabe K, Ushiku Y and Harada T. 2018. Maximum classifier discrepancy for unsupervised domain adaptation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3723-3732 [DOI: 10.1109/CVPR.2018.00392]
- Sun B C and Saenko K. 2016. Deep CORAL: correlation alignment for deep domain adaptation//Proceedings of European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 443-550 [DOI: 10.1007/978-3-319-49409-8_35]
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 1-9 [DOI: 10.1109/CVPR.2015.7298594]
- Tzeng E, Hoffman J, Zhang N, Saenko K and Darrell T. 2014. Deep domain confusion: Maximizing for domain invariance [EB/OL]. [2020-08-04]. <https://arxiv.org/pdf/1412.3474.pdf>
- Wang J D, Feng W J, Chen Y Q, Yu H, Huang M Y and Yu P S. 2018. Visual domain adaptation with manifold embedded distribution alignment//Proceedings of the 26th ACM international conference on Multimedia. Seoul, Korea(South): ACM: 402-410 [DOI: 10.1145/3240508.3240512]
- Zeiler M D. 2012. ADADELTA: an adaptive learning rate method [EB/OL]. [2020-08-04]. <https://arxiv.org/pdf/1212.5701.pdf>
- Zhang H and Patel V M. 2018. Densely connected pyramid dehazing network//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3194-3203 [DOI: 10.1109/CVPR.2018.00337]
- Zhu J Y, Park T, Isola P and Efros A A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks//Proceedings of 2017 IEEE International Conference on Computer Vision.

Venice, Italy: IEEE 532-542 [DOI: 10.1109/ICCV.2017.244]

作者简介



陈硕, 1994 年生, 男, 硕士, 主要研究方向为深度学习、计算机视觉、迁移学习。

E-mail: shuowoshishui@163.com



钟汇才, 通信作者, 男, 研究员, 主要研究方向为大数据与人工智能。

E-mail: zhonghuicai@ime.ac.cn

李勇周, 男, 研究员, 主要研究方向为大数据与人工智能。

E-mail: liyongzhou@ime.ac.cn

王师峥, 男, 副研究员, 主要研究方向为面向极端环境和已压缩视频的鲁棒性计算机视觉算法。

E-mail: wangshizheng@ime.ac.cn

杨建刚, 男, 博士研究生, 主要研究方向为计算机视觉和迁移学习。E-mail: yangjiangang@ime.ac.cn