

基于YOLOv7-Tiny的密集行人检测模型

赵世礼, 莫红, 杨澳男, 欧阳玉琦
(长沙理工大学电气与信息工程学院, 湖南 长沙 410114)

摘要: 针对现有密集行人检测模型的参数量大及精度低的问题, 提出了一种基于YOLOv7-Tiny的跨级优化YOLOv7 (you only look once version 7-tiny partial minimum cross, YOLOv7-TPMC) 模型。首先, 提出了结合部分卷积 (partial convolution, PConv) 与逐点卷积 (pointwise convolution, PWConv) 的部分点卷积高效层聚合网络 (efficient layer aggregation networks-partial and pointwise, ELAN-PP) 模块。其次, 设计了一种跨级融合特征金字塔结构, 通过浅层特征复用提升密集场景下行人检测的准确性。最后, 为解决完整交并比 (complete intersection over union, CIoU) 损失函数在某些条件下定位不准确的问题, 引入了最小焦点距交并比 (minimum point distance and focaler-intersection over union, MF-IOU) 损失函数。在WiderPerson和CrowdHuman数据集上进行实验, YOLOv7-TPMC相比于基准模型, 参数量减少了20.5%, 帧率 (frame per second, FPS) 提升了12.2%, mAP@0.5的值在WiderPerson数据集上增加了1.2%, 在CrowdHuman数据集上增加了2.0%, 能够较好地应用于密集行人检测。

关键词: 行人检测; YOLOv7-Tiny; 特征金字塔; 部分卷积; 注意力机制

中图分类号: TP391.4

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202526

Dense pedestrian detection model based on improved YOLOv7-Tiny

ZHAO Shili, MO Hong, YANG Aonan, OUYANG Yuqi

School of Electrical & Information Engineering, Changsha University of Science & Technology, Changsha 410114, China

Abstract: In response to the issues of large parameter quantity and low accuracy in the existing dense pedestrian detection models, YOLOv7-TPMC (you only look once version 7-tiny partial minimum cross) based on YOLOv7-Tiny was proposed. Firstly, the ELAN-PP (efficient layer aggregation networks-partial and pointwise) module combining PConv (partial convolution) and PWConv (pointwise convolution) was proposed. Secondly, a cross-level fusion feature pyramid structure was designed to improve the accuracy of pedestrian detection in dense scenes through shallow feature reuse. Finally, in order to solve the problem that the CIoU (complete intersection over union) loss function was not accurate in positioning under certain conditions, the MF-IOU (minimum point distance and focaler-intersection over union) loss function was introduced. Experiments were carried out on the WiderPerson and CrowdHuman datasets. Compared with the baseline model, YOLOv7-TPMC had the number of parameters reduced by 20.5%, FPS (frame per second) improved by 12.2%, and the value of mAP@0.5 increased by 1.2% on the WiderPerson dataset. It increased by 2.0% on the CrowdHuman dataset, which can be well applied to dense pedestrian detection.

Key words: pedestrian detection, YOLOv7-Tiny, FPN, partial convolution, attention

0 引言

目标检测是计算机视觉领域的一个重要任务,

主要是确定特定目标在图像中的类别及位置^[1]。行人检测作为目标检测的一个重要分支, 不仅为行人姿态估计和行人跟踪提供技术基础, 而且被广泛应

用于自动驾驶、智能机器人、图像字幕等领域^[2]。由于密集行人场景存在遮挡、目标小、衣着色彩和行为姿态不一等问题，行人检测模型的鲁棒性仍有待提升。传统的行人检测方法^[3-6]局限于使用固定的模板对目标进行建模和匹配，尤其是在跟踪过程中，目标的外观特征可能会因为光照、视角等因素的变化而出现变化，这往往会导致传统目标跟踪方法的跟踪性能不是很理想。针对低分辨率行人难以检测的问题，Akyon等^[7]提出了切片辅助超推理（slicing aided hyper inference）方法，首先使用滑动窗口将图像分割，然后将得到的小图像与原图分别进行预测，最后将两者预测结果进行合并处理。这种方法有效地提高了部分低分辨率行人目标检测的精度，但是对于低分辨率条件下多个行人聚集或重叠的识别能力没有明显改善。近年来，基于深度学习的一阶段目标检测器因具有较好的实时性得到了广泛的应用，包括只需要看一次（you only look once, YOLO）系列^[8]和Liu等^[9]提出的SSD算法等。传统的卷积神经网络算法提取输入图像的特征，通过端到端网络直接输出分类和目标位置，简单快速。但由于密集行人检测任务中存在小目标、重叠目标以及复杂背景中的低分辨率目标，这些网络难以分离前景和背景，对复杂行人目标的定位能力差。苑朝等^[10]采用基于通道注意力机制的MobileNet网络（efficient channel attention and squeeze excitation-mobile network, ES-MobileNet）替换第四代只需要看一次（you only look once version 4, YOLOv4）的主干网络，并引入协调注意力（coordinate attention, CA）和挤压激励通道注意力（squeeze and excitation-efficient channel attention, SE-ECA）增强重要特征通道，虽然在模型轻量化方面表现较好，但改进后YOLOv4模型对于多色彩的密集目标识别效果有待提升。Han等^[11]基于微型YOLOv4（YOLOv4-Tiny）算法，引入幽灵卷积（ghost convolution）模块和扩张卷积降低模型的体积。然而，这种方法限制了模型提取高级特征的能力，对小尺度目标适应性较弱。Ma等^[12]用混洗网络（shuffle network, ShuffleNet）改进YOLOv4的结构，并使用深度可分离卷积减小计算量，但是其特征提取能力有所降低。史健婷等^[13]基于YOLOv4对输入图像进行自适应预处理以增强行人特征，但是在处理与背景重合度较高的小尺度行人以及部分被遮挡行人时效果不明显。高正中等^[14]将Transformer结构

与YOLOv5s的C3模块结合，提高了模型特征融合能力，但也增加了模型的参数量和计算量，其实时性有所降低。Zhu等^[15]针对小目标基于YOLOv5与Transformer引入了新的预测头，可以获得更丰富的语义信息，但这种额外增加预测头的设计将大大增加模型计算量以及对硬件资源的要求。YOLOv7^[16]网络采用新的标签分配方式，更加高效。YOLOv7-Tiny是YOLOv7的轻量化版本，更符合密集行人检测任务的实际部署需求。然而，YOLOv7-Tiny是通用的网络结构，参数量相对较小，但是面对密集行人检测任务中的复杂状况，难以提取有效的目标信息。

为解决上述问题，对YOLOv7-Tiny结构进行优化，设计了一种基于轻量级卷积运算与浅层特征复用的YOLOv7-TPMC网络。首先，基于部分卷积与逐点卷积设计了ELAN-PP模块，替换骨干网络中的高效层聚合网络（efficient layer aggregation network, ELAN）模块，使骨干网络参数量减少的同时保持其特征提取能力。其次，对YOLOv7-Tiny的特征金字塔结构进行改进，设计了一种跨级融合特征金字塔网络（cross-level feature-fusion pyramid network, CFPN）结构，通过浅层特征复用提高模型对复杂行人目标的敏感程度。最后，引入综合考虑预测框与标注框重叠区域的MF-IOU损失函数^[17]。

1 YOLOv7-TPMC算法

1.1 YOLOv7-TPMC结构

YOLOv7系列涵盖多个版本，YOLOv7-Tiny因具有较低的参数与计算量，在检测速度及网络复杂度方面有显著优势，避免了对边缘设备性能的高要求。

本文聚焦于实现模型复杂度与检测精度的平衡，选定YOLOv7-Tiny作为基准模型进行改进实验。YOLOv7-TPMC结构如图1所示^[18]，该网络由输入预处理、主干特征提取、颈部特征融合及检测头输出4个部分构成^[19]。图像数据首先经输入端调整至640×640分辨率，随后进入骨干网络，经历一系列ELAN-PP模块、最大池化（max pooling, MP）及ELAN模块，通过卷积激活模块（convolution batch normalization leakyrelu, CBL）的密集连接强化特征提取处理，生成涵盖浅层细节与深层语义的丰富特征图。颈部网络则采用改进特征金字塔有效融合多尺度特征，进一步精炼特征表示，最终由检测头进行多尺度目标预测。

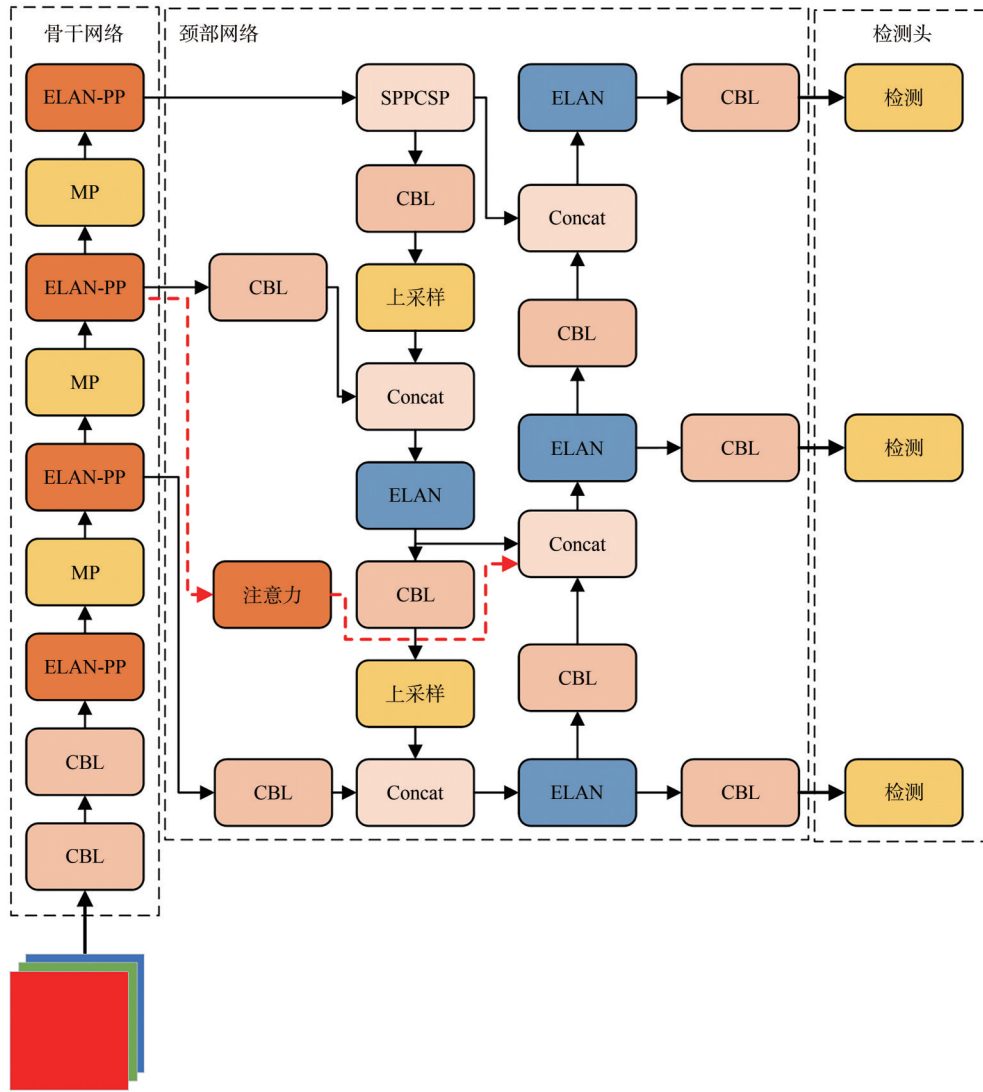


图1 YOLOv7-TPMC结构

1.2 ELAN-PP模块设计

快速网络 (faster network, FasterNet) 中引入了部分卷积技术^[20], 这一特殊卷积在处理图像特征时, 巧妙地利用了特征图内的冗余, 仅对选定部分输入通道执行标准卷积操作, 而保持其余通道不变。部分卷积通常选取特征图最前或最后的连续层, 选定输入输出通道数一致且不丧失一般性。PConv与常见卷积的计算差异如图2所示。设输入数据为 $I \in R^{c \times h \times w}$, 卷积核为 $K \in R^{k \times k}$, 输出数据为 $I \in R^{c \times h \times w}$, Conv、PConv、深度可分离卷积 (depth-wise convolution, DWConv) 计算量如式 (1) 所示:

$$\begin{cases} F_{\text{Conv}} = h \times w \times k^2 \times c^2 \\ F_{\text{PConv}} = h \times w \times k^2 \times c_n^2 \\ F_{\text{DWConv}} = h \times w \times k^2 \times c \end{cases} \quad (1)$$

由式 (1) 可知, 常规卷积 (Conv) 的计算量远高于 PConv; DWConv 通过逐通道卷积降低计算量, 但通道独立操作导致计算密度低, 在图形处理器 (graphics processing unit, GPU) 等并行设备上难以充分利用硬件资源; PConv 对部分通道执行常规卷积, 其余通道进行保留, 既减少了计算量, 又保留了较高的计算密度, 更符合 GPU 的并行计算特性。

PConv 仅对部分通道进行卷积操作, 虽然降低了计算量, 但未处理的通道可能包含重要特征。因此, 需要在 PConv 之后引入逐点卷积, 避免关键信息丢失。PConv 与 PWConv 结合后计算过程如图3所示, 用于替代 CBL 模块的常规卷积, 在减少参数数量的同时, 确保一定的模型精度。

ELAN-PP 结构如图4所示, PConv+PWConv

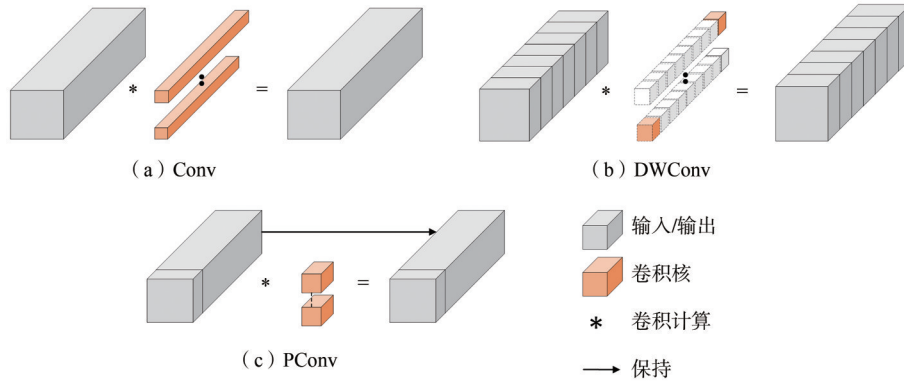


图2 Conv、DWConv和PConv计算示意图

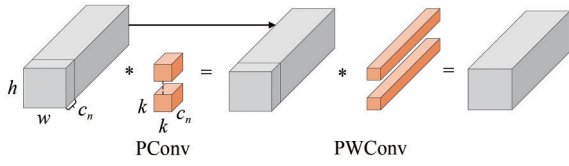


图3 PConv+PWConv计算示意图

仅对部分特征进行提取,可能会忽视某些细节,所以对单个CBL模块分支进行保留,仅用部分点卷积激活模块(partial pointwise batch normalization leaky_relu, PPBL)代替另一分支的3个CBL模块,以在减少计算量和保证每个分支的特征提取能力之间取得平衡。

1.3 优化特征金字塔结构

在YOLOv7-Tiny中,随着骨干网络层数增加,提取的特征图具有越来越丰富的语义信息,而图像细节信息逐渐减少。通过构建自底向上与自顶向下相结合的双向路径,路径聚合网络^[21](path aggregation network, PANet)有效地扩大了浅层特征图的感受野,使这些特征能够包含更丰富的上下文信息。然而,在密集行人场景中,小尺度行人在图像中的像素数量本就有限,随着特征图在自顶向下融合过程中经历上采样和逐层融合,其原本就稀疏的特征表示可能会因分辨率的降低而更不明显。尽管PANet、加权双向特征金字塔网络^[22](bi-directional

feature pyramid network, BiFPN)及渐进特征金字塔网络(asymptotic feature pyramid network, AFPN)^[23]在自然场景的大尺度和中尺度目标检测中表现出色,但它们的设计初衷和特征融合方式并不完全契合密集行人图像中小目标和多遮挡目标的检测需求。因此,在将PANet应用于密集行人检测时,需要对其进行针对性的改进和优化,以更好地适应这一特殊场景下的检测任务。对于上述问题,本文基于BiFPN结构中的跳跃连接思想,提出了CFPN结构。

1.3.1 CBAM模块

空间通道注意力模块(convolutional block attention module, CBAM)^[24]由通道注意力模块和空间注意力模块两个部分组成,其结构如图5^[25]所示。通道注意力模块压缩特征图的空间维度,保持通道维度不变,只关注特征图中的行人色彩、纹理等信息,使特征图经过并行的平均池化和最大池化,更高效地得到通道维度的权重分配,计算式为:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgP}(F)) + \text{MLP}(\text{MaxP}(F))) \quad (2)$$

其中,MLP是多层感知器, σ 是Sigmoid激活函数, F 是输入特征,AvgP是平均池化操作,MaxP是最大池化操作。空间注意力模块压缩通道维度,保持空间维度不变,关注的是行人的位置特征。

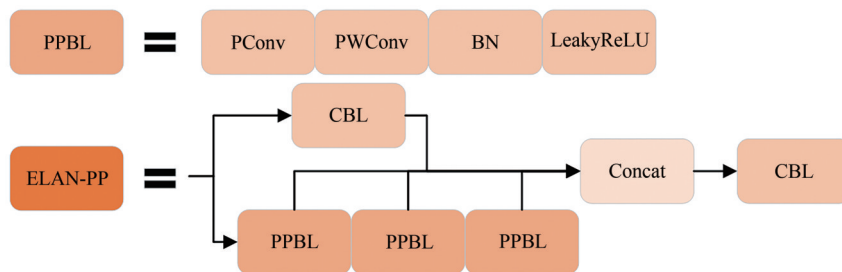


图4 ELAN-PP结构

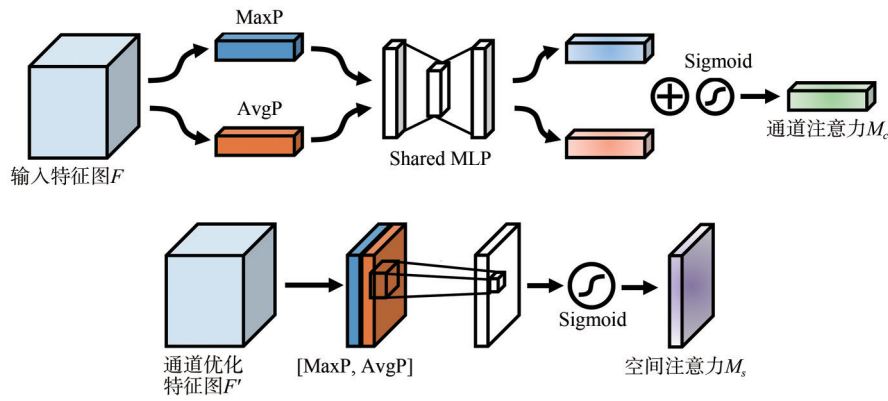


图5 CBAM结构

相较于通道注意力模块,空间注意力模块在池化操作后要与 7×7 卷积核进行计算变为1通道的特征图,其计算式为:

$$M_s(F') = \sigma(f^{7 \times 7}([AvgP(F'); MaxP(F')])) \quad (3)$$

其中, $f^{7 \times 7}$ 表示卷积操作。空间通道注意力机制聚焦于输入特征中的重要空间位置和通道,能够动态增强关键行人特征表征的能力,不仅可以提高模型的性能,还增强了模型的可解释性,使模型决策过程更加透明。

1.3.2 跨级融合特征金字塔

CFPN结构如图6所示,通过原网络中的上采样模块恢复的特征图(N_1 、 N_2),不可避免地存在语义信息和纹理信息损失的问题。这种信息损失在深层网络(P_2)融合时尤为显著,直接影响了模型对遮挡、尺度差别大或多色彩行人目标的识别精度。为了解决这一问题,CFPN将浅层网络中未进行上采样处理的特征图,以跨层级的方式直接嵌入深层网络中。此做法旨在减少由多次下采样与上采样操作带来的信息衰减,确保更多的细节信息(如边

缘、纹理等)能够直接参与到深层特征的构建中,从而提升目标边界框回归的精确度,增强模型对目标位置的精确定位能力。然而,模型实际使用过程中输入的图像质量不一,甚至中低等质量图像占比较大,其中包含的遮挡现象、边界模糊等问题,为直接跨层级融合带来了挑战^[26]。直接进行跨级融合可能将无关紧要的噪声或干扰特征一并引入深层网络,导致模型检测性能的下降。因此,在跨层级融合之前,引入CBAM注意力机制,如图6中的注意力模块所示。浅层特征图经过通道注意力处理后,行人色彩、纹理等信息比重增大,得到优化后的特征图,随后进入空间注意力模块进行处理,提高行人位置信息比重。在这个过程中,背景噪声和干扰因素得到抑制,进一步提升模型在复杂场景下的检测精度和效率。

1.4 损失函数改进

YOLOv7-Tiny中的CIoU损失函数在低质量样本对回归损失产生影响时,对优质样本难以进行进一步优化,从而导致正负样本分配失衡。针对这

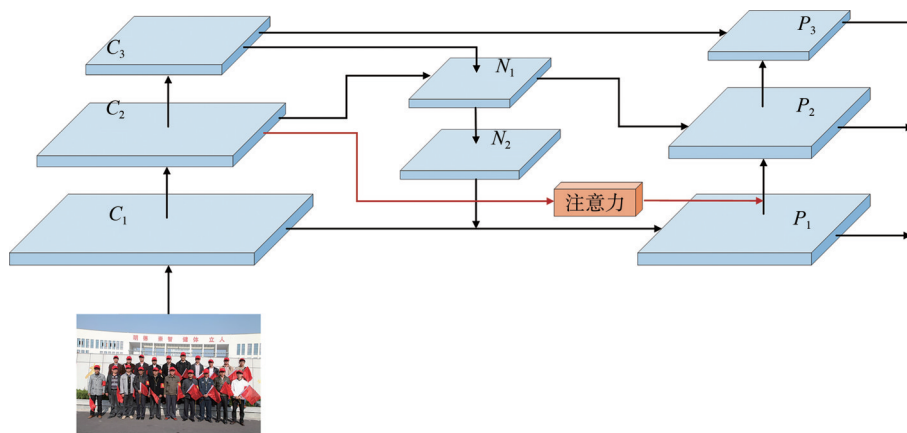


图6 CFPPN结构

一问题，引入MF-IOU替换CIOU，该损失函数由最小点距交并比^[27]（minimum point distance intersection over union, MPDIOU）与焦点交并比^[28]（focaler-intersection over union, Focaler-IOU）构成，如式（4）所示：

$$L_{\text{Focaler-MPDIOU}} = L_{\text{MPDIOU}} + \text{IOU} - \text{IOU}^{\text{Focaler}} \quad (4)$$

MPDIOU能更精细地区分相同宽高比但尺寸或位置不同的边界框，在预测框与标注框部分重叠或尺寸接近时，MPDIOU的损失计算会放大局部偏差，如式（5）~（7）所示：

$$\text{MPDIOU} = \text{IOU} - \frac{d_1^2}{h_0^2 + w_0^2} - \frac{d_2^2}{h_0^2 + w_0^2} \quad (5)$$

$$\begin{cases} d_1^2 = (x_1 - x_1^{\text{gt}})^2 + (y_1 - y_1^{\text{gt}})^2 \\ d_2^2 = (x_2 - x_2^{\text{gt}})^2 + (y_2 - y_2^{\text{gt}})^2 \end{cases} \quad (6)$$

$$L_{\text{MPDIOU}} = 1 - \text{MPDIOU} \quad (7)$$

其中， w_0 与 h_0 代表输入图片的宽高， (x_1, y_1) 与 (x_2, y_2) 代表预测框左上顶点与右下顶点坐标， $(x_1^{\text{gt}}, y_1^{\text{gt}})$ 与 $(x_2^{\text{gt}}, y_2^{\text{gt}})$ 代表标注框左上顶点与右下顶点坐标。

Focaler-IOU通过线性区间映射对IOU进行重构，具备一定难易样本自适应调节能力，通过对简单样本损失的适度抑制有效平衡模型对数据集中简单样本与困难样本的关注程度，从而改善模型在复杂场景下的收敛速度，计算方法如式（8）与（9）所示：

$$\text{IOU}^{\text{Focaler}} = \begin{cases} 0, & \text{IOU} < d \\ \frac{\text{IOU} - d}{u - d}, & d \leq \text{IOU} \leq u \\ 1, & \text{IOU} > u \end{cases} \quad (8)$$

$$L_{\text{Focaler-IOU}} = 1 - \text{IOU}^{\text{Focaler}} \quad (9)$$

其中，IOU为IOU实际值； d 和 u 是位于 $[0, 1]$ 区间的预设参数，调整 d 和 u 的值可调整模型对困难样本的关注度。

MF-IOU损失函数结合MPDIOU与Focaler-IOU两种损失函数的优势，通过综合考虑边界框的重叠区域、中心点偏移以及宽高比差异等信息，采用更高效的计算方式，无须额外增设惩罚项，简化计算流程。

2 实验验证

2.1 实验数据集与实验环境

实验在Windows系统上运行，硬件为Intel Core

i5-12400F和单张NVIDIA GTX2070s显卡，开发环境基于PyTorch 1.2.0框架，在官方版VSCODE中采用Python 3.6.13。YOLOv7-Tiny作为基准网络，批量大小（batchsize）设置为8，使用Adam优化器，训练阶段均设置了200个epoch，其余参数均设置为默认值，在后续实验中不再对超参数进行修改。

WiderPerson数据集^[29]是一个针对户外密集行人检测的数据集，共包含13 382张图像。这些图像是从广泛的场景中选择的，不再局限于传统的交通场景。因为此数据集标注未公开，所以在实验前需要对冗余图片进行处理，将数据集按照8:1:1的比例划分为训练集（7 200张图片）、验证集（900张图片）和测试集（900张图片）。CrowdHuman数据集^[30]是一个专注于行人检测的大规模数据集。该数据集包含了24 370张图像，平均每张图像有约20个行人实例。在实验中将已公开标注的19 368张图片重新分配为训练集15 494张、验证集1 938张和测试集1 936张。

2.2 评价指标

实验采用mAP@0.5、mAP@0.5:0.95、参数量与FPS作为行人检测模型的评价指标。其中，均值平均精度（mean of average precision, mAP）是一种用于评估检测模型对特定目标类别检测能力的指标，mAP@0.5是IOU=0.5时模型的平均检测精度，mAP@0.5:0.95是IOU从0.5到0.95、步长为0.05时mAP的平均值。mAP的计算过程如式（10）~式（12）所示。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

$$\text{mAP} = \int_0^1 P(R) dR \quad (12)$$

其中，TP为正样本中被标记为正样本的数量，FN为正样本中被标记为负样本的数量，FP为负样本中被标记为正样本的数量。

2.3 损失函数参数验证

为了找到比较适合密集行人检测任务的MF-IOU损失函数的调节因子参数，实验中仅替换YOLOv7-Tiny的CIOU损失函数，在WiderPerson数据集上以步长0.05设置了A~E 5个对照实验，即（ d , u ）的分别设置为（0, 0.95）、（0.05, 0.90）、（0.1,

0.85)、(0.15, 0.80) 和 (0.2, 0.75) 5组, 实验结果见表1, 调节因子对比实验结果如图7所示。

表1 调节因子对比实验

对照实验组	d	u	mAP@0.5
A	0	0.95	71.8%
B	0.05	0.90	72.1%
C	0.1	0.85	72.3%
D	0.15	0.80	72.8%
E	0.2	0.75	72.5%

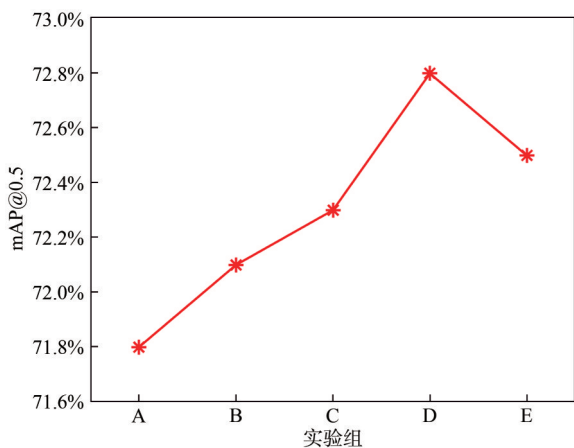


图7 调节因子对比实验结果

由表1可知, 当调节因子 (d , u) 设置为 (0.15, 0.8) 时, 损失函数对行人检测任务的适应性较好, 后续实验中均采用此参数。

2.4 损失函数对比实验

为验证引入的改进损失函数的有效性, 基于 YOLOv7-Tiny 模型, 将原模型中的 CIOU 损失函数分别替换为 MF-IOU、形状感知交并比^[31] (shape-aware intersection over union, SIOU) 和 Focaler-IOU, 保持模型其他部分不变。随后, 在 WiderPerson 数据集上进行了对比实验, 训练过程的损失变化如图8所示, 实验结果见表2。

在图8中, 替换 MF-IOU 后, 模型损失值更低。从表2可以看出, 采用 MF-IOU 定位损失函数后模型的 mAP 值上升较为明显, 证明了改进损失函数的有效性。

2.5 跨级强化特征金字塔消融实验

为验证 CFPN 对小目标行人和部分被遮挡行人检测效果的影响, 以及添加注意力机制的必要性, 在 WiderPerson 数据集上进行了 CFPN 消融实验, 基准模型为 YOLOv7-Tiny, 仅修改特征金字塔结构, 其余部分保持不变, 结果见表3。

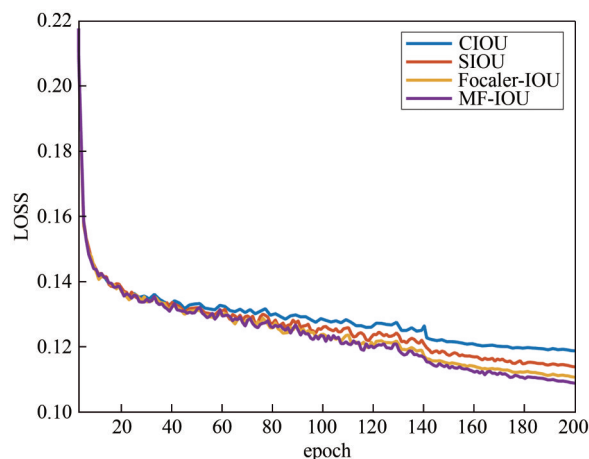


图8 训练过程的损失变化

表2 IOU 对比实验

IOU 损失函数	mAP@0.5	mAP@0.5:0.95
CIOU	72.2%	43.0%
SIOU	72.5%	43.3%
Focaler-IOU	72.7%	43.4%
MF-IOU	72.8%	43.7%

由表3可知, 仅加入跨级金字塔结构不加入 CBAM 注意力模块时相较于 YOLOv7-Tiny, 模型 mAP@0.5 提高了 0.3%, mAP@0.5:0.95 基本不变, 参数量略有上升, 对模型检测速度几乎没有影响。额外加入注意力机制后, 参数量上升并不明显, 但是检测精度得到有效提升。

表3 CFPN 消融实验

模型	mAP@0.5	mAP@0.5:0.95	参数量/ $\times 10^6$	FPS
YOLOv7-Tiny	72.2%	43.0%	6.23	98
YOLOv7-Tiny+CFPN-CBAM	72.5%	43.0%	6.26	97
YOLOv7-Tiny+CFPN	73.2%	43.3%	6.26	97

为更直观地验证注意力机制对 CFPN 结构的影响, 分别对模型使用 CFPN 前后和添加注意力机制前后的推理结果进行分析, 如图9所示。图9(a)与图9(d)是没有使用 CFPN 和 CBAM 的结果, 图9(b)与图9(e)是使用 CFPN 但不添加 CBAM 结果, 图9(c)与图9(f)是使用了 CFPN 添加 CBAM 的结果。

从图9(a)、图9(b)与图9(c)可以看出, 使用了 CFPN 而不使用 CBAM 的模型在识别部分被遮挡行人目标和部分小尺度行人有所提升, 如图9(b)左侧标记所示, 但是在该图右侧发生了信息丢失。

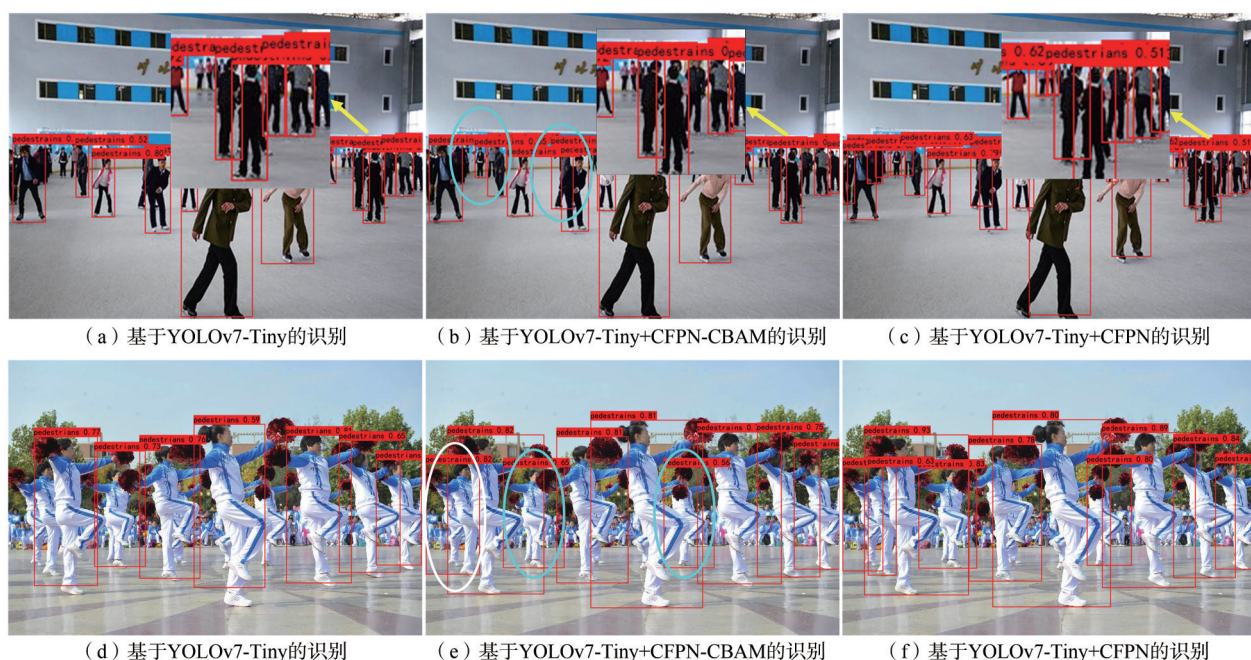


图9 CFPN消融实验效果

添加CBAM后干扰信息得到抑制,信息丢失情况有所改善,图9(c)左右两侧均识别到了更多的行人信息并且在该图标记部分,小目标行人区域获得了较多的关注;从图9(d)、图9(e)与图9(f)可以看出,由于图片主体为大尺度行人,不加入CBAM时图9(e)没有出现信息丢失,但是模型没有识别到该图左侧标记所示的被遮挡目标,在引入CBAM后,图9(f)中此目标被识别。实验结果表明,在加入CBAM后,浅层特征图的噪声和干扰得以抑制,特征信息得到增强,模型更专注于复杂行人的特征区域,能有效提升模型的感知能力和表达能力。

2.6 改进模型消融实验

为全面评估改进策略的有效性,在WiderPerson数据集和CrowdHuman数据集上进行了消融实验,实验结果见表4和表5,表中√代表该组实验采用的改进方法。

从表4的第1组中可以看出,在骨干网络引入轻量化的ELAN-PP模块后,模型参数量都有较为明显的下降,FPS值有所提升。表4的第2组与第3组实验在引入MF-IOU与CFPN后平衡了模型的检测速度与精度。由表4和表5的第6组实验可知,结合3种改进方法之后的YOLOv7-TPMC模型,相比基准模型,参数量减少了20.5%,同时mAP@0.5、mAP@0.5:0.95的值在WiderPerson数据集上分

表4 WiderPerson数据集上消融实验结果

组别	ELAN-PP	MF-IOU	CFPN	mAP@0.5	mAP@0.5:0.95	参数量/ $\times 10^6$	FPS
基准	—	—	—	72.2%	43.0%	6.23	98
1	√	—	—	71.7%	40.4%	4.93	112
2	—	√	—	72.8%	43.7%	6.23	98
3	—	—	√	73.2%	43.3%	6.26	97
4	√	√	—	72.5%	41.6%	4.93	112
5	√	—	√	72.8%	43.0%	4.95	110
6	√	√	√	73.4%	43.9%	4.95	110

表5 CrowdHuman数据集上消融实验结果

组别	ELAN-PP	MF-IOU	CFPN	mAP@0.5	mAP@0.5:0.95	参数量/ $\times 10^6$	FPS
基准	—	—	—	83.1%	52.3%	6.23	98
1	√	—	—	82.2%	49.2%	4.93	112
2	—	√	—	84.9%	54.8%	6.23	98
3	—	—	√	84.8%	54.6%	6.26	97
4	√	√	—	84.1%	51.1%	4.93	112
5	√	—	√	83.7%	50.8%	4.95	110
6	√	√	√	85.1%	54.6%	4.95	110

别增加了1.2%和0.9%,在CrowdHuman数据集上分别增加了2.0%和2.3%,FPS提升了12.2%。消融实验综合验证了3种改进方法的有效性以及YOLOv7-TPMC模型的泛化能力。

由表4与表5的第2组和第3组实验对比可知,在WiderPerson数据集上替换MF-IOU损失函数相

比于替换CFPN结构对模型的精度提升较低,在CrowdHuman数据集上则相反,这可能是由于WiderPerson数据集具有多样化的行人姿态、多尺度分布及复杂背景特性,而CrowdHuman数据集中高密度行人分布和严重遮挡场景为主要部分,这一差异导致MF-IOU在边界框的精细调整和关注难样本方面的能力不同程度地体现。

2.7 对比实验

为进一步验证改进算法的有效性,将YOLOv7-TPMC与SSD、YOLOv3-Tiny^[32]、YOLOv4-Tiny、YOLOv5s、YOLOv7-Tiny与YOLOv8n^[33]等主流算法在WiderPerson数据集与CrowdHuman数据集上进行了对比实验,实验结果见表6与表7。

表6 WiderPerson数据集上对比实验结果

模型名称	mAP@0.5	mAP@0.5:0.95	参数量/ $\times 10^6$	FPS
SSD	57.4%	29.7%	26.29	29
YOLOv3-Tiny	46.9%	23.1%	8.67	103
YOLOv4-Tiny	64.6%	35.5%	6.06	121
YOLOv5s	70.1%	40.6%	7.01	62
YOLOv7-Tiny	72.2%	43.8%	6.23	98
YOLOv8n	71.6%	43.5%	3.0	58
YOLOv7-TPMC	73.4%	43.9%	4.95	110

由表6与表7可知,YOLOv4-Tiny与YOLOv8n在两个密集行人数据集上表现出了较好的检测速度,FPS值分别达到了121与124,但是相较于YOLOv7-TPMC模型mAP@0.5与mAP@0.5:0.95的值均有所

表7 CrowdHuman数据集上对比实验结果

模型名称	mAP@0.5	mAP@0.5:0.95	参数量/ $\times 10^6$	FPS
SSD	69.8%	35.5%	26.29	29
YOLOv3-Tiny	63.1%	33.6%	8.67	103
YOLOv4-Tiny	75.7%	43.5%	6.06	121
YOLOv5s	81.9%	49.8%	7.01	62
YOLOv7-Tiny	83.1%	52.3%	6.23	98
YOLOv8n	82.0%	51.7%	3.0	124
YOLOv7-TPMC	85.1%	54.6%	4.95	110

降低。YOLOv7-TPMC相较于经典的SSD,mAP@0.5的值在WiderPerson数据集和CrowdHuman数据集上分别提升了16%、15.3%,参数量仅为其18.8%。相较于以往的轻量级YOLO,改进后的模型在准确率与模型简化方面有较好的提升。在两个数据集上的对比实验结果验证了改进方法的有效性,YOLOv7-TPMC在检测精度与检测速度之间取得了较好的平衡,更符合实际密集行人检测任务的要求。

为更加直观地展示YOLOv7-TPMC在密集行人检测方面的有效性,在WiderPerson数据集和CrowdHuman数据集的测试集中分别选取两张图片进行模型推理,推理结果如图10和图11所示。

由图10(a)与图10(b)对比可得,改进的模型实现了对单个行人的准确框选,并提高了对被遮挡行人的识别能力;图10(d)相比于图10(c)识别到了更多的小目标行人。在图11(a)图11(b)中,低光照条件下的行人特征变得更模糊,但改进后的模型对边缘特征不明显的遮挡目标也保持了良

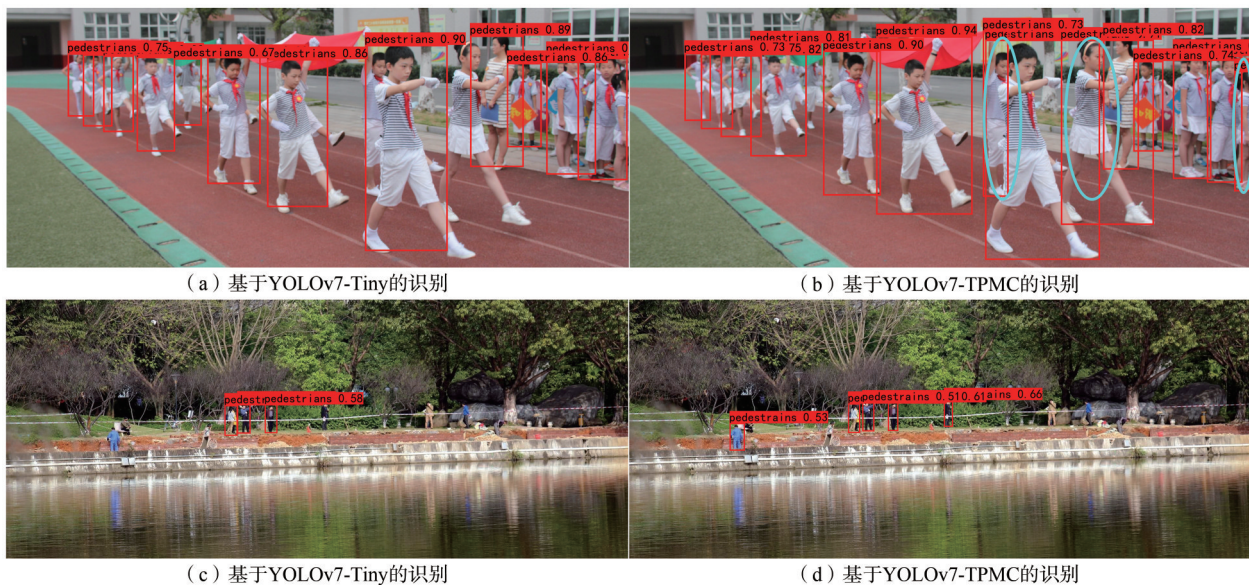


图10 WiderPerson测试集图片推理结果

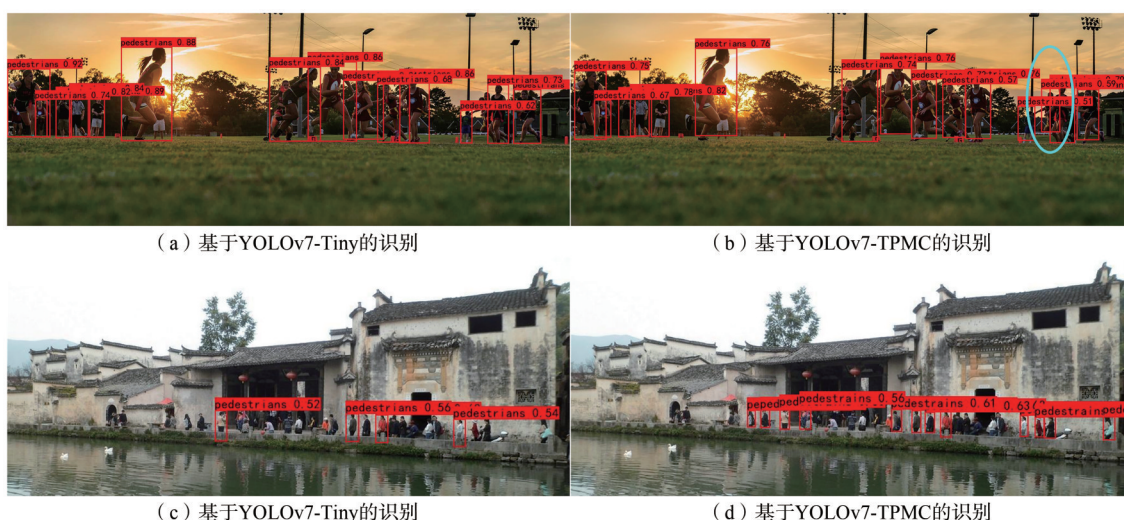


图11 CrowdHuman测试集图片推理结果

好的识别能力，验证了模型的鲁棒性；在图11(c)与图11(d)中，YOLOv7-TPMC相较于原模型更关注小尺度行人区域，模型对小尺度行人的识别效果得到提高。

通过对比4组图片可以看出，YOLOv7-TPMC在面对部分遮挡及小目标行人时，表现出较强的检测能力，验证了改进策略在密集行人场景方面的有效性。

3 结束语

本文基于YOLOv7-Tiny模型，提出了基于ELAN-PP模块、跨级融合特征金字塔与MF-IOU的YOLOv7-TPMC模型，旨在适应密集行人检测任务，降低模型的漏检、误检概率和模型复杂度。经实验验证，3种改进方法有效提升了原模型的检测性能。改进后的模型在小目标行人和部分被遮挡行人检测方面有较好的改善，模型轻量化与检测速度较原模型提升明显。实验所用WiderPerson数据集与CrowdHuman数据集，行人姿态偏向于直立行人，在行人做出弯腰及蹲起等动作时模型可能产生漏检。其他的公开数据集也都存在行人姿态不够丰富的问题，未来的研究可使用更丰富的数据集进行训练，也需要探索模型进一步轻量化的方法。

参考文献：

- [1] ZOU Z X, CHEN K Y, SHI Z W, et al. Object detection in 20 years: a survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.
- [2] 章博闻. 基于深度学习的有锚框行人检测方法综述[J]. 传感器世界, 2024, 30(1): 7-12.
- [3] ZHANG B W. A review of anchor-based pedestrian detection methods based on deep learning[J]. Sensor World, 2024, 30(1): 7-12.
- [3] HARRIS C, STEPHENS M. A combined corner and edge detector[C]//Proceedings of the Alvey Vision Conference 1988. Alvey Vision Club. United Kingdom: The Plessey Company, 1988: 147-152.
- [4] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Piscataway: IEEE Press, 2005: 886-893.
- [6] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [7] AKYON F C, ONUR ALTINUC S, TEMIZEL A. Slicing aided hyper inference and fine-tuning for small object detection[C]//Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2022: 966-970.
- [8] 李欣健, 张大胜, 孙利雷, 等. 复杂场景下基于CNN的轻量火焰检测方法[J]. 模式识别与人工智能, 2021, 34(5): 415-422.
- [8] LI X J, ZHANG D S, SUN L L, et al. CNN-based lightweight flame detection method in complex scenes[J]. Pattern Recognition and Artificial Intelligence, 2021, 34(5): 415-422.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Computer Vision-ECCV 2016: Proceedings of the 14th European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.
- [10] 苑朝, 赵亚冬, 张耀, 等. 基于YOLO轻量化的多模态行人检测算法[J]. 图学学报, 2024, 45(1): 35-46.
- [10] YUAN C, ZHAO Y D, ZHANG Y, et al. Lightweight multi-modal pedestrian detection algorithm based on YOLO[J]. Journal of Graphics, 2024, 45(1): 35-46.
- [11] HAN K, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 1577-1586.
- [12] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 122-138.
- [13] 史健婷, 张贵强, 陶金, 等. 改进的YOLOv4红外图像行人检测算法[J].

- 智能计算机与应用, 2021, 11(8): 31-34, 41.
- SHI J T, ZHANG G Q, TAO J, et al. Pedestrian detection algorithm in infrared image based on improved YOLOv4[J]. Intelligent Computer and Applications, 2021, 11(8): 31-34, 41.
- [14] 高正中, 于明沆, 孟晗, 等. 基于改进YOLOv5算法的红外图像行人目标检测[J]. 中国科技论文, 2024, 19(2): 209-214.
- GAO Z Z, YU M H, MENG H, et al. Improved YOLOv5-based infrared pedestrian target detection[J]. China Sciencepaper, 2024, 19(2): 209-214.
- [15] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway: IEEE Press, 2021: 2778-2788.
- [16] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 7464-7475.
- [17] 杨慧聪, 陈慈发, 张上. 基于通道剪枝的YOLOv8n印刷电路板缺陷检测[J]. 无线电工程, 2025, 55(2): 271-280.
- YANG H C, CHEN C F, ZHANG S. PCB defect detection using YOLOv8n based on channel pruning[J]. Radio Engineering, 2025, 55(2): 271-280.
- [18] 涂智荣, 凌海英, 李幅, 等. 基于改进YOLOv7-Tiny的轻量化百香果检测方法[J]. 广西师范大学学报(自然科学版), 2024, 42(5): 79-90.
- TU Z R, LING H Y, LI G, et al. Lightweight passion fruit detection method based on improved YOLOv7-tiny[J]. Journal of Guangxi Normal University (Natural Science Edition), 2024, 42(5): 79-90.
- [19] MA L, ZHAO L Y, WANG Z X, et al. Detection and counting of small target apples under complicated environments by using improved YOLOv7-tiny[J]. Agronomy, 2023, 13(5): 1419.
- [20] CHEN J R, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 12021-12031.
- [21] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759-8768.
- [22] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 10781-10790.
- [23] YANG G Y, LEI J, ZHU Z K, et al. AFPN: asymptotic feature pyramid network for object detection[C]//Proceedings of the 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Piscataway: IEEE Press, 2023: 2184-2189.
- [24] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the 15th European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 3-19.
- [25] 方祯祺, 李雪, 莫红. 基于改进EfficientNet的乳腺肿瘤诊断[J]. 智能科学与技术学报, 2023, 5(4): 505-514.
- FANG Z Q, LI X, MO H. Diagnostic of breast tumors based on improved EfficientNet[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(4): 505-514.
- [26] XU S B, ZHANG M H, SONG W, et al. A systematic review and analysis of deep learning-based underwater object detection[J]. Neurocomputing, 2023, 527: 204-232.
- [27] OU J Y, SHEN Y J. Underwater target detection based on improved YOLOv7 algorithm with BiFusion neck structure and MPDIoU loss function[J]. IEEE Access, 2024, 12: 105165-105177.
- [28] ZHANG H, ZHANG S J. Focaler-IoU: more focused intersection over union loss[EB]. arXiv preprint, 2024, arXiv: 2401.10525.
- [29] ZHANG S F, XIE Y L, WAN J, et al. WiderPerson: a diverse dataset for dense pedestrian detection in the wild[J]. IEEE Transactions on Multimedia, 2020, 22(2): 380-393.
- [30] HE H Y, LIN W Y, ZHANG Y A, et al. Toward accurate and robust pedestrian detection via variational inference[J]. International Journal of Computer Vision, 2025, 133(2): 867-889.
- [31] ZHANG L J, QIU Y, FAN J W, et al. Underwater fish detection and counting using image segmentation[J]. Aquaculture International, 2024, 32(4): 4799-4817.]
- [32] LI D Y, WANG G F, ZHANG Y, et al. Coal gangue detection and recognition algorithm based on deformable convolution YOLOv3[J]. IET Image Processing, 2022, 16(1): 134-144.
- [33] HE Z Y, TONG M J. LT-YOLO: a lightweight network for detecting tomato leaf diseases[J]. Computers, Materials & Continua, 2025, 82(3): 4301-4317.

[作者简介]



赵世礼 (2000-), 男, 长沙理工大学电气与信息工程学院硕士生, 主要研究方向为数字图像与机器视觉。



莫红 (1972-), 女, 长沙理工大学电气与信息工程学院教授, 主要研究方向为智慧医疗、模糊AI和复杂系统管理与控制。



杨澳男 (2000-), 男, 长沙理工大学电气与信息工程学院硕士生, 主要研究方向为数字图像与机器视觉。



欧阳玉琦 (2000-), 男, 长沙理工大学电气与信息工程学院硕士生, 主要研究方向为数字图像与机器视觉。