

基于Transformer的道路病害高精度算法

孙润生, 张鲁豫, 张 恭

[济南市市政工程设计研究院(集团)有限责任公司, 山东 济南 250003]

摘要: 为解决目前道路病害检测算法精度低、鲁棒性差的问题, 提出了Bi-Swin Transformer结构。其主干特征提取网络采用四层Swin Transformer Block来捕获全局依赖关系, 在特征融合阶段使用BiFPN融合网络实现特征在不同尺度之间的双向交流。在来自7个不同国家的道路病害数据库上进行训练, 并应用数据增强方式。结果表明, Bi-Swin Transformer有着比Transformer模型及卷积神经网络更为优秀的性能, 更适用于道路病害检测, 平均准确率达到了0.539。其在小目标道路病害中的检测精度仍然为最优, 高于当前先进的检测模型。通过实际检测表明, Bi-Swin Transformer模型不仅具有高精度, 而且具有强鲁棒性。

关键词: 深度学习; 无损检测; 神经网络; 计算机视觉

中图分类号: TP391

文献标志码: A

文章编号: 1009-7716(2025)01-0225-04

High Precision Algorithm of Road Disease Based on Transformer

SUN Runsheng, ZHANG Luyu, ZHANG Gong

[Jinan Municipal Engineering Design and Research Institute (Group) Co., Ltd., Jinan 250003, China]

Abstract: To solve the problems of low precision and poor robustness of current road disease detection algorithm, a Bi-Swin Transformer structure is proposed. Its backbone feature extraction network employs a 4-layer Swin Transformer Block to capture the global dependencies. In the feature fusion stage, a BiFPN fusion network is used to enable the bidirectional communication of features among the different scales. A data enhancement approach is utilized by training on road disease databases from seven different countries. The results indicate that the Bi-Swin Transformer outperforms both the Transformer model and the Convolutional Neural Network, which is more suitable for the detection of road disease. Its average accuracy is 0.539. Its detection accuracy for small-target road diseases is still optimal and higher than the current advanced detection models. The Bi-Swin Transformer model has not only high precision, but also strong robustness through practical detection.

Keywords: deep learning; nondestructive testing; neural network; computer vision

0 引言

随着我国公路网的快速发展, 公路网的通行效率与安全性也开始被人们所关注。道路作为交通基础设施的一部分, 其表面的坑槽、裂缝等病害会对行车安全、通行效率造成较大的影响。随着道路被频繁地使用, 这些病害还会逐渐恶化, 发展得更为严重。因此及时地发现路表病害, 并进行修补, 可以最大限度地减少经济安全损失。目前常用的人工巡检手段费时费力, 难以满足大规模道路病害检测的需求。开发高效率、高精度的检测设备是当下道路病害巡检急需的。路表图像中有着病害丰富的特征, 因此目前学者多研究基于计算机视觉的道路病害智

能检测算法。然而, 实际场景中的道路病害图像十分复杂, 因此需要研究鲁棒性更强的检测算法。通过引入提取特征能力更强的Transformer结构来创建强鲁棒性的道路病害检测模型。

1 道路病害检测研究现状

人工检测道路病害具有精度高的特点, 但在对道路病害进行检测时, 需要封闭道路, 且在大规模检测时动用的劳动力也十分巨大。因此开始出现其他智能检测方法, 多功能道路病害检测车就是其中之一, 其可以高精度地检测路面、路基病害。多功能道路病害检测车可以实现80 km/h的巡检速度, 具有智能化程度高、不干扰实时交通的特点。然而, 高精度道路病害检测车使用成本较为昂贵, 难以适用于大规模场景。

随着深度学习技术的发展, 越来越多人将深度

收稿日期: 2024-02-27

作者简介: 孙润生(1992—), 男, 硕士, 工程师, 从事道路病害智能检测工作。

学习技术用于路表病害的智能检测。周秋红^[1]以SSD模型作为基础算法,实现了不规则标注框的道路病害,相较于规则标注框,精度有了进一步提升。田杨等^[2]通过自行构建道路病害数据库,并基于Fast-RCNN网络进行病害检测,使精度达到了84.9%。喻露等^[3]使用空洞卷积改进YOLO v5网络,从而加强模型提取不同尺度特征的能力。张莹超^[4]提出了YOLO v5-DCN网络,并将其部署到实际检测活动中,并开发了配套的检测软件,但实际检测精度仍较低。为进一步提高精度,张欢欢^[5]将注意力机制引入到YOLO v5模型中,提高模型对重要区域的关注程度,降低对非重要区域的关注程度,从而进一步提升了道路病害的检测精度。虽然基于卷积神经网络的道路病害检测算法已经取得了较高的精度,但实际场景中的各种极端情况,如路面潮湿、阴影等,会极大地影响检测算法的精度。这也是道路病害智能检测技术发展缓慢的原因之一。因此,开发一些鲁棒性强、精度更高的道路病害检测算法是十分必要的。

随着Transformer结构在自然语言处理领域取得了巨大的成功,越来越多的学者开始研究Transformer结构在道路病害检测领域中的应用。视觉Transformer结构^[6]具有精度高、鲁棒性强的特点,可以适用于复杂环境下的目标检测。然而目前将Transformer结构用于道路表面缺陷检测领域的研究较少,因此将视觉Transformer模型应用于道路病害检测领域,提出了强鲁棒性的道路病害检测网络Bi-Swin Transformer模型。

2 网络结构

2.1 主干特征提取网络

主干特征提取网络是指模型从输入图像中提取待检测目标特征的网络部分。采用Swin Transformer的主干特征提取网络^[7],如图1所示。该网络的输入为 $[H, W, 3]$,其中 H 和 W 分别为输入图像的高和宽。Swin Transformer结构首先将输入图像划分成若干个块,然后输入到4个阶段中分别提取不同分辨率的特征。在每一个阶段中,该网络采用滑动窗口的方式限制了自注意力机制的范围,这可以显著地减少参数量。Swin Transformer结构融合了卷积神经网络与Transformer结构各自的优点,既有广泛的上下文信息,又有类似于卷积神经网络的多层次特征提取方式。因此,该网络在视觉检测领域非常出色。采

用其作为主干特征提取网络。

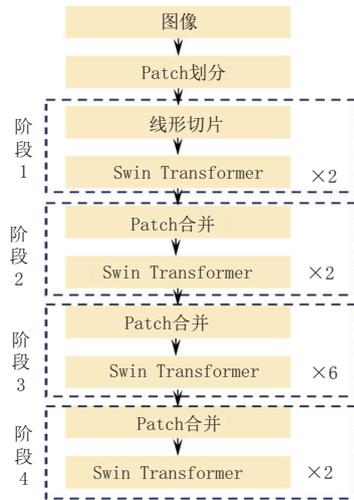


图1 主干特征提取网络

2.2 特征融合网络

常见的道路病害主要分为横向裂缝、纵向裂缝、网状裂缝以及坑槽。横向裂缝与纵向裂缝属于长条状的极其不规则的目标;坑槽一般目标尺寸较小,在目标检测领域中属于小目标,为道路病害检测带来了巨大的挑战。为更好地利用主干特征提取网络提取出的特征,采用一种新型的特征融合网络BiFPN (Bidirectional Feature Pyramid Network),其结构如图2所示。

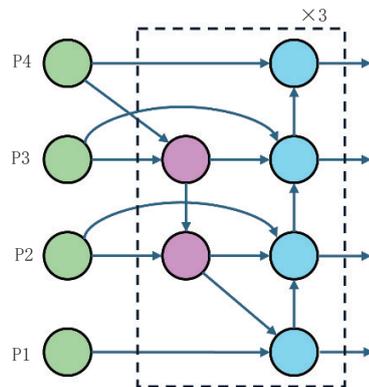


图2 BiFPN特征融合网络

2.3 损失函数

损失函数在训练过程中起着至关重要的作用,用于评估模型的预测结果与实际标签之间的差异。损失函数由交叉熵损失函数、边界框回归损失函数(Smooth L1 Loss)及DIoU损失函数组成。

3 数据集与实验设计

3.1 数据集

道路病害检测数据集采用了Arya等^[8]于2022年提出的RDD 2022数据集。RDD 2022数据集包括来

自中国、智利、印度、日本、挪威及美国收集的道路病害图像,其中各个路表病害类别的种类及数量见表1。各符号分别表示:D00(横向裂缝)、D10(纵向裂缝)、D20(网状裂缝)、D40(坑槽)。在RDD 2022数据集中,横向裂缝的数量最多,坑槽的数量最少,存在着数据分布不均衡的问题。因此在训练前,将采用数据增强的方法,减小数据分布不均衡这一问题带来的影响。

表1 RDD 2022数据集

类别	D00	D10	D20	D40
数量	23 504	10 708	9 532	5 837

3.2 实验设计

所有的实验均在CPU为i7-13700、GPU为GeForce RTX 4090、运行内存为128G的工作站中运行。在进行模型训练时,模型的输入被固定为640×640的大小,网络均被训练100个轮次,初始学习率为0.001,采用SGD随机梯度下降优化器更新模型的权重,学习率每轮次衰减为原来的0.937倍。

3.3 对比模型

本研究采用几个较为著名的卷积神经网络进行对比,包括Faster RCNN^[9]、YOLO F^[10]、YOLO X^[11]、CenterNet^[12]和Swin Transformer结构^[7]。Faster RCNN使用了区域建议网络(Region Proposal Network, RPN)来生成候选区域,然后再继续检测,是一种二阶段检测算法。这一算法可以大幅提高目标检测的精度。YOLO F模型则是依赖强大的主干特征提取网络来提高特征提取能力,最后利用简化的检测头来达到高效的检测性能。YOLO X是对YOLO系列算法的进一步改进。依据模型参数数量,将模型分为了 n 、 s 、 1 、 x 等,分别提供高精度或高速度的检测网络,可以适用于不同的应用场景。CenterNet是一种不使用Anchor的目标检测算法,通过预测物体中心点的位置,来确定目标检测框的位置。这一操作剔除了预先设定的Anchor的操作,极大地简化了检测网络的框架,因此具有较高的准确率。

4 结果与讨论

4.1 评价指标

为综合评价各算法的优劣,采用COCO数据集中的评价指标平均准确率(Average Precision, AP)。其中AP[0.5]表示IoU取0.5时的AP值;AP[0.75]表示IoU取0.75时的AP值;AP[0.5:0.95]表示IoU从0.5开始,步长为0.05值时的各AP平均值;AP[s]表示像素

面积小于32×32目标的AP值;AP[m]表示像素面积大于32×32但小于96×96目标的AP值;AP[l]表示像素面积大于96×96目标的AP值。

4.2 实验结果

实验结果见表2。

表2 检测结果

模型	AP[0.5]	AP[0.75]	AP[0.5:0.95]
Faster RCNN	0.430	0.100	0.171
YOLO F	0.319	0.069	0.124
YOLO X	0.455	0.162	0.208
CenterNet	0.324	0.073	0.123
Swin Transformer	0.525	0.183	0.238
Bi-Swin Transformer	0.539	0.194	0.244

从表2中可以看出,Faster RCNN和YOLO X在AP[0.5]上表现较好,表明它们在较宽松的IoU阈值下能够产生较高质量的预测框。Swin Transformer和Bi-Swin Transformer的AP[0.5]更高。这是因为Transformer模型能够利用全局信息更好地理解图像内容,从而产生更准确的预测边界框。Bi-Swin Transformer相比于Swin Transformer表现更优,是因为采用的BiFPN模块提供了更有效的特征融合能力,加强了多尺度特征的整合,这对于检测不同大小的物体尤其重要。

在较严格的IoU阈值(0.75)下,所有模型的表现都有所下降。这是因为要求预测框与真实框的重叠度更高、难度增大。YOLO X、Swin Transformer及Bi-Swin Transformer在AP[0.75]上相比其他模型表现要好,尤其是Bi-Swin Transformer。这表明它在精确检测上具有更好的性能,再次强调了BiFPN在精确定位方面的贡献。

AP[0.5:0.95]是一个更全面的指标,因为它考虑了从宽松到严格的一系列IoU阈值。Bi-Swin Transformer在此指标上得分最高,表明它在不同的IoU阈值下都保持了稳定的性能。Bi-Swin Transformer在各个IoU阈值下都展现出了更优越的检测性能,这在很大程度上得益于其强大的特征融合能力和全局信息提取能力。

为分析模型对不同大小目标的检测能力,将道路病害划分为小目标($S < 32 \times 32 \text{ pixel}^2$)、中目标($96 \times 96 \text{ pixel}^2 < S < 32 \times 32 \text{ pixel}^2$)和大目标($S < 96 \times 96 \text{ pixel}^2$),检测结果见表3。在小目标检测(AP[s])中,Bi-Swin Transformer模型表现最为出色,其次是Swin Transformer,其中Bi-Swin Transformer达到了0.129的AP值。这是因为Transformer结构的自注意

力机制可以捕获全局依赖关系,使得模型能够整合来自整个图像的信息来增强目标特征。这对于小目标特别有利,因为它们的局部特征可能不足以支撑准确检测。BiFPN则可以通过多尺度特征融合增强模型捕捉不同尺度信息的能力。对于小目标而言,这意味着即便在高层次的特征图中目标尺寸较小,依然可以通过与低层次的详细特征图进行融合而增强其表示。BiFPN提供了针对尺度变化的鲁棒性,而Transformer提供了长距离依赖的建模能力。这种互补关系意味着模型不仅能够识别细粒度的局部特征,还能够理解它们在全局上下文中的含义。相对地,YOLO F在小目标上的表现较差,这可能是由于其特征提取能力在捕捉小物体的细节方面有所不足。此外,Faster RCNN和YOLO X之间的差距进一步表明了在小目标检测任务上,模型架构和特征融合机制对AP值的显著影响。

综上所述,Bi-Swin Transformer的综合性能表明,其通过BiFPN和Transformer结构的结合,为各种尺寸的目标检测提供了强大的特征表示能力,在小目标和中目标上的表现尤为突出。

表3 不同尺寸目标检测结果

模型	AP[s]	AP[m]	AP[l]
Faster RCNN	0.062	0.128	0.191
YOLO F	0.034	0.089	0.134
YOLO X	0.098	0.167	0.219
CenterNet	0.048	0.109	0.124
Swin Transformer	0.121	0.188	0.270
Bi-Swin Transformer	0.129	0.195	0.279

5 结 语

为解决目前道路病害算法精度低、鲁棒性差的问题,提出了一种基于Transformer结构的Bi-Swin Transformer道路病害检测模型。该模型通过结合

Swin Transformer Block的四层特征提取能力与BiFPN的特征融合策略,优化了模型对全局依赖关系的捕获及不同尺度特征间的双向交流。在分析了人工检测和现有深度学习技术的局限性后,本研究通过利用视觉Transformer结构,并结合数据增强技术在国际多个数据库上进行训练,证实了Bi-Swin Transformer在道路病害检测中具有更优的性能,尤其是在小目标检测方面。实验结果显示平均准确率达到0.539,表明该模型在处理复杂环境下的道路病害检测问题时,比现有模型更加有效。这一进步有望推动道路病害智能检测领域的发展,为后续相关技术的研究和应用奠定基础。

参考文献:

- [1] 周秋红. 基于改进SSD模型的道路病害检测研究[J]. 黑龙江交通科技, 2023, 46(4): 30-32.
- [2] 田杨, 邱玲. 一种基于Fast-RCNN的道路病害检测算法[J]. 市政技术, 2020, 38(增刊1): 123-127.
- [3] 喻露, 戴甜杰, 余丽华. 基于改进YOLOv5的道路病害智能检测[J]. 福建工程学院学报, 2023, 21(4): 332-337.
- [4] 张营超. 基于深度学习的路面状况指数检测系统[D]. 济南: 山东大学, 2023.
- [5] 张欢欢. 基于YOLOv5的道路病害检测模型改进方法的研究与应用[D]. 上海: 华东师范大学, 2024.
- [6] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale[J]. arXiv, 2021.
- [7] Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows[J]. arXiv, 2021.
- [8] Arya D, Maeda H, Ghosh S K, et al. RDD2022: A multi-national image dataset for automatic Road Damage Detection[J]. arXiv, 2022.
- [9] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. arXiv, 2015.
- [10] Chen Q, Wang Y, Yang T, et al. You Only Look One-level Feature[J]. arXiv, 2021.
- [11] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding YOLO Series in 2021 [J]. arXiv, 2021.
- [12] Duan K, Bai S, Xie L, et al. CenterNet: Keypoint Triplets for Object Detection[J]. arXiv, 2019.

《城市道桥与防洪》杂志

是您合作的伙伴,为您提供平台,携手共同发展!

欢迎新老读者订阅期刊 欢迎新老客户刊登广告

官方网址: <http://www.csdqyfh.com> 电话: 021-55008850 联系邮箱: roadfloodbridge@163.com

