DOI:10.19651/j.cnki.emt.2211795

面向密集场景结合 TC-YOLOX 的小目标检测方法*

李翔宇 王 伟 王峰萍 韩岩江
(西安工程大学 西安 710048)

摘 要: 密集场景下小目标的高效精确检测是目标检测领域的关键问题。为了解决环境的多样性和小目标自身复杂性存在着特征难以提取、检测精度低等问题,提出一种面向密集场景结合 TC-YOLOX 的小目标检测方法。首先,通过在 CSPNet 中引入 Transformer Encode 模块,不断更新目标权重实现增强目标特征信息,提高网络的特征提取能力;其次,在特征金字塔网络中增加卷积注意力机制模块,关注重要特征并抑制不必要特征,提高不同尺度目标的检测准确度;然后,采用 CIoU 代替 IoU 作为回归损失函数,使得模型训练过程中网络收敛更快,性能更好;最后在 PASCAL VOC 2007 数据集上验证。实验结果表明,所设计的 TC-YOLOX 模型能够有效的检测出多样化场景中正常、密集、稀疏、黑暗条件下的小目标物体,mAP 和检测速度可以达到 94.6%和 38 fps,与原始模型相比提升了 10.9%和 1 fps,对多种密集场景下的小目标检测任务均具有较好的适用性。

关键词: 小目标检测;YOLOX;卷积注意力机制模块;Transformer Encode;CIoU 回归损失函数
中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Small target detection method for dense scenes combined with TC-YOLOX

Li Xiangyu Wang Wei Wang Fengping Han Yanjiang
(Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: Efficient and accurate detection of small targets in dense scenes is a key problem in the field of target detection. In order to solve the problems of diversity of environments and complexity of small targets, such as difficult feature extraction and low detection accuracy, a small target detection method for dense scenes combined with TC-YOLOX is proposed. Firstly, by introducing Transformer Encode module into CSPNet, the target weight is continuously updated to enhance the target feature information and improve the network feature extraction capability. Secondly, the convolutional attention mechanism module is added to the feature pyramid network to focus on important features and suppress unnecessary features, so as to improve the detection accuracy of targets of different scales. Then, CIoU is used to replace IoU as the regression loss function, which makes the network converge faster and has better performance in the process of model training. Finally, it is verified on PASCAL VOC 2007 dataset. The experimental results show that the designed TC-YOLOX model can effectively detect small target objects under normal, dense, sparse and dark conditions in diversified scenes. The mAP and detection speed can reach 94.6% and 38 fps, which is 10.9% and 1 fps higher than the original model. It has good applicability to small target detection tasks in multiple dense scenes.

Keywords: small object detection; YOLOX; CBAM; Transformer Encode; CIoU regression loss function

0 引 言

密集场景下小目标检测在计算机视觉领域中应用非常广泛,可应用于工业生产、卫星遥感、目标跟踪等领域。然而,在这些密集场景中都存在一系列共同的特点,如目标繁多且易互相遮挡、目标可能出现在图片任何位置且尺寸动态变化、同一类物体在不同图片中的角度和姿态不同等问题,这都为目标检测带来极大的挑战[1]。

传统的目标检测[2]分为 3 个步骤:1)区域选择,预先设置不同的尺度,不同的长宽比,然后采用滑动窗口的方法对

收稿日期:2022-10-23
* 基金项目:2021 年中国高校产学研创新基金(2021ALA02002)、2021 年"纺织之光"中国纺织工业联合会高等教育教学改革研究项目(2021BKJGLX004)、西安工程大学 2020 年高等教育研究项目(20GJ05)资助

整幅图像进行遍历；2）特征提取，常用的算法有 Haar[3]、SIFT[4]、HOG[5] 等；3）分类，分类算法有 SVM[6]、Adaboost[7]等。传统方法用于密集场景的目标检测简单易行，但没有针对性选择滑动窗口区域的策略导致学习效率较低，人工设计的特征对于多样化环境没有很好的鲁棒性，适应性较弱。

近年来，随着 GPU 算力的不断增强和人工智能技术的迅速发展，深度学习方法在计算机视觉任务中表现优异[8]。基于深度学习的目标检测方法分为两阶段检测和单阶段检测。两阶段检测由候选区域提取、候选区域分类和候选区域坐标修正 3 个步骤组成。典型的两阶段检测算法有 R-CNN[9]、Fast R-CNN[10]、Faster R-CNN[11]。单阶段检测算法是通过一个神经网络模型可直接输出检测结果。典型的单阶段检测算法有 YOLO[12] 系列和 SSD[13]。YOLO 结构简单、计算效率高，能够方便地进行端到端的训练，在实时目标检测领域中有很大的应用潜力。

目前，许多研究者针对小目标检测存在的问题进行深入的研究。Zhang 等[14]提出 YOLSO 目标检测算法，由于小目标外观信息提出不足和周围背景干扰量大，无法合理表达小目标的特征，为了解决这个问题，提出了 HSSC 和 FPE 模块来改善小目标的特征信息，但是和主流模型相比并没有取得较大的领先。徐晓光等[15]提出利用多尺度特征融合的方法来解决 YOLO 定位不精准的问题，但是主干网络深度比较浅，精度提升不是很大。Qu 等[16]提出了结合空洞卷积与特征融合，通过融合高分辨率的低级特征图和语义信息丰富的高级特征图，在一定程度上提升了遥感小目标的检测效果。Redmon 等[17]提出 YOLO 检测算法，在检测精度和速度上都取得了不错的效果并且迁移性很强，可以灵活的运用到其他新的领域。但对于两个非常靠近的目标以及很小的群体检测效果不是很好。Benjumea 等[18]提出一种小目标检测的算法 YOLO-Z，该算法使用 ResNet50[19]替换原始网络，并且将 PANet[20] 替换成 Bi-FPN[21]，提高整体的检测准确度。但 YOLO-Z 算法只能在特定的场景下有较好的效果，泛化能力较差，主要应用在自动驾驶的场景。Fang 等[22]提出一种基于 ViT 改进的 YOLOS 算法，该算法在图片上添加 100 个可以学习的 DET token 用于目标检测，并且将 ViT 中的图像分类损失替换为二分匹配损失，按照 DETR[23] 的预测方式去进行目标检测，实现 2D 目标检测可以以纯序列到序列的方式完成。但 YOLOS 是为了在目标检测中更好地显示 Transformer 编码器[24]的可行性，和其他模型相比 YOLOS 模型的性能提升并不是很大，而且 YOLOS 模型需要庞大的数据集进行训练才能达到比较好的检测效果。

基于此，针对环境的多样性和小目标自身复杂性存在着特征难以提取、检测精度低等问题，提出一种面向密集场景结合 TC-YOLOX 的小目标检测方法。主要贡献如下：1）在 CSPNet 中引入编码器（transformer encode，TE）模块，更好地提取特征信息，增加网络的特征提取能力；2）将卷积注意力机制模块（convolutional block attention module，CBAM）集成到特征融合网络，便于在特征融合后的特征图像中找到感兴趣的区域。3）使用 CIoU 回归损失函数代替 IoU 回归损失函数，使得预测框更加符合真实框，在训练过程中收敛更快，性能更好。在 PASCAL VOC 2007 数据集上进行实验，实验结果表明，改进后的模型在密集场景下 mAP 有明显提升，具有较高的鲁棒性和适用性。

1　YOLOX 算法

YOLOX[25]结合 YOLO 系列的部分经验形成一个的高性能目标检测模型，主要包括主干特征提取网络（Backbone）、特征金字塔网络（feature pyramid networks，FPN）或路径聚合网络（path aggregation network，PANet）、预测头（prediction head），其结构如图 1 所示。

网络深度和网络宽度的不同，YOLOX 有 4 种不同型号的 YOLO 模型在 COCO 数据集中性能对比，如表 1 所示。

1.1　Backbone

YOLOX 使用 CSPDarknet53 作为其主干网络，它具有 5 个重要特点：1）残差网络，缓解网络中增加深度带来的梯度消失问题；2）CSPNet 网络结构，分为左右两部分：主干部分进行残差块的堆叠，另一部分进行少量的处理直接连接到最后；3）Focus 网络结构可以减少下采样带来的信息缺失，并且减少网络层数和参数；4）SiLU[26]激活函数，具备无上界有下界、平滑、非单调的特性，在深层模型上的效果优于 ReLU；5）SPP 结构，提高网络的感受野。

1.2　PANet

PANet 是目标检测框架中的关键环节，可以更好地利用 Backbone 提取出的特征进行特征融合。从主干网络提取出的 3 个有效特征层进入 PANet 进行特征融合，这样可以融合不同尺度的特征信息。通常，特征金字塔是从下到上进行上采样、特征融合、特征提取，再从上到下进行下采样、特征融合、特征提取，最后会输出 3 个有效特征层。虽然特征金字塔能提高一些准确度，但是在检测小目标的时候还是不尽人意。

1.3　Prediction head

Prediction head 就是将 PANet 特征融合后的 3 个有效特征层进行判断，在 YOLOX 中，Prediction head 把分类和回归分成两部分，每一个特征层可以获取 3 个预测结果，分别是特征点的回归参数，特征点是否包含物体，判断特征点包含的物品种类，最后将这 3 个预测结果融合到一起。通过 Prediction head 得到每个特征层的预测框的位置和类别，再对预测框进行 NMS[27]筛选得到最终的框。

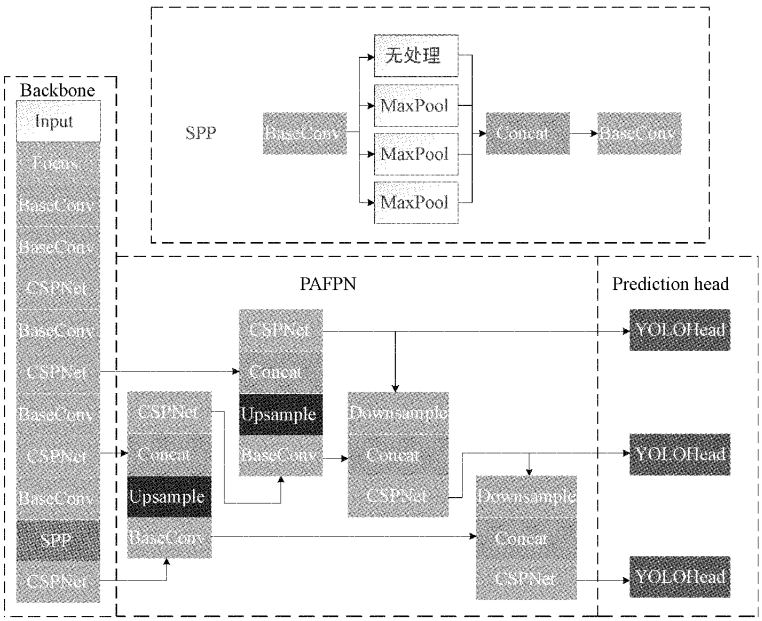


图 1 YOLOX 网络结构

表 1 不同型号的 YOLO 模型性能对比

模型	YOLOX_s	YOLOX_m	YOLOX_l	YOLOX_x
Size	640×640	640×640	640×640	640×640
mAP0.5:0.95	40.5	47.2	50.1	51.5
Speed/ms	9.8	12.3	14.5	17.3
Params/M	9.0	25.3	54.2	99.1

2 面向密集场景的小目标检测

2.1 TC-YOLOX 网络模型

为了解决密集场景下的小目标检测存在的特征难以提取、检测精度低等问题,设计了面向密集场景的 TC-YOLOX 小目标检测模型,如图 2 所示。

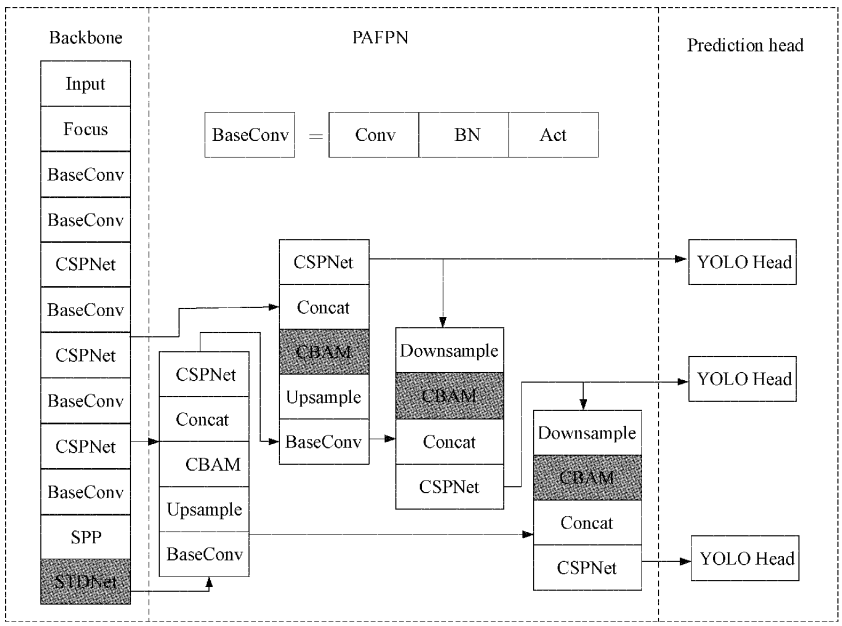


图 2 面向密集场景的 TC-YOLOX 小目标检测模型

该模型主要包括以下 3 个部分:

1)在 CSPNet 中引入 TE 模块并将其使用在 Backbone 中,提高模型获取全局信息和丰富的上下文信息的能力。

2)将 CBAM 集成到 YOLOX 中的 PANet,帮助网络在具有大区域覆盖的图像中找到感兴趣的区域,能够在密集场景中精确定位对象,提高检测的准确度。

3)采用 CIoU 回归损失函数,解决重叠面积、中心点距离和长宽比等问题,使得预测框更加符合真实框,在训练过程中收敛更快,性能更好。

2.2　基于 STDNet 的特征提取

密集场景中的大多数目标尺寸较小,传统特征提取模块很难充分的提取到目标特征,导致模型的小目标检测性能不理想[28]。为了提高模型的特征提取能力,在原模型的 backbone 中引入了 STDNet 模块。所设计的 STDNet 模块由卷积层、残差块和 TE 模块组成的。与原模型的主干特征提取模块相比,其可以充分的获取全局信息和丰富的上下文信息,STDNet 模块的结构如图 3 所示。

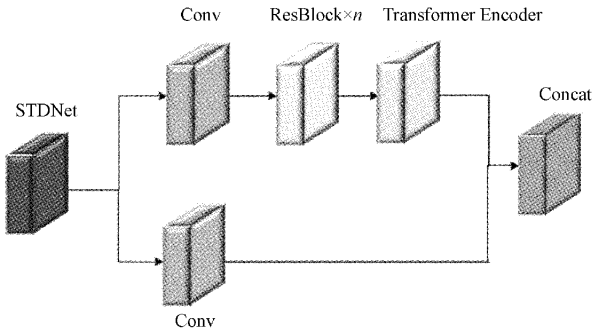


图 3　STDNet 模块结构

TE 模块结构如图 4 所示,包含多头自注意力(multi-head self-attention,MHS)和全连接层两个子层。每个子层之间使用残差连接,LayerNorm 和 Dropout 层有助于网络更好的融合,防止网络过拟合。TE 模块提高获取不同特征信息的能力,它还可以利用 MHS 不断更新目标权重实现增强特征提取的信息。

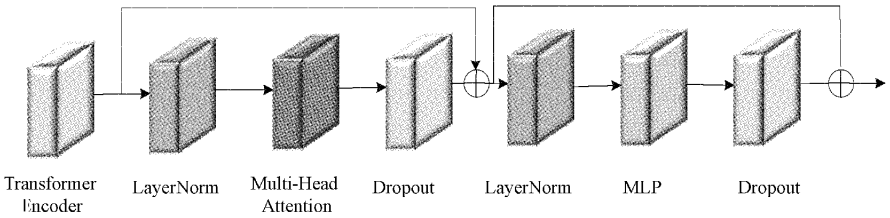


图 4　TE 模块结构

MHS 的原理如图 5 所示,通过把一张完整的图片分成几个 *patches*,利用 *LayerNorm* 得到 $x_1,x_2,x_3,\cdots,x_T$,然后分别使用 3 个不同矩阵 $\boldsymbol{W}^Q,\boldsymbol{W}^K,\boldsymbol{W}^V$ 分别与之相乘,得到 $q_i,k_i,v_i,i\in(1,2,3,\cdots,T)$。利用 q_1 分别与 $k_1,k_2,k_3,\cdots,k_T$ 做向量点积,再通过 Softmax 进行处理,使其的注意力权重值均在 0～1,得到 $H_{11},H_{12},H_{13},\cdots,H_{1T}$。$\boldsymbol{Q}=(q_1,q_2,q_3,\cdots,q_T),\boldsymbol{K}=(k_1,k_2,k_3,\cdots,k_T),\boldsymbol{V}=(v_1,v_2,v_3,\cdots,v_T)$,如式(1)所示。

$$head_i=Attention(\boldsymbol{QW}_i^Q,\boldsymbol{KW}_i^K,\boldsymbol{VW}_i^V) \tag{1}$$

将上一步得到的值继续与对应位置的 $v_1,v_2,v_3,\cdots,v_T$ 相乘得到与输入的 x_1 所对应的输出 y_1,$\sqrt{d_k}$ 为缩放因子:

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=softmax\left(\frac{\boldsymbol{QK}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{2}$$

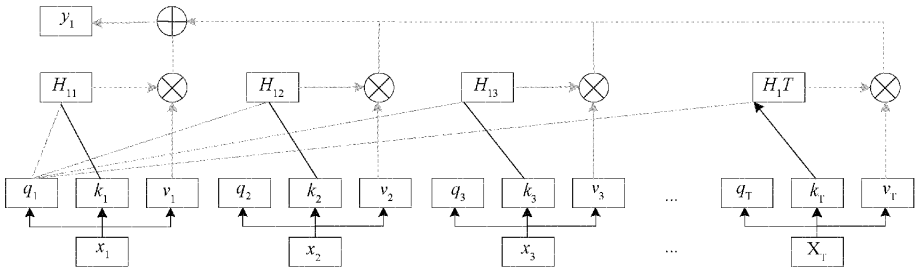


图 5　MHS 原理

其他的 $q_2,q_3,\cdots,q_T$ 按以上的步骤以此类推,得到 $y_1,y_2,y_3,\cdots,y_T$,最后求和输出结果。如下:

$$MultiHead(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=Concat(y_1,y_2,y_3,\cdots,y_T) \tag{3}$$

2.3　结合 CBAM 的 PANet

多尺度特征提取不充足等原因导致在小目标检测过程中效果不是很理想。CBAM 能够快速、准确地分析复杂场景信息,关注重要特征并抑制不必要特征,有效的提取注意力区域的特征,帮助 YOLOX 抵抗混乱的信息,因此将 CBAM[29]集成到 PANet 中,可以充分的提取多尺度特征,如图 6 所示。

CBAM 是一个简单且有效的轻量级注意力模块。当输入一个单个特征图,CBAM 沿着通道和空间两个独立的维度依次推断注意图,然后将注意图与输入特征图相乘,以执行自适应特征细化。CBAM 模块的结构如图 7 所示。

CBAM 由两部分组成:通道注意力机制(channel attention module,CAM)和空间注意力机制(spatial attention module,SAM)。CAM 是将特征图在空间维度上

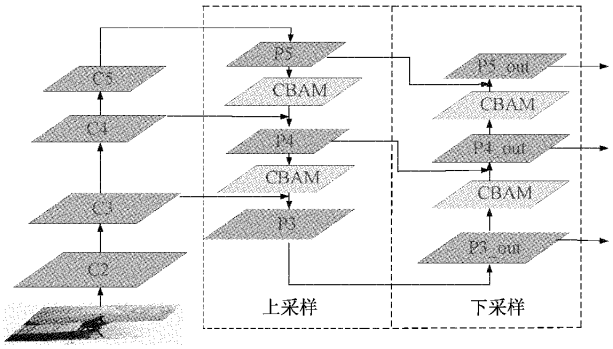


图 6 结合 CBAM 的 PANet

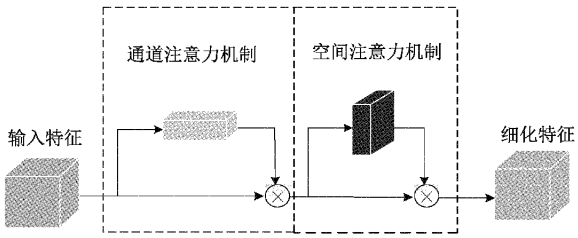


图 7 CBAM 结构

进行压缩,得到一个一维矢量后再进行操作。SAM 是对通道进行压缩,在通道维度分别进行平均值池化和最大值池化。CBAM 效果明显比单一的 CAM 或者单一的 SAM 好很多。

2.4 损失函数

IoU 有两个缺点:1)当两个框不相交时,IoU=0,Loss=1−IoU=1,除非网络要进行多轮迭代,才有可能两个框有交集;2)Loss 仅仅和交并比有关。这时,需要考虑非重合区域占比指标问题,并且增加一些限制指标。因此采用 CIoU 代替 IoU 作为回归损失函数。

CIoU 损失函数通过对角线距离把检测框和预测框的中心距离归一化来改善非重合区域占比问题并且增加长宽比的限制项,使得提高准确度。CIoU 公式如式(4)所示。

$$R_{CIoU}=\frac{\rho^2(B,B^{gt})}{c^2}+\alpha v \tag{4}$$

其中,B、B^{gt} 分别代表预测框和真实框的中心点,ρ 是两个中心点之间的欧氏距离,c 是能够同时包含预测框和真实框的最小闭包区域的对角线如图 8 所示。

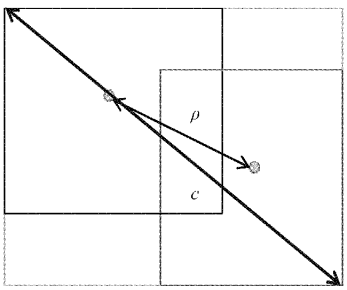


图 8 B、B^{gt} 两点欧氏距离和最小闭包区域对角线

v 是衡量长宽比一致性的如式(5)所示。

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right)^2 \tag{5}$$

IoU 计算公式如式(6)所示。

$$IoU=\frac{|B\cap B^{gt}|}{|B\cup B^{gt}|} \tag{6}$$

α 通过 v 和 IoU 所示计算出的参数公式:

$$\alpha=\frac{v}{(1-IoU)+v} \tag{7}$$

最后,$CIoU$[30]回归损失函数为:

$$L_{CIoU}=1-IoU+\frac{\rho^2(B,B^{gt})}{c^2}+\alpha v \tag{8}$$

3 实验结果分析

3.1 数据集介绍

PASCAL VOC 2007 公开数据集,可以用于构建和评估图像分类、目标检测和图像分割的算法。该数据集包含 9 963 张标注过的图片,共标注出 24 640 个物体,共有 20 类。实验将数据集按照 9∶1 划分为训练集和验证集。

3.2 实验环境

实验采用 8 G 显存的 3070Ti Laptop、Window11 操作系统、Python 3.6、torch 1.7.1+cu101。在 PASCAL VOC 2007 数据集上训练 300 个 epochs。采用的是随机梯度下降(SGD)进行训练,动量参数为 0.9。根据判断设置的 batch_size,自适应调整学习率,初始的 lr=0.01,并且采用余弦衰减,权重衰减系数为 0.000 5。

3.3 评价指标

实验使用平均精度(average precision,AP)、平均精度均值(mean average precision,mAP)和帧率(frame per second,FPS)检验模型的有效性。

AP 以准确率(Percision)为纵坐标,以召回率(Recall)为横坐标画出的曲线,再通过积分的方式求这个曲线的面积计算 AP。具体公式为:

$$\text{Precision}=\frac{TP}{TP+FP} \tag{9}$$

$$\text{Recall}=\frac{TP}{TP+FN} \tag{10}$$

$$AP=\int_0^1 p(r)\mathrm{d}r \tag{11}$$

其中,TP 是正确预测框的数量,FP 是错误预测框的数量,FN 是没有检测到的预测框数量。

$$\text{mAP}=\frac{\sum_{i=1}^{j}AP_i}{j} \tag{12}$$

其中,j 是总类数。

3.4 实验结果展示

为了证明 TC-YOLOX 模型的检测效果,进行了大量的实验,并且设计了一个目标检测系统对本文进行说明。

选取了车辆、船舶、行人等多样化场景，验证所提出模型在正常、密集、稀疏、黑暗等条件下的检测效果。

目标检测系统如图 9 所示，首先上传所需要检测的图片，其次通过拉动 IoU 和 FPS 按钮调整参数会有不同效果，IoU 越小，目标的识别率就越高，FPS 越大，检测是速度就越快，最后点击检测按钮就可以完成检测任务。

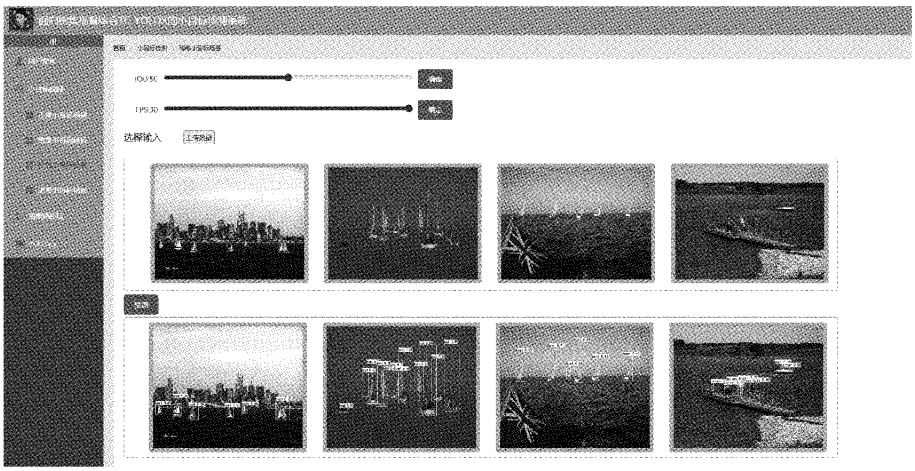
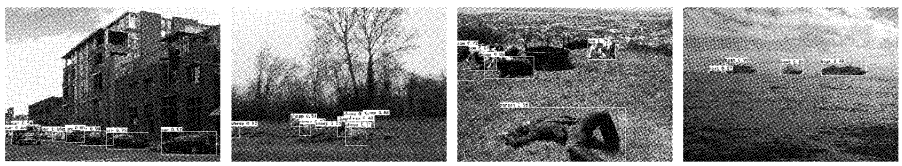
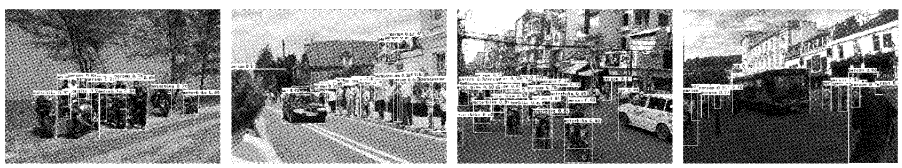


图 9　目标检测系统

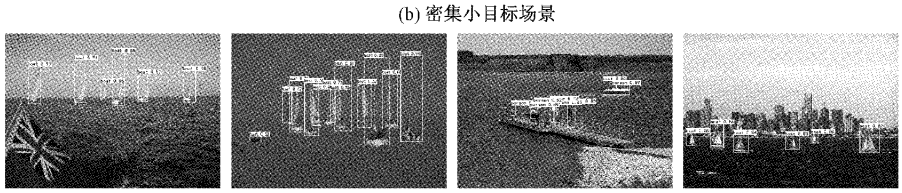
如图 10 所示，可以看出在不同场景下，图 10(a)中远处的车、羊和船等，均可高效的检测出来；图 10(b)中在密集街道的场景中行人和骑摩托的人高度重叠，依然可以准确的检测出人和摩托车；图 10(c)中帆船的帆虽然很长，但可以很准确的确定预测框位置；图 10(d)中无论在明亮还是黑暗的环境下，小目标检测依旧很准确。由此可见，本文模型在不同场景下的小目标检测均取得了良好的检测结果。



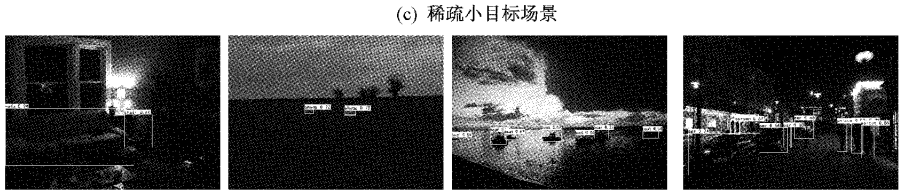
(a) 正常小目标场景



(b) 密集小目标场景



(c) 稀疏小目标场景



(d) 黑暗小目标场景

图 10　不同场景下的小目标检测结果

3.5　对比实验

为了验证 STDNet 模块在模型中的效果，在 YOLOv5s 和 YOLOX-s 相同位置中添加 STDNet 模块进行对比。如表 2 所示，添加 STDNet 模块的模型每秒浮点运算次数(FLOPs)增加，计算量变大，但是在 mAP 指标上均高于 YOLOv5 和 YOLOX，能达到 86.1%。

表 2 不同模型添加 STDNet 性能对比

模型	mAP/%	FLOPs/10^9
YOLOv5	75.1	17.1
YOLOv5+STD	76.2	19.6
YOLOX	83.7	26.8
YOLOX+STD	86.1	29.3

在主流模型 YOLOv5 和 YOLOX 相同位置中添加改进的特征融合网络进行对比,如表 3 所示。可以看出,对不同模型添加 CBAM,其浮点运算次数和参数量几乎没有变化,在不增加开销的情况下提升网络的检测性能,相比较原模型 mAP 提高 5.4%。

表 3 不同模型添加 CBAM 性能对比

模型	mAP/%	FLOPs/10^9	Params/M
YOLOv5	75.1	17.1	7.3
YOLOv5+CBAM	79.0	17.1	7.3
YOLOX	83.7	26.8	8.9
YOLOX+CBAM	89.1	26.8	8.9

TC-YOLOX 模型和目前主流的目标检测算法进行对比如表 4 所示,可以看到 YOLOv6 参数量和浮点数运算较大,YOLOv5s 和 YOLOv6 帧率较低。与此同时,TC-YOLOX 模型虽然没有从各方面都优于其他主流模型,但是在保证 FPS 不受影响的同时有效的提高目标的准确度。由于使用 CBAM 和 STDNet 模块导致模型大小增加 3.3 M、FPS 提升 2 帧,模型的 mAP 提升 10.9%。TC-YOLOX 模型在 mAP 值、帧率和参数量都优于 YOLOv5s 和 YOLOv6 等主流模型,且检测速度能够达到实时检测的标准。

表 4 TC-YOLOX 模型与其他主流模型性能对比

模型	mAP/%	FPS	Params/M	FLOPs/10^9
YOLOv5s	75.1	28	7.3	17.1
YOLOX-s	83.7	37	8.9	26.8
YOLOv6[31]	94.6	32	17.2	44.2
PP-YOLOE-s	89.6	62	7.9	17.4
TC-YOLOX	94.6	38	12.2	29.3

TC-YOLOX 模型与 YOLOv5 和 YOLOX 等主流模型相比较不同类别的 AP 值如表 5 所示,TC-YOLOX 模型的 AP 提升的比较明显,20 种类别的目标检测精度均有一定提高,尤其是 bus 这个类别已经达到 99.22%的 AP 值。虽然 pottedplant 的 AP 值是最低的,但是精度是提升最明显的,精度提升约 25%。总体看,在 PASCAL VOC 2007 公开数据集中,TC-YOLOX 模型相比较原模型在识别准确度方面表现的更出色。

表 5 不同模型不同类别的 AP 对比

模型	plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sheep	sofa	train	tv	mAP
YOLOv5	88.9	80.7	77.6	59.5	77.0	87.5	74.0	84.9	60.6	76.3	50.8	81.3	82.2	82.4	89.7	42.9	77.6	59.2	86.8	80.8	75.0
YOLOX	93.7	88.9	86.7	74.5	67.1	96.6	89.1	85.5	71.3	88.3	62.5	87.8	95.7	89.3	92.1	58.0	90.1	81.6	90.6	85.5	83.7
本文	98.7	95.6	94.9	94.2	90.9	99.2	95.0	98.3	84.9	96.5	92.4	98.5	97.8	94.4	95.3	83.0	96.3	94.1	95.9	96.1	94.6

如图 11 所示,通过折线图更加直观的看出改进的 YOLOX 算法与其他目标检测算法在 PASCAL VOC 2007 测试数据集上的每一类的准确度的变化。由此可见,TC-YOLOX 算法在处理密集小目标检测任务中具有很大的优势,其精度取得了很大的提升。

为了验证模型的有效性,从测试集中选取部分图片,

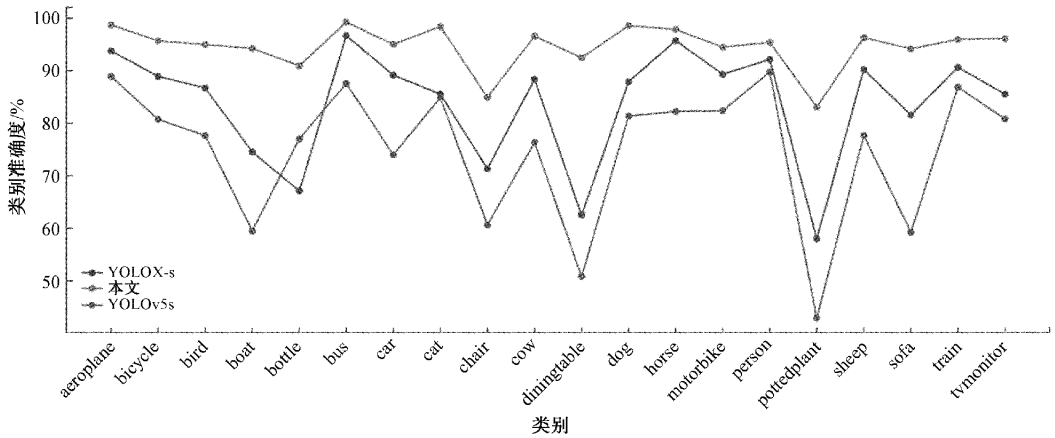
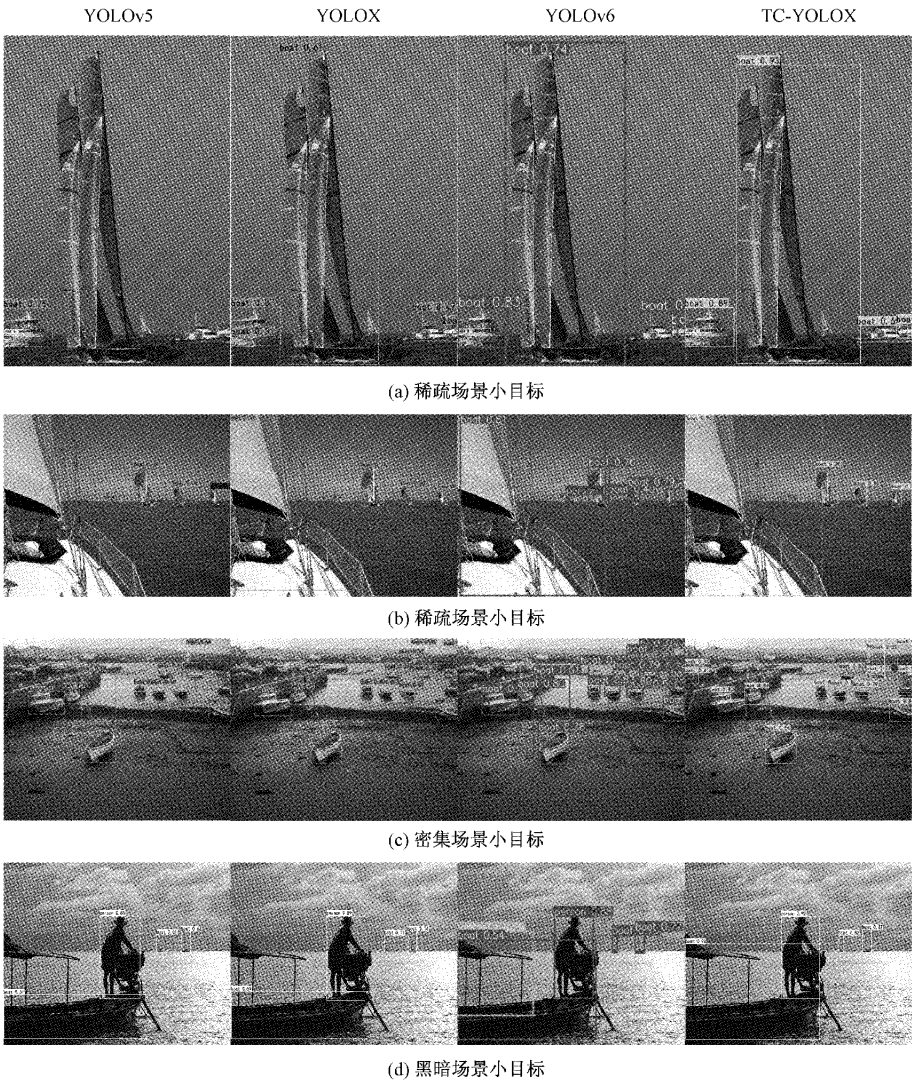


图 11 各模型不同类别 AP 的变化

分别使用 YOLOv5s、YOLOX-s、YOLOv6 和 TC-YOLOX 模型进行小目标检测。不同场景下不同模型的小目标检测结果对比如图 12 所示，如图 12(a)所示可以看出中 YOLOv5 效果是有些不理想的，其他模型准确度相对较低，如图 12(b)所示中 YOLOv5 和 YOLOv6 都出现分类错误，如图 12(c)所示中小目标识别率相对较低，如图 12(d)所示中预测框不是很准确。TC-YOLOX 相比较主流模型具有较好的检测效果，在稀疏场景小目标检测中分类准确且准确率高，在密集场景小目标检测中相比其他模型能够识别更多的小目标，在黑暗场景小目标检测预测框更加精准。与原模型相比，改进后的模型在准确度和识别率均有提升，同时预测框更加精准。



(a) 稀疏场景小目标

(b) 稀疏场景小目标

(c) 密集场景小目标

(d) 黑暗场景小目标

图 12　不同模型不同场景下的小目标检测

通过对比实验结果可以看出，在密集小目标场景下，TC-YOLOX 模型在小目标检测方面的性能高于其他主流模型，与原始模型 YOLOX 相比，对小目标的检测效果更加优秀，提升小目标检测性能，且对于小目标检测准确度提升明显，同时对预测框不准确等问题有所改善，更适合在密集小目标场景下进行应用。

3.6　消融实验

通过消融实验，验证 TC-YOLOX 中各模块的有效性，在 PASCAL VOC 2007 数据集上做消融实验，以 YOLOX 为基础进行对比，以 mAP、推理时间、FPS、模型体积和 FLOPs 作为评价指标。消融实验结果如表 6 所示。消融实验结果表明，TC-YOLOX 模型的 FPS 和推理时间几乎没有变化的情况下，保证实时检测能力，mAP 提升 10.9% 且推理时间仅为 1.48 ms。在使用 STDNet 模块以后使得模型增加对目标重要特征的关注，mAP 提升 2.4%，但是模型的参数量增加 3.1 M；使用 CBAM 对特征融合网络进行改进，减少特征损失改善网络对小目标的检测能力，在保证参数量、FPS 和推理时间基本没有变化的情况下，mAP 提升 5.4%；替换 CIoU 损失函数后，mAP 较使用 IoU 时候相比提升 2%，并且回归的预测框位置也十分的

表 6　消融实验

模型	STDNet	CBAM	CIoU	mAP%	t/ms	FPS	Params/M	FLOPs/10^9
YOLOX-s	×	×	×	83.7	1.47	37	8.9	26.8
改进方法 1	√	×	×	86.1	1.48	36	12.0	29.3
改进方法 2	×	√	×	89.1	1.47	37	9.0	26.8
改进方法 3	×	×	√	85.7	1.47	37	9.0	26.8
TC-YOLOX	√	√	√	94.6	1.48	38	12.2	29.3

准确,网络模型收敛能力更好。

如图 13 所示不同模型在训练过程中 mAP 的变化曲线和如图 14 所示不同模型在训练过程中损失函数的变化曲线,横坐标为训练的轮次,纵坐标为 IoU 为 0.5 时的 mAP 和 loss 值。模型在训练初期,准确度平稳上升,损失函数下降的较为平稳。在 180 轮后,准确度提升明显,损失函数断崖式下降。随着模型不断改进,模型的收敛速度和识别的准确度都有明显的提升,改进网络结构在准确度提升显著,而损失函数变化不大,改完损失函数后,损失函数的数值下降明显,准确度也得到提升。

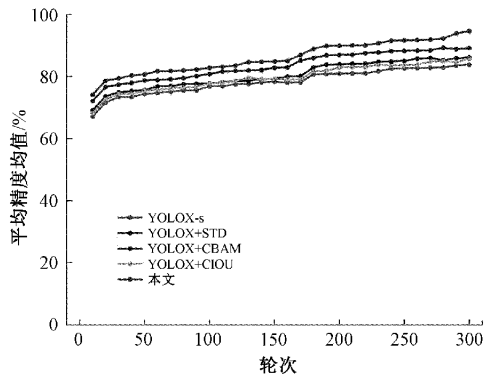


图 13　不同模型训练过程中 mAP 变化曲线

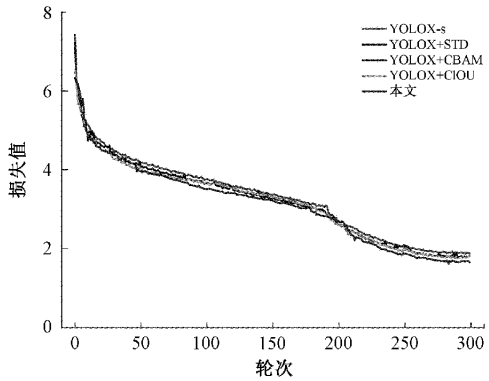


图 14　不同模型训练过程中损失函数变化曲线

4　结　　论

本文提出一种面向密集场景结合 TC-YOLOX 的小目标检测方法。为了解决环境的多样性和小目标自身复杂性存在着特征难以提取、检测精度低等问题,引入 TE 模块对 CSPNet 进行改进,提出 STDNet 模块,可以更加充分的提取特征信息,增强重要特征,削弱一般特征;在 PANet 中添加 CBAM 模块,特征再提取时可以更好的提取感兴趣的目标;使用 CIoU 代替 IoU 回归损失函数,提高预测框的准确度,使得网络收敛更快,性能更好。

TC-YOLOX 与 YOLOv5、YOLOX 和 YOLOv6 进行对比实验,实验结果表明,STDNet 的提取能力相比较 CSPNet 有一定的提升,增加 CBAM 的 PANet 增强检测小目标能力,替换成 CIoU 回归损失函数使预测框和真实框更加匹配。TC-YOLOX 可以有效的提高小目标的识别率,大大提高检测的准确度,检测速率变化不大,能够满足小目标检测。本实验的检测平均准确度达到 94.6%,检测速率达到 37.8 帧。

本文方法有效的提高小目标的准确度和识别率,实验结果表现也比较良好,但由于对原始图像清晰度要求较高,针对一些场景检测效果不稳定等问题,未来将会对模型进一步优化和改进。

参考文献

[1]　彭豪, 李晓明. 基于改进 Faster R-CNN 的小目标检测模型[J]. 电子测量技术,2021,44(24):122-127.

[2]　方路平, 何杭江, 周国民. 目标检测算法研究综述[J]. 计算机工程与应用, 2018, 54(13): 11-18.

[3]　AI J, TIAN R, LUO Q, et al. Multi-scale rotation-invariant Haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(12): 10070-10087.

[4]　GUPTA S, THAKUR K, KUMAR M. 2D-human face recognition using SIFT and SURF descriptors of face's feature regions[J]. The Visual Computer, 2021, 37(3): 447-456.

[5]　WANG Y, ZHU X, WU B. Automatic detection of individual oil palm trees from UAV images using HOG features and an SVM classifier[J]. International Journal of Remote Sensing, 2019, 40(19): 7356-7370.

[6]　XIAO J. SVM and KNN ensemble learning for traffic incident detection [J]. Physica A: Statistical Mechanics and its Applications, 2019, 517: 29-35.

[7]　WANG W, SUN D. The improved AdaBoost

algorithms for imbalanced data classification [J]. Information Sciences, 2021, 563: 358-374.
[8] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
[9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
[10] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
[11] CAO C, WANG B, ZHANG W, et al. An improved faster R-CNN for small object detection[J]. IEEE Access, 2019, 7: 106838-106846.
[12] 李鹏飞,刘瑶,李珣,等. YOLO9000 模型的车辆多目标视频检测系统研究[J]. 计算机测量与控制,2019,27(8):21-24.
[13] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
[14] ZHANG J, ZHANG L, LIU T, et al. YOLSO: You only look small object [J]. Journal of Visual Communication and Image Representation, 2021, 81: 103348.
[15] 徐晓光,李海. 多尺度特征在 YOLO 算法中的应用研究[J]. 电子测量与仪器学报,2021, 35,246(6):96-101.
[16] QU J, SU C, ZHANG Z, et al. Dilated convolution and feature fusion SSD network for small object detection in remote sensing images[J]. IEEE Access, 2020, 8: 82832-82843.
[17] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
[18] BENJUMEA A, TEETI I, CUZZOLIN F, et al. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles [J]. ArXiv Preprint,2021, ArXiv:2112.11798.
[19] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
[20] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
[21] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
[22] FANG Y, LIAO B, WANG X, et al. You only look at one sequence: Rethinking transformer in vision through object detection [J]. Advances in Neural Information Processing Systems, 2021, 34: 26183-26197.
[23] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision. Springer, Cham, 2020: 213-229.
[24] BEAL J, KIM E, TZENG E, et al. Toward transformer-based object detection [J]. ArXiv Preprint,2020, ArXiv:2012.09958.
[25] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [J]. ArXiv Preprint, 2021, ArXiv:2107.08430.
[26] ELFWING S, UCHIBE E, DOYA K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. Neural Networks, 2018, 107: 3-11.
[27] LIU S, HUANG D, WANG Y. Adaptive NMS: Refining pedestrian detection in a crowd [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 6459-6468.
[28] 李壮飞, 杨风暴, 郝岳强. 一种基于残差网络优化的航拍小目标检测算法[J]. 国外电子测量技术, 2022, 41(8):27-33.
[29] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 3-19.
[30] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
[31] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv Preprint,2022, ArXiv:2209.02976.

作者简介

李翔宇(通信作者),硕士研究生,主要研究方向为计算机视觉、目标检测。
E-mail:749848734@qq.com

王伟,博士,副教授,主要研究方向为网络信息安全、网络智能化应用。
E-mail:174456430@qq.com

王峰萍,博士,讲师,主要研究方向为图形图像处理、人工智能与模式识别、交通信息处理。
E-mail:467202822@qq.com

韩岩江,硕士研究生,主要研究方向为目标检测。
E-mail:2621232368@qq.com