



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

中国图象图形学报

2021
11
VOL.26

ISSN1006-8961
CN11-3758/TB





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

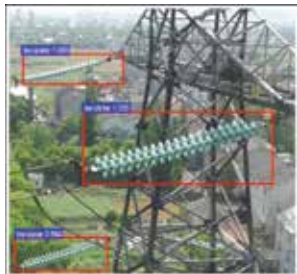
Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

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

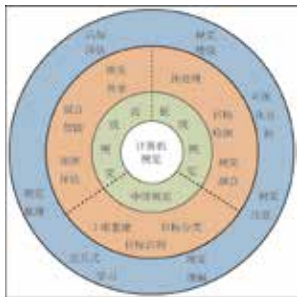
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

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, 威银城, 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

威银城, 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

威银城, 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

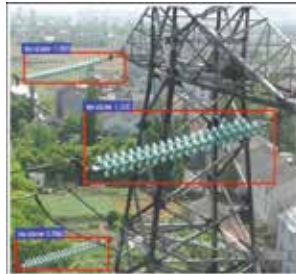
李彤, 张钧萍 2732

融合特征的卫星视频车辆单目标跟踪

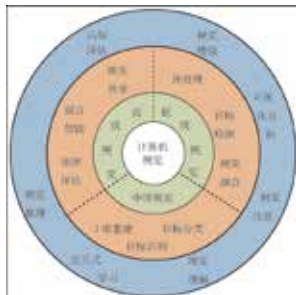
韩鸣飞, 李盛阳, 万雪, 轩诗宇, 赵子飞, 谭洪, 张万峰 2741

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: TP399 文献标识码: A 文章编号: 1006-8961(2021)11-2732-09

论文引用格式: Li T and Zhang J P. 2021. Remote sensing image scene recognition based on adversarial learning. Journal of Image and Graphics, 26(11): 2732-2740 (李彤, 张钧萍. 2021. 对抗学习遥感图像场景识别. 中国图象图形学报, 26(11): 2732-2740) [DOI: 10.11834/jig.200419]

对抗学习遥感图像场景识别

李彤^{1,2}, 张钧萍^{1*}

1. 哈尔滨工业大学电子与信息工程学院, 哈尔滨 150001; 2. 上海卫星工程研究所, 上海 201109

摘要: **目的** 在高分辨率遥感图像场景识别问题中,经典的监督机器学习算法大多需要充足的标记样本训练模型,而获取遥感图像的标注费时费力。为解决遥感图像场景识别中标记样本缺乏且不同数据集无法共享标记样本问题,提出一种结合对抗学习与变分自动编码器的迁移学习网络。**方法** 利用变分自动编码器(variational auto-encoders, VAE)在源域数据集上进行训练,分别获得编码器和分类器网络参数,并用源域编码器网络参数初始化目标域编码器。采用对抗学习的思想,引入判别网络,交替训练并更新目标域编码器与判别网络参数,使目标域与源域编码器提取的特征尽量相似,从而实现遥感图像源域到目标域的特征迁移。**结果** 利用两个遥感场景识别数据集进行实验,验证特征迁移算法的有效性,同时尝试利用 SUN397 自然场景数据集与遥感场景间的迁移识别,采用相关性对齐以及均衡分布适应两种迁移学习方法作为对比。两组遥感场景数据集间的实验中,相比于仅利用源域样本训练的网络,经过迁移学习后的网络场景识别精度提升约 10%,利用少量目标域标记样本后提升更为明显;与对照实验结果相比,利用少量目标域标记样本时提出方法的识别精度提升均在 3% 之上,仅利用源域标记样本时提出方法场景识别精度提升了 10%~40%;利用自然场景数据集时,方法仍能在一定程度上提升场景识别精度。**结论** 本文提出的对抗迁移学习网络可以在目标域样本缺乏的条件下,充分利用其他数据集中的样本信息,实现不同场景图像数据集间的特征迁移及场景识别,有效提升遥感图像的场景识别精度。

关键词: 场景识别;遥感图像;对抗学习;迁移学习;变分自动编码器(VAE)

Remote sensing image scene recognition based on adversarial learning

Li Tong^{1,2}, Zhang Junping^{1*}1. School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China;
2. Shanghai Institute of Satellite Engineering, Shanghai 201109, 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

收稿日期: 2020-07-28; 修回日期: 2020-10-29; 预印本日期: 2020-11-05

* 通信作者: 张钧萍 zhangjp@hit.edu.cn

基金项目: 国家自然科学基金项目(61871150)

Supported by: National Natural Science Foundation of China (61871150)

between different datasets due to different sensors and complex light conditions, deep learning architecture and adversarial learning are investigated. A feature transfer method based on adversarial variational autoencoder (VAE) is proposed.

Method Feature transfer architecture can be divided into three parts. The first part is the pretrain module. Given the limited samples with scene labels, the unsupervised learning model, 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 VAE-based 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: scene recognition; remote sensing image; adversarial learning; transfer learning; variational autoencoder (VAE)

0 引言

场景是多类地物、目标及其背景上下文信息的有机组合。遥感图像场景识别是通过分析图像场景中包含的地物和目标的组成及统计关系实现对图像场景主题的描述,分为基于中层特征的方法和深度

学习的方法。基于中层特征的方法包括视觉词包模型(bag-of-visual-words, BOVW)及其改进(Zhang等, 2016)、Fisher 矢量编码(Fisher vector coding, FVC)(Zhao等, 2016)、局部特征聚合描述概率主题模型(vector of locally aggregated descriptors, VLAD)(Wang等, 2017a)、结合稀疏表示(Zhu等, 2018)等。近年来,深度学习的方法在遥感场景识别中取

得了巨大突破。Chaib 等人(2017)提出基于 VGG-Net(Visual Geometry Group net)与判别相关分析的融合方法,有效提升了场景识别精度。Cheng 等人(2018)将深度学习与度量学习相结合,提出判别卷积神经网络,很好地解决了遥感场景类间相似性高、类内差异大的问题。Wang 等人(2019)和 Tong 等人(2020)将注意机制与深度学习模型相结合,一定程度上解决了深度学习模型在遥感图像处理中的过拟合问题。然而适用这些监督机器学习算法的最主要前提是训练样本和测试样本必须处于同一特征空间,服从相同的分布。同时,算法需要足够的标记样本,才能使模型得到充分训练,尤其是近年来在图像分类识别任务中表现出色的深度学习算法,更是需要大量的标记本来训练网络参数。对于遥感图像而言,获取大量类别标签既耗时又费力,且不同传感器获取的多时相遥感图像由于光学器件、成像链路、光照条件等的不同,无法共享同类别的标记样本。因此,若能实现有效的知识迁移,便可显著提高机器学习模型的效果,同时避免代价高昂的人工标注。

迁移学习以其在处理数据分布差异、充分利用原有标注数据方面的有效性,在遥感图像的分类识别(Wu 等,2019)等实际应用中颇受关注。基于迁移学习的高分图像场景级分类识别的研究多为利用计算机视觉领域中的深度学习网络的参数迁移。李冠东等人(2019)利用 ImageNet 数据集对卷积神经网络进行预训练,而后利用少量有标签的遥感场景微调网络,实现网络参数的迁移。Hu 等人(2015)在预训练卷积神经网络的基础上提取多尺度遥感图像场景的特征并进行全局特征编码。Yao 等人(2017)利用预训练的深度学习网络对遥感场景进行特征提取,采用随机森林分类器实现遥感图像的场景识别。Gong 等人(2018)将预训练的卷积神经网络与视觉显著性区域提取相结合,同时在场景中引入随机噪声,提升网络模型的抗噪性能。然而这类利用网络参数迁移的方法,若利用前人提供的网络参数则网络结构无法自由调整,若搭建网络自行训练则十分耗时。因此,本文考虑从特征迁移的角度实现不同数据集间标记样本的共享。

光学遥感领域中,大多数特征迁移算法均针对高光谱图像像素级分类,典型算法包括联合分布适应(Long 等,2013)、迁移成分分析(Matasci 等,2015)、相关性对齐(Sun 等,2016)和均衡分布适应

(Wang 等,2017b)等。这些特征迁移算法的共同点是寻找源域到目标域的变换,使变换后源域和目标域样本特征具有相同的概率分布或较大的相关性。出于同样的目的,本文提出基于对抗学习的遥感图像场景识别方法,结合对抗学习思想,引入判别网络约束变分自动编码器(variational autoencoders, VAE)的训练过程,使判别网络无法区分变分自动编码器提取的源域场景特征和目标域场景特征,从而实现不同场景数据集间的特征迁移。

1 变分自动编码器

在深度学习的研究中,生成式模型由于可以自动学习原始输入数据的数据分布并生成相似的模拟数据而成为热点。变分自动编码器(Kingma 和 Welling,2014)是自动编码器(Hinton 和 Salakhutdinov,2006)的隐变量模型的扩展形式,属于典型的生成模型,其将变分推断与自动编码器思想相结合,在编码器与解码器之间引入隐变量 z ,输入 x 通过编码器 $p(z|x)$ 得到隐变量 z 分布 $p(z)$ 的均值和方差,通过对该分布采样得到新的特征 z' ,再通过解码器 $p(x|z)$ 对输入进行重建,网络训练完成后,只需对 z 进行采样,再通过解码器,即可生成与输入 x 同分布的新样本 $\hat{x}$ 。网络结构如图1所示。

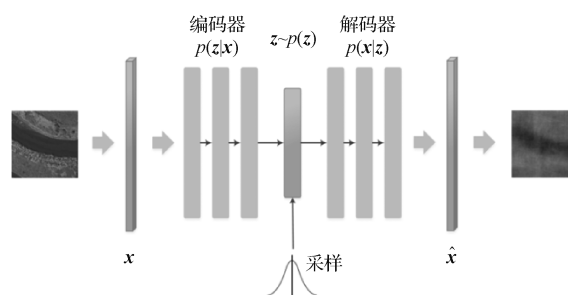


图1 变分自动编码器网络结构

Fig. 1 The graphical model of variational autoencoders

采用极大似然估计训练隐变量模型,则需计算

$$\arg \max_{\theta} E_{x \sim p} \log p_{\theta}(x) \quad (1)$$

式中, $E(\cdot)$ 表示分布函数的期望, θ 为模型参数。

由贝叶斯公式 $p_{\theta}(x) = \frac{p_{\theta}(x, z)}{p_{\theta}(z|x)}$,可将问题转化为求 $p_{\theta}(z|x)$ 。对此,引入变分推断思想,在变分函数组中寻找与后验分布 $p_{\theta}(z|x)$ 最接近的分布 $q_{\phi}(z|x)$, ϕ 为变分拟合参数。采用KL(Kullback-

Leibler) 散度 D_{KL} 来描述两个概率分布之间的差异, 即

$$\begin{aligned} D_{KL}(q_\phi(z|x) \| p_\theta(z|x)) = \\ E_{z \sim q_\phi(z|x)} (\log q_\phi(z|x) - \log p_\theta(z|x)) = \\ E_{q_\phi(z|x)} (\log q_\phi(z|x) - \log p_\theta(x, z) + \log p_\theta(x)) \end{aligned} \quad (2)$$

式中, $\log p_\theta(x)$ 与 z 无关, 因此有

$$\begin{aligned} \log p_\theta(x) = D_{KL}(q_\phi(z|x) \| p_\theta(z|x)) + \\ E_{q_\phi(z|x)} (\log p_\theta(x, z) - \log q_\phi(z|x)) \end{aligned} \quad (3)$$

式中, 第2项称为证据下界 (evidence lower bound, ELBO) 或变分下界 (variational lower bound), 记作 $L(\theta, \phi; x)$ 。由于需要令 $q(z|x)$ 与 $p(z|x)$ 尽可能相似, 即式(2)最小, 由 KL 散度性质可知

$$D_{KL}(q(z|x) \| p(z|x)) > 0 \quad (4)$$

同时, 在 z 的空间中, $\log p_\theta(x)$ 固定, 要令 KL 散度最小, 因此问题进一步转化为求 $L(\theta, \phi; x)$ 的最大值。ELBO 可变换为

$$\begin{aligned} f_{ELBO} = -D_{KL}(q_\phi(z|x) \| p_\theta(z)) + \\ E_{q_\phi(z|x)} (\log p_\theta(x|z)) \end{aligned} \quad (5)$$

变分自动编码器的损失函数 L_{VAE} 的定义为

$$\begin{aligned} L_{VAE} = -E_{q_\phi(z|x)} (\log p_\theta(x|z)) + \\ D_{KL}(q_\phi(z|x) \| p_\theta(z)) \end{aligned} \quad (6)$$

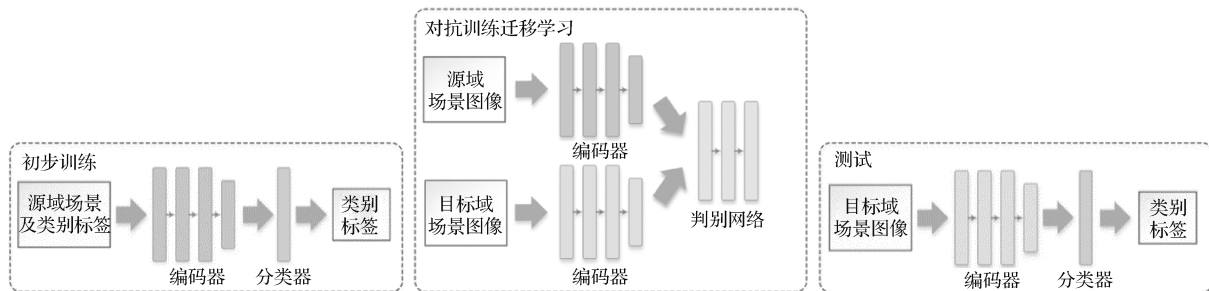


图2 基于对抗迁移学习的高分遥感图像场景识别流程图

Fig. 2 The diagram of remote sensing images scene recognition based on adversarial transfer learning

从图2可以看出, 依据式(6), 利用源域场景图像 I_s 充分训练变分自动编码器网络后, 保留编码器进行特征映射 M_s , 继而输入分类器 C , 利用相应的场景类别标签 Y_s 进行整个网络的训练, 获得源域场景的特征映射 M_s 和分类器 C , 完成网络的初始训练。初始训练的目标函数 L_{MC} 为

$$L_{MC}(I_s, Y_s) = -E_{(I_s, Y_s)} \sum_{k=1}^K 1_{k=y_s} \log C(M_s(I_s^k)) \quad (7)$$

然后进行对抗学习, 固定源域场景特征映射 M_s

2 基于对抗学习的特征迁移

2.1 对抗学习思想

对抗学习的思想源于 Goodfellow 等人(2014)提出的生成对抗网络 (generative adversarial network, GAN)。对抗学习思想由生成网络和判别网络构成, 利用生成的随机分布的数据通过生成网络生成模拟数据, 将真实数据与模拟数据同时输入判别网络, 将网络输出反馈回生成网络更新网络参数。称之为对抗学习是因为生成网络与判别网络的目标是相互冲突的, 即生成网络的目标是令生成的模拟数据与真实数据尽可能接近, 使得判别网络无法判断输入是否为真实数据; 而判别网络的目标是不断优化网络, 使判别网络能够分辨真实数据和模拟数据。

2.2 对抗变分自动编码器

本文提出的基于对抗变分自动编码器的迁移学习方法 (adversarial autoencoder transfer, AAET) 借鉴了对抗学习的思想, 令源域场景图像提取的特征与目标域场景图像提取的特征尽量相似, 流程如图2所示。

并初始化目标域场景图像特征映射参数 M_{RS} , 令 $M_{RS} = M_s$, 将源域场景图像 I_s 和目标域场景图像 I_{RS} 分别进行特征映射 M_s 和 M_{RS} , 并输入判别网络 D , 交替更新特征映射 M_{RS} 和判别网络 D 。

判别网络的目标函数 L_D 为

$$\begin{aligned} L_D(I_s, I_{RS}, M_s, M_{RS}) = -E_{I \sim I_s} [\log D(M_s(I_s))] - \\ E_{I \sim I_{RS}} [\log(1 - D(M_{RS}(I_{RS})))] \end{aligned} \quad (8)$$

目标域编码器的目标函数 L_M 为

$$L_M(I_{RS}, Y_{RS}, D) = -E_{I \sim I_{RS}} [\log D(M_{RS}(I_{RS}))] \quad (9)$$

利用源域场景图像特征映射的网络参数初始化目标域场景图像特征映射网络,以此为基础继续调节网络参数,即

$$\psi_{l_i}(\mathbf{M}_S^{l_i}, \mathbf{M}_{RS}^{l_i}) = (\mathbf{M}_S^{l_i} = \mathbf{M}_{RS}^{l_i}) \quad (10)$$

式中, $\psi(\cdot)$ 表示源域与目标域映射间的约束关系, l_i 为第 i 层网络参数。网络整体的损失函数可记为

$$\begin{aligned} & \min_{\mathbf{M}_S, C} L_{MC}(\mathbf{I}_S, Y_S) \\ & \min_D L_D(\mathbf{I}_S, \mathbf{I}_{RS}, \mathbf{M}_S, \mathbf{M}_{RS}) \\ & \min_{\mathbf{M}_S, \mathbf{M}_{RS}} L_M(\mathbf{I}_{RS}, Y_{RS}, D) \\ \text{s. t. } & \psi_{l_i}(\mathbf{M}_S^{l_i}, \mathbf{M}_{RS}^{l_i}) = (\mathbf{M}_S^{l_i} = \mathbf{M}_{RS}^{l_i}) \quad (11) \end{aligned}$$

得到目标域场景图像的特征映射后,利用少量有标签的目标域场景图像对特征映射 $\mathbf{M}_{RS}$ 和分类器 C 进行简单微调,最终利用特征映射 $\mathbf{M}_{RS}$ 和分类器 C 实现对目标域的遥感场景图像识别。

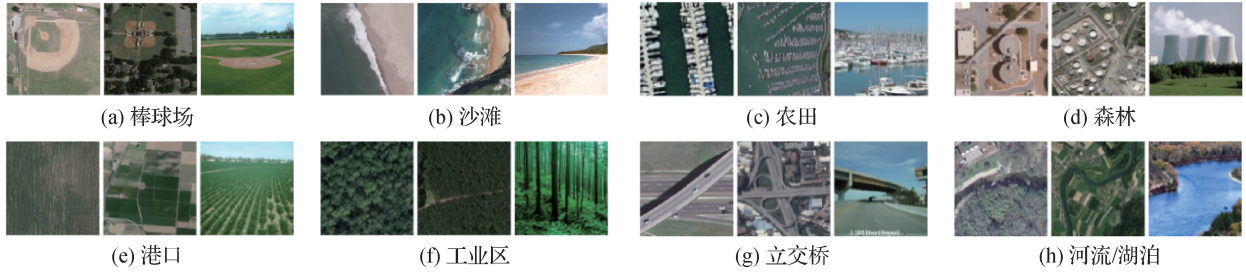


图3 实验数据

Fig. 3 Experiment data

((a) baseball diamond; (b) beach; (c) farmland; (d) forest; (e) harbor; (f) industrial area; (g) overpass; (h) river/lake)

3.2 NWPU-RESISC45 到 UCMerced-21 场景迁移识别

实验中首先利用特征提取与图像重建,选择卷积与反卷积训练变分自动编码器网络,编码器的网络结构与参数设置如图4所示,编码器输出层节点数设为500,解码器的网络结构与编码器对称,变分自动编码器网络训练完成后,舍弃解码器部分,在编码器后添加分类器,利用源域数据集中的样本与类别标签微调编码器与分类器的网络参数,训练次数设为200,完成网络的初步训练。

在对抗训练迁移学习部分,用源域的编码器网络参数初始化目标域的编码器,判别网络包含2个隐含层,隐含层节点数均设为500,输出层节点数为2,用于判别输入特征为源域特征或目标域特征。首先训练判别网络,使其能够区分源域与目标域特征,

3 实验结果及分析

3.1 实验数据

实验在 UCMerced-21 (Yang 和 Newsam, 2010) 与 NWPU-RESISC45 (Cheng 等, 2017) 两组遥感场景识别数据集与 SUN397 (Xiao 等, 2016) 自然场景图像数据集上进行,主要考察遥感场景数据集间的特征迁移,尝试视角差异较大的自然场景与遥感场景间的特征迁移。选取数据集中共同的8种场景类别,即棒球场、沙滩、农田、森林、港口、工业区、立交桥和河流/湖泊,UCMerced-21 数据集中图像空间分辨率均为0.3 m, NWPU-RESISC45 数据集中图像空间分辨率为0.2 m~30 m不等。3组数据集中各场景类别的图像对比如图3所示,图中每组场景类别左侧为 UCMerced-21 遥感场景图像,中间为 NWPU-RESISC45 遥感场景图像,右侧为 SUN397 自然场景图像。

源域为1,目标域为0;然后固定判别网络参数,训练目标域编码器,使判别网络的输出为1,如此循环训练,实现对抗学习;最后用少量目标域的样本微调网络。采用相关性对齐 (correlation alignment, CORAL) 以及均衡分布适应 (balanced distribution adaptation, BDA) 作为对比实验,选择总体分类精度 (overall accuracy, OA) 和 Kappa 系数作为评价指标,实验结果如表1所示。表1中 AE (auto-encoder) 表示不利用源域样本和迁移学习方法,而直接利用目标域样本训练自动编码器网络后,利用少量有标签的目标域样本微调自动编码器网络与分类器的场景识别结果;AAET-C (AAET-classifier) 表示仅利用少量目标域的样本微调分类器,而不对编码器参数进行调整的场景识别结果;AAET-E (AAET-encoder) 表示利用少量目标域样本同时调整目标域编码器与分类器

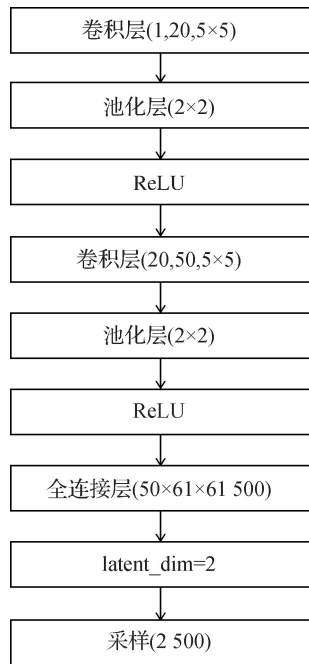


图4 源域与目标域编码器网络结构

Fig. 4 The architecture of encoders for source and target domain

表1 NWPU-RESISC45 迁移到 UCMerced-21 的场景识别精度和 Kappa 系数

Table 1 Scene recognition results and Kappa transferring NWPU-RESISC45 to UCMerced-21 dataset

方法	无迁移		无监督		目标域样本数					
					5		10		15	
	OA/%	Kappa	OA/%	Kappa	OA/%	Kappa	OA/%	Kappa	OA/%	Kappa
AE	-	-	-	-	40.00	0.314	47.38	0.399	54.75	0.483
CORAL	-	-	12.63	0.001	58.29	0.523	68.19	0.637	73.53	0.697
BDA	-	-	46.25	0.001	62.50	0.523	70.14	0.637	74.12	0.697
AAET-C	45.75	0.38	57.13	0.51	63.63	0.584	72.25	0.683	75.63	0.721
AAET-E	45.75	0.38	57.13	0.51	63.88	0.587	73.38	0.696	77.13	0.739

注:加粗字体表示各列最优结果,“-”表示未计算该项精度。

每类场景选择 15 个训练样本微调网络时, UCMerced-21 场景识别的混淆矩阵如图 5 所示。由混淆矩阵可见,工业区仍然是较难识别区分的场景类别,主要是由于工业区场景中的主要目标地物(如油库)的样式、大小以及目标地物的空间分布较为多变,图 3(f)可印证这一点,这种不规律为模型的建立制造了一定障碍。立交桥场景也有同样的问题,相较之下,不同情况下一致性较好的沙滩、农田、森林、港口等场景则可以获得较为满意的识别结果。

为验证算法的稳定性,依次选取 5~8 个场景类别进行特征迁移和场景识别实验。本文提出的 AAET 方法的识别精度如图 6 所示。由图中结果可

参数的场景识别结果;无迁移表示利用源域编码器与分类器直接对目标域实验图像进行场景识别的结果;无监督表示经过对抗学习后,不利用目标与样本对网络进行调整的场景识别结果;后续依次为利用目标域数据集中每类 5 幅、10 幅、15 幅有标签的场景图像作为训练样本调整网络时的场景识别结果。从表 1 可以看出,利用 NWPU-RESISC45 数据集特征迁移后的识别结果均明显优于仅利用少量目标域 UCMerced-21 数据集中的标记样本的识别结果,且经过对抗学习特征迁移后对 UCMerced-21 数据集中的场景识别精度相比直接利用 NWPU-RESISC45 数据集的编码器与分类器提升了约 12%,而后仅使用少量 UCMerced-21 数据集中的样本即可以显著提高场景识别精度。与 CORAL 和 BDA 方法相比,对抗学习的特征迁移方法在利用少量目标域样本参与训练时识别精度有所提升,并在无监督情况下具有明显优势。

见,场景类别数对无迁移学习情况下的场景识别精度影响较大;经过迁移学习后,无论是否采用少量目标域样本对网络参数进行微调,场景识别精度均可在一定范围内保持稳定。

3.3 UCMerced-21 到 NWPU-RESISC45 场景迁移识别

从 UCMerced-21 数据集到 NWPU-RESISC45 数据集的特征迁移采用了相同的网络结构和参数设置,微调网络参数时,分别选择 NWPU-RESISC45 数据集中 1%、2% 和 3% 的场景图像作为训练样本进行场景识别实验,结果如表 2 所示。可以看出,基于对抗学习的变分自动编码器网络虽然如从 NWPU-

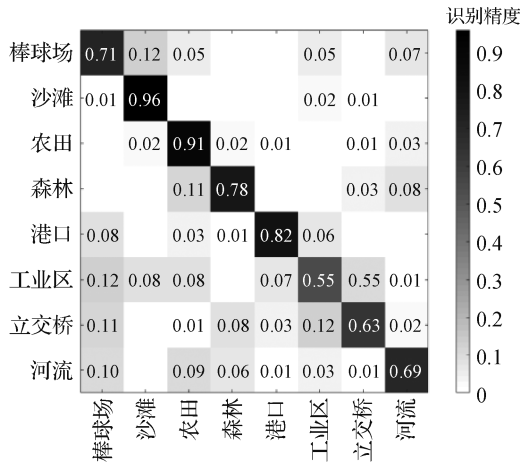


图5 NWPU-RESISC45 迁移到 UCMerced-21

场景识别混淆矩阵

Fig. 5 The confusion matrix of UCMerced-21 transferring NWPU-RESISC45 dataset

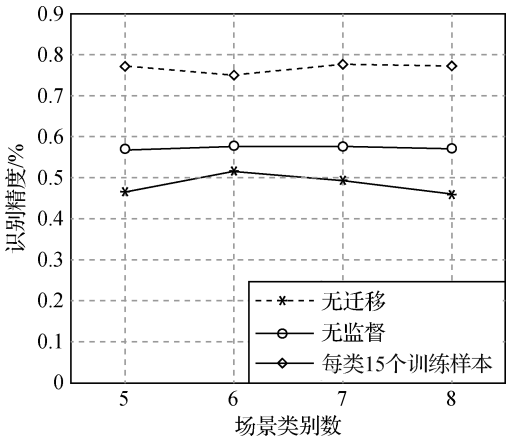


图6 不同场景类别数下场景识别精度变化

Fig. 6 Scene recognition accuracies of different numbers of scene categories

表2 UCMerced-21 迁移到 NWPU-RESISC45 的场景识别精度和 Kappa 系数

Table 2 Scene recognition results and Kappa transferring UCMerced-21 to NWPU-RESISC45 dataset

方法	无迁移		无监督		目标域样本比例					
					1%		2%		3%	
	OA/%	Kappa	OA/%	Kappa	OA/%	Kappa	OA/%	Kappa	OA/%	Kappa
AE	—	—	—	—	25.29	0.146	28.46	0.182	31.43	0.216
CORAL	—	—	12.59	0.001	42.37	0.341	46.81	0.392	50.40	0.433
BDA	—	—	15.14	0.03	44.21	0.362	47.08	0.395	49.56	0.424
AAET-C	33.75	0.243	42.77	0.346	45.52	0.377	47.88	0.404	51.39	0.444
AAET-E	33.75	0.243	42.77	0.346	46.89	0.393	49.46	0.422	53.07	0.464

注:加粗字体表示各列最优结果,“—”表示未计算该项精度。

RESISC45 数据集迁移到 UCMerced-21 数据集一样大幅提升了场景识别精度,但识别结果远不及其,原因主要为 UCMerced-21 数据集中图像的空间分辨率均为 0.3 m,故同一类别的场景中各类地物的组分大致相近,而 NWPU-RESISC45 数据集中图像的空间分辨率从亚米级到米级差异较大,使得特征迁移时,NWPU-RESISC45 数据集可以涵盖 UCMerced-21 中大部分图像场景的特征,而从 UCMerced-21 中提取的场景特征却不能对 NWPU-RESISC45 中的图像场景进行完整描述,但相较于仅利用目标域数据集的标记样本时,场景识别精度提升在 20% 以上,同时优于对比方法的实验结果。

每类场景选择 3% 的图像作为训练样本微调网络时,对 NWPU-RESISC45 数据集进行场景识别的混淆矩阵如图 7 所示。可以看出,与基于光谱信息

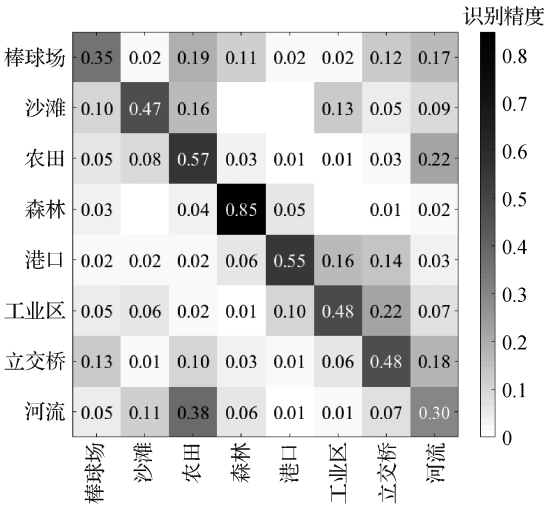


图7 UCMerced-21 迁移到 NWPU-RESISC45

场景识别混淆矩阵

Fig. 7 The confusion matrix of NWPU-RESISC45 transferring UCMerced-21 dataset

的地物像素分类中水体的光谱及其他地物光谱的良好可分性不同,河流场景间的差异较大,主要体现在河流的形状走势各异,且两岸地物组成也相差较大,可能为峡谷或植被;此外,实验中图像空间分辨率的不同对河流场景的识别精度影响也较大。

3.4 SUN397 到 UCMerced-21 场景迁移识别

尝试自然场景与遥感场景数据集间的迁移学习,实现从水平视角的 SUN397 数据集到鸟瞰视角的 UCMerced-21 数据集的特征迁移。由于 SUN397 自然场

景识别数据集中的图像大小大多不一致,故将二者的图像统一变换为 256×256 像素后进行后续实验,场景识别精度如表 3 所示。虽然在无迁移学习的情况下总体分类精度约为 23%,但通过分析混淆矩阵可知,网络将遥感图像场景几乎全部分为第 2 类和第 4 类,并没有很好地提取到遥感图像场景的特征,而随着对抗学习对网络参数的调整,这种情况才有所改善。但由于两组数据集的视角差异,基于对抗自动编码机的方法对于这种视角差异较大的情况仍有待改进。

表 3 SUN397 迁移到 UCMerced-21 的场景识别精度及 Kappa 系数
Table 3 Scene recognition results and Kappa transferring SUN397 to UCMerced-21 dataset

方法	无迁移		无监督		目标域样本数					
					5		10		15	
	OA/%	Kappa	OA/%	Kappa	OA/%	Kappa	OA/%	Kappa	OA/%	Kappa
AE	-	-	-	-	40.00	0.314	47.38	0.399	54.75	0.483
AAET-E	23.38	0.124	29.63	0.196	45.13	0.373	54.75	0.483	59.38	0.536

注:加粗字体表示各列最优结果,“-”表示未计算该项精度。

4 结 论

本文针对遥感图像场景识别中标记样本缺乏且不同数据集无法共享标记样本问题,提出基于对抗变分自动编码机的遥感场景识别方法。该方法一方面利用无监督学习的变分自动编码器,充分利用源域场景图像信息;另一方面引入对抗学习思想,使编码器网络提取的目标域特征与源域特征尽量相似,实现对目标域基于特征迁移的场景识别。实验表明,本文方法有效提升了目标域标记样本较少时的场景识别精度,证明了利用特征迁移实现遥感场景识别的有效性。但是,对于水平视角的自然场景到鸟瞰视角的遥感场景的特征迁移,本文方法仅能在一定程度上实现场景识别精度的提升,仍需进一步研究以更好地实现视角差异较大情况下的特征迁移。

参考文献 (References)

Chaib S, Liu H, Gu Y F and Yao H X. 2017. Deep feature fusion for VHR remote sensing scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8): 4775-4784 [DOI: 10.1109/TGRS.2017.2700322]

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]

Cheng G, Yang C Y, Yao X W, Guo L and Han J W. 2018. When deep learning meets metric learning: remote sensing image classification via learning discriminative CNNs. *IEEE Transactions on Geoscience and Remote Sensing*, 56(5): 2811-2821 [DOI: 10.1109/TGRS.2017.2783902]

Gong X, Xie Z, Liu Y Y, Shi X G and Zheng Z. 2018. Deep salient feature based anti-noise transfer network for scene classification of remote sensing imagery. *Remote Sensing*, 10(3): #410 [DOI: 10.3390/rs10030410]

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial nets// *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Montreal, Canada: MIT Press: 2672-2680

Hinton G E and Salakhutdinov R R. 2006. Reducing the dimensionality of data with neural networks. *Science*, 313(5786): 504-507 [DOI: 10.1126/science.1127647]

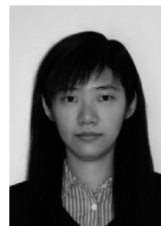
Hu F, Xia G S, Hu J W and Zhang L P. 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, 7(11): 14680-14707 [DOI: 10.3390/rs71114680]

Kingma D P and Welling M. 2014. Auto-encoding Variational Bayes [EB/OL]. [2020-07-28]. <http://arxiv.org/pdf/1312.6114.pdf>

Li G D, Zhang C J, Wang M K, Zhang X Y and Gao F. 2019. Transfer

- learning using convolutional neural network for scene classification within high resolution remote sensing image. *Science of Surveying and Mapping*, 44(4): 116-123, 174 (李冠东, 张春菊, 王铭恺, 张雪英, 高飞. 2019. 卷积神经网络迁移的高分影像场景分类学习. *测绘科学*, 44(4): 116-123, 174) [DOI: 10.16251/j.cnki.1009-2307.2019.04.018]
- Long M S, Wang J M, Ding G G, Sun J G and Yu P S. 2013. Transfer feature learning with joint distribution adaptation//*Proceedings of 2013 IEEE International Conference on Computer Vision*. Sydney, Australia: IEEE: 2200-2207 [DOI: 10.1109/ICCV.2013.274]
- Matasci G, Volpi M, Kanevski M, Bruzzone L and Tuia D. 2015. Semisupervised transfer component analysis for domain adaptation in remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7): 3550-3564 [DOI: 10.1109/TGRS.2014.2377785]
- Sun B C, Feng J S and Saenko K. 2016. Return of frustratingly easy domain adaptation//*Proceedings of the 30th AAAI Conference on Artificial Intelligence*. Phoenix, USA: AAAI Press: 2058-2065
- Tong W, Chen W, Han W, Li X and Wang L. 2020. Channel-attention-based densenet network for remote sensing image scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 4121-4132 [DOI: 10.1109/JSTARS.2020.3009352]
- Wang G L, Fan B, Xiang S M and Pan C H. 2017a. Aggregating rich hierarchical features for scene classification in remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(9): 4104-4115 [DOI: 10.1109/JSTARS.2017.2705419]
- Wang J D, Chen Y Q, Hao S J, Feng W J and Shen Z Q. 2017b. Balanced distribution adaptation for transfer learning//*Proceedings of 2017 IEEE International Conference on Data Mining (ICDM)*. New Orleans, USA: IEEE: 1129-1134 [DOI: 10.1109/ICDM.2017.150]
- Wang Q, Liu S T, Chanussot J and Li X L. 2019. Scene classification with recurrent attention of VHR remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2): 1155-1167 [DOI: 10.1109/TGRS.2018.2864987]
- Wu R, Li Y, Han H, Chen X and Lin Y. 2019. Remote sensing image analysis based on transfer learning: a survey//*Proceedings of International Conference on Advanced Hybrid Information Processing*. Nanjing, China: Springer: 408-415 [DOI: 10.1007/978-3-030-19086-6_45]
- Xiao J X, Ehinger K A, Hays J, Torralba A and Oliva A. 2016. SUN database: exploring a large collection of scene categories. *International Journal of Computer Vision*, 119(1): 3-22 [DOI: 10.1007/s11263-014-0748-y]
- Yang Y and Newsam S. 2010. Bag-of-visual-words and spatial extensions for land-use classification//*Proceeding of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. San Jose, USA: Association for Computing Machinery: 270-279 [DOI: 10.1145/1869790.1869829]
- Yao Y, Liang H, Li X, Zhang J and He J. 2017. Sensing urban land-use patterns by integrating Google Tensorflow and scene-classification models//*The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XLII-2/W7: 981-988 [DOI: 10.5194/isprs-archives-XLII-2-W7-981-2017]
- Zhang J P, Li T, Lu X C and Cheng Z. 2016. Semantic classification of high-resolution remote-sensing images based on mid-level features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6): 2343-2353 [DOI: 10.1109/JSTARS.2016.2536943]
- Zhao B, Zhong Y F, Zhang L P and Huang B. 2016. The Fisher kernel coding framework for high spatial resolution scene classification. *Remote Sensing*, 8(2): #157 [DOI: 10.3390/rs8020157]
- Zhu Q Q, Zhong Y F, Wu S Q, Zhang L P and Li D R. 2018. Scene classification based on the sparse homogeneous-heterogeneous topic feature model. *IEEE Transactions on Geoscience and Remote Sensing*, 56(5): 2689-2703 [DOI: 10.1109/TGRS.2017.2781712]

作者简介



李彤, 1990年生, 女, 博士研究生, 主要研究方向为高分遥感图像解译。

E-mail: ltong@hit.edu.cn



张钧萍, 通信作者, 女, 教授, 博士生导师, 主要研究方向为遥感图像处理、信息融合。

E-mail: zhangjp@hit.edu.cn