

# 基于多尺度特征融合的轻量化道路损伤检测算法<sup>\*</sup>

武 兵 田 莹

(辽宁科技大学计算机与软件工程学院 鞍山 114000)

**摘 要:**为了提高现阶段道路损伤检测方法在复杂环境下检测困难、细节纹理丢失严重、效率低等问题,提出了多尺度特征融合的轻量化 YOLO 算法(MSL-YOLO)。首先,在 YOLO11n 的基础上进行改进,针对损伤目标特征表达能力弱,设计特征融合通道注意力(feature fusion channel attention, FFCA)模块提高损伤信息的权重,加强特征信息的提取,减少冗余信息;为了在复杂环境下更好地捕捉不同尺寸的损伤目标,设计了一种多尺度特征增强(multi-scale feature enhancement, MSFE)模块提升模型的多尺度特征融合能力,进一步提高检测性能;为实现模型轻量化和检测实时化,在 Neck 部分引入了轻量级网络(lightweight network, LNet)来减轻模型的计算复杂度,方便模型的部署和应用。实验结果表明,在 RDD2022 道路裂缝数据集上,所提方法检测平均精度为 52.5%,模型参数数量为  $2.3 \times 10^6$ ,相较于 YOLO11n 算法平均精度提升了 1.8%,参数量下降了 11.5%。不仅能满足对道路损伤检测的高精度、高速度、轻量化的要求,且具有较强的鲁棒性和实时性。

**关键词:**道路损伤检测;YOLO11;多尺度特征增强;通道注意力;轻量化

**中图分类号:** TP391;TN911.73

**文献标识码:** A

**国家标准学科分类代码:** 520.2060

## Lightweight road damage detection algorithm based on multi-scale feature fusion

Wu Bing Tian Ying

(School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan 114000, China)

**Abstract:** In order to improve the current road damage detection methods in complex environment detection difficulties, serious detail texture loss, low efficiency, a multi-scale feature fusion lightweight YOLO (MSL-YOLO) method is proposed. Firstly, based on the improvement of YOLO11n, the Feature fusion channel attention (FFCA) module is designed to improve the weight of damage information, strengthen the extraction of feature information, and reduce redundant information. In order to better capture damage targets of different sizes in complex environments, a multi-scale feature enhancement (MSFE) module is designed to enhance the multi-scale feature fusion capability of the model and further improve the detection performance. In order to realize the Lightweight model and real-time detection, lightweight network (LNet) is introduced in Neck to reduce the computational complexity of the model and facilitate the deployment and application of the model. The experimental results show that on the RDD2022 road crack dataset, the proposed method has an average detection accuracy of 52.5%, and the number of model parameters is  $2.3 \times 10^6$ , which is 1.8% higher than that of YOLO11n algorithm, and the number of parameters is 11.5% lower. It can not only meet the requirements of high precision, high speed and lightweight for road damage detection, but also has strong robustness and real-time.

**Keywords:** road damage detection; YOLO11; multi-scale feature enhancement; channel attention; light weight

## 0 引 言

道路是交通基础设施的重要组成部分,承载着社会

进步与经济建设的重要使命。随着车流量和使用年限的增加,以及受到雨雪天气和昼夜温差变化等多种因素的影响,许多道路面临着不同程度的损伤问题,如变形、裂缝、沉陷等。这些损伤不仅增加了交通事故的风险,还对

交通效率 and 经济发展造成了不利影响。因此,及时且有效的道路损伤检测与维护已成为交通管理部门面临的重要挑战。

传统的道路损伤检测长期以人工巡检为主,通过人工对路面进行视觉检查和记录道路损伤。光学成像设备的优化后,图像采集加人工标注成为主流,通过车载摄像头连续拍摄高清路面图像,后期再由人工在图像中标记损坏位置和类型。以上方式耗时耗力且检测结果往往受主观因素的影响,缺乏一致性,已无法满足当下日益增长的检测需求。近年来,计算机视觉快速发展,基于深度学习的自动化检测技术使损伤目标识别的准确性和效率得到了显著提升。

基于深度学习的目标检测技术<sup>[1]</sup>分为 two-stage 检测算法和 one-stage 检测算法。two-stage 检测算法是基于候选区域进行目标的分类与回归,代表算法有 R-CNN<sup>[2]</sup>、Fast R-CNN<sup>[3]</sup>、Faster R-CNN<sup>[4]</sup>、U-net<sup>[5]</sup>等。于海洋等<sup>[6]</sup>提出了一种基于卷积块注意力模块(CBAM)机制和残差网络的 U-net 改进模型,高效提取了更多道路裂缝的细节信息,提高了检测能力。Sekar 等<sup>[7]</sup>在 Faster R-CNN 网络上面结合了全局平均池化和感兴趣区域对齐的方式,减少了关键裂缝信息丢失的问题。然而这类检测方法过于依赖候选区域的生成,如果生成的候选区域不准确或不全面,后续的分类和回归结果会受到影响。在复杂场景中,背景噪声、遮挡、部分目标的缺失等因素会降低候选区域质量导致检测性能下降。one-stage 检测算法是直接图像中同时进行目标检测和分类,代表算法有 SSD<sup>[8]</sup> (single shot multibox detector) 算法和 YOLO<sup>[9-12]</sup> (you only look once) 算法。在目标检测任务中,YOLO 系列模型具有较快的检测速度,较高检测精度和广泛的应用场景,能够快速且精准的识别图像中的物体。任安虎等<sup>[13]</sup>提出了改进 YOLOv5s 的裂缝检测算法,将深度可分离卷积结合到全局注意力(global attention mechanism, GAM)后引入主干网络,在降低注意力复杂度的同时获得更丰富的特征信息,增强了模型对裂缝的识别能力,然而该模型参数量和计算量增加较多,推理速度下降,难于满足当下道路检测的实时性需求。李松等<sup>[14]</sup>在 YOLOv8 的基础上使用 Ghost 模块重构 C2f 模块,结合注意力机制和全局特征信息提取模块来适应裂缝目标的大跨度和细长特征,该算法具有较小的模型体积和较低的计算成本,但对细小曲折的裂缝检测精度较低。

与传统道路损伤检测方法相比,上述道路损伤检测算法在检测性能上有了较大的提升,但是上述算法模型架构相对复杂,较大的计算量和参数量增加了计算成本以及内存消耗,推理速度较慢,不利于模型的部署和应用。并且忽略了深层特征与浅层特征的内部联系。针对

道路损伤的形状、大小、方向和深度具有高度多样性,损伤裂缝与道路边界模糊以及光照条件和天气变化等复杂环境下对损伤目标产生的影响导致检测精度较低和计算资源消耗高的问题。以 YOLO11n 算法作为基础进行改进,提出了一种多尺度特征融合的轻量化 YOLO 算法(multi-scale lightweight you only look once, MSL-YOLO)。

针对损伤目标的在背景干扰下导致部分细节信息丢失和特征表达能力不足的问题,本文设计了一种特征融合通道注意力(feature fusion channel attention, FFCA)模块提取重要特征信息,提高检测精度。为能够更精确地在复杂环境下捕捉不同大小和形状的裂缝,设计了一种多尺度特征增强(multi-scale feature enhancement, MSFE)模块提供更为丰富和细致的多尺度特征信息。为实现模型轻量化和实时化,在保持检测精度的前提下 Neck 部分引入了轻量级网络(lightweight network, LNet)以减轻模型的复杂度。

在 YOLO11n 算法基础上改进的新模型,能显著的降低检测过程中的计算复杂度,实现更加轻量化的设计。同时,建立了各个尺度特征的内部联系,增强对多尺度目标特征的聚焦,提高检测性能和效率。

## 1 YOLO11 模型

YOLO11<sup>[15]</sup>是 Ultralytics 团队<sup>[16]</sup>在 2024 年提出的实时目标检测算法,相较于其他主流算法,YOLO11 算法检测精度更高,速度更快。该算法按规格大小分为 n、s、m、l、x 5 种模型,YOLO11 模型的网络结构主要分为 4 部分,Input、Backbone、Neck 和 Head 部分。其中 Input 为图像输入层,对图像进行数据增强等数据预处理操作;Backbone 为特征提取层,负责提取输入图像的特征。主要由卷积模块(Conv)、特征提取模块(C3k2)、空间金字塔池化模块(spatial pyramid pooling-faster, SPPF)和具有并行空间注意的卷积块(C2PSA)组成;Neck 为特征融合层,通过特征金字塔或上采样操作对来自 Backbone 的特征进一步提炼与整合;Head 为输出层,由 3 个不同尺度的检测头组成,分别用于大、中、小目标的最终预测和分类标签。

## 2 MSL-YOLO 检测模型

MSL-YOLO 模型网络结构如图 1 所示,在 Neck 部分通过 FFCA 模块动态调整通道权重,丰富特征详细信息,增强对目标特征的聚焦;使用 MSFE 模块实现 P2、P3 和 P4 特征层的特征融合,将网络传输中的浅层细节特征和深层抽象特征相融合,增强模型的多尺度特征融合能力,

进一步提升模型检测精度;在 YOLO11n 的 Neck 基础上设计了 LNet,其中 GSconv<sup>[17]</sup>通过常规卷积和深度卷积组

合的方式,实现轻量化网络结构,提高网络检测速度。

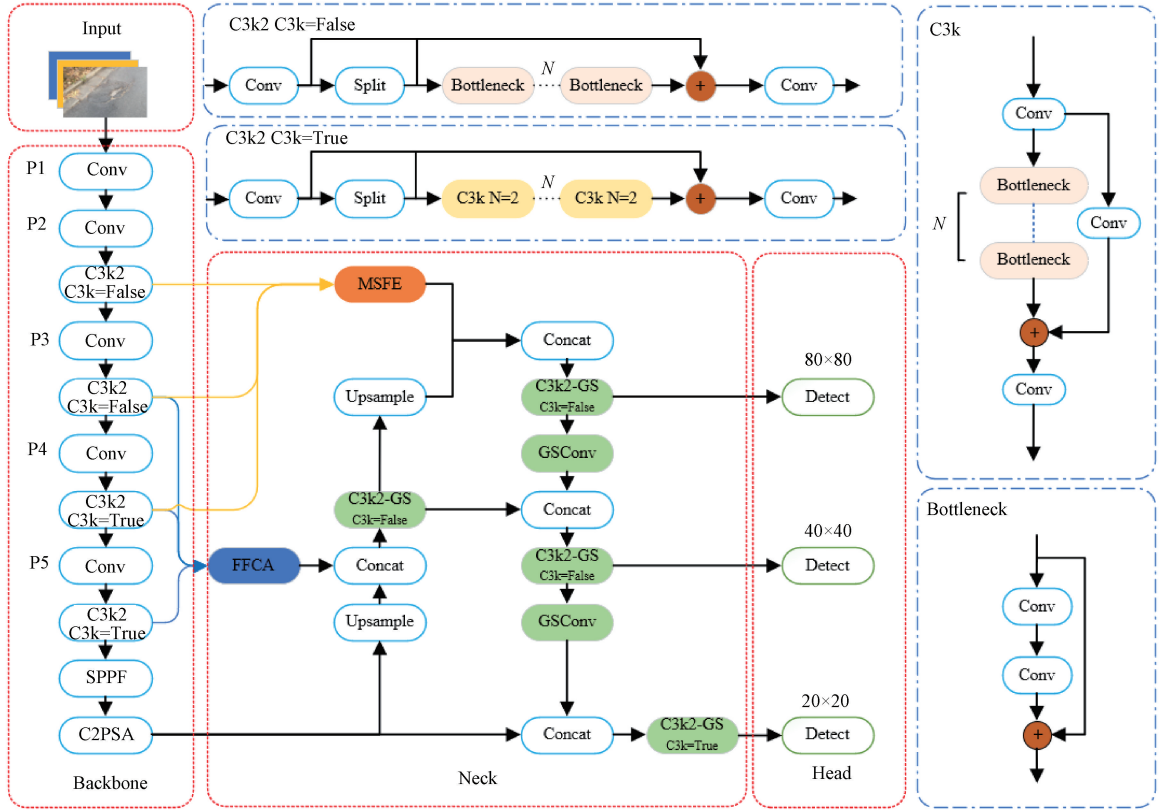


图 1 MSL-YOLO 模型结构

Fig. 1 MSL-YOLO model structure diagram

## 2.1 FFCA 模块

卷积神经网络在提取特征的过程中随着网络层次的增加常常伴随着目标的细节丢失,缺乏对细节的敏感性,常规的特征融合机制<sup>[18]</sup>仅通过小尺寸特征图线性上采样后拼接到上一层特征上进行特征融合,忽略了浅层大尺寸特征图具有的详细信息,设计了多层特征信息融合后在通道编码的 FFCA,将多层特征融合后的特征图通过增加通道注意力机制能够帮助模型更好地从数据中提取有价值的特征。FFCA 模块结构如图 2 所示。

为了帮助模型更好的理解上下文信息和细节,将多层特征进行融合,首先调整特征通道的大小,对 P3 层特征图  $F_3$  进行全局平均池化和最大池化操作相结合的方式下采样,调整特征图尺寸为 1S,可以提取全局上下文信息,以保留高分辨率特征的多样性;将 P5 层特征图  $F_5$  进行的最近临插值方法调整尺寸为 1S,保证特征图在上采样的过程中局部特征的丰富性;然后将 P3 和 P5 得到的特征图在通道上与 P4 层特征拼  $F_4$  接生成尺寸为 1S 通道为 3C 的新特征图  $F'$ 。然后对新特征图通过通道级全局平均池化操作减少空间维度,保留重要特征信

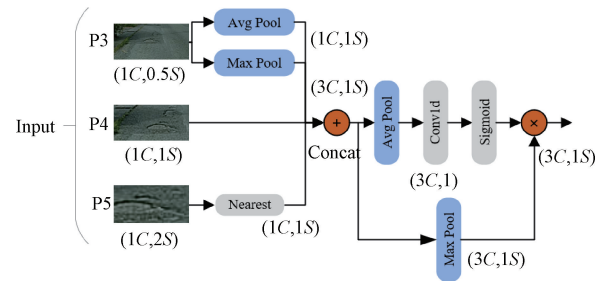


图 2 FFCA 模型结构

Fig. 2 FFCA model structure diagram

息,使用 1D 卷积可以在较大通道数的特征层中更好地捕获局部跨通道交互,有效捕捉输入序列中的局部特征。考虑到模型能够突出显著特征,避免重要信息的损失,最后将 sigmoid 函数生成的通道重要性权重应用于经过最大池化后的拼接特征图。FFCA 可以通过式 (1) ~ (4) 将损伤表面的细节信息进行有效的提取和整合。

$$F_3' = \text{MaxPool}(F_3) + \text{AvgPool}(F_3) \quad (1)$$

$$F_5' = \text{Nearest}(F_5) \quad (2)$$

$$\mathbf{F}' = \text{Concat}(\mathbf{F}_3', \mathbf{F}_4, \mathbf{F}_5') \quad (3)$$

$$\mathbf{F} = \text{Sigmoid}(\text{Conv1d}(\text{AvgPool}(\mathbf{F}')) \times \text{MaxPool}(\mathbf{F}')) \quad (4)$$

式中:  $\mathbf{F}_3$ 、 $\mathbf{F}_4$  和  $\mathbf{F}_5$  分别代表 P3、P4 和 P5 层的特征图;  $\text{Concat}()$  表示拼接;  $\mathbf{F}'$  由  $\mathbf{F}_3$ 、 $\mathbf{F}_4$  和  $\mathbf{F}_5$  拼接得到;  $\mathbf{F}'$  和  $\mathbf{F}_4$  具有相同的尺寸,但通道数是  $\mathbf{F}_4$  的 3 倍;  $\text{Sigmoid}()$  表示激活函数;  $\text{Conv1d}()$  表示一维卷积;  $\text{MaxPool}()$  和  $\text{AvgPool}()$  分别表示全局最大池化和全局平均池化;  $\text{Nearest}()$  表示最近邻插值;  $\mathbf{F}$  表示最终的输出特征。

## 2.2 MSFE 模块

特征提取过程中浅层网络更关注局部低级特征,比如损伤目标的边缘和纹理信息等,但在复杂环境下特征的表达能力较弱,在深层网络中更倾向于提取抽象的全局语义特征,无法捕捉到细致的浅层目标局部特征,在最

终的检测中容易导致细小裂缝的漏检或误检。受 3D 卷积操作在视频数据连续帧中可以通过更少的层数提取更复杂特征的启发,为了提高复杂环境下多尺度目标的检测精度,将不同尺度的特征图在 P3 层融合,通过 3D 卷积提取特征。设计的 MSFE 模块如图 3 所示。

首先使用轻量级的卷积模块 GSCnv 调整特征通道数,对浅层特征图 P2 进行池化组合调整尺寸为  $H \times W$ ,对深层特征图 P4 使用最近邻插值法调整尺寸为  $H \times W$ ,使用  $\text{unsqueeze}$  方法增加每个特征图的维度为  $1 \times H \times W$ ,然后在深度维度上将 P2、P3、P4 水平堆叠形成一个新的 3D 特征图,使用 3D 卷积,批量归一化和 ReLU 激活函数完成多尺度特征信息的融合和提取,最后通过最大池化操作进一步提炼这些特征,再使用  $\text{squeeze}$  方法降维后生成输出特征图 P3'。

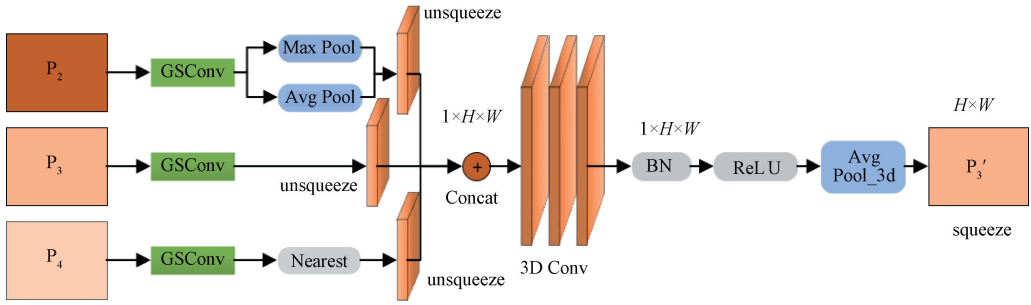


图 3 MSFE 模型结构

Fig. 3 MSFE model structure diagram

该模块通过独特的三维卷积策略,有效的将浅层目标的细致特征与深层目标的抽象特征相融合,增强对全局上下文的理解,捕捉更多的重要特征,提高了模型在处理多尺度问题时的性能,尤其是在复杂环境下的检测场景,能够更精确地捕捉不同大小和形状的裂缝损伤。

## 2.3 轻量级网络

### 1) GSCnv

尽管 YOLO11n 模型的计算量与参数量相较之前版本的模型已有了明显的减少,为实现更快的推理速度,达到检测实时化的目标,方便下一步在移动设备和嵌入式系统中运行,因此在 Neck 部分引入了轻量级卷积 GSCnv 来减轻模型复杂度并保持检测准确性。

GSCnv 结构如图 4 所示,首先对输入特征图使用常规卷积进行下采样,然后使用 DWConv<sup>[19]</sup> 深度卷积生成另一个特征图,最后将两个不同卷积生成的特征拼接后使用 shuffle 混洗这种均匀混合策略,将常规卷积生成的信息渗透到 DWConv 生成的所有信息中,生成最终特征图。GSCnv 卷积特征提取能力与传统卷积相仿,但是计算量仅为传统卷积的 1/2 左右,更符合轻量化设计理念<sup>[20]</sup>。

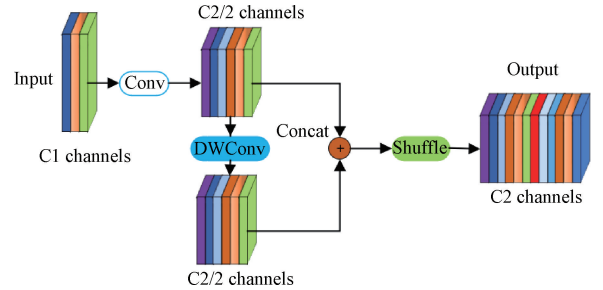


图 4 GSCnv 模型结构

Fig. 4 GSCnv model structure diagram

在常规卷积中,假设使用的卷积核为  $K \times K$  卷积生成的计算量为  $S_1$ ;常见的轻量化 DWConv 深度卷积针对每个输入通道使用单独的卷积核进行卷积操作,其计算量为  $S_2$ ;将常规卷积核与深度卷积结合的 GSCnv 的计算量为  $S_3$ ;GSCnv 产生的计算量远小于常规卷积产生的计算量,其计算公式如下:

$$S_1 = H \times W \times K \times K \times C_1 \times C_2 \quad (5)$$

$$S_2 = H \times W \times K \times K \times C_1 \quad (6)$$

$$S_3 = H \times W \times K \times K \times C_2 \times (C_1 + 1)/2 \quad (7)$$

式中:  $H$  和  $W$  为输入特征图的高度和宽度;  $C_1$  为输入特



征图的通道数;  $C_2$  为输出特征图的通道数。

## 2) C3k2-GS

特征提取模块 C3k2 在浅层网络中 C3k 参数为 False, 该模块将直接调用常规的瓶颈结构 Bottleneck 提取特征, 使用了适应性较强的轻量级卷积 GSConv 设计了 Bottleneck-GS 模块代替 C3k2 的 Bottleneck, Bottleneck-GS 模块由 2 层 GSConv 顺序连接, 如图 5 所示。

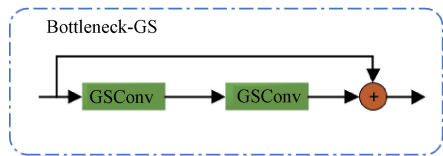


图 5 Bottleneck-GS 模型结构

Fig. 5 Bottleneck-GS model structure diagram

由卷积层 Conv、分离层 Split 和 Bottleneck-GS 组成的全新 C3k2-GS C3k=False 模块如图 6 所示。

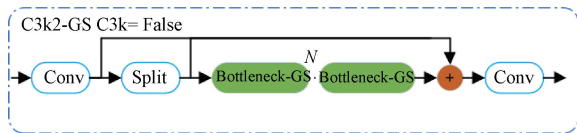


图 6 C3k2-GS C3k=False 模型结构

Fig. 6 C3k2-GS C3k=False model structure diagram

特征提取模块 C3k2 在深层网络中 C3k 参数为 True, 模块调用多尺度可调整大小的 C3k 模块, 其中 C3k 模块在使用 Bottleneck 的同时结合了更大的卷积核, 扩展了模型的感受野, 能够捕捉到更广泛的上下文信息, 尤其适合大目标检测和复杂多变的场景。使用 Bottleneck-GS 模块构建了 C3k-GS 模块以及 C3k2-GS C3k=True 模块如图 7 所示。

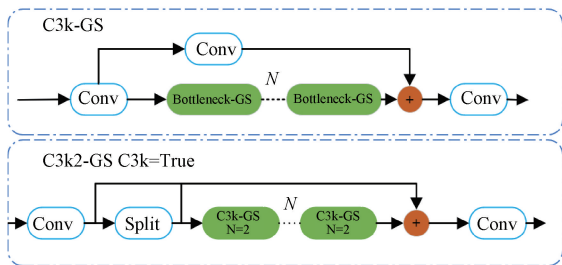


图 7 C3k2-GS C3k=True 模型结构

Fig. 7 C3k2-GS C3k=True model structure diagram

模型在 Neck 部分使用 GSConv 和 C3k2-GS 组合而成的轻量化网络 LNet 能够在不同尺度上对输入特征图进行高效率采样, 既能实现了更加轻量化的设计, 又能有效的提取特征信息。

## 3 实验结果与分析

### 3.1 实验环境及参数

实验使用 Windows 10 版本的服务器, 搭载着 NVIDIA GeForce RTX 4070ti 显存大小为 32 GB 的显卡。编程语言为 Python3. 9. 18, Pytorch 版本为 1. 13, 采用 CUDA 11. 7 对训练进行加速, 输入图像尺寸设置为 640×640, 训练周期 epoch 为 150 轮, 批量大小 batchsize 为 16, 在 140 轮后关闭数据 mosaic 增强。

### 3.2 实验数据集

实验数据集采用了多国道路损伤图像数据集 RDD2022<sup>[21]</sup>, 由印度理工学院罗凯里分校交通系统中心等机构创建, 被使用于基于 Crowdsensing 的道路损坏检测挑战赛。该数据集包含来自 6 个国家的 47 420 张道路图像, 标注了超过 55 000 个道路损伤实例, 数据集中捕获了 4 种类型的道路损伤, 纵向裂纹 (D00)、横向裂纹 (D10)、交叉裂缝 (D20) 和坑洞 (D40), 旨在通过深度学习方法自动检测和分类道路损伤。由于计算资源有限, 部分国家图像尺寸较大, 难以用于训练, 在该数据集中清洗出了 Japan 所有带标签的图像共 8 900 张, 并按照 7 : 1 : 2 的比例随机划分为训练集 (5 530 张图像)、验证集 (800 张图像) 和测试集 (1 570 张图像)。

### 3.3 实验评估指标

评估目标检测实验有效性的主要指标包括精准率 (precision, P)、召回率 (recall, R)、平均精度均值 (mean average precision, mAP)、参数量 (Params) 和计算量 (GFLOPs)。计算公式如下:

$$P = TP / (TP + FP) \quad (8)$$

$$R = TP / (TP + FN) \quad (9)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n \int_0^1 P(R) dR \quad (10)$$

式中:  $TP$  为检测结果中正确目标的数量;  $FP$  为检测结果中错误目标的数量;  $FN$  分别为检测结果中缺失目标的数量。

### 3.4 实验结果与分析

#### 1) 特征融合通道注意力对比试验

在实验过程中, 将提出的通道注意力对特征融合前后的特征图进行对比试验, 分别为单独的 P4 层特征图, 常规 FPN 特征融合操作中的 P5 层上采样拼接到 P4 层的特征图和本文提出的 P3、P4、P5 相融合的特征图。实验结果如表 1 所示 (加粗表示此项评价指标中的最优值, 下同), 仅对 P4 层进行通道注意力操作后, 各个标签的检测精度都有轻微的上涨, mAP@0. 5 上涨了 0. 3%; P4

和 P5 层融合的通道注意力操作后,其中 D40 标签有了明显的上涨,为 1.1%。mAP@0.5 和 mAP@0.5:0.95 上涨幅度较小;提出的方法进一步融合了浅层细节特征信息,D00、D10 和 D20 等裂缝的检测精度分别提高了 0.9%、0.7%和 1.0%。mAP@0.5 和 mAP@0.5:0.95 提高了 0.9%和 0.3%。

表 1 特征融合通道注意力对比试验

Table 1 Feature fusion channel attention comparison test ( % )

算法	D00	D10	D20	D40	mAP@0.5	mAP@0.5:0.95
YOLO11	46.6	45.5	63.6	52.4	50.7	21.9
P4	46.8	45.6	64.0	52.9	51.0	21.9
P4+P5	47.0	45.9	64.0	53.5	51.2	22.0
P3+P4+P5	47.5	46.2	64.6	53.6	51.6	22.2

2) 多种卷积对比实验

在 Neck 部分进行了轻量化卷积对比试验,实验过程中,将常规卷积和 DWConv 深度卷积分别放到 GSConv 同样的位置进行对比,实验结果如表 2 所示。

表 2 多种卷积对比实验

Table 2 Multiple convolution contrast experiments

算法	mAP@0.5/ %	mAP@0.5: 0.95/%	Params/ ( $\times 10^6$ )	计算量/ GFLOPs
Conv	50.7	21.9	2.6	6.5
DWConv	50.4	21.8	2.3	6.0
GSConv	50.8	21.9	2.4	6.2

从表 2 可以看出,YOLO11 模型网络 Neck 部分的常规卷积替换为 DWConv 后,参数量与计算量有明显的减少,但 DWConv 在每个通道使用单独卷积核进行卷积操作导致精度下降明显;而常规卷积与 DWConv 结合的 GSConv 相较于常规卷积,不但使检测精度有微小的提升,而且降低了模型的计算成本,更好的平衡了模型的准确性和检测速度。

3) 轻量级网络不同位置对比实验

在不影响检测精度的同时,更好的满足网络模型的网络轻量化,检测实时化,将 GSConv 和 C3k2-GS 组合的轻量级网络 LNet 分别替换 Backbone 和 Neck 的 Conv 和 C3k2 模块进行对比实验,实验结果如表 3 所示。

表 3 轻量级网络不同位置对比实验

Table 3 Lightweight network different position comparison experiment

算法	mAP@0.5/ %	mAP@0.5: 0.95/%	Params/ ( $\times 10^6$ )	计算量/ GFLOPs
YOLO11n	50.7	21.9	2.6	6.5
Backbone	50.3	21.5	2.0	4.8
Neck	50.7	21.8	2.1	5.3

从表 3 可以看出,在 Backbone 中使用轻量级网络 LNet 时,参数量和计算量有明显的下降,但由于输入图像的空间信息在 Backbone 中正在逐步向通道传输,常规卷积的密集卷积计算最大限度的保留了每个通道之间的连接,而 GSConv 的稀疏卷积操作导致空间信息的丢失,导致最终检测精度下降明显。在 Neck 中使用 LNet 时,Neck 部分网络通道维度已经达到最大,稀疏卷积操作对通道影响最小,模型参数量与计算量下降的同时准确性非常接近原始模型。因此在 Neck 部分使用 LNet 能有效提取特征信息的同时实现更加轻量化的设计。

4) 消融实验

为了直观地体现各模块对网络模型性能的影响,验证提出的模块对算法改进的有效性,即在 YOLO11 n 网络模型上添加 FFCA 模块和 MSFE 模块以及在 Neck 中使用 LNet,进行了 7 组消融实验,其结果如表 4 所示。

表 4 消融实验结果

Table 4 Result of ablation experiment

YOLO11n	FFCA	MSFE	LNet	mAP@0.5/%	mAP@0.5:0.95/%	Params/( $\times 10^6$ )	计算量/GFLOPs
√				50.7	21.9	2.6	6.5
√	√			51.6	22.2	2.6	6.6
√		√		51.5	22.1	2.7	6.7
√			√	50.7	22.0	2.1	5.3
√	√	√		52.4	22.7	2.8	6.8
√	√		√	52.1	22.3	2.2	5.5
√	√	√	√	52.5	22.8	2.3	5.7

由表 4 可以看出,模型添加 FFCA 模块后平均检测精度 mAP@0.5 从 50.7%提升至 51.6%,与此同时,模型

的推理速度几乎没有造成太多的不良影响;添加 MSFE 模块后,多尺度特征融合后使用 3D 卷积提炼特征信息,

由于 3D 卷积需要处理更多维度的信息,该模块导致模型计算量的增加,但  $mAP@0.5$  从 50.7% 提升至 51.5%;使用 LNet 后,模型的精度保持不变,参数量和计算量分别下降 19.2% 和 18.4%。综合以上实验,同时使用 FFCA 模块,MSFE 模块和 LNet 模块,经过 150 轮迭代训练,最后得到的 MSL-YOLO 模型  $mAP@0.5$  提升了 1.8%,参数量和计算量下降了 11.5% 和 12.3%,在提升精度的同时满足了模型的轻量化。

5) 可视化分析

为了更直观验证 MSL-YOLO 模型的损伤检测效果,选取 RDD2022 数据集部分图片进行对比实验,结果如图 8、9 所示。图 8 为环境较为单一下的检测对比实验,图 9 为环境复杂下的检测对比实验,图 8(a) 和 9(a) 为数据集原始图,图 8(b) 和 9(b) 为 YOLO11n 模型检测结果,图 8(c) 和 9(c) 为 MSL-YOLO 模型检测结果。

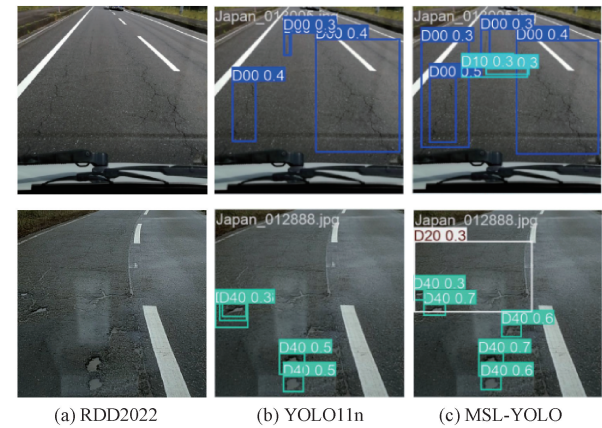


图 8 单一环境对比

Fig. 8 Comparison diagram of a single environmental

图 8 中道路环境单一,由于 YOLO11n 模型方法在下采样时会丢失大量的细节信息,在 D10 和 D20 目标上有明显的漏检,改进后的模型保留了大量细节信息,检测精度更高的同时解决了漏检的问题。图 9 中道路环境相对复杂,光照,阴影等客观条件对损伤目标特征识别造成了巨大的困扰,图中明显的看到 YOLO11n 模型的检测精度低,目标捕捉困难导致漏检率高,经过改进的 MSL-YOLO 模型融合多尺度信息后大幅减少了漏检的问题。

精度-召回率曲线可以直观看出模型对所有标签类别的平均精度的影响,精度-召回率曲线下方围成的面积即精度值,通常根据曲线下方面积大小进行实验比较,图 10 所示为 YOLO11n 与 MSL-YOLO 的精度-召回率曲线对比,图 10(a) 为 YOLO11n 模型的精度-召回率曲线,图 10(b) 为 MSL-YOLO 的精度-召回率曲线,明显的看出 MSL-YOLO 的精度-召回率曲线下方面积更大,综合考虑召回率和准确率的情况下,MSL-YOLO 模型的性能优于 YOLO11n。



图 9 复杂环境对比

Fig. 9 Comparison diagram of complex environment

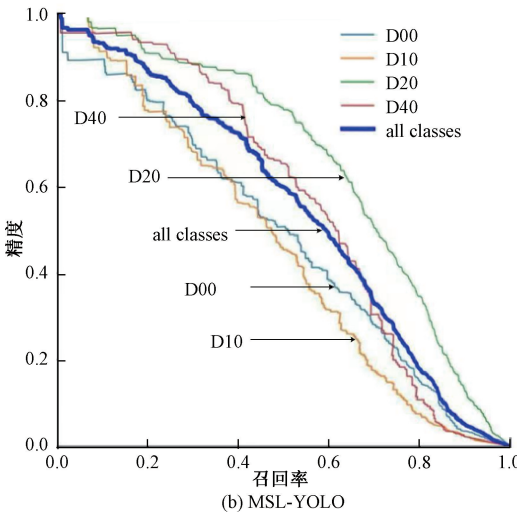
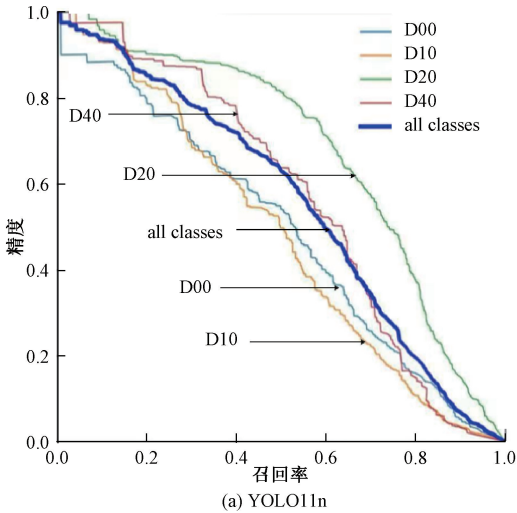


图 10 PR 曲线对比图

Fig. 10 PR curve comparison chart

6) 对比实验

表 5 为 MSL-YOLO 模型与多种现有方法的比较,包



括单阶段检测算法中其他经典 YOLO 算法,两阶段检测算法中的 Faster-FCNN 算法和基于 Transformer<sup>[22]</sup> 架构的计算机视觉模型 Swin Transformer 算法等。

表 5 对比实验结果  
Table 5 Result of comparative experiment

算法	mAP@ 0. 5/%	mAP@ 0. 5; 0. 95/%	Params/ (×10 <sup>6</sup> )	计算量/ GFLOPs
Faster-RCNN	47. 7	19. 7	137. 5	370. 3
YOLOv5s	47. 6	19. 9	7. 0	16. 0
YOLOv7tiny	48. 4	20. 4	6. 0	13. 2
YOLOv8n	50. 1	21. 0	3. 0	8. 2
YOLOv9s	50. 2	21. 0	7. 9	28. 1
YOLOv10n	50. 4	21. 6	2. 8	8. 4
YOLO11n	50. 7	21. 9	2. 6	6. 5
RT-DETR	50. 7	22. 0	20. 1	58. 3
Swin Transformer	50. 8	22. 1	96. 8	17. 1
MSL-YOLO	<b>52. 5</b>	<b>22. 8</b>	<b>2. 3</b>	<b>5. 7</b>

结果表明,相较于 Faster-FCNN 算法,MSL-YOLO 模型参数量和计算量更少,mAP@ 0. 5 和 mAP@ 0. 5:0. 95 指标上分别提高了 4. 8%和 3. 1%,相较于 one-stage 检测算法中轻量级 YOLOv5s、YOLOv7tiny、YOLOv8n、

YOLOv9s、YOLOv10n 和 YOLO11n 算法,MSL-YOLO 模型在参数量和计算量更少的情况下 mAP@ 0. 5 分别提高了 4. 9%、4. 1%、2. 4%、2. 3%、2. 1%和 1. 8%,mAP@ 0. 5:0. 95 分别提高了 2. 9%、2. 4%、1. 8%、1. 8%、1. 2%和 0. 9%。相较于 Swin Transformer 算法和 RT-DETR<sup>[23]</sup> 算法,MSL-YOLO 模型精度更高的同时模型体积远远小于以上两种模型。综上所述,改进后的算法提高检测精度的同时具有更少的参数量和实时的检测速度,使用成本更低,具有高效性和优越性。

7) 泛化实验

为了评估提出的 MSL-YOLO 模型在实际应用中的性能,另选道路损伤数据集 Road Damage 进行泛化实验,实验结果如表 6 所示。该数据集是使用手机拍摄的 3 321 张道路损伤图像,其标签分类与 RDD2022 数据集一致。实验环境和配置采用 3. 1 节所用环境,经过 150 轮训练后,MSL-YOLO 模型与 YOLO11n 相比,mAP@ 0. 5 和 mAP@ 0. 5:0. 95 分别提升了 1. 4%和 1. 6%,参数量和计算量均有所减少,在提高检测精度的同时依旧达到了轻量化模型。综上实验,MSL-YOLO 模型具有一定的泛化能力。

表 6 泛化能力实验  
Table 6 Result of generalization ability experiment

算法	D00/%	D10/%	D20/%	D40/%	mAP@ 0. 5/%	mAP@ 0. 5:0. 95/%	Params/(×10 <sup>6</sup> )	计算量/GFLOPs
YOLO11	47. 9	46. 3	63. 8	52. 9	52. 2	22. 5	2. 6	6. 5
MSL-YOLO	<b>48. 5</b>	<b>46. 9</b>	<b>64. 8</b>	<b>54. 2</b>	<b>53. 6</b>	<b>24. 1</b>	<b>2. 3</b>	<b>5. 6</b>

4 结 论

根据当前道路损伤检测方法精度高、速度快、易部署的要求,提出了一种多尺度特征融合的轻量化道路损伤检测方法 MSL-YOLO。针对道路损伤目标的在背景干扰下细节信息丢失,特征表达能力不足,提出了特征融合通道注意力 FFCA 模块提高损伤信息的权重,抑制复杂背景信息的干扰,增强模型提取损伤特征的能力;通过设计多尺度特征增强 MSFE 模块使模型能够在复杂环境下更好的捕捉损伤目标,进一步提高道路损伤检测性能;通过引入轻量级网络 LNet,使模型提高检测效率,实现轻量化和实时化。实验结果表明,在道路损伤数据集 RDD2022 上,mAP@ 0. 5 和 mAP@ 0. 5:0. 95 分别达到了 52. 5%和 22. 8%,相较于 YOLO11n,分别提高了 1. 8%和 0. 9%,参数量和计算量下降了 11. 5%和 12. 3%。基本满足了道路损伤检测的要求。

参考文献

[ 1 ] 曾志超,徐玥,王景玉,等. 基于 SOE-YOLO 轻量化

的水面目标检测算法[J]. 图学学报,2024,45(4): 736-744.

ZENG ZH CH, XU Y, WANG J Y, et al. A water surface target detection algorithm based on SOE-YOLO lightweight network [J]. Journal of Graphics, 2024, 45(4):736-744.

[ 2 ] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Newyork: IEEE Press, 2014: 580-587.

[ 3 ] GIRSHICK R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision. Newyork: IEEE Press, 2015: 1440-1448.

[ 4 ] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.



- [ 5 ] 刘子龙, 罗晨, 周怡君, 等. 基于圆卷积神经网络的粘连导电粒子检测 [J]. 光学精密工程, 2024, 32(11): 1788-1800.
- LIU Z L, LUO CH, ZHOU Y J, et al. Detection of conductive multi-particles based on circular convolutional neural network [J]. Optics and Precision Engineering, 2024, 32(11): 1788-1800.
- [ 6 ] 于海洋, 景鹏, 张文涛, 等. 基于残差与注意力机制的道路裂缝检测 U-Net 改进模型 [J]. 计算机工程, 2023, 49(6): 265-273.
- YU H Y, JING P, ZHANG W T. Improved U-Net model for road crack detection based on residual and attention mechanism [J]. Computer Engineering, 2023, 49(6): 265-273.
- [ 7 ] SEKAR A, PERUMAL V. Automatic road crack detection and classification using multi-tasking faster RCNN [J]. Journal of Intelligent & Fuzzy Systems, 2021, 41(6): 6615-6628.
- [ 8 ] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector [C]. LEIBE B, MATAS J, SEBE N, et al, eds. Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [ 9 ] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 779-788.
- [ 10 ] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2017: 6517-6525.
- [ 11 ] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. ArXiv preprint arXiv: 1804. 02767, 2018.
- [ 12 ] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. ArXiv preprint arXiv:2004. 10934, 2020.
- [ 13 ] 任安虎, 姜子渊, 马晨浩. 基于改进 YOLOv5s 的道路裂缝检测算法 [J]. 激光杂志, 2024, 45(4): 88-94.
- REN AN H, JIANG Z Y, MA CH H. Road crack detection algorithm based on improved YOLOv5s [J]. Laser Journal, 2024, 45(4): 88-94.
- [ 14 ] 李松, 史涛, 井方科. 改进 YOLOv8 的道路损伤检测算法 [J]. 计算机工程与应用, 2023, 59(23): 165-174.
- LI S, SHI T, JING F K. Improved YOLOv8 road damage detection algorithm [J]. Journal of Computer Engineering & Applications, 2023, 59(23): 165-174.
- [ 15 ] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements [J]. ArXiv preprint arXiv:2410. 17725, 2024.
- [ 16 ] SAPKOTA R, ULTRALYTICS K M. YOLO evolution: An overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 object detectors for computer vision and pattern recognition [J]. ArXiv preprint arXiv: 2510. 09653, 2025.
- [ 17 ] LI H L, LI J, WEI H B, et al. Slim-neck by GSConv: A lightweight-design for real-time detector architectures [J]. Journal of Real-Time Image Processing, 2024, 21(3): 62.
- [ 18 ] 张家波, 何阿娟, 唐上松. 语义增强和自适应多尺度特征融合的人体姿态估计 [J/OL]. 计算机工程与应用, 1-13 [2025-01-17]. <http://kns.cnki.net/kcms/detail/11.2127.tp.20241230.1059.004.html>.
- ZHANG J B, HE A J, TANG SH S. Semantic enhancements and adaptive multi-scale feature fusion of the human body posture estimation [J/OL]. Computer engineering and application, 1-13 [2025-01-17]. <http://kns.cnki.net/kcms/detail/11.2127.tp.20241230.1059.004.html>.
- [ 19 ] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.
- [ 20 ] 杨宇龙, 张银胜, 段修贤, 等. 基于轻量级卷积网络的风力发电机表面缺陷检测研究 [J]. 电子测量与仪器学报, 2024, 38(8): 36-45.
- YANG Y L, ZHANG Y SH, DUAN X X, et al. Wind generator surface defect detection based on lightweight convolution network study [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(8): 36-45.
- [ 21 ] ARYA D, MAEDA H, GHOSH S K, et al. RDD2022: A multi-national image dataset for automatic road damage detection [J]. Geoscience Data Journal, 2024, 11(4): 846-862.
- [ 22 ] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. ArXiv preprint arXiv: 1706. 03762, 2017.
- [ 23 ] ZHAO Y, LV W, XU S, et al. Detsr beat yolos on real-time object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

2024; 16965-16974.

作者简介



武兵, 2023 年于辽宁科技大学获得学士学位, 现为辽宁科技大学硕士研究生, 主要研究方向为计算机视觉、深度学习。

E-mail: 2258860606@qq.com

**Wu Bing** received his B. Sc. degree from University of Science and Technology Liaoning

in 2023. He is now a M. Sc. candidate at University of Science and Technology Liaoning. His main research interests include computer vision and deep learning.



田莹 (通信作者), 2008 年于沈阳工业大学获得博士学位, 现为辽宁科技大学教授, 主要研究方向为计算机视觉、深度学习、模式识别等。

E-mail: t\_tianying@126.com

**Tian Ying** ( Corresponding author )

received her Ph. D. degree from Shenyang University of Technology in 2008. She is now a professor at University of Science and Technology Liaoning. Her main research interests include computer vision and deep learning, and pattern recognition, etc.