



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

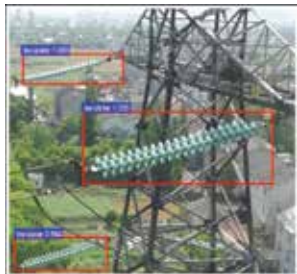
Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

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

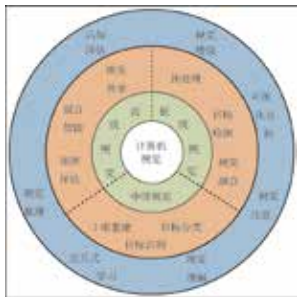
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

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, 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

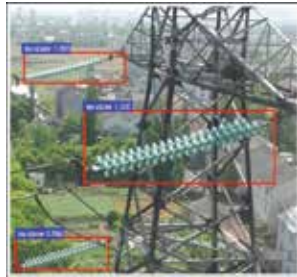
李彤, 张钧萍 2732

融合特征的卫星视频车辆单目标跟踪

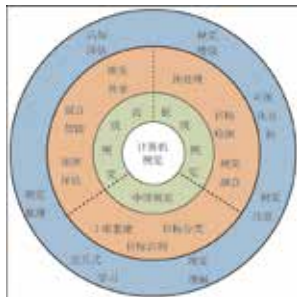
韩鸣飞, 李盛阳, 万雪, 轩诗宇, 赵子飞, 谭洪, 张万峰 2741

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: TP301.6 文献标识码: A 文章编号: 1006-8961(2021)11-2582-12

论文引用格式: Zhang K, He Y X, Zhao K, Feng X H, Zhao Z B and Ma Z Y. 2021. Multi-label classification method of bolt attributes based on deformable NTS-Net. Journal of Image and Graphics, 26(11): 2582-2593 (张珂, 何颖宣, 赵凯, 冯晓晗, 赵振兵, 马占宇. 2021. 可变形 NTS-Net 的螺栓属性多标签分类. 中国图象图形学报, 26(11): 2582-2593) [DOI:10.11834/jig.200703]

可变形 NTS-Net 的螺栓属性多标签分类

张珂^{1,2*}, 何颖宣¹, 赵凯¹, 冯晓晗¹, 赵振兵^{1,2}, 马占宇³

1. 华北电力大学电子与通信工程系, 保定 071003; 2. 华北电力大学河北省电力物联网技术重点实验室, 保定 071003;
3. 北京邮电大学人工智能学院, 北京 100086

摘要: **目的** 螺栓是确保输电线路安全牢靠的基石, 螺栓缺销、松动和锈蚀等缺陷是造成输电线路损坏甚至重大事故的重要原因之一。本文针对螺栓缺陷存在视觉不可分的问题, 提出了一种改进 NTS-Net (navigator-teacher-scrutinizer network) 的螺栓属性多标签分类方法。 **方法** 为了增强模型对不规则轮廓的处理能力, 采用可变形卷积的 ResNet-50 网络作为特征提取网络从原图中提取全局特征。利用 NTS-Net 学习得到图像信息量最大的判别性局部区域。为了考虑不同局部特征对不同属性标签的不同影响, 在局部特征与全局特征融合时引入通道注意力机制, 提取特征的通道权重, 获取关键通道特征以改善多标签分类效果。 **结果** 实验结果表明, 本文在螺栓多属性分类数据集上的平均分类精确率为 84.5%, 比采用传统的多标签分类精确率提升了 10%~20%。 **结论** 本文通过可变形卷积提升网络的特征提取能力以及引入通道注意力机制实现了对 NTS-Net 提供的局部特征的高效利用, 为解决螺栓多属性分类中存在的问题提供了一种新的思路。

关键词: 螺栓缺陷; 可变形卷积; NTS-Net 网络; 多标签分类; 通道注意力

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

Zhang Ke^{1,2*}, He Yingxuan¹, Zhao Kai¹, Feng Xiaohan¹, Zhao Zhenbing^{1,2}, Ma Zhanyu³

1. Department of Electronic and Communication Engineering, North China Electric Power University, Baoding 071003, China;
2. Hebei Key Laboratory of Power Internet of Things Technology, North China Electric Power University, Baoding 071003, China;
3. School of Artificial Intelligence, Beijing University of Posts and Telecommunication, Beijing 100086, China

Abstract: **Objective** Unmanned aerial vehicle (UAV)-based transmission line inspection technology has achieved long-term progress and development. The use of computer vision technology to automatically and accurately locate line equipment, such as wires, insulators, and bolts, from aerial inspection images under complex natural backgrounds and accurately detect their defects has become an important technical issue. The defects inspected by transmission lines mainly include tower, wire, insulator, and metal fitting defects. Given the large size of metal fittings and insulators, their defects are obvious and easy to identify. By contrast, numerous bolts are present in poles, insulators, and metal fittings. Bolts change easily

收稿日期: 2020-12-15; 修回日期: 2021-03-01; 预印本日期: 2021-03-08

* 通信作者: 张珂 zhangkeit@ncepu.edu.cn

基金项目: 国家自然科学基金项目 (62076093, 61871182, 61922015, 61773071, 61302163); 河北省自然科学基金项目 (F2020502009, F2015502062, F2016502062); 北京市自然科学基金 (4192055); 中央高校基本科研业务费专项资金资助 (2020YJ006, 2020MS099)

Supported by: National Natural Science Foundation of China (62076093, 61871182, 61922015, 61773071, 61302163); Natural Science Foundation of Hebei Province, China (F2020502009, F2015502062, F2016502062); Beijing Municipal Natural Science Foundation (4192055); Fundamental Research Funds for the Central Universities (2020YJ006, 2020MS099)

from the normal state to the defect state due to the large number of bolts and complex stress conditions. The use of deep learning has achieved good results in visual detection, identification, and classification of tower, wire, insulator, and metal fitting defects, but only a few studies have been conducted on bolt defects. In addition, bolt defects are not completely visually separable problems; they are visually inseparable, and they can not be solved by object detection algorithms alone. Thus, we believe that the bolt defect detection problem is not only an object detection problem, but also an image classification problem. Multi-label classification of bolts must be implemented efficiently and quickly to provide a basis for defect detection. The convolutional neural network (CNN) is inherently limited by model geometry transformation due to its fixed geometric structure. An offset variable must be added to the position of each sampling point in the convolution kernel to weaken this limitation and improve the feature extraction capability of bolts. By adding these variables, the convolution kernel is given random sampling near the current position and is no longer limited to the previous regular grid points. The convolution operation after expansion is called deformable convolution. Deformable convolution changes the sampling position of the standard convolution kernel by adding an additional offset to the sampling point. The compensation obtained can be learned through training without additional supervision. **Method** The object to be inspected in the transmission line bolt multi-label classification task has similar overall characteristics as those of the object in the general image multi-label classification task. The classification model needs to capture the key local features that can distinguish the attributes of different categories. The idea of using local regions to assist in classification belongs to fine-grained classification. Several studies on fine-grained classification algorithms used detailed local area labels to train the model so that the model can accurately locate the regions containing detailed semantic information. However, this approach requires a huge amount of work in the production of labels. In other studies, unsupervised learning was used to locate key areas. Although this strategy eliminates tedious label-making work, the accuracy of the model in locating key details can not be ensured. The multi-label classification method proposed in this study is mainly divided into three steps. First, navigator-teacher-scrutinizer network (NTS-Net) is used as the basic network, and the feature extraction network is improved into a deformable ResNet-50 network in accordance with the various properties of the bolt target shape. Second, the navigator network in NTS-Net continuously learns and provides k regions with the most information under the guidance of the teacher network to obtain the discriminative region of the bolt target. Lastly, to make the model use discriminant features effectively, the input features of the k regions receiving the most information from the navigator network are extracted, and corresponding feature vectors are generated and connected to the feature vectors of the entire input image. Afterward, the features need to be passed through the channel attention module, which can enhance the feature with a large weight and suppress the feature with a small weight. **Result** This study uses the bolt multi-attribute classification dataset to evaluate the model. The bolt defect images are from samples obtained by UAV line inspection. The data sample has a total of 2 000 pictures, of which 1 500 are used as training samples and 500 are used as test samples. The bolt defect attributes are divided into six categories based the idea of visual separability. Each bolt defect image contains one or more defect attributes, which can be divided into the following six categories: a pin hole is present, shim is present, a nut is present, rust is present, the nut is loose, and the pin is loose; they are labeled 0—5 respectively. In the multi-label classification task in this study, a 1×6 matrix is constructed for each picture as the label of the picture. If the corresponding attribute category exists, the value is set to 1 and vice versa. Experimental results show that the mean average precision of the proposed method in the bolt multi-attribute classification dataset is 84.5%, which is 10%~20% higher than the accuracy of multi-label classification using traditional networks. **Conclusion** The feature extraction capability of the network is improved through deformable convolution, and the channel attention mechanism is introduced to realize the efficient utilization of the local features provided by NTS-Net. Experimental results show that the proposed method performs better than the traditional method in the bolt multi-attribute classification dataset. The proposed method provides a new idea for applying multi-attribute information to bolt defect reasoning and realizing bolt defect detection.

Key words: bolt defect; deformable convolution; NTS-Net network; multi-label classification; channel attention

0 引言

近年来,中国基于无人机的输电线路巡检技术取得了快速进步与发展。对于如何在复杂的自然背景下,利用计算机视觉技术,从航拍巡检图像中自动精确地定位线路设备(如导线、绝缘子和螺栓等),并准确检测其缺陷,已成为一项关键技术问题(赵振兵等,2019)。输电线路中的缺陷主要包括杆塔缺陷、导地线缺陷、绝缘子缺陷和金具缺陷。其中,因为金具和绝缘子等器件体积较大,缺陷明显,易于识别和检测。相比之下,螺栓数量众多,广泛存在于杆塔、绝缘子和金具之中,且螺栓体积较小、缺陷类型复杂,导致对于螺栓缺陷的识别和检测困难。螺栓在输电线路分布广泛,起到固定部件、维持输电线路结构稳定的关键作用。但由于其庞大的数量和受力复杂等条件,螺栓容易从正常状态转变为缺陷状态。一般缺陷如螺母松动、销子松动等会导致部件松动,影响输电线路结构稳定;危急缺陷如销子缺失、螺母缺失等,可能导致大面积线路故障,严重威胁电网安全运行。因此,螺栓缺陷的检测是确保输电线路正常运行的必要工作。

目前,深度学习技术在杆塔缺陷、导地线缺陷、绝缘子缺陷和金具缺陷的视觉检测、识别和分类的研究中均取得了很好的效果,但在螺栓缺陷检测方面的研究相对较少。Chen 等人(2018)提出基于 SSD(single shot multibox detector)与 YOLO(you only look once)网络构建三级缺陷检测结构,首先在定位支撑装置后对螺栓进行检测,随后单独构建深度神经网络用于缺陷的识别,以完成接触网支撑装置上螺栓缺陷的检测。Zhong 等人(2019)在缺销缺陷上展开更深入的研究,通过构建优化的 PVANET(performance vs accuracy net)结构以统一不同深度神经网络的结构,将最后一级的识别改为对多个局部区域的检测,综合多个区域的视觉信息来判断开口缺销的缺陷,得到很好的效果。Zhao 等人(2020)提出了一种针对缺销缺陷自动检测模型 AVSCNet,该模型提出一种可视化螺栓形状的无监督聚类方法,并应用该方法构建一种能够学习视觉形状差异的缺陷检测模型。然后,在模型中使用 3 种深度卷积神经网络优化方法对区域特征进行回归计算和分类,得到缺陷检测结果。薛阳等人(2020)提出了一种基于改进 Faster

R-CNN(region convolutional neural network)模型的检测方法,首先对图像数据进行翻转、旋转和平移等方式进行数据扩增,再通过增加网络深度的方式提高模型的特征提取能力以解决螺栓体积小问题。

虽然目标检测算法在螺栓缺陷检测方面取得一定的成果,但是螺栓缺陷中的一些如销子缺失、垫片缺失和螺母缺失等属于视觉不可分的范畴,因为螺栓在某些电力部件上不需要销子、垫片或者螺母,故仅依靠基于目标检测算法无法解决螺栓缺陷检测问题。而螺栓是否有销孔、是否有垫片和是否有螺母等属性等属于视觉可分的类别,所以本文将螺栓识别作为多属性分类问题,采用多标签分类方法获取准确的螺栓多属性信息,为螺栓缺陷检测奠定基础。

自 Krizhevsky 等人(2012)提出 AlexNet 后,越来越多的深度卷积神经网络相继被提出,这些网络的深度不断提升。残差卷积神经网络(residual networks, ResNets)(He 等,2016)通过越层连接将输入直接传到输出,故底层可以直接接收到顶层的信息,缓解了梯度消失的问题。为进一步提升 ResNets 的性能,研究人员提出诸多残差网络变体,如宽残差网络(wide residual networks, WRNs)(Zagoruyko 和 Komodakis,2016)、ResNeXt(Xie 等,2017)、多级残差神经网络(residual networks of residual networks, RoR)(Zhang 等,2018)和金字塔残差网络(pyramidal RoR, PProR)(Zhang 等,2019)等。这些具有强大表征能力的卷积神经网络的快速发展,使其在单标签图像分类任务中取得了突破性的进展,为多标签图像分类提供了一定的思路。然而,由于图像中存在目标不显著、目标差异较小以及背景复杂等问题,基于深度卷积神经网络的单标签图像分类方法不适合直接应用于多标签图像分类任务中。

处理多标签图像分类任务的常见方法是将其转换为多个单标签图像分类任务。Razavian 等人(2014)首先使用大规模数据集 ImageNet(Deng 等,2009)预训练网络模型,随后将该网络模型的参数迁移到多标签网络模型,并使用网络模型输出的图像特征来训练每个标签的支持向量机分类器,取得了一定的分类效果。Wei 等人(2016)提出了 HCP(hypotheses-CNN-pooling)网络模型,该模型首先利用 BING(binarily normed gradients)(Cheng 等,2014)算法提取一系列的候选块,并假设每一块候选块都包含单个目标,随后通过聚类方式从所有候选块

中挑选一定量的候选块作为网络的输入,网络会对每个输入候选块输出一个分类结果,最终通过类别最大池化的方式进行融合得到多标签预测结果。另一方面,研究人员利用递归神经网络(recurrent neural network, RNN)对标签之间的依赖性进行建模(Wang 等, 2016)。但是,无论是直接将多标签分类任务转化为单标签分类任务,还是利用 RNN 建模标签的相关性,上述方法在构建多标签分类模型时都不加区别地分析整个图像,以便平等地考虑无用和多余的信息。

与一般的图像多标签分类任务相比,输电线路螺栓多属性分类任务中待检目标具有相似的整体特征,分类模型需要对可以区分不同类别属性的关键局部特征进行抓取,属于细粒度分类的范畴。部分细粒度分类算法的研究利用细致的局部区域标签来训练模型,使得模型能够对包含细节语义信息的区域准确定位(Shih 等, 2015)。这类方法虽然在识别准确率上获得了较好的表现,但往往需要在标签的制作上耗费巨大的工作量;另一些研究则采用无监督学习的方法来定位关键区域(Fu 等, 2017),这种策略虽然摆脱了烦琐的标签制作工作,但难以保证模型对关键细节区域定位的准确性。

针对上述问题,本文提出了一种改进 NTS-Net

的螺栓属性多标签分类方法,本文模型利用 NTS-Net(navigator-teacher-scrutinizer network)(Yang 等, 2018)作为基本框架,通过引入可变形卷积实现对几何变换幅度较大的螺栓目标关键局部区域准确定位,并在最后将局部特征与全局特征融合后引入通道注意力机制,最终实现螺栓多属性分类。

1 本文方法

本文方法网络模型结构如图 1 所示,所提出的多标签分类方法主要分为 3 个步骤:首先采用 NTS-Net 作为主干网络,并针对螺栓目标形状变化多样的属性将特征提取网络改进为可变形 ResNet-50 网络;随后,为了获取螺栓目标的判别性区域,NTS-Net 中的导航器网络在教师网络的指导下不断学习并提供信息量最多的 k 个区域;最后,为了让模型能够更加有效地利用判别性特征,在将从导航器网络接收到的信息量最多的 k 个区域输入特征提取网络生成对应的特征向量并与整个输入图像的特征向量相连接后,将其经过通道注意力模块以获取通道权重,增强权重大即更重要的特征并抑制权重小即更不重要的特征。图 1 中,⊗表示矩阵相乘,⊕表示特征相加。

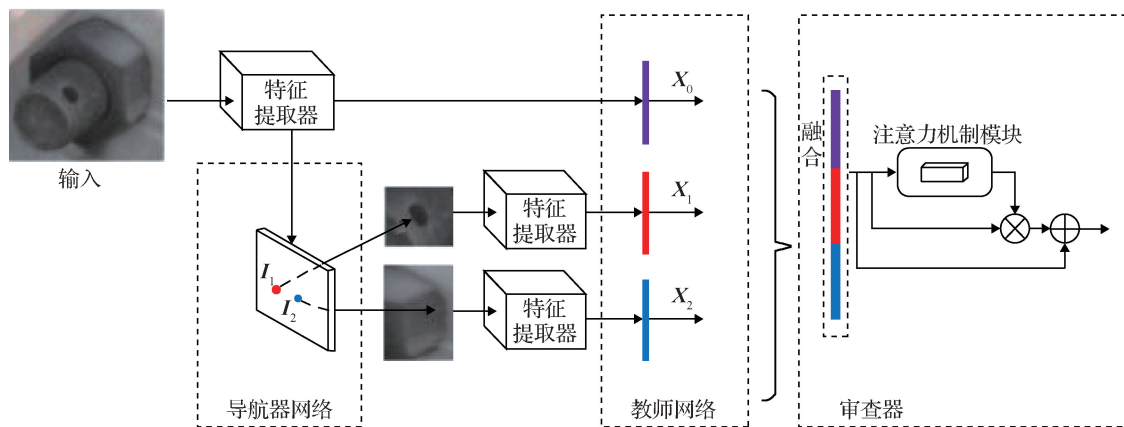


图1 本文模型的具体实现过程

Fig. 1 The concrete realization process of our model

1.1 可变形卷积以及特征提取网络

深度卷积神经网络由于其固定的几何结构而固有限制于模型几何变换。针对此问题,Dai 等人(2017)提出了可变形卷积模块,基本思想是利用带偏移的采样代替原来的固定位置采样,最终实现以自适应的方式学习目标的几何形变。因此,本文为了减少螺栓几何形变的限制,引入了可变形卷积,对

卷积核中每个采样点的位置都增加一个偏移的变量,故卷积核在当前位置附近可随意地采样而不再局限于之前的规则格点。

可变形卷积通过在采样点添加额外的偏移量来更改标准卷积内核的采样位置,获得的补偿可以通过训练来学习而不需要额外的监督。图 2 显示了标准卷积核与可变形卷积核之间的区别。图 2 中左图

显示标准卷积的采样网络是规则的,右图是将可学习的偏移量添加到标准卷积核之后的可变形卷积核。可以看出,可变形卷积的采样位置是不规则的。

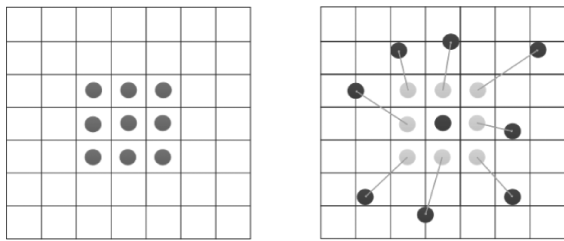


图2 标准卷积与可变形卷积

Fig. 2 Standard convolution and deformable convolution

图3展示了可变形卷积层的实现过程,在输入层中原本固定采样区域分别加入偏移向量,然后在卷积核中通过双线性插值算法将带有偏移向量的采样点集中在一起,并且使得带偏移向量的输出特征图与输入的特征图具有一致的空间分辨率,最终传递给下一个网络层。这样,在不改变空间分辨率的情况下,实现对不规则图像特征的提取。

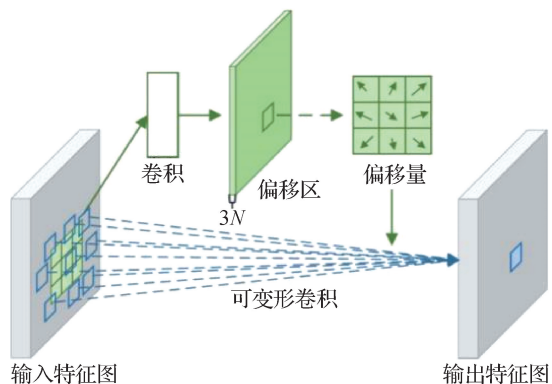


图3 可变形卷积层的实现过程

Fig. 3 Implementation process of deformable convolutional layer

本文提出的网络模型以 ResNet-50 作为特征提取网络,残差网络的关键是引入了残差块,在网络中增加了跃层连接,允许原始输入信息直接传到后面的层。本文设计的可变形 ResNet-50 网络如图4所示,设计将 conv5_x 层中的3个 3×3 传统卷积替换为可变形卷积,构成可变形卷积的 ResNet-50 网络。所以,经过改进后,原来规则网络 R 会附加一个偏移量 $\{\Delta p_s | s = 1, \dots, A\}$, $A = |R|$,对于输出特征图 U 中的每个位置 p_0 , $U(p_0)$ 计算为

$$U(p_0) = \sum_{p_s \in R} w(p_s) \cdot I(p_0 + p_s + \Delta p_s) \quad (1)$$

式中, w 表示权值, $I(\cdot)$ 表示输入特征图。根据可

变形卷积的原理,首先在输入特征图的每一个位置生成一个偏移量,并将其前向传播;在后向传播过程中,新增的偏移位置 $p_s + \Delta p_s$ 会被同时学习,然后经过双线性插值得到整数的空间位置,进而执行可变形卷积。由此可知,可变形卷积的 ResNet-50 网络能够提取到更加丰富的判别性特征,增强网络对目标形变的检测能力。

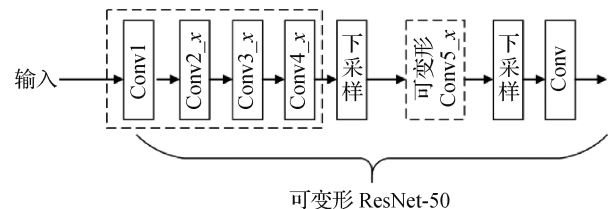


图4 特征提取网络

Fig. 4 Feature extraction network

1.2 导航器网络与教师网络

本文方法所基于的假设是:信息丰富的区域有助于更好地表达对象,融合信息丰富区域和全图将获得更好的效果。因此,目标在于定位信息最丰富的区域。假设所有区域都是矩形,定义 A 来表示给定图像中所有信息区域的集合。定义信息函数 $\mathcal{J}: A \rightarrow (-\infty, \infty)$ 以评估区域 $R \in A$ 的信息值,定义置信函数 $\mathcal{C}: A \rightarrow [0, 1]$ 作为分类器以评估该区域属于真实类别的置信度。因此,对于信息更加丰富的区域应该具有更高的置信度,即对任何 $R_1, R_2 \in A$, 如果 $\mathcal{C}(R_1) > \mathcal{C}(R_2)$, 则 $\mathcal{J}(R_1) > \mathcal{J}(R_2)$ 成立。

在本文方法中,分别使用导航器网络和教师网络来近似信息函数 $\mathcal{J}$ 以及置信度函数 $\mathcal{C}$ 。在区域空间 A 中选择 M 个区域 A_M , 对于每个区域 $R_i \in A_M$, 通过导航器网络评估其信息量 $\mathcal{J}(R_i)$, 教师网络评估其置信度 $\mathcal{C}(R_i)$ 。

导航器网络的任务是找到可能提供信息的区域,如何导航到信息区域可以视为区域建议问题,已获得广泛研究。导航器网络在对原图中所有局部区域预测信息量时需要提取每个区域的特征向量,为了减少重复计算,导航器网络首先对全图提取全局特征,再由全图到特征图的映射关系在全局特征图上抠取局部区域的局部特征。受到锚点的启发,导航器网络在不同尺度特征图上设置了在固定位置上固定尺度和不同比例的锚点框,锚点框的作用是更好地确定局部区域的范围以抽取局部特征来进行

信息量的预测。也说明了导航器网络的作用是通过预测信息量来选取其中可能包含最多语义信息的区域。导航器网络能够生成一系列矩形区域 $\{R'_1, R'_2, \dots, R'_A\}$, 每个区域都有对应的分数代表该区域的信息量, 随后生成一个列表来表示所有锚点的信息量。按照式(2)对信息列表进行排序, 即

$$\mathcal{J}(R_1) \geq \mathcal{J}(R_2) \geq \dots \geq \mathcal{J}(R_A) \quad (2)$$

为了减少区域冗余, 本文基于区域的信息量对区域采用非极大值抑制 (non-maximum suppression, NMS), 然后选择前 M 个信息区域 $\{R_1, R_2, \dots, R_M\}$ 输入教师网络。教师网络接收到从导航器网络产生的信息量最大的 M 个局部区域后, 通过调整这些局部区域的尺度为固定大小, 并送入特征提取网络以得到局部特征图, 在对每个局部区域的特征图完成全局池化后送入全连接层预测置信度, 得到对应的置信度 $\{\mathcal{C}(R_1), \mathcal{C}(R_2), \dots, \mathcal{C}(R_M)\}$ 。

在本文的训练过程中, 利用网络模型预测排序和正确排序计算损失函数, 并通过优化损失函数以使网络模型预测的排序更加接近正确排序。导航器网络提取信息量最大的 M 个局部区域表示为 $R = \{R_1, R_2, \dots, R_M\}$, 其对应的信息量集合为 $\mathcal{J} = \{\mathcal{J}(R_1), \mathcal{J}(R_2), \dots, \mathcal{J}(R_M)\}$, 由教师网络预测得到的置信度集合为 $\mathcal{C} = \{\mathcal{C}(R_1), \mathcal{C}(R_2), \dots, \mathcal{C}(R_M)\}$ 。其中 $\mathcal{J}$ 和 $\mathcal{C}$ 分别对应于网络的预测排序

和正确排序。则导航器网络损失函数可以定义为

$$L_N(\mathcal{J}, \mathcal{C}) = \sum_{(i,j): \mathcal{C}(R_i) > \mathcal{C}(R_j)} g(\mathcal{J}(R_i) - \mathcal{J}(R_j)) \quad (3)$$

式中, f 函数为合页损失函数 $f(x) = \max\{1 - x, 0\}$ 。若导航器网络提取某两个局部区域的信息量和教师网络对该两个局部区域预测的置信度大小一致, 则该损失函数对导航器网络产生激励; 反之, 损失函数则对网络进行惩罚。

教师网络预测的置信度指标是由真实样本类和教师网络所预测置信度之间的交叉熵损失函数来优化, 其表达式为

$$L_T = - \sum_{i=1}^M [P_d \cdot \log \mathcal{C}(R_i) + (1 - P_d) \cdot \log(1 - \mathcal{C}(R_i))] \quad (4)$$

式中, $\mathcal{C}$ 值表示教师网络预测的置信度, P_d 表示原图是否包含该类别的标注。

1.3 审查网络与通道注意力模块

随着导航器网络逐渐收敛, 审查网络将接收到的信息丰富排名前 k 的区域输入特征提取器以生成它们的特征向量, 并将此 k 个特征与输入图像的全局特征相连接。若将连接后的特征直接输入全连接层中, 这会对图像所有特征平均地处理。为了有选择性地加强包含有用信息的特征并抑制无用特征, 本文引入了如图5所示的通道注意力机制模块。

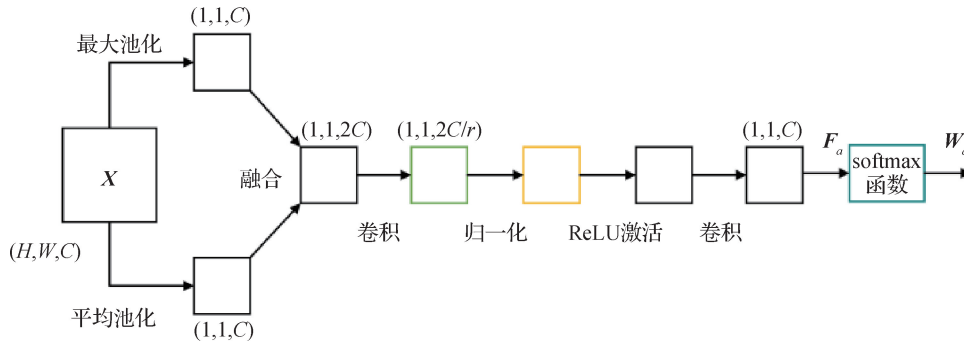


图5 通道注意力机制模块

Fig.5 Channel attention mechanism module

传统算法压缩激活网络 SENet (Hu 等, 2017)、BAM (bottleneck attention module) (Park 等, 2018) 在实现通道注意力时采用平均池化对空间维度进行压缩, 不能有效提取纹理特征; CBAM (convolutional block attention module) (Woo 等, 2018) 将平均池化结果与最大池化结果直接相加, 结合方式过于简单。

为了充分保留背景和纹理信息, 本文采取将两个池化结果进行拼接的方法, 如图5所示。

假设输入特征 X 的维度为 (B, H, W, C) , 在维度 C 上对 X 中每幅特征图的重要程度进行权重分配。具体做法是: 为了去除空间位置信息的干扰, 需要先对原有特征图的空间进行压缩, 将其从空间

(B, H, W, C) 映射到空间 $(B, 1, 1, C)$ 当中;随后分别使用全局最大池化和全局平均池化的方法,将二者的池化结果进行拼接,得到的特征图维度为 $(B, 1, 1, 2C)$ 。由于原始输入特征图的通道数为 C ,需要通过两个 1×1 大小的卷积核来实现通道数的降维,以进一步提取通道特征。通道压缩率表示为 r ,在本文实验中取 $r = 16$ 。上述过程可表示为

$$\begin{aligned} F_a = & \text{Conv}(\text{ReLU}(\text{BN} \\ & (\text{Conv}(\text{concat}(\text{maxpool}, \text{avgpool})))))) \quad (5) \end{aligned}$$

式中, F_a 代表通道特征矩阵, BN 为归一化, ReLU 为激活函数, Conv 表示卷积, concat 为拼接函数, maxpool 为最大池化操作, avgpool 为平均池化操作。 $F_a \in \mathbf{R}^{(B, 1, 1, C)}$, 经过 softmax 后, 得到相应的权重矩阵 $W_a \in \mathbf{R}^{(B, 1, 1, C)}$ 。然后通过元素相乘逐通道加权到原先的特征图上, 从而完成通道维度的注意力机制。

随后, 将经过上述处理得到的特征接入具有 $(k + 1)$ 个神经元的全连接层, 在通过最小化交叉熵损失对整个网络模型进行端到端训练之后, 网络模型通过对比概率值与阈值来实现对输入图像的多标签分类预测。

2 定量实验结果与分析

2.1 数据集

本文使用螺栓多属性分类数据集对模型进行评估, 螺栓缺陷图像均由无人机线路巡检拍摄的样本截取而来, 数据样本一共有 2 000 幅图像, 其中 1 500 幅为训练样本, 500 幅为测试样本。利用视觉可分性的思想, 将螺栓缺陷属性分为 6 类, 每幅螺栓缺陷图像包含 1 类或以上缺陷属性。螺栓样本的属性及其对应标签如表 1 所示, 具体可分为以下 6 类: 是

表 1 6 种螺栓属性及其标签

Table 1 Six kinds of bolt attributes and their labels

标签	属性
0	是否有销孔
1	是否有垫片
2	是否有螺母
3	是否锈蚀
4	螺母是否松动
5	销子是否松动

否有销孔、是否有垫片、是否有螺母、是否锈蚀、螺母是否松动、销子是否松动, 并将其分别设置为标签 0 ~ 5。

在本文多标签分类任务中, 可以为每幅图像构建一个 1×6 的矩阵作为图像的标签, 其中对应的属性类别如果存在, 则置 1, 反之置 0。例如以下编号为 501 的螺栓图像中存在销孔和螺母, 其对应的标签矩阵为 $(1, 0, 1, 0, 0, 0)$ 。一些螺栓缺陷图像样本及其对应标签矩阵如图 6 所示。

编号	图像	标签: 0	1	2	3	4	5
500		0	1	1	0	1	0
501		1	0	1	0	0	0
502		0	0	1	0	0	1
503		0	0	1	1	0	0

图 6 螺栓缺陷图像样本及其对应标签矩阵
Fig. 6 Picture samples and corresponding label matrix of bolt defects

2.2 实验细节

本文实验使用 PyTorch 框架实现, 并使用单个 NVIDIA Geforce GTX Titan X GPU 对本文方法进行训练和测试。在所有实验中, 将超参数 k 固定为 2, 即使用 2 个区域来训练导航器网络。本文使用可变形卷积的 ResNet-50 网络作为特征提取器, 对输入图像进行随机水平翻转, 并使用批处理规范化作为正则化。模型训练的初始学习率设置为 0.001, 且在每 60 个周期降低为之前的 $1/10$, 权重衰减为 0.000 5, NMS 的阈值设置为 0.25, 使用动量梯度下降法进行优化。

2.3 实验结果

为了充分验证本文提出模型的性能, 将螺栓多属性分类数据集在 VGG (Visual Geometry Group)、ResNets 等其他基础网络上进行多标签分类, 记录其结果作为对比数据。表 2 显示了本文提出模型与其他网络模型之间的精确率比较。从表中可以看出, ResNet-50 是效果较好的基准, 本身就可以达到

77.2% 的平均精确率。另外, 使用基于 ResNet-50 的 NTS-Net 模型进行多标签分类时, 能够获得 82.6% 的平均精确率, 相较于基准获得了 5.4% 的提升, 而本文提出的模型能达到 84.5% 的平均精确率, 明显优于 NTS-Net 近 2 个百分点。而且从表 2 中可以看出, 在螺栓多属性分类数据集的 6 个标签的精确率中, 所有标签的精确率都相较于以往的方法有着更好的表现, 这说明通过本文的改进方法能够有效提升对螺栓多属性分类数据集的多标签分类效果。

表 2 螺栓多属性分类数据集分类结果

Table 2 Classification result of bolt multi-attribute classification dataset

方法	0	1	2	3	4	5	mAP
AlexNet	72.5	66.5	69.1	80.5	67.6	58.5	68.9
VGG16	74.3	69.4	71.5	81.8	71.1	60.1	71.3
ResNet18	77.8	71.2	74.2	83.2	72.1	63.4	74.7
ResNet34	78.6	71.9	75.3	84.4	73.2	63.5	75.4
ResNet50	80.0	74.3	77.4	88.3	74.8	64.8	77.2
NTS-Net	83.5	79.9	82.8	91.1	82.7	72.3	82.6
本文	84.7	81.1	84.0	93.5	84.5	74.3	84.5

注: 加粗字体表示每列最优结果。

2.4 消融实验

为验证所提出模型改进方法的有效性, 本文在螺栓多属性分类数据集上分别设计了不同模块的对比实验。

1) 为了验证在不同位置嵌入可变形卷积对于特征提取效果的影响, 本文设计了分别使用标准卷积的 ResNet-50 网络与不同位置嵌入可变形卷积的 ResNet-50 网络作为特征提取网络的 NTS-Net 对螺栓多属性分类数据集多标签分类的对比实验, 实验结果如表 3 所示。实验结果表明, 可变形卷积增强了网络对于几何变换的建模能力, 平均分类精确率获得了不同幅度的提升。另外, 将 ResNet-50 不同位置的 3×3 传统卷积替换为可变形卷积对实验结果产生不同的影响, 位置越靠后对于模型性能的提升幅度越大。因此, 在本文其他实验中将 Conv5_x 层中的 3 个 3×3 传统卷积替换为可变形卷积, 构成可变形卷积的 ResNet-50 网络。

表 3 各模块对图像分类效果的影响

Table 3 The influence of each model on image classification

方法	0	1	2	3	4	5	mAP
ResNet-50	83.5	79.9	82.8	91.1	82.7	72.3	82.6
dcn-Conv2	83.7	80.2	82.7	91.1	83.2	73.1	83.1
dcn-Conv3	83.7	80.3	83.2	91.3	83.0	72.9	83.2
dcn-Conv4	83.5	80.5	83.4	91.3	83.7	73.1	83.5
dcn-Conv5	83.9	80.5	83.5	92.1	83.9	73.5	83.7

注: 加粗字体表示各列最优结果, dcn 为可变形卷积。

2) 为分析 NTS-Net 模型中不同组件以及超参数 k 的作用, 本文设计了以下对比实验并将结果记录在表 4 中。实验结果表明, 当没有教师网络的指导 (即 NS-Net), 仅让导航器网络提出区域建议时, 模型准确率下降为 76.8%, 这说明没有受到指导时, 导航器随意地提出区域建议是不利于分类的。另外, 为研究超参数 k (即使用多少个局部区域来帮助分类), 本文分别设置 k 的不同取值来进行对比实验。由表 4 可知, 当用来辅助分类的局部区域数 k 从 0 增加到 2 时, 平均精确率提升了 2.1%。当 k 再增加时, 平均精确率的提升几乎可以忽略, 而特征维度几乎增加了一倍, 所以本文设置超参数 k 的值为 2。

表 4 超参数 k 对实验结果的影响Table 4 The influence of hyperparameter k on experimental

方法	0	1	2	3	4	5	mAP
ResNet-50	80.0	74.3	77.4	88.3	74.8	64.8	77.2
NS-Net	79.2	74.1	76.9	88.2	74.5	64.6	76.8
本文 ($k=0$)	83.5	79.7	82.7	91.0	82.6	72.5	82.4
本文 ($k=1$)	83.8	80.6	83.3	91.5	83.2	73.1	83.2
本文 ($k=2$)	84.7	81.1	84.0	93.5	84.5	74.3	84.5
本文 ($k=3$)	84.8	81.0	84.2	93.1	84.7	74.6	84.7
本文 ($k=4$)	84.7	81.3	84.4	93.3	84.5	74.7	84.8

注: 加粗字体表示每列最优结果。

3) 为验证本文通道注意力机制的有效性, 分别在使用标准卷积和可变形卷积的 ResNet-50 网络作为特征提取网络的 NTS-Net 上加入通道注意力机制。从表 5 可以看出, 在加入通道注意力机制后 (在表 5 中以 att 表示), 最终的 mAP 为 83.0% 和

84.5%,分别获得了0.4%和0.8%的提升,说明了本文提出的通道注意力机制的有效性。

表 5 通道注意力机制对分类效果的影响
Table 5 The influence of channel attention on classification

方法	/%						
	0	1	2	3	4	5	mAP
ResNet-50	83.5	79.9	82.8	91.1	82.7	72.3	82.6
ResNet-50 + att	83.7	80.1	82.8	91.3	83.1	72.7	83.0
dcn-ResNet-50	83.9	80.5	83.5	92.1	83.9	73.5	83.7
dcn-ResNet-50 + att	84.7	81.1	84.0	93.5	84.5	74.3	84.5

注:加粗字体表示每列最优结果。方法中“att”表示加入注意力机制。

另外,为了验证本文采用的通道注意力机制方法相较于 SENet、BAM 以及 CBAM 具有优越性,本文设计在使用可变形卷积的 ResNet-50 网络作为特征提取网络的 NTS-Net 上分别加入 4 种注意力机制进行对比实验,实验结果如表 6 所示。从表中可以看出,本文方法在各类精确率和平均精确率上都能够获得更好的表现,验证了本文通道注意力机制的优越性。

表 6 不同通道注意力机制对分类效果的影响
Table 6 The influence of different channel attention on classification

方法	%						
	0	1	2	3	4	5	mAP
SENet	83.9	80.7	83.7	92.4	83.1	73.7	83.8
BAM	83.9	80.7	83.5	92.8	83.1	73.9	83.9
CBAM	84.3	80.9	83.7	93.4	83.9	74.1	84.2
本文	84.7	81.1	84.0	93.5	84.5	74.3	84.5

注:加粗字体表示每列最优结果。

2.5 通用性实验

本文所提出的模型是基于螺栓目标的多标签分类任务,也适用于其他场景,为了验证本模型的通用性,使用包含 20 个类别和 9 963 幅图像的 PASCAL VOC2007 (pattern analysis, statistical modeling and computational learning visual object classes) 数据集进行测试,分别对 5 011 幅图像进行训练和 4 952 幅图像进行测试。结果如表 7 所示。从表 7 可以看出,本文方法相较于 ResNet-50 基础网络在 PASCAL VOC2007 数据集上的效果有一定程度的提升,验证了本文模型的通用性。

表 7 PASCAL VOC2007 数据集分类结果

Table 7 Experimental results of the PASCAL VOC2007 dataset

/%	
方法	mAP
ResNet-50	88.3
本文	93.7

注:加粗字体表示各列最优结果。

3 定性实验结果与分析

第 2 节中,通过设置不同对比实验定量分析了本文模型在螺栓多属性分类数据集上的有效性。为了直观感受本文模型改进所带来的效果上的提升,本文设置了如下定性实验。

1) 为了分析本文模型的导航器网络所提供的判别性区域,将 k 取值为 2 和 3 时,网络所提供的判别性区域绘制在图 7 中。其中,第 1 行是原始图像;第 2 行是 $k=2$ 时的判别性区域;第 3 行是 $k=3$ 时的判别性区域;并用红色、橙色、绿色矩形分别表示网络提供的信息量前 3 的区域。可以看出,当 $k=2$ 时,信息量较大的局部区域能够对螺栓不同属性的分类提供有效信息,如第 2 行第 2 幅图像中,重点关注了销子和螺母是否松动的区域,这与人类的感知行为是类似的。然而,当 $k=3$ 时,绿色框提供的判别性区域出现重叠冗余,对分类效率的提升不大,这验证了定量分析中 k 取值大于 2 时,分类精确率提升有限的问题。

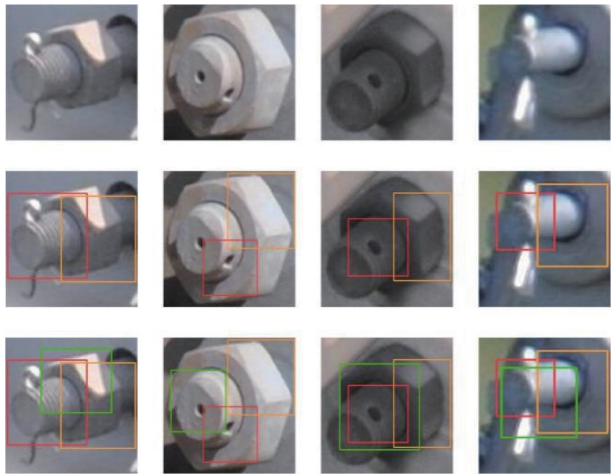


图 7 导航器网络提供的判别性区域

Fig. 7 The most informative regions proposed by navigator network

2) 为了证明不同判别性区域对不同属性分类决策的影响, 本文还进行了可视化实验, 如图 8 所示。采用 Grad-CAM (gradient-weighted class activation mapping) 算法 (Selvaraju 等, 2017), 该算法将判别性区域图像 (第 2 行和第 4 行图像) 作为输入, 通过将热力图和反向传播进行逐元素相乘, 得到 Grad-CAM 可视化图 (第 3 行和第 5 行图像)。从图中可以看出, 当 $k=2$ 时, 导航器网络提供的两个判别性区域上注意到的信息是不一样的, 如第 1 列, 两个判别性区域分别关注销孔和螺母区域。

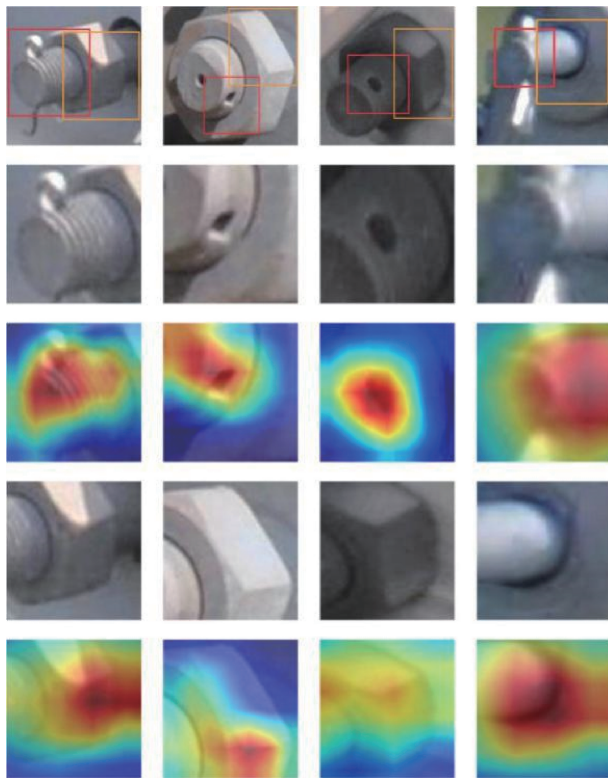


图 8 判别性区域特征图可视化

Fig. 8 Visualization of discriminative regional feature map

3) 本文利用 Grad-CAM 算法可视化了模型改进前后的特征图注意力区域, 如图 9 所示。其中, 第 1 行图像为原始图像, 第 2 行图像为模型改进前特征图注意力区域的可视化效果, 第 3 行图像为经过本文模型改进后的特征图注意力区域的可视化效果。从图中可以看出, 第 3 行图像与第 2 行图像相比, 网络能够重点关注更具显著性和判别性的特征, 同时能够有效去除冗余信息, 效果提升较为明显。如第 3 列的第 3 行图片相比第 2 行图像, 模型关注区域更加集中在销孔位置。

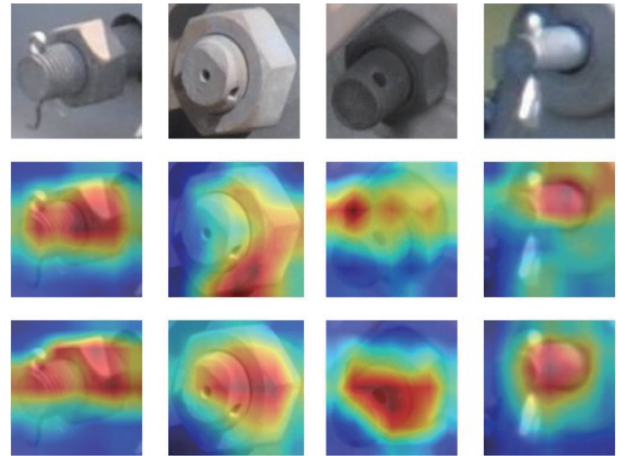


图 9 特征图注意力区域可视化

Fig. 9 Visualization of areas of attention in feature map

4 结 论

本文针对目前螺栓缺陷检测准确率较低的问题, 提出了一种改进 NTS-Net 网络的螺栓属性多标签分类方法。传统的卷积方式无法有效地模拟几何变换, 本文引入的可变形卷积方法对螺栓目标的空间变换有着较强的自适应能力。为了更好地辅助分类, 本文采用 NTS-Net 作为基本网络框架, 能够提供信息量较大的判别性局部区域。另外, 在最后将局部特征与全局特征融合时引入通道注意力机制, 提取特征的通道权重, 获取关键通道特征, 从而更好地利用判别性区域信息。实验表明, 本文方法在螺栓多属性分类数据集上的表现优于传统方法, 为后续将多属性信息用于螺栓缺陷推理, 实现更精确的螺栓缺陷检测提供了新的思路。

但本文提出模型在螺栓缺陷数据集上对不同缺陷属性依然存在分类效果优劣相差较大的问题, 这是由于在数据集上不同属性缺陷的数据量差别较大。为此, 在今后的研究中需要进一步改进模型, 研究为不同属性类别设置合适的权重系数, 解决分类效果差距明显的问题。

参考文献 (References)

- Chen J W, Liu Z G, Wang H R, Núñez A and Han Z W. 2018. Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. IEEE Transactions on Instrumentation and Measurement, 67 (2): 257-269 [DOI: 10.

- 1109/tim. 2017. 2775345]
- Cheng M M, Zhang Z M, Lin W Y and Torr P. 2014. BING: binarized normed gradients for objectness estimation at 300 fps//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA; IEEE; 3286-3293 [DOI: 10.1109/CVPR. 2014. 414]
- Dai J F, Qi H Z, Xiong Y W, Li Y, Zhang G D, Hu H and Wei Y C. 2017. Deformable convolutional networks//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy; IEEE; 764-773 [DOI: 10.1109/ICCV. 2017. 89]
- Deng J, Dong W, Socher R, Li L J, Li K and Li F F. 2009. ImageNet: a large-scale hierarchical image database//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA; IEEE; 248-255 [DOI: 10.1109/CVPR. 2009. 5206848]
- Fu J L, Zheng H L and Mei T. 2017. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE; 4476-4484 [DOI: 10.1109/CVPR. 2017. 476]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA; IEEE; 770-778 [DOI: 10.1109/CVPR. 2016. 90]
- Hu J, Shen L, Albanie S, Sun G and Wu E H. 2017. Squeeze-and-Excitation networks[EB/OL]. [2020-10-11]. <https://arxiv.org/pdf/1709.01507.pdf>
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, USA; Curran Associates Inc.; 1097-1105 [DOI: 10.5555/2999134.2999257]
- Park J, Woo S, Lee J Y and Kweon I S. 2018. BAM: Bottleneck attention module [EB/OL]. [2020-10-11]. <https://arxiv.org/pdf/1807.06514.pdf>
- Razavian A S, Azizpour H, Sullivan J and Carlsson S. 2014. CNN features off-the-shelf: an astounding baseline for recognition//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, USA; IEEE; 512-519 [DOI: 10.1109/cvprw. 2014. 131]
- Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy; IEEE; 618-626 [DOI: 10.1109/ICCV. 2017. 74]
- Shih K J, Mallya A, Singh S and Hoiem D. 2015. Part localization using multi-proposal consensus for fine-grained categorization//Proceedings of the British Machine Vision Conference (BMVC). Guildford, UK; BMVA Press; #128 [DOI: 10.5244/c.29.128]
- Wang J, Yang Y, Mao J H, Huang Z H, Huang C and Xu W. 2016. CNN-RNN: a unified framework for multi-label image classification//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE; 2285-2294 [DOI: 10.1109/CVPR. 2016. 251]
- Wei Y C, Xia W, Lin M, Huang J S, Ni B B, Dong J, Zhao Y and Yan S C. 2016. HCP: a flexible CNN framework for multi-label image classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(9): 1901-1907 [DOI: 10.1109/TPAMI. 2015. 2491929]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: Convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany; Springer; 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Xie S N, Girshick R, Dollár P, Tu Z W and He K M. 2017. Aggregated residual transformations for deep neural networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE; 5987-5995 [DOI: 10.1109/CVPR. 2017. 634]
- Xue Y, Wu H D, Zhang N, Yu Z C, Ye X K and Hua X. 2020. Detection of insulation piercing connectors and bolts on the transmission line using improved faster R-CNN. Laser and Optoelectronics Progress, 57(8): #081008 (薛阳, 吴海东, 张宁, 俞志程, 叶晓康, 华茜. 2020. 基于改进 Faster R-CNN 输电线路穿刺线夹及螺栓的检测. 激光与光电子学进展, 57(8): #081008) [DOI: 10.3788/LOP57.081008]
- Yang Z, Luo T G, Wang D, Hu Z Q, Gao J and Wang L W. 2018. Learning to navigate for fine-grained classification [EB/OL]. [2020-10-11]. <https://arxiv.org/pdf/1809.00287v1.pdf>
- Zagoruyko S and Komodakis N. 2016. Wide residual networks//Proceedings of British Machine Vision Conference 2016. Guildford, UK; BMVA Press; 87. [DOI: 10.5244/C.30.87]
- Zhang K, Guo L R, Gao C and Zhao Z B. 2019. Pyramidal RoR for image classification. Cluster Computing, 22(S2): 5115-5125 [DOI: 10.1007/s10586-017-1443-x]
- Zhang K, Sun M, Han T X, Yuan X F, Guo L R and Liu T. 2018. Residual networks of residual networks: multilevel residual networks. IEEE Transactions on Circuits and Systems for Video Technology, 28(6): 1303-1314 [DOI: 10.1109/TCSVT. 2017. 2654543]
- Zhao Z B, Qi H Y and Nie L Q. 2019. Research overview on visual detection of transmission lines based on deep learning. Guangdong Electric Power, 32(9): 11-23 (赵振兵, 齐鸿雨, 聂礼强. 2019. 基于深度学习的输电线路视觉检测研究综述. 广东电力, 32(9): 11-23) [DOI: 10.3969/j.issn.1007-290X. 2019. 009. 002]
- Zhao Z B, Qi H Y, Qi Y C, Zhang K, Zhai Y J and Zhao W Q. 2020. Detection method based on automatic visual shape clustering for pin-missing defect in transmission lines. IEEE Transactions on Instru-

mentation and Measurement, 69(9): 6080-6091 [DOI: 10.1109/tim.2020.2969057]

Zhong J P, Liu Z G, Han Z W, Han Y and Zhang W X. 2019. A CNN-based defect inspection method for catenary split pins in high-speed railway. IEEE Transactions on Instrumentation and Measurement, 68(8): 2849-2860 [DOI: 10.1109/tim.2018.2871353]

作者简介



张珂, 1980年生, 男, 教授, 博士生导师, 主要研究方向为深度学习、计算机视觉、生物特征识别、电力视觉。

E-mail: zhangkeit@ncepu.edu.cn

何颖宣, 男, 硕士研究生, 主要研究方向为深度学习和图像分类。E-mail: 809205509@qq.com

赵凯, 男, 硕士研究生, 主要研究方向为深度学习和电力视觉。E-mail: 2321834855@qq.com

冯晓晗, 女, 硕士研究生, 主要研究方向为深度学习和图像分类。E-mail: 1127504514@qq.com

赵振兵, 男, 教授, 主要研究方向为电力视觉。

E-mail: zhaozhenbing@ncepu.edu.cn

马占宇, 男, 教授, 主要研究方向为模式识别与机器学习基础理论与方法。E-mail: mazhanyu@bupt.edu.cn