

# 基于改进ViT-YOLOv5的高速公路表面裂缝检测研究

王鑫<sup>1</sup>, 廖明军<sup>2\*</sup>, 王凯英<sup>2</sup>

<sup>1</sup>盐城工学院机械工程学院, 江苏 盐城

<sup>2</sup>盐城工学院交通工程系, 江苏 盐城

收稿日期: 2023年9月11日; 录用日期: 2023年11月1日; 发布日期: 2023年11月7日

## 摘要

道路裂缝会对行车安全构成威胁甚至引发交通事故, 现有的检测方法在精度和准确率上仍然有改进的空间。因此提出了一种基于改进的ViT-YOLOv5 (Transformer-You Only Look Once Version 5) 的高速公路表面裂缝检测研究。首先引入了双向特征金字塔网络(BiFPN)特征融合结构, 并引入了自注意力机制和改进的Squeeze-and-Excitation (SE)注意力机制, 动态地调整每个通道的特征权重, 对输入特征图的不同通道给予不同的重视程度。其次借鉴了BiFPN的双向跨尺度连接和加权特征融合结构, 改进了YOLOv5的头部网络, 更有效地实现了多尺度特征融合。最后采用Transformer的核心思想, 设计了Transformer层模块, 引入了自注意力机制, 增强了模型关注信息的不同方面的能力。实验结果表明, 改进后的目标检测算法在高速公路表面裂缝检测中的平均精度mAP@0.5, 与传统YOLOv5算法相比提升了28个百分点。改进后的算法能够快速、准确地提高高速公路表面裂缝检测精度。

## 关键词

高速公路, 裂缝检测, YOLOv5, Transformer, 注意力机制, 特征融合

# Research on Freeway Surface Crack Detection Based on Improved ViT-YOLOv5

Xin Wang<sup>1</sup>, Mingjun Liao<sup>2\*</sup>, Kaiying Wang<sup>2</sup>

<sup>1</sup>School of Mechanical Engineering, Yancheng Institute of Technology, Yancheng Jiangsu

<sup>2</sup>Department of Traffic Engineering, Yancheng Institute of Technology, Yancheng Jiangsu

Received: Sep. 11<sup>th</sup>, 2023; accepted: Nov. 1<sup>st</sup>, 2023; published: Nov. 7<sup>th</sup>, 2023

\*通讯作者。

文章引用: 王鑫, 廖明军, 王凯英. 基于改进 ViT-YOLOv5 的高速公路表面裂缝检测研究[J]. 建模与仿真, 2023, 12(6): 5114-5127. DOI: 10.12677/mos.2023.126465

## Abstract

Serious freeway pavement crack can pose a threat to driving safety and even cause traffic accidents. Existing detection methods still have room for improvement in precision and accuracy. Therefore a method on freeway surface crack detection based on the improved ViT-YOLOv5 (Transformer-You Only Look Once Version 5) is proposed. Firstly, the feature fusion structure of bidirectional feature pyramid network (BiFPN) feature is introduced, and the self-attention mechanism and the improved Squeeze-and-Excitation (SE) attention mechanism are introduced to dynamically adjust the feature weight of each channel and give different attention to different channels of the input feature map. Secondly, by drawing on the bidirectional cross-scale connection and weighted feature fusion structure of BiFPN, the head network of YOLOv5 is improved to realize multi-scale feature fusion more effectively. Finally, the Transformer layer module is designed, based on the core idea of Transformer, and the self-attention mechanism is introduced, which enhances the ability of the model to pay attention to different aspects of information. The results show that the average accuracy of the improved object detection algorithm in freeway surface crack detection is mAP@0.5, which is 28 percentage more than the traditional YOLOv5 algorithm. The improved algorithm can quickly and accurately improve the accuracy of highway surface crack detection.

## Keywords

Freeway, Crack Detection, YOLOv5, Transformer, Attention Mechanisms, Feature Fusion

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

高速公路作为全球交通网络的重要组成部分,对于现代社会和经济的正常运行具有重要作用[1]。然而,路面裂缝对驾驶者的舒适度和行车安全构成严重威胁[2]。传统的人工检测方法无法满足大规模、高效率的公路裂缝检测需求[3]。因此,有效、准确地检测高速公路表面裂缝,对于维护道路状况和预防交通事故具有重要意义[4]。

许多研究人员对此进行了努力, Lekshmipathy 等人[5]和 Zhang 等人[6]利用加速度传感器、陀螺仪和磁传感器收集的路面状态数据,设计人工神经网络,实现了对桥头堡裂缝、坑洼和车辆跳跃等路面条件的高精度分类。但是,这个方法只能探测到轮子经过的地方,而不能探测到轮子外面的地方。

形态学操纵被广泛应用于公路病害的诊断[7] [8]。Amhaz 等人[9]利用裂纹相邻像素之间的相干性,实现了二维图像中的裂纹自动识别。Peng et al. [10]对图像中的局部裂纹斑块进行了处理。虽然形态学操作能去除不相干的结构,使得计算机能更好的对图像进行识别和理解,但算法精度低、自适应差。

近年来,基于深度卷积神经网络的建筑裂纹检测与分类方法得到了广泛的应用[11] [12]。Mao [13]等提出了一种以 VGG-16 [14]网络模型为基础的道路裂缝分类网络。Opara [15]等基于 YOLOv3 对裂纹、凹坑进行了检测。Arya 等[16]利用轻型移动网(MobileNet)络完成了公路裂缝探测任务[17]。Hedge et al. [18]采用了 YOLOv5 对象探测模型和数据增强策略。Doshi 和 Yilmaz [19]采用基于 YOLOv4 [20]的模型。Pei 等[21]利用 CascadeR-CNN 模型[22]和各种数据增强方法,得到 F1 评分 0.629。

虽然上述方法优于传统模型，但是卷积神经网络感受野的大小受限于卷积层的层数，由于计算量的限制，感受野的大小往往没有达到最佳值。受近年 Transformer 在 CV 领域的火热的启发，本文提出了一种基于改进的 ViT-YOLO 检测算法。工作的贡献如下：

- 1) 针对形状不规则和小目标的高速公路表面裂缝问题，提出了一种基于改进的 ViT-YOLO 结构的裂缝目标检测算法。
- 2) 采用双向特征金字塔网络(BiFPN)特征融合结构，并通过学习得到权重参数来融合不同层的特征映射。
- 3) 采用了 Transformer 的核心思想，设计了自注意力模块，增加有用特征的权重，抑制无效特征的权重。同时还引入了改进的 Squeeze-and-Excitation (SE)注意力机制，使模型能够对输入特征图的不同通道给予不同的重视程度。

## 2. YOLOv5 网络及原理

YOLOv5 骨干网将 Focus 架构与 CSP 架构相结合，实现了对输入样本量的有效提取。YOLOv5 采用 PANet 结构对特征层进行融合，并对特征层进行了三种尺度的预测。PANet 是一个以 FPN 为基础的自底向上的增强型网络，它取代了原有的单向网络，实现了双向网络的融合。本文从精度、效率和规模三个方面综合考量，提出一种基于 YOLOv5 的高速公路地表裂纹识别网络优化方案。图 1 所示为最初的 YOLOv5 中的网络架构。

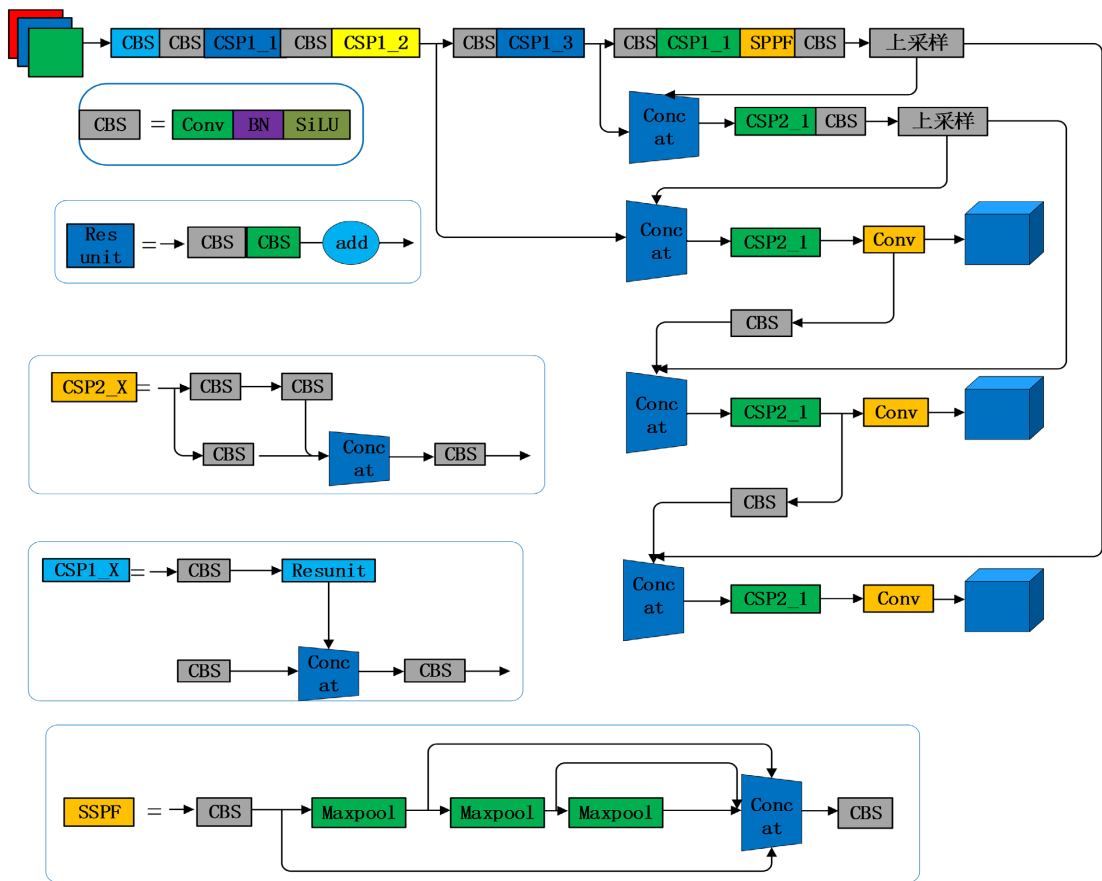


Figure 1. Original YOLOv5 network structure

图 1. 原始 YOLOv5 网络结构

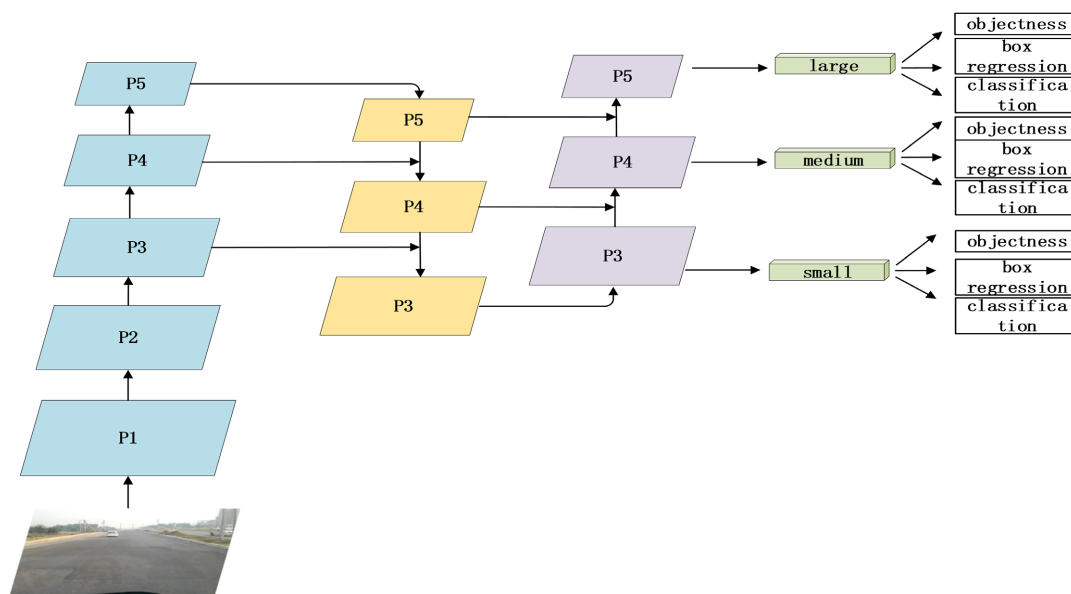


Figure 2. Network structure of the PANet header

图 2. PANet 头部网络结构

从图 2 可以看出, 一个 PANet 包含了一个自上而下的融合路线和一个自下而上的融合路线。该算法先用 2 次上采样来校正最近的特性图 X, 然后再用  $1 \times 1 \times 1$  的卷积方法把该特性图 Y 加入到预分层特性图 Y 中。

### 3. 改进的网络结构

#### 3.1. BiFPN 头部特征融合网络

与 FPN 和 PAN 等其他特征融合结构相比, 双向特征金字塔网络结构(BiFPN)可以实现更高效的双向跨尺度连接和加权特征融合。FPN 最开始应用于 YOLOv3, 但此特征融合结构本质上受到单向信息流的限制, 该结构如图 3(a)所示, FPN 自上而下传输特征信息, 并通过低级别特征融合高级别特征, 以获得用于预测的特征图, 由于 FPN 对位置信息的丢失。Yolov4 和 Yolov5 采用 PAN 特征融合结构在 FPN 的基础上增加了一个额外的自下而上的路径聚合网络。该结构如图 3(b)所示, 它将浅层定位信息传输到深层, 在多个尺度上增强了定位能力。图 3(c)的 NAS 结构广泛运用于强化学习, 旨在搜索出合适的网络结构, 虽然该模型效果好, 但是搜索空间大耗时长。BiFPN 特征融合结构如图 3(d)所示。将双向(自下而上和自上而下)路径视为特征网络层, 并多次重复同一层以实现更高级别的特征融合。

Google Brain 在 FPN 结构的基础上提出了名为 EfficientDet 的目标检测算法, 和同时期的其他目标检测算法如 YOLOv3、MaskRCNN、NAS-FPN 等相比, 在准确率和检测速度和运算量上 EfficientDet 均是一枝独秀, 如图 4 可以看到详细的实验数据对比图。虽然随着 YOLO 算法的发展, EfficientDet 算法的性能已经比不上 YOLOv4 和 YOLOv5, 但是它提出的 BiFPN 特征融合结构却值得参考和借鉴。本文将 BiFPN 特征融合结构引入 YOLOv5 算法, 改进后的 YOLOv5 的模型网络结构如图 5 所示。

头部网络结构在改进为 BiFPN 结构的同时保持了骨干层的网络结构, 例如, 输入  $640 \times 640$  像素的图片, 通过骨干层后, 输出为 1024 个大小  $4 \times 4$  的特征图。在第 6 层中, 通过 ATCSP 层, 采用三倍层叠, 通过多次卷积等操作, 输出 512 个尺寸为  $8 \times 8$  的特征图。在第 10 层中, 使用  $1 \times 1$  的卷积核运算, 获得了 512 个大小  $4 \times 4$  的特征图样。在第 11 层, 利用上采样操作, 可以获得大小  $8 \times 8$  的特征图。在第 12 层, 将从第 6 层得到的 512 个特征图样与之相结合, 可以获得 1024 个大小  $8 \times 8$  的特征图。CSP 层的内部结构如图 6 所示。

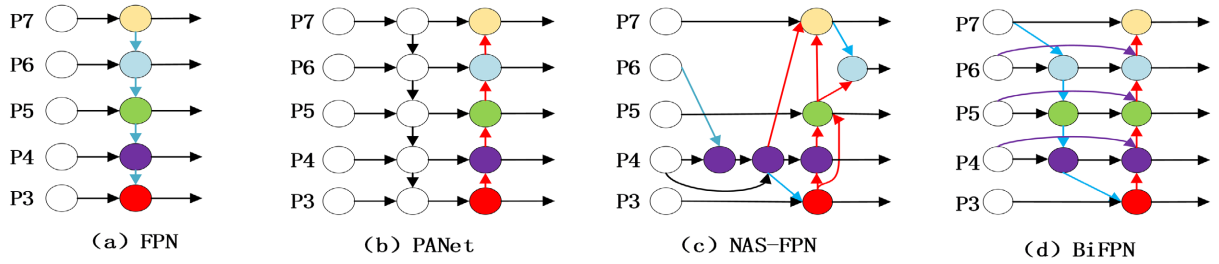


Figure 3. Feature fusion network structure  
图 3. 特征融合网络结构

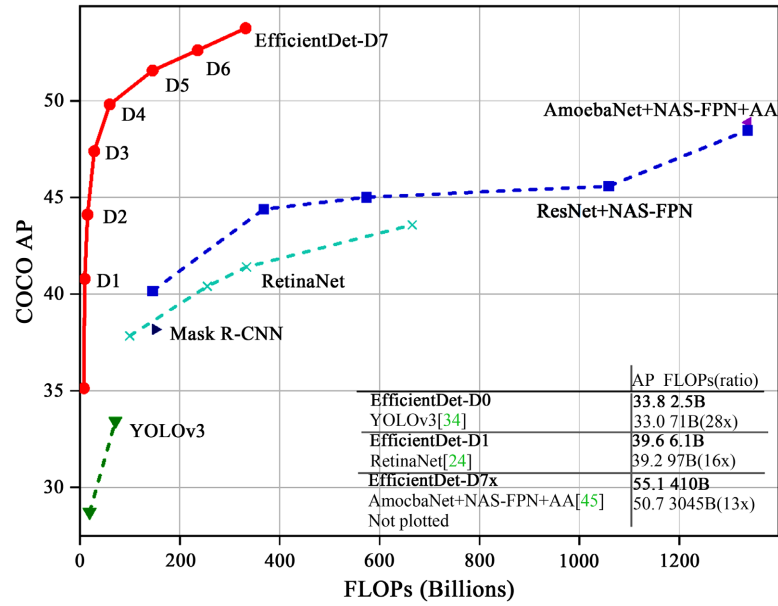


Figure 4. EfficientDet performance compared with other algorithms  
图 4. EfficientDet 与其他算法的性能对比图

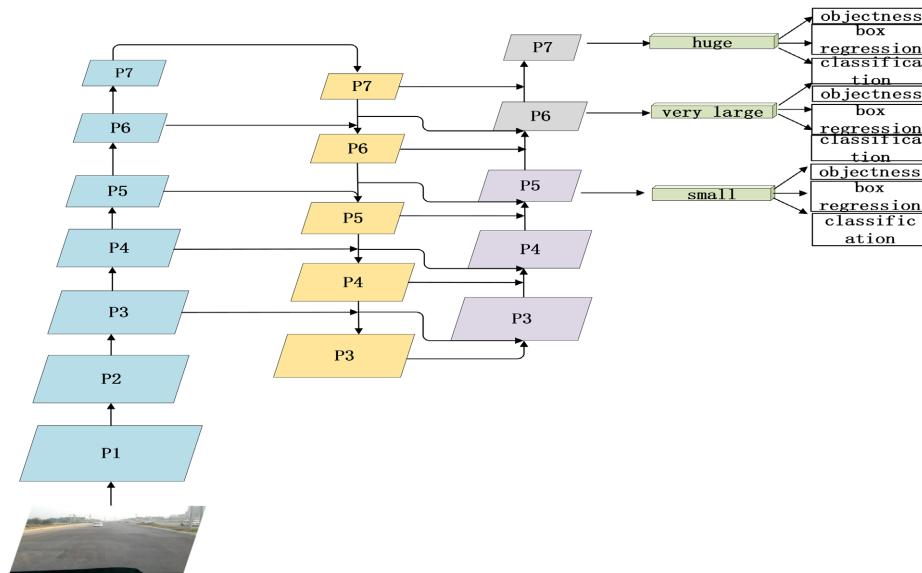


Figure 5. Improved header network structure  
图 5. 改进后的头部网络结构

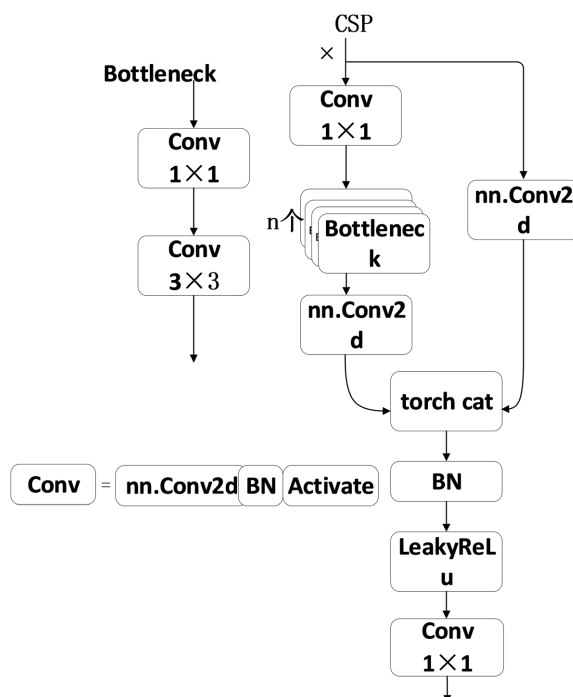


Figure 6. Internal structure of CSP layer

图 6. CSP 层的内部结构

14 至 17 层使用与 10 至 13 层相同的网状构造，在 14 层中 1024 个大小为  $8 \times 8$  的特性图形产生 256 个大小  $16 \times 16$  的特性图形，并将其应用于第 17 层。在第 18 层，使用  $3 \times 3$  的卷积核心，进行连续的两个卷积层运算，可获得大小为  $8 \times 8$  的 256 个本征图样。在第 19 层，将 BiFPN 的特征融合模式的理念与之相结合，实现了双向的跨尺度连接。

对特征图尺寸同为  $8 \times 8$  的第 19 层、第 14 层和第 6 层进行多尺度特征融合，得到 1024 个尺寸为  $8 \times 8$  的特征图。在第 20 层通过 CSP 层，采用三倍层叠，通过多次卷积等操作输出 512 个尺寸为  $8 \times 8$  的特征图。同理，在 21 层到 23 层采用与 18 层到 20 层一样的网络结构，其中，在 22 层采用双向跨尺度连接，对特征图尺寸同为  $4 \times 4$  的第 22 层、第 10 层和第 9 层进行多尺度特征融合，得到 2048 个尺寸为  $4 \times 4$  的特征图。最终，在 23 层输出 1024 个尺寸为  $4 \times 4$  的特征图，检测层也保持与原有的 YOLOv5 一致的结构。改进后的模型用于苹果叶部初次检测，由于该模型保留了原始 YOLOv5 模型的骨干网络，方便了对成熟的苹果叶部检测的 YOLOv5 模型进行迁移学习。

### 3.2. 自注意力机制

Transformer 机制最早被用在了很多 NLP 任务上。近年来，有些学者改进 Transformer 机制，使得能应用到计算机视觉任务中。该算法仅对目标像素的局部邻域进行自动关注，未对其整体进行处理。本文将使用一种可扩充的逼近方法，来获得一种具有全局自注意力的稀疏转换方法。本研究借鉴 Transformer 的核心理念引入自注意力机制，构建传输器结构模型。图 7 显示了 Transformer 层的结构。

图 7 中 Q, K, V 分别代表 query (请求)、key (键值)、value (值)，其中，V 表示训练好的模型中数据 K 对应的值，Q 代表输入 K 中的需要预测的数据。后续通过计算比较输入的 Q 值与 K 的相似度，然后将相似度进行概率化，将概率化后的值作为各自的权重，实现自注意功能。具体关系如下图 8 所示。

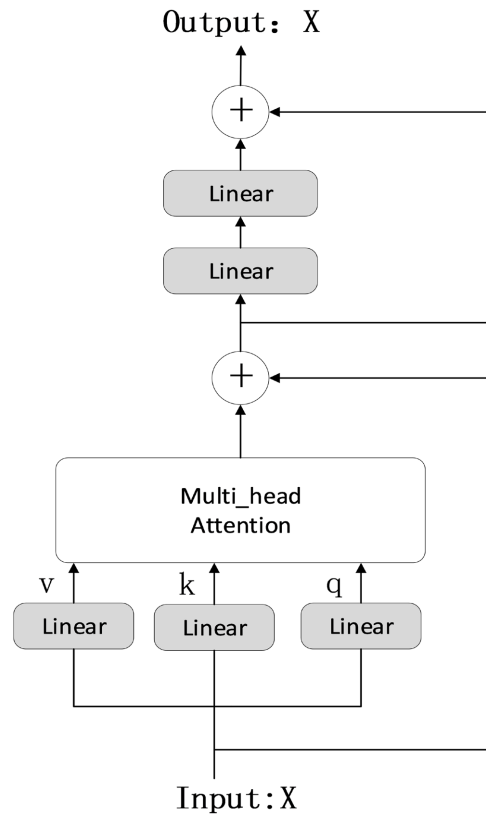


Figure 7. Structure of the TRANSFORMER layer  
图 7. Transformer 层的结构

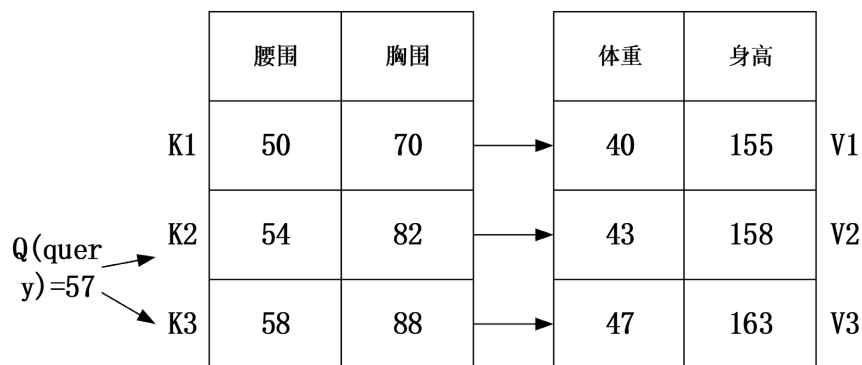


Figure 8. Parametric diagram  
图 8. 参数关系图

上图中，输入 Q 的值位于 K2、K3 之间，通过计算 Q 与 K1、K2、K3 相似度赋予不同的权重，实现 Q 的 V 值预测。

图 9 显示了多头注意力模块的结构，多头自注意力模块是设计特征融合网络的基本组成部分。注意力机制计算公式(1)如下：

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

式中； $\mathbf{Q}$ ， $\mathbf{K}$ ， $\mathbf{V}$  表示多维情况下的输入矩阵，除以  $\sqrt{d_k}$  是为了缓解梯度消失。

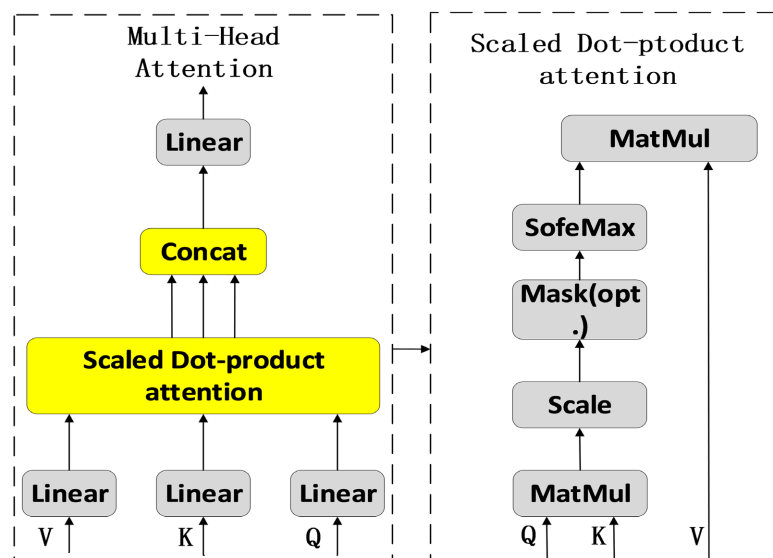


Figure 9. Structure of multi-head attention module

图 9. 多头注意力模块的结构

将注意机制扩展到多个头部，使该机制能够考虑不同的注意分布，并使模型关注信息的不同方面。多头注意机制见下公式(2)、(3)：

$$\text{MultiHead}(Q, K, V) = \text{Concat}(H_1, \dots, H_n)W^0 \quad (2)$$

$$H_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

式(2)中，乘以  $W^0$  是为了降低维度，式(3)中  $W_i^Q$ 、 $W_i^K$ 、 $W_i^V$  是没有经过线性变换的  $Q$ 、 $K$ 、 $V$  参数矩阵。

高速公路表面损伤的种类和大小多种多样，同一种类的损伤大小也可能不固定。

纵向裂缝和横向裂缝都呈现独特的特征。纵向裂缝通常颜色较深并从路面突出，而横向裂缝则通常呈现浅色，损坏的地方会出现凹陷。而在有些情况下，这些裂缝可能呈现灰色，并在内部具有网状的纹路。

鳄鱼裂缝的特征是连续的裂缝成簇出现并集中在某个局部区域。这类裂缝通常形状复杂，呈现出网格形状，呈浅灰色，影响区域较大。

而对于坑洞，这类损伤通常会在整个车道上产生明显的影响。坑洞的表面覆盖着明显的颜色变化，深度较大的坑洞是黑色的，较浅的坑洞呈现黄色。

### 3.3. SE 注意力机制

在本研究中，引入了一种改进的 Squeeze-and-Excitation (SE)注意力机制，并将其集成到 YOLOv5 模型中，以提高道路裂缝检测系统的性能。SE 注意力机制的基本原理是通过动态调整每个通道的特征权重，以显式建模模型的注意力，使模型能够对输入特征图的不同通道给予不同的重视程度。

然而，针对复杂的道路裂缝特征，传统的 SE 注意力机制可能存在捕捉信息不充分的问题。为了解决这个问题，本文对 SE 注意力机制进行了以下改进。

在 YOLOv5 的网络结构中，选择了一些关键的位置，在这些卷积层之后插入了改进的 SE 模块。这些位置主要是一些降采样层之后的位置，因为这些位置的特征图包含了更丰富、更高层次的信息。本文的改进策略主要是通过 SE 模块动态地调整这些高级特征的权重，以优化模型的表现。在这种情况下，特征权重调整可以被表示为：



$$\sum w_i = \text{Sigmoid}\left(\text{FC}\left(\text{ReLU}\left(\text{FC}\left(\text{avgpool}(F)\right)\right)\right)\right) \quad (4)$$

式中： $F$  代表原始输入的特征图， $\text{avgpool}$  代表全局平均池化， $\text{FC}$  代表全连接层， $\text{ReLU}$  是非线性激活函数， $\text{Sigmoid}$  是将权重映射到 0~1 之间的激活函数， $w_i$  为对应的权重。

本文在 SE 模块中进行了进一步的改进。首先调整了全连接层的大小，使其更适合处理道路裂缝检测任务的特性。然后，在 SE 模块中添加了一些额外的层，例如归一化层，以进一步增强模型的性能。本文的改进策略主要是通过增加模型的深度，以改进模型的特征提取能力。

在训练过程中，采用了端到端的训练策略，使得整个网络能够进行联合优化。这样，网络不仅能够优化单个卷积层或 SE 模块，而且能够优化整个特征表示，从而在整个系统上提高道路裂缝检测的性能。在这种情况下，整个网络的优化目标可以表示为：

$$\sum \text{minimize } L = \Sigma L_{cls} + \lambda 1 * \Sigma L_{loc} + \lambda 2 * \Sigma L_{SE} \quad (5)$$

其中  $L_{cls}$  为分类损失， $L_{loc}$  为位置回归损失， $L_{SE}$  为 SE 模块的损失， $\lambda 1$  和  $\lambda 2$  是权重系数，用于控制不同损失之间的相对重要性。

总的来说，通过在 YOLOv5 中集成改进的 SE 注意力机制，期望模型能够更好地捕捉和利用道路裂缝的特性，从而提高道路裂缝检测的准确性和鲁棒性。这种结合深度学习和注意力机制的策略，为进一步提高道路裂缝检测系统的性能提供了一种有效的途径。

## 4. 实验结果与讨论

### 4.1. RDD2020 道路裂缝数据集

RDD2020 数据集包括 26,336 幅来自印度，日本，捷克共和国的公路受损情况的图片，并对其进行了标签。其中包含纵裂裂纹、横裂裂纹、鳄鱼裂纹、凹坑四个类别。数据集中除了包含损坏的道路还包含一部分完好的道路的图片，被用来探测虚警。

### 4.2. 数据增强

由于数据集受到室外光照的影响，在拍摄过程中，会使图像整体亮度分布不均匀，在图像部分位置亮度过亮或过暗，因此需要调整图像的对比度，实现对图像的增强。

#### (1) 直方图均衡化

图像直方图是表示图像中亮度分布的直方图，描述了图像中每个亮度值的像素数，直方图均衡化是通过改变图像直方图来改变图像中各像素的灰值，把原始图像的直方图变换为均衡的形式，对像素个数较多的灰度值进行扩展，对像素个数较少的灰度值进行并归，从而增大对比度，达到对图像增强的效果。直方图均衡化的数学表达式如式(6)所示：

$$S_k = T(r_k) = \sum_{j=0}^k \frac{n_j}{n}, \quad k=0,1,\dots,L-1 \quad (6)$$

式中， $n_j$  为  $r_k$  的像素数， $n$  为像素总数， $k$  为灰度级数， $L$  为可能出现的灰度级总数。

虽然直方图均衡化后的图像整体亮度提升，但对于图像中地面上过亮的区域更亮，所在过暗的区域没有起到明显的增强效果。

#### (2) 改进直方图均衡化

针对上述问题，利用自适应阈值的方法对图像直方图均衡化，通过对图像局部对比度的调整使整幅图像的对比度得到增强。

首先将图像分成多个小块，分别计算每个小块图像的直方图，然后根据每个小块图像直方图的峰值，计算每个小块图像的阈值，同时为了防止每个小块图像的对比度过高或者过低，图像子块之间的对比度差距过大从而引入新的噪声。对相邻块的阈值进行插值，使得整幅图像的阈值低于某个限制值。通过动态调整找到图像直方图的某个阈值  $X$ ，使过高幅值的小块图像直方图降至阈值  $X$ ，同时对于幅值过低的区域进行填补，通过对整幅图像局部阈值的调整，从而解决图像过亮或过暗的区域对比度没有明显增强的问题。

在自适应直方图均衡化的基础上，对训练集原始裂缝图像进行旋转、随机亮度增强、随机色度增强、随机对比度增强和锐化。经过上述步骤能有效的解决图像过亮或过暗的区域对比度没有明显增强的问题，减小后续图像特征提取的计算量。

### 4.3. 训练环境和评价指标

为了衡量改进的算法相比原始算法的性能改善情况，本文采用  $mAP@0.5$ 、 $mAP@0.5: 0.95$  和召回率( $R$ )作为评价指标。

精度( $P$ )为所有的预测出的目标中预测正确的目标所占的比例，数学计算表达式如下：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

召回率( $R$ )为所有标注为正确的目标中预测正确的目标所占的比值，数学计算表达式如下：

$$\text{Recall} = \frac{TP}{TR + FN} \quad (8)$$

其中，上述两个公式中的参数  $TP$  表示预测正确的正面示例数， $FP$  表示预测错误的反面示例数， $FN$  表示预测错误的正面示例数。

平均精度( $AP$ )是衡量训练出的网络模型在单个类别上的优劣的重要指标，数学计算表达式如下：

$$AP = \int_0^1 P(R) dR \quad (9)$$

式中  $P(R)$  表示， $P$  (Precision)和  $R$  (Recall)形成的曲线。

所有类别的平均精度( $mAP$ )是衡量训练出的网络模型在每个类别上的平均优劣的重要指标， $mAP$  就是取所有类别上  $AP$  的平均值。它通常分为  $mAP@0.5$  和  $mAP@0.5: 0.95$ ，其中， $mAP@0.5$  指将  $IoU$  设为 0.5 时，所有类别图片的  $AP$  平均值。 $mAP@0.5: 0.95$  指  $IoU$  阈值从 0.5 到 0.95，步长 0.05 上的平均  $mAP$ 。 $IoU$  是指预测框和真实框的交集部分的面积与预测框和真实框的并集部分的面积比值。 $IoU$  和  $mAP$  的数学计算表达式如下：

$$IoU = \frac{S_{\text{overlap}}}{S_{\text{union}}} \quad (10)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (11)$$

### 4.4. 实验结果分析

本研究中的道路裂缝检测应同时考虑精确度和召回率，原始网络 YOLOv5s 训练参数如图 10，改进后的网络训练结果如图 11。

本小节利用  $mAP$ 、召回率等参数来验证模型的性能，并对检测结果进行评价。为了确认改进后的 YOLOv5 网络的各组件对最终性能的影响。通过 ATCSP 模块的整合， $mAP@0.5$  从 0.076516 提升到

0.35236, 提高了 0.275844, 召回率从 0.13881 提升到 0.45422, 提高了 0.31541。这表明改进后的 YOLOv5 可以过滤其他级别的特征, 并将有用的信息保持在该级别, 以提高对小裂缝的检测精度。然后, 采用 Transformer 机制设计了 Transformer 层模块。它可以增加有用特征的权重, 抑制无效特征的权重, 利用多头注意结构, 提高目标检测的精度。

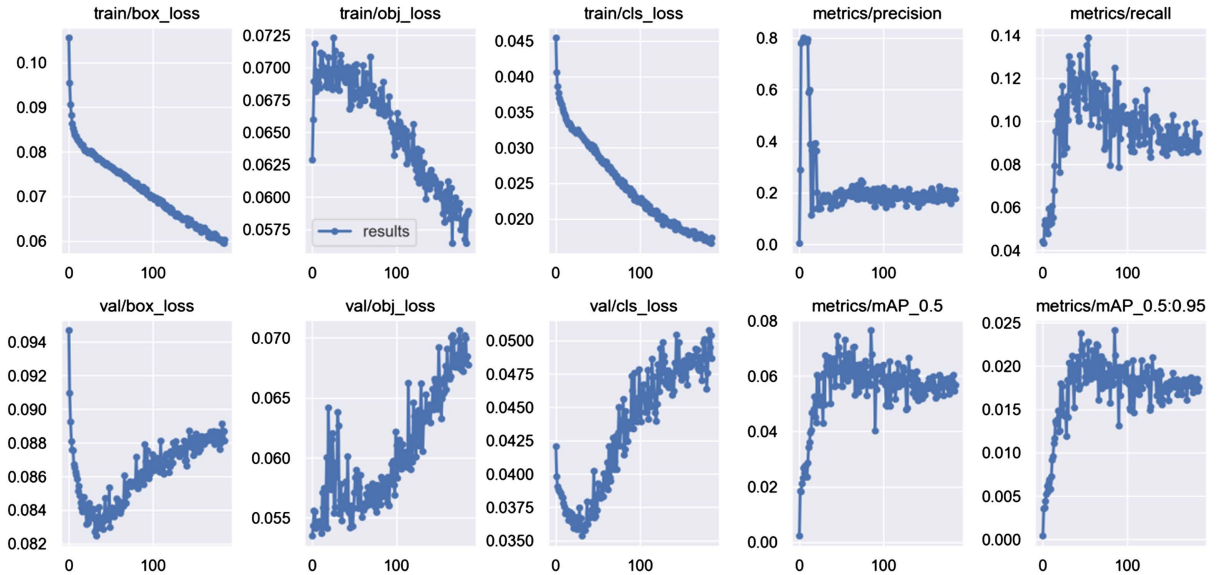


Figure 10. Original network training parameter diagram

图 10. 原始网络训练参数图

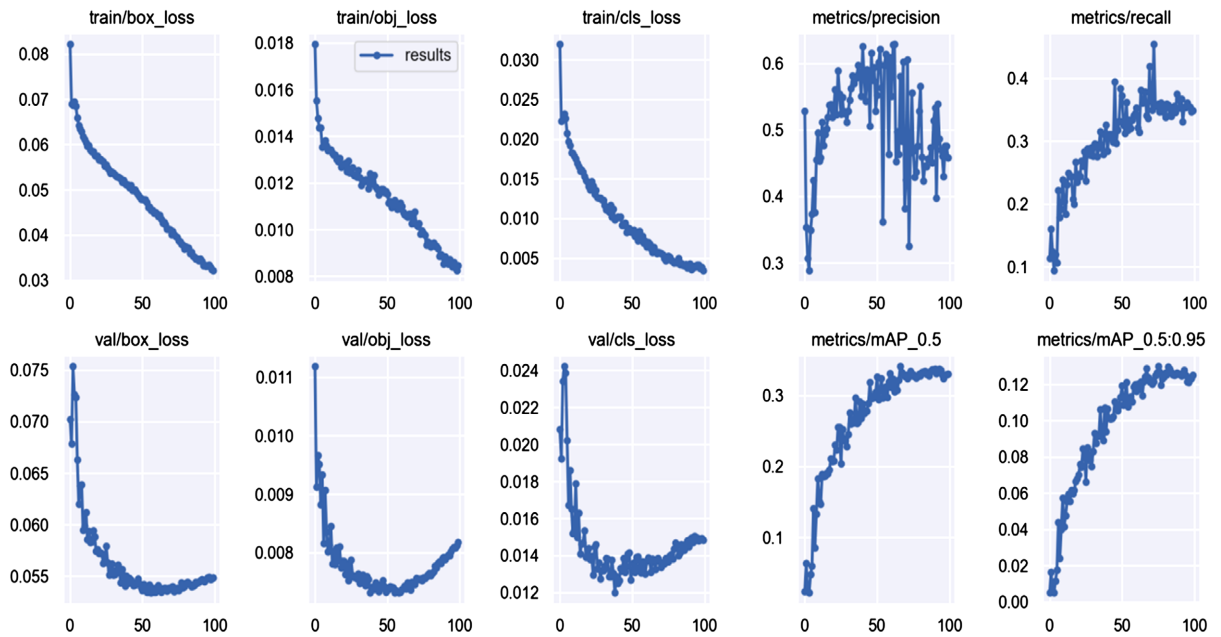


Figure 11. Improved network training parameter graph

图 11. 改进后网络训练参数图

本文给出了每种裂缝的 F1 评分曲线, 以便进行更全面的模型评估。F1 分数是精度和召回率的总和平均值, 在评估模型时将两者考虑在内。图 12 显示, 当置信水平在 0.2 和 0.4 之间时, 模型对所有裂缝

的平均 F1 得分都很高,而在置信水平为 0.38 时,F1 平均得分最高,为 0.277。结果表明,改进后的 YOLOv5 能够较好地对 4 种道路裂缝进行分类,满足实际应用场景的检测需求。

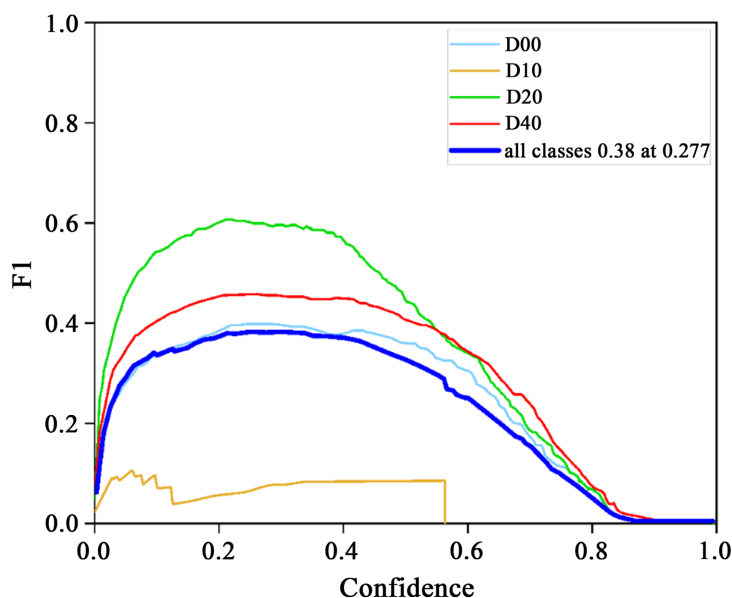


Figure 12. F1 curve

图 12. F1 曲线

同时做了其他经典单阶段算法的 mAP\_0.5 对比,如 SSD、FasterRCNN,结果如下图 13 所示,效果远不如改进后的 YOLOv5 算法。这些改进意味着所提出的 ATCSP 可以在复杂的背景下检测道路裂缝。综上所述,本文提出的改进的 YOLOv5 算法可以在满足实时检测要求的同时,保持较高的精度水平。实验结果表明,改进的 ViT-YOLO 网络在高速公路损伤检测方面具有显著优势。FasterRCNN 和 SSD 仅在模型尺寸方面优于所提出的模型,但在精度方面更差,并且具有较低的最终验证精度。此外,ViT-YOLO 在其他比较模型中实现了最高的精度和 FPS,综上所述,表明复杂度相对较低,所提模型检测速度快,易于在实践中实现。与目前最先进的检测模型相比,该模型的平均精度更高,更适合高速公路图像中的路面损伤检测。

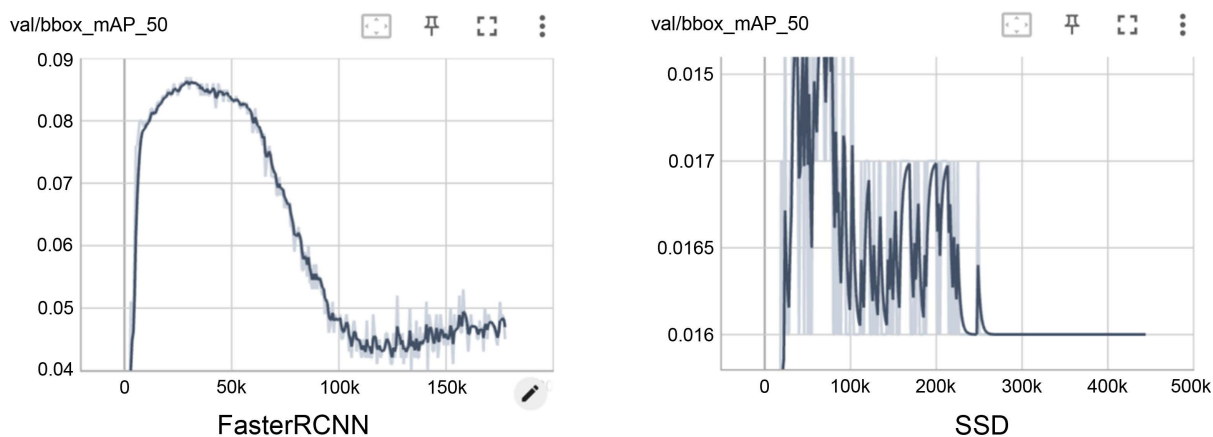


Figure 13. Experimental results of FasterRCNN and SSD

图 13. FasterRCNN 和 SSD 的实验结果

## 5. 结论

本研究基于高速公路表面裂缝检测的需求,对YOLOv5网络的头部网络进行了改进。引入了双向特征金字塔网络(BiFPN)特征融合结构,自注意力机制和改进的Squeeze-and-Excitation(SE)注意力机制,对原始网络进行了优化。BiFPN相比于FPN及FPN改进型结构,考虑了输入特征不同的分辨率对特征融合的贡献并引入了相应的权值,同时反复进行多尺度特征融合。这种改进在提升召回率的同时,有效地提升了平均精度。引入了Transformer的核心思想,设计了Transformer层模块,增强了模型关注信息的不同方面的能力。改进的SE注意力机制通过动态地调整每个通道的特征权重,对模型的注意力进行显式建模,使模型能够对输入特征图的不同通道给予不同的重视程度。在网络的关键位置插入改进的自注意力模块,有效地调整了这些高级特征的权重,从而在整个系统上提高道路裂缝检测的性能。

通过以上一系列的改进,模型在高速公路表面裂缝检测中的平均精度mAP@0.5提升了0.28。这项成果为高速公路表面裂缝检测的实际应用提供了更高效、更准确的算法方案。在未来的研究中,将进一步优化和完善模型的结构,以期在道路损伤检测方面取得更好的效果。同时也将探索更多新的注意力机制,以提升模型的性能和鲁棒性。

## 基金项目

盐城市创新创业领军人才项目(yccx2019219);盐城工学院研究生科研与实践创新计划项目(SICX21\_XY028);盐城工学院校级科研项目资助(xjr2019029)。

## 参考文献

- [1] 常州大学. 一种基于改进YOLO的道路裂缝检测方法及系统[P]. 中国, CN202210854022.9. 2022-10-18.
- [2] 王骛. 基于图像特征的路面破损检测方法及其装备技术[D]: [硕士学位论文]. 武汉: 武汉工程大学, 2016.
- [3] 王勇. 基于车载360°环视影像和深度神经网络的路面坑塘与裂缝检测[J]. 电气自动化, 2020, 42(4): 99-102.
- [4] 张志华, 温亚楠, 慕号伟, 等. 结合双注意力机制的道路裂缝检测[J]. 中国图象图形学报, 2022, 27(7): 2240-2250.
- [5] Lekshmiathy, J., Samuel, N.M. and Velayudhan, S. (2020) Vibration vs. Vision: Best Approach for Automated Pavement Distress Detection. *International Journal of Pavement Research and Technology*, **13**, 402-410. <https://doi.org/10.1007/s42947-020-0302-y>
- [6] Zhang, Y., Ma, Z., Song, X., Wu, J., Liu, S., Chen, X. and Guo, X. (2021) Road Surface Defects Detection Based on IMU Sensor. *IEEE Sensors Journal*, **22**, 2711-2721. <https://doi.org/10.1109/JSEN.2021.3135388>
- [7] Dorafshan, S., Thomas, R.J. and Maguire, M. (2018) Comparison of Deep Convolutional Neural Networks and Edge Detectors for Image-Based Crack Detection in Concrete. *Construction and Building Materials*, **186**, 1031-1045. <https://doi.org/10.1016/j.conbuildmat.2018.08.011>
- [8] Peng, C., Yang, M.Q., Zheng, Q.H., Zhang, J., Wang, D.Q., Yan, R.Y., Wang, J.X. and Li, B.J. (2020) A Triple-Thresholds Pavement Crack Detection Method Leveraging Random Structured Forest. *Construction and Building Materials*, **263**, Article ID: 120080. <https://doi.org/10.1016/j.conbuildmat.2020.120080>
- [9] Amhaz, R., Chambon, S., Idier, J. and Baltazart, V. (2016) Automatic Crack Detection on Two-Dimensional Pavement Images: An Algorithm Based on Minimal Path Selection. *IEEE Transactions on Intelligent Transportation Systems*, **17**, 2718-2729. <https://doi.org/10.1109/TITS.2015.2477675>
- [10] F. Roadware. <https://www.fugro.com/our-services/asset-integrity/roadware>
- [11] Park, S.E., Eem, S.-H. and Jeon, H. (2020) Concrete Crack Detection and Quantification Using Deep Learning and Structured Light. *Construction and Building Materials*, **252**, Article ID: 119096. <https://doi.org/10.1016/j.conbuildmat.2020.119096>
- [12] Tong, Z., Gao, J. and Zhang, H. (2018) Innovative Method for Recognizing Subgrade Defects Based on a Convolutional Neural Network. *Construction and Building Materials*, **169**, 69-82. <https://doi.org/10.1016/j.conbuildmat.2018.02.081>
- [13] Mao, Z.J., Zhao, C.H., Zheng, Y.F., Mao, Y., Li, H., Hua, L.R. and Liu, Y. (2020) Research on Detection Method of

- 
- Pavement Diseases Based on Unmanned Aerial Vehicle (UAV). *International Conference on Image, Video Processing and Artificial Intelligence*, Volume 11584, 127-132. <https://doi.org/10.1117/12.2580285>
- [14] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. 1-14.
- [15] Opara, J.N., Thein, A.B.B., Izumi, S., Yasuhara, H. and Chun, P.J. (2021) Defect Detection on Asphalt Pavement by Deep Learning. *International Journal of GEOMATE*, **21**, 87-94. <https://doi.org/10.21660/2021.83.6153>
- [16] Arya, D., Maeda, H., Ghosh, S.K., Toshniwal, D., Mraz, A., Kashiyama, T. and Sekimoto, Y. (2021) Deep Learning-Based Road Damage Detection and Classification for Multiple Countries. *Automation in Construction*, **132**, Article ID: 103935 <https://doi.org/10.1016/j.autcon.2021.103935>
- [17] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. 1-9.
- [18] Hegde, V., Trivedi, D., Alfarrarjeh, A., Deepak, A., Kim, S.H. and Shahabi, C. (2020) Yet Another Deep Learning Approach for Road Damage Detection Using Ensemble Learning. *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, 10-13 December 2020, 5553-5558. <https://doi.org/10.1109/BigData50022.2020.9377833>
- [19] Doshi, K. and Yilmaz, Y. (2020) Road Damage Detection Using Deep Ensemble Learning. *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, 10-13 December 2020, 5540-5544. <https://doi.org/10.1109/BigData50022.2020.9377774>
- [20] Bochkovskiy, A., Wang, C.-Y. and Liao, H.-Y.M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. 1-17.
- [21] Pei, Z., Lin, R., Zhang, X., Shen, H., Tang, J. and Yang, Y. (2020) CFM: A Consistency Filtering Mechanism for Road Damage Detection. *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, 10-13 December 2020, 5584-5591. <https://doi.org/10.1109/BigData50022.2020.9377911>
- [22] Cai, Z. and Vasconcelos, N. (2019) CascadeR-CNN: High Quality Object Detection and Instance Segmentation. <https://doi.org/10.1109/CVPR.2018.00644>