

# 基于改进 YOLOv5 的道路缺陷检测与分类研究

蒋大伟<sup>1</sup> 吴正平<sup>1</sup> 景思伟<sup>1</sup>

JIANG Dawei WU Zhengping JING Siwei

## 摘要

针对道路缺陷识别中识别准确度低, 容易出现漏检情况等问题, 提出了一种改进 YOLOv5 网络的图像识别算法用于道路缺陷识别。首先, 在 YOLOv5 网络的颈部网络中引入集中注意力机制 (CBAM) 与原有的 CSP 结构结合, 提高对道路缺陷特征的关注能力。然后, 在特征融合层引入自适应特征融合机制 (ASFF), 改善主干网络的特征提取能力, 提升算法检测精度。最后, 道路缺陷识别消融实验和对比实验表明, 使用改进后的 YOLOv5s 算法的平均准确率达到 77.2%, 相较于原 YOLOv5s 算法提高 2.3%, 召回率达到 74.9%, 提高 2.5%, 其他各项指标均有提升, 验证了改进方法的有效性。

## 关键词

道路缺陷检测; YOLO 网络; 自适应特征融合机制; 集中注意力机制

doi: 10.3969/j.issn.1672-9528.2024.02.007

## 0 引言

近年来, 随着我国的交通网络建设日益完善, 公路建设蓬勃发展。交通运输部 2022 年交通行业发展统计报告显示, 截止到 2022 年年末, 我国公路总里程已达 535.48 万公里。同时, 我国对公路养护的需求也随之增大, 这就要求对道路缺陷进行及时发现与修复。如何快速准确地在地面信息中得出检测结果成为发现道路缺陷的关键所在。

随着机器视觉的快速发展, 将机器视觉检测应用于道路缺陷检测成为可能。近些年来, 基于深度学习方法的图像检测技术<sup>[1]</sup>迅速发展, 逐步应用于各个领域。

基于深度学习的图像检测算法在提取图像特征方面具有显著优势, 其鲁棒性与泛化能力更好, 使用场景更广。基于深度学习的图像检测方法在近两年来涌现出了一批优秀的深度学习检测算法, 可以分为二阶段和一阶段两类。二阶段是指先使用网络提取目标区域, 然后再对区域采用卷积神经网络 (CNN) 进行识别, 如 R-CNN<sup>[2]</sup>、Fast R-CNN<sup>[3]</sup> 均为二阶段网络。一阶段目标检测方法能够做到一次提取特征就可以实现目标检测, 代表性算法有 YOLO<sup>[4]</sup>、SSD<sup>[5]</sup> 等。赵珊珊等人基于 CNN 网络提出了一种裂缝检测方法<sup>[6]</sup>, 但由于道路背景干扰, 检测精度不高; 晏班夫等人基于 Faster R-CNN 对道路裂缝进行检测, 能够提取裂缝形态及大小, 但误检率偏高, 难以实际应用<sup>[7]</sup>。

一阶段网络只需一次提取特征即可实现目标检测, 在精度略微下降的同时, 大幅提高了训练速度, 使实时检测成为可能。王博等人提出了一种基于 SSD 的轻量级的道路裂缝检

测算法, 在提高了检测速度的同时, 略微提高了检测精度<sup>[8]</sup>, 但对于部分种类的道路裂缝检测准确率有所降低。石丽等人基于 RetinaNet 模型的基础上进行改进, 引入 SE Net, 增强特征图的特征表达能力, 完成对路面露骨病害图像的检测<sup>[9]</sup>, 但检测速度有所降低。何铁军在 YOLOv5 中引入 Space-to-depth 应对图像分辨率低的问题, 同时添加 SPPFCSPC 实现多尺度特征的提取, 有更强的特征提取能力, 准确率更高<sup>[10]</sup>, 但因为大幅增加了模型的参数导致模型的检测速度降低。

以上在道路缺陷检测中所使用的方法都在实际应用中取得了较好的效果, 但仍然存在检测难度大、泛化能力弱等缺陷, 在检测精度上仍然有提升空间。YOLOv5 算法在 YOLOv4 的基础上增加了 Focus 结构并将 CSP 应用到了 neck 层中, 同时增加了自适应锚定框, 对比 YOLOv4 算法, YOLOv5 网络更小, 检测速度更快, 灵活性更高。

通过对现有方法进行深入分析后, 本文提出了一种基于 YOLOv5 的道路缺陷检测改进方案。在 YOLOv5s 的颈部 (neck) 网络中引入集中注意力机制 (convolutional block attention module, CBAM), 同时从通道与空间维度提升网络对道路裂缝特征信息的关注度, 提升模型的特征提取能力; 在特征融合处引入自适应特征融合机制 (ASFF) 进行优化, 提升网络的特征融合效果, 进而提升识别精度。通过实验验证, 改进后的算法能够显著提升对道路裂缝的检测精度, 降低漏检率, 验证了改进方法的有效性。

## 1 YOLOv5 网络概述

YOLOv5 是一种单阶段目标检测算法, 是目前比较成熟的深度学习目标检测模型。根据其在宽度 (width) 和深度 (depth) 方面的不同配置, 分为四种版本: YOLOv5s、

1. 三峡大学电气与新能源学院 湖北宜昌 443002

YOLOv5m、YOLOv5l、YOLOv5n。四个模型中，YOLOv5s 模型最小，同时识别速度最快，但其检测精度还有提升空间。因此本文选择 YOLOv5s 为基础模型并对其进行改进。

YOLOv5 的网络结构可以分为四个部分，分别是输入端、backbone 结构、颈部 neck 网络以及 head 输出端。

输入端在目标检测中负责对输入的图片进行多项数据增强和预处理工作，以准备图像用于后续的目标检测算法，提升模型训练速度和精度。backbone 部分是 YOLOv5 网络模型的主干部分，主要结构有 Conv 模块、CSP 模块和 SPPF 模块，选用相对轻量化的 CSPDarknet53 作为骨干网络，在保证较高精度的同时，能够尽可能地减少计算量。其中 Conv 模块主要通过卷积、归一化最终通过激活函数实现对输入特征的转换和提取，用于后续的检测与分类。CSP 模块可以通过增加网络的深度和感受野提高网络的特征提取能力。SPPF 模块通过池化、连接、全连接三个步骤将不同大小的感受野应用于同一张图像，实现对不同尺度特征信息的捕捉。neck 网络中采取特征金字塔结构 (FPN) 将不同层次特征图融合在一起，进一步提高特征的多样性和鲁棒性。Head 输出端主要负责进行多尺度目标检测，得出检测结果。

## 2 改进的 YOLOv5 算法

由于正常环境下道路背景复杂，目标有所差异，难以有效地对道路等缺陷进行检测，因此需要对模型进行改进和提升。接下来探究 YOLOv5 加入集中注意力机制，自适应空间特征融合的方法。

### 2.1 集中注意力机制

CBAM 是一种新提出的集中注意力机制，其中包含通道注意力模块 (channel attention module) 和空间注意力模块 (spatial attention module)，这两个模块分别进行通道与空间维度上的注意力特征融合<sup>[11]</sup>。通道注意力机制计算跨空间维度的特征图的加权平均值，而空间注意力机制计算跨通道维度的特征图的加权和。将两种注意力机制结合后得到了 CBAM 模块。CBAM 模块结构图如图 1 所示。

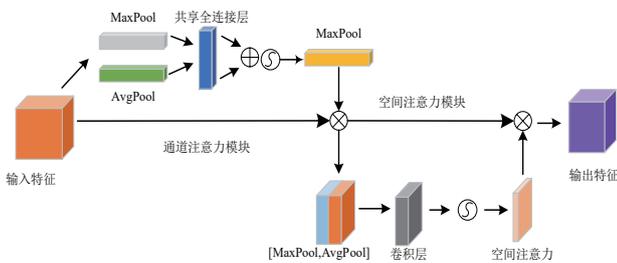


图 1 CBAM 注意力机制

具体来说，通道注意力模块使用全局最大池化和全局平均池化层对输入进行降维，得到两个二维特征图，然后通过两个全连接层得到通道注意力系数，将其与输入进行乘积操

作，得到加权的输出。空间注意力模块则进行基于特征图的全局最大池化和全局平均池化层处理，得到两个二维特征图，将特征图拼接后通过卷积层后得到空间注意力特征图。

由于道路缺陷的背景复杂，缺陷目标在图像中一般所占面积小。将 CBAM 模块引入到 YOLOv5s 的结构中，网络可以学习关注相关特征并抑制不相关特征，从而提高目标检测任务的准确性。

通过多次训练，发现将 CBAM 模块与 YOLOv5s 网络 neck 中的 CSP 结构结合之后，效果最好。

### 2.2 自适应空间融合

ASFF (adaptively spatial feature fusion) 是一种新的特征融合方式，与 YOLOv5 中原本的 FPN 特征融合方式相比，它能够更好更充分地利用不同尺度的特征<sup>[12]</sup>。在路面缺陷特征图像中，由于道路缺陷目标图像占比较少且背景复杂，因此目标特征信息往往处于网络底中层卷积后的特征层中。YOLOv5 中为了能够充分利用高层语义信息和底层细粒度融合，采用了 FPN 架构，因此无法充分利用不同尺度的特征信息。为充分利用不同尺度的特征，在网络中引入自适应空间特征融合方式 (ASFF)，提高网络的检测精度。ASFF 结构如图 2 所示。

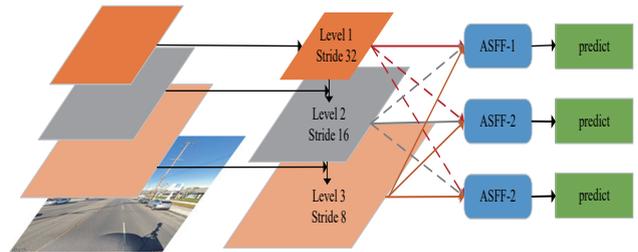


图 2 ASFF 模块结构

该结构中 Level1、Level2、Level3 会将各自的特征发送到每个 ASFF 层中，将各个特征与相应的权重参数  $\alpha^l$ 、 $\beta^l$ 、 $\gamma^l$  相乘后相加，便能够得到新的融合特征。新的特征公式为：

$$y'_{ij} = \alpha^l_{ij} \cdot x^{1 \rightarrow l}_{ij} + \beta^l_{ij} \cdot x^{2 \rightarrow l}_{ij} + \gamma^l_{ij} \cdot x^{3 \rightarrow l}_{ij} \quad (1)$$

式中： $y'_{ij}$  是指输出特征的  $y^l$  的第  $(i, j)$  个向量， $\alpha^l_{ij}$ 、 $\beta^l_{ij}$ 、 $\gamma^l_{ij}$  为来自不同特征层的权重参数， $x^{1 \rightarrow l}_{ij}$  为来自 Level1 ~ 3 相应层中的特征。

由于计算融合特征时采取相加的方式，因此不同 ASFF 层计算时需要调整 Level1 ~ 3 输出特征大小，保持相应计算时特征大小相同。同时，权重参数是调整尺寸后通过 Level1 ~ 3 的特征图经过  $1 \times 1$  卷积后得到权重参数  $\alpha$ 、 $\beta$ 、 $\gamma$ ，之后通过 concat 拼接后再通过 softmax 使其范围在  $[0,1]$  内并且和为 1。其中  $\alpha^l_{ij}$  计算公式为：

$$\alpha^l_{ij} = \frac{e^{\lambda^l_{\alpha_{i,j}}}}{e^{\lambda^l_{\alpha_{i,j}}} + e^{\lambda^l_{\beta_{i,j}}} + e^{\lambda^l_{\gamma_{i,j}}}} \quad (2)$$

改进后的网络结构如图 3 所示。

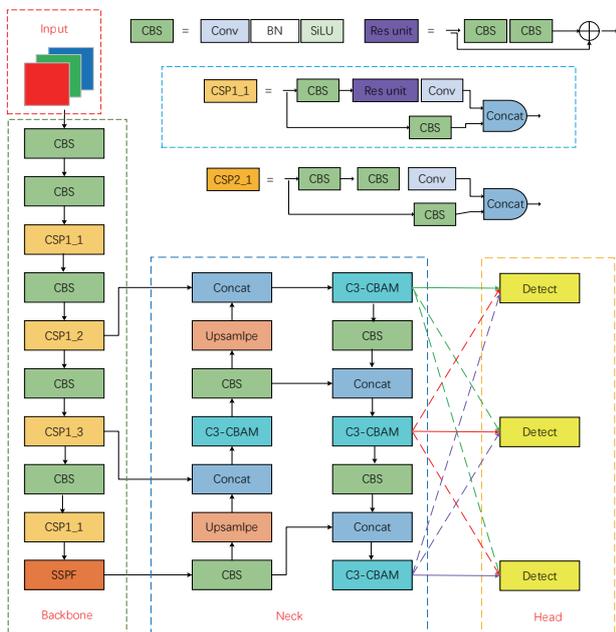


图 3 改进后的 YOLOv5 模型

### 3 实验与分析

#### 3.1 数据准备

本文所使用的数据集来自 RDD2022 开源数据集。RDD2022 数据集是一个用于道路缺陷检测的公开数据集，由中国科学院自动化研究所发布。该数据集收集了大量的道路图像和视频，并提供了多种道路缺陷的标注，包括裂缝、坑洼、路面变形等。本文选取该数据集中的部分图片，共 6782 张图片，包含路面上常见的 4 种缺陷以及缺陷修补后的图像，分别是 D00（纵向裂缝）、D10（横向裂缝）、D20（龟裂）、D40（坑洞）、repair（已修复），使用随机划分方法将其划分为训练集和验证集，比例为 8:2，最终训练集中有 5426 张图片，验证集中有 1356 张图片。

#### 3.2 实验环境

本实验环境使用的计算机硬件配置为 Intel Core i9-12900H CPU，NVIDIA GeForce RTX3060 Laptop GPU，6 GB 显存，使用的操作系统为 Window11，16 GB 内存，采用 Python3.7，PyTorch1.12.0 深度学习框架，CUDA11.6GPU 加速库。

采用 YOLOv5 系列中模型最小的 YOLOv5 为初始模型，整个训练过程使用随机梯度法进行训练，共训练 150 个 epoch，batch size 设置为 4，初始学习率为 0.01，权重衰减为 0.0005，图片输入尺寸设置为 640×640。

#### 3.3 评估指标

本实验选择召回率（recall,  $R$ ）、精确率（precision,  $P$ ）、各类别准确率均值（mAP@0.5）以及平衡分数 F1 作为模型的评价指标。其中召回率是指被正确检测到的缺陷数

量（ $T_p$ ）占数据集中总缺陷数量（ $T_p+F_p$ ）的比例。精确率是指被正确检测到的缺陷数量（ $T_p$ ）占所有预测为缺陷目标的比例（ $T_p+F_N$ ）。mAP@0.5 是指将 IOU 设置为 0.5 时，计算出所有类别的准确率，然后所有类别求平均，即可得到 mAP@0.5。 $F_1$  为召回率与精确率的调和平均。各指标的计算方法如公式（3）~（5）所示。

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (3)$$

$$R = \frac{T_p}{T_p + F_N} \times 100\% \quad (4)$$

$$P_{mAP} = \frac{\sum_{i=1}^n P_{APi}}{n} \times 100\% \quad (5)$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \times 100\% \quad (6)$$

式中： $T_p$  为能够被正确检测到的缺陷， $F_p$  为被错误检测为缺陷的背景数量， $F_N$  为数据集中没有被检测出来的缺陷样本数量， $n$  为数据集中标注出来的样本种类。

#### 3.4 实验结果分析

训练模型过程中所得到的 mAP 变化曲线图如图 4 所示。从图 5 所示的结果中可以看出改进后的模型拥有更高的 mAP。改进后的模型在迭代 130 次左右时稳定在 77.3% 左右，同时改进前的模型也稳定在了 74.9% 左右。由此可见，相比于 YOLOv5 算法，本文在平均精度均值 mAP 方面有了一定的提升。

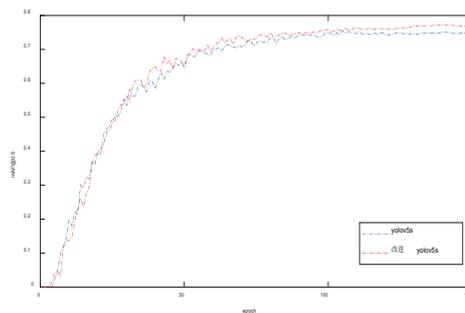
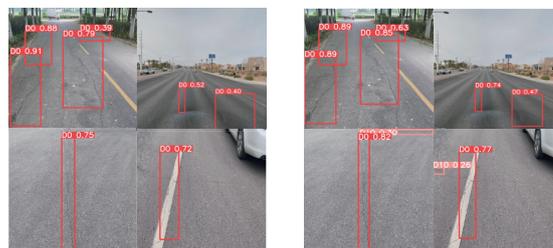


图 4 改进前后 mAP@0.5 对比曲线



(a) 改进前

(b) 改进后

图 5 YOLOv5 和改进 YOLOv5 检测结果对比

为了进一步验证本文改进方法的有效性，分别对添加了自适应空间特征融合（ASFF）、集中注意力机制的 YOLOv5

模型进行实验验证,即进行消融实验。实验结果如表 1 所示。

表 1 不同模型在测试集上的效能评估

模型	mAP/%	R/%	P/%	F <sub>1</sub> /%
YOLOv5	74.9	72.4	76.8	74.5
YOLOv5+CBAM	75.4	73.7	76.5	75.1
YOLOv5+ASFF	76.6	72.8	76.7	74.6
YOLOv5+CBAM+ASFF	77.2	74.9	78.5	76.7

由表 1 可知,改进后的 YOLOv5 模型算法对于道路缺陷的识别准确率均值 mAP@50 达到了 77.2%,提高了 2.3%;召回率达到了 74.9%,提升了 2.5%;精确率达到了 78.5%,提升了 1.7%;平衡分数 F<sub>1</sub> 达到 76.7%,提升了 3.2%,说明本文的改进方法是有效的,改进后的 YOLOv5 模型能够更准确地识别道路上的缺陷。

为了验证改进后算法的优势,选取了部分缺陷路面分别进行改进前后模型验证。图 5 所示为算法改进前后对相同路面缺陷的检测结果。从图 5 中第一组检测结果可以看出,改进后的算法对路面缺陷检测精度更高;在第二组实验中可以看出,由于路面缺陷不明显,导致原 YOLOv5 算法出现了漏检现象,而改进后的算法检测出来原模型未检测出的目标,能够有效降低漏检率。因此,改进后的算法能够更加有效地对道路缺陷目标进行检测。

为了更好地显示改进 YOLOv5 的优势,在相同的硬件条件下,将改进后的 YOLOv5 模型与其他深度学习模型进行对比,对比结果如表 2 所示。可以看出本文算法各方面性能均表现更好,验证了本文改进模型的有效性。

表 2 各种模型性能指标

方法	mAP/%	R/%	P/%	F <sub>1</sub> /%
YOLOv3	65.1	68.1	61.2	64.5
YOLOv4	42.1	36.0	32.0	34.0
YOLOv5s	74.9	72.4	76.8	74.5
文中方法	77.2	74.9	78.5	76.7

#### 4 结语

为了解决目前在复杂道路环境下难以准确检测路面缺陷的问题,本文提出了一种基于 YOLOv5s 的改进算法。针对复杂道路环境下缺陷的多样性和复杂性,对 YOLOv5s 的颈部(neck)网络进行了优化,引入了集中注意力机制(CBAM)来增强网络对重要特征信息的关注度,并通过自适应特征融合机制(ASFF)进行优化,提高了特征融合效果,从而提高网络对道路缺陷的检测能力。在本文使用的道路缺陷数据集上进行实验,改进后的算法相比于原始的 YOLOv5 模型取得了显著的性能提升,改进后的网络模型在召回率以及 mAP 上均有较大提升,分别提升了 2.5% 和 2.3%,同时精确率也提升了 1.7%,优于 YOLOv5 算法,但对横向裂缝的检测精度

仍然偏低,需要进一步提高。后续将进一步研究如何提高对裂缝的检测精度,同时尝试采用更多的路面图片数据训练模型,提高算法的实用性。

#### 参考文献:

- [1] 马荣贵,徐琨,刘飞飞.高速公路路面裂缝识别算法研究[J].交通信息与安全,2014,32(2):90-94.
- [2] GIRSHICK R B,DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[EB/OL].(2013-11-11)[2023-08-25].https://arxiv.org/abs/1311.2524.
- [3] REN S, HE K, GIRSHICK R,et al. Faster R-CNN: towards real-time object detection with region proposal networks[EB/OL].(2020-04-23)[2023-08-05].IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6):1137-1149.
- [4] BOCHKOVSKIY A, WANG C, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL].(2020-04-23)[2023-08-17].https://arxiv.org/abs/2004.10934.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[EB/OL].(2016-11-29)[2023-09-01].https://arxiv.org/abs/1512.02325.
- [6] 赵珊珊,何宁.基于卷积神经网络的路面裂缝检测[J].传感器与微系统,2017, 36(11): 135-138.
- [7] 晏班夫,徐观亚,栾健,等.基于 Faster R-CNN 与形态法的路面病害识别[J].中国公路学报,2021,34(9):181-193.
- [8] 王博,李齐,刘皎.一种轻量级的 SSD 道路裂缝检测算法[J].商洛学院学报,2022, 36(4): 83-90.
- [9] 石丽,裴莉莉,陈昊,等.改进 RetinaNet 的水泥路面露骨病害检测[J].计算机系统应用,2022, 31(4): 352-359.
- [10] 何铁军,李华思.基于改进 YOLOv5 的路面病害检测模型[J/OL].土木工程学报,1-12[2024-02-04].https://doi.org/10.15951/j.tmgxcb.22101073.
- [11] 刘凯,罗素云.基于改进 YOLOX-S 的交通标志识别[J].电子测量技术,2023, 46(1): 112-119.
- [12] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection[EB/OL].(2019-11-21)[2023-08-22]. https://arxiv.org/abs/1911.09516.
- [13] 李鑫,李香蓉,汪诚,等.基于改进 YOLOv5 的航空发动机表面缺陷检测模型[J].激光与光电子学进展,2023, 60(16): 304-313.

#### 【作者简介】

蒋大伟(1997—),男,安徽阜阳人,硕士研究生,研究方向:图像采集与识别。

吴正平(1978—),男,湖南宁乡人,博士,教授,硕士生导师,研究方向:计算机视觉、图像处理、模式识别。

(收稿日期:2023-11-02)