http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0003

基于改进 YOLOv5-s 的交通场景小目标检测算法

王坤*，冯康威
(中国民航大学 电子信息与自动化学院，天津 300300)

摘 要：针对交通标志和交通灯等交通场景小目标特征不明显导致检测困难的问题，提出基于改进 YOLOv5-s 的交通场景小目标检测算法。设计特征补充模块 (FSM)，通过进一步获取浅层细节信息对相邻的深层检测层进行特征补充，有效提高了小目标的检测效果，并通过相邻层间的矩阵运算避免了特征冗余；设计有效融合模块 (EFM)，分别处理特征金字塔融合时的横向浅层特征和上采样特征，缓解二者之间的特征冲突，使其更有效的融合；提出超级增强交并比 (SEIOU) 损失计算方式，通过添加真实框和预测框主对角之间的距离度量，改善回归效果，提升检测精度。在 CCTSDB、S2TLD、TLD 和 PASCAL VOC 数据集上进行实验，结果表明：所提算法在精度上分别提升了 2.54%、3.62%、4.33% 和 2.01%，检测速度达到了 113 帧/s，适用于实际交通场景下的检测任务。

关 键 词：YOLOv5-s 算法；小目标检测；特征补充；特征融合；损失函数
中图分类号：TP391.4
文献标志码：A **文章编号**：1001-5965(2026)04-1015-13

随着科学技术的不断进步，近年来自动驾驶技术飞速发展，使人们的出行、生活方式更加智能化。自动驾驶的核心可以概括为感知、规划和控制 3 个部分，感知是指自动驾驶系统从环境中收集提取相关知识的能力，而环境感知特指对于环境的场景理解能力[1]。为确保对环境的理解和把握，车辆需要获取大量的周围环境信息，在自动驾驶时做出合理行为规划。交通标志和交通灯是自动驾驶过程中的重要环境信息，此类场景目标的检测是高级驾驶辅助系统的重要组成部分。

传统的交通标志和交通灯检测算法主要根据目标的颜色和形状等特征进行处理，易受到光照、车辆尾灯等因素干扰[2-3]。相比于传统的检测算法，基于深度学习的目标检测算法具有更好的实时性和准确性，可以帮助自动驾驶车辆识别交通场景中的交通标志、交通灯等环境信息，通过对信息进行分析从而自主地控制车辆实现自动驾驶。目前，基于深度学习的目标检测方法主要分为 2 类。一类是两阶段检测算法，将候选框提取与分类任务分开，其代表算法主要有快速区域卷积神经网络(faster region-convolutional neural networks, Faster R-CNN)算法[4]、掩码 R-CNN(Mask R-CNN)算法[5]、空间金字塔池化网络(spatial pyramid pooling-networks, SPP-Net)算法[6] 等。另一类是一阶段检测算法，将目标的检测与分类任务以端到端方式完成，其代表算法主要有 YOLO 系列[7-9] 检测算法、一阶段单次多边框检测器 (single shot multibox detector, SSD) 算法[10] 等。

交通标志和交通灯的检测数据有其自身的特点，车载摄像头在采集自然场景下的检测图像时，会受到道路环境、天气、透视变化等因素的影响，此外，交通标志和交通灯的尺度在原始图像中占

收稿日期：2024-01-03；**录用日期**：2024-02-25；**网络出版时间**：2024-03-12 20:21
网络出版地址：link.cnki.net/urlid/11.2625.V.20240312.1540.004
基金项目：国家自然科学基金（62173331）
*通信作者. E-mail：yogo_w@163.com

引用格式：王坤，冯康威. 基于改进 YOLOv5-s 的交通场景小目标检测算法 [J]. 北京航空航天大学学报，2026，52（4）：1015-1027.
WANG K，FENG K W. Small target detection algorithm for traffic scenes based on improved YOLOv5-s [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（4）：1015-1027（in Chinese）.

比较小，特征信息较弱，具有天然的细粒度特征，从而使实时检测更具挑战性。在现有的交通标志和交通灯检测的研究中，井方科等[11]针对小目标交通标志容易误检和漏检的问题，设计一种双向自适应特征金字塔网络（feature pyramid networks, FPN），引入跳跃连接和自适应特征融合因子，增强多尺度特征融合，提升小目标检测精度。钱伍等[12]设计了记忆性特征融合网络，高效利用了高级语义信息和底层特征，增加模型对小目标的学习能力。Yao等[13]提出了一种基于自适应特征金字塔网络的特征融合方法，能够更好地融合骨干网络2个尺度特征层的输出结果，融合后的特征具有更多的语义信息和位置信息。Shi等[14]针对交通标志小目标的检测，考虑到主干网络下采样会导致低级特征信息融合不够，小目标部分特征信息丢失，提出了一种参数量少、特征融合能力强的密集路径聚合网络（dense path aggregation networks, Dense-PAN），可以有效地融合浅层细节信息和深层语义信息。上述文献提出的特征融合方法一定程度上提高了模型的检测精度，解决了小目标信息缺失，特征难区分等问题，但不同的检测通道对应的目标尺度不同，不同检测层在融合过程中不可避免存在一定程度的特征冲突，在融合过程中存在特征冗余并影响检测效果。此外，在基于YOLO系列检测算法的相关研究中，通常采取增加浅层检测通道的方法提高小目标的检测能力，Li等[15]和Mou等[16]都采用该方法来对特征融合网络进行扩展。但新增浅层小目标检测通道，需要重新聚类生成对应的锚框，同时，在原来基础上继续上、下采样操作融合浅层特征，最后，添加对应的检测头执行分类回归操作，这势必会大幅度增加计算量和网络复杂程度。

针对上述问题，本文提出了基于改进YOLOv5-s的交通场景小目标检测算法，通过改进特征融合结构和融合方式，以轻微的参数量增长为代价，解决了检测通道中细节特征不足和下采样过程中的信息丢失问题，并且避免了融合过程中的层间特征冲突问题，提高了小目标的检测精度。本文主要创新有：

1）在YOLOv5-s特征融合结构的基础上，提出特征补充模块（feature supplement module, FSM），该模块通过提取浅层的丰富细节特征，对相邻的深层特征层进行特征补充，有效的弥补了深层网络细节特征的不足，完善了小目标检测机制。此外通过对相邻2层进行矩阵乘法运算突出关键信息，避免对深层造成特征冗余。

2）在FPN的层间融合部分，设计了有效融合模块（effective fusion module, EFM），对来自主干的横向浅层细节特征和来自深层的上采样特征进行处理，缓解融合时的层间特征冲突，使深层抽象语义信息与浅层细粒度信息更有效融合。

3）提出超级增强交并比（super enhanced intersection over union, SEIOU）损失计算方式，解决了完全交并比（complete intersection over union, CIOU）长宽比的模糊定义，并添加真实框和预测框的左上角及右下角之间的距离度量，综合了2框之间的距离和长宽关系，有效避免特殊情况下惩罚项失效的问题，使网络有更快的收敛速度和更好的定位结果。

4）考虑到检测的快速性，本文在推理阶段将卷积层（Conv）与批量归一化（batch normalization, BN）层相融合，融合后的Conv包含BN层的特性，在不损失精度的情况下，推理速度得到显著提升。

1 YOLOv5原理

YOLOv5的结构可以分为主干网络、颈部及检测头3个部分。YOLOv5采用CSPDarknet53[17]作为主干特征提取网络，对目标图像进行特征提取，提取到的特征称作特征层，在主干部分定义像素尺寸为80×80、40×40和20×20大小的3个特征层为有效特征层，提供不同尺度的特征信息。颈部作为特征融合部分，对来自主干的3个有效特征层的多尺度信息进行融合，路径聚合网络（path aggregation networks, PANet）[18]结构先经上采样操作传递深层整体特征和语义信息，之后进行下采样传递浅层细节信息，同时结合横向连接最终实现不同层级特征间的有效融合。其中，横向特征和上、下采样特征采用拼接（Concat）方式进行融合，对特征矩阵在通道维度上直接叠加，增加描述图像本身的特征数。通过主干网络提取特征并进行特征融合后获得3个加强的有效特征层，每个特征层都有宽、高和通道数，将特征图看作多个特征点的集合，每个特征点都有通道数个特征，最后，通过检测头（Head）对特征点进行决策，判断特征点是否包含物体，并通过回归参数对先验信息进行调整从而获取最终的检测结果。YOLOv5的整体网络结构如图1所示，其中，SPP表示空间金字塔池化，CBS模块包含卷积层、批归一化层和Silu激活函数。

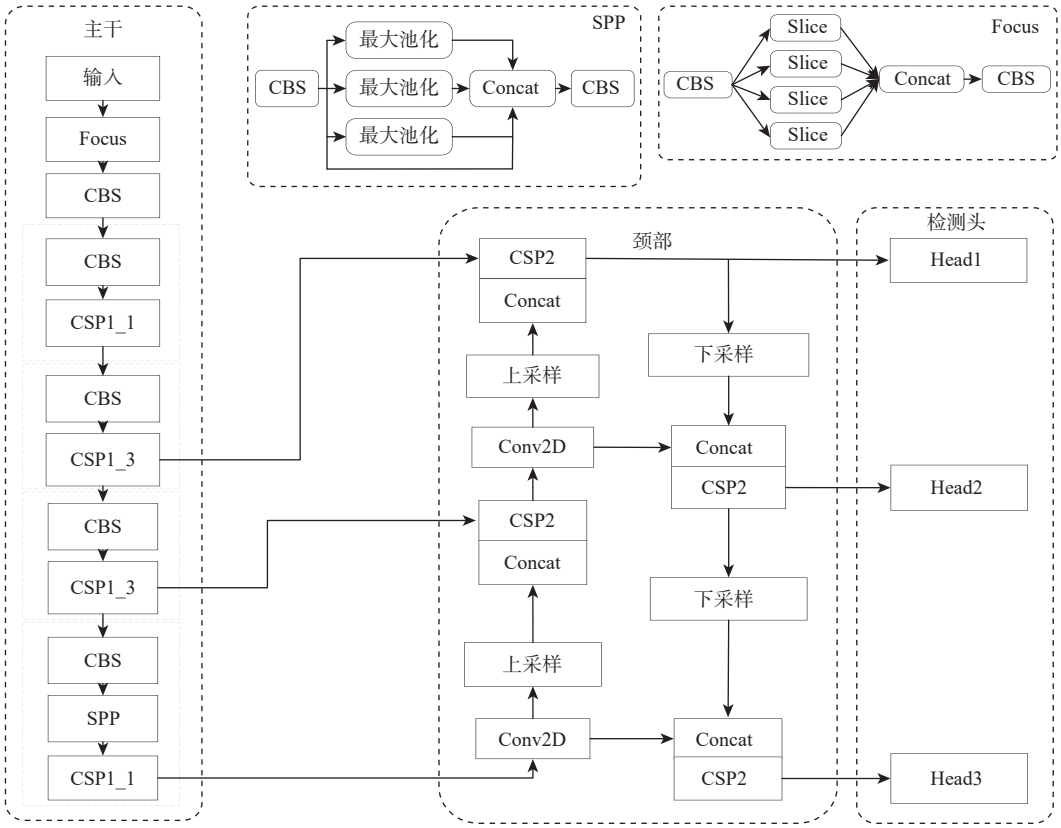


图 1　YOLOv5 整体网络结构

Fig. 1　Overall networks structure of YOLOv5

2　交通场景小目标检测算法

本文在比较网络检测的准确性、效率和模型大小的基础上, 考虑到 YOLOv5-s 网络结构简单, 易于训练, 方便部署, 有着广泛的应用背景, 从而选取 YOLOv5-s 作为基准网络来完成交通场景小目标的检测任务。针对具体的检测需求, 在 YOLOv5-s 基础上提出 FSM、EFM、SEIOU 等改进方案, 改进后的网络结构如图 2 所示, 其中, $\boldsymbol{F}_1$、$\boldsymbol{F}_2$、$\boldsymbol{F}_3$为特征输出。YOLOv5-s 的颈部仅通过 PANet 对来自主干的 3 个尺度的特征信息进行融合, 改进后的网络在此基础上添加 FSM, 不仅从主干网络更浅层获取更多的细节信息, 而且通过感受野注意力模块 (receptive field attention module, RFAM) 对图像特征进行加权处理, 扩大感受野, 获取更多的上下文信息。此外, 采用了浅层信息对相邻深层进行特征补充的融合策略, 弥补细节特征的不足, 并通过矩阵运算避免了融合带来的特征冗余。EFM 是对 PANet 融合深层上采样特征和横向浅层特征时 Concat 的替换。Concat 仅在通道维度上对特征矩阵直接叠加, 增加描述图像的特征数, 没有考虑不同层级间特征存在冲突, 采用 EFM 模块对上采样特征和横向浅层特征进一步处理能缓解这一问题, 使二者更有效融合。

2.1　特征补充模块

本文主要针对交通标志和交通灯等道路场景小目标进行检测, 由于小目标具有分辨率低、视觉信息少的特点, 相比大目标, 小目标的检测需要更多的浅层细节特征。考虑到从更浅层获取附加细节特征能实现更多尺度的特征融合, 此外, 空洞卷积可以扩大小目标感受野, 获取更多的上下文特征作为额外信息来指导小目标的检测, 因此, 在 YOLOv5-s 特征融合的基础上提出 FSM 补充浅层细节特征和上下文特征来改善小目标检测效果。首先, 从图 2 的 D_2 部分开始引出有效特征层以获取更浅层细粒度特征; 然后, 经如图 3 所示的 RFAM 进行处理, 将图像特征抽象为注意权值添加到原始的空间和信道特征中, 并将处理后的加权特征图和原始特征输入$\boldsymbol{F}_{\text{in}}$相加以保留更多的细粒度信息, 之后将特征输出$\boldsymbol{F}_{\text{out}}$经下采样操作后与相邻深层进行特征融合, 使小目标的特征信息得到补充。

对于 RFAM, 给定$\boldsymbol{F}_{\text{in}} \in \mathbf{R}^{C\times H\times W}$作为输入, RFAM 依次推出通道注意力 M_{C}和空间注意力 M_{S}, 整个过程描述为

$$\boldsymbol{F}_{\text{C}} = M_{\text{C}}(\boldsymbol{F}_{\text{in}}) \otimes \boldsymbol{F}_{\text{in}} \tag{1}$$

$$\boldsymbol{F}_{\text{S}} = M_{\text{S}}(\boldsymbol{F}_{\text{C}}) \otimes \boldsymbol{F}_{\text{C}} \tag{2}$$

$$\boldsymbol{F}_{\text{out}} = \boldsymbol{F}_{\text{in}} \oplus \boldsymbol{F}_{\text{S}} \tag{3}$$

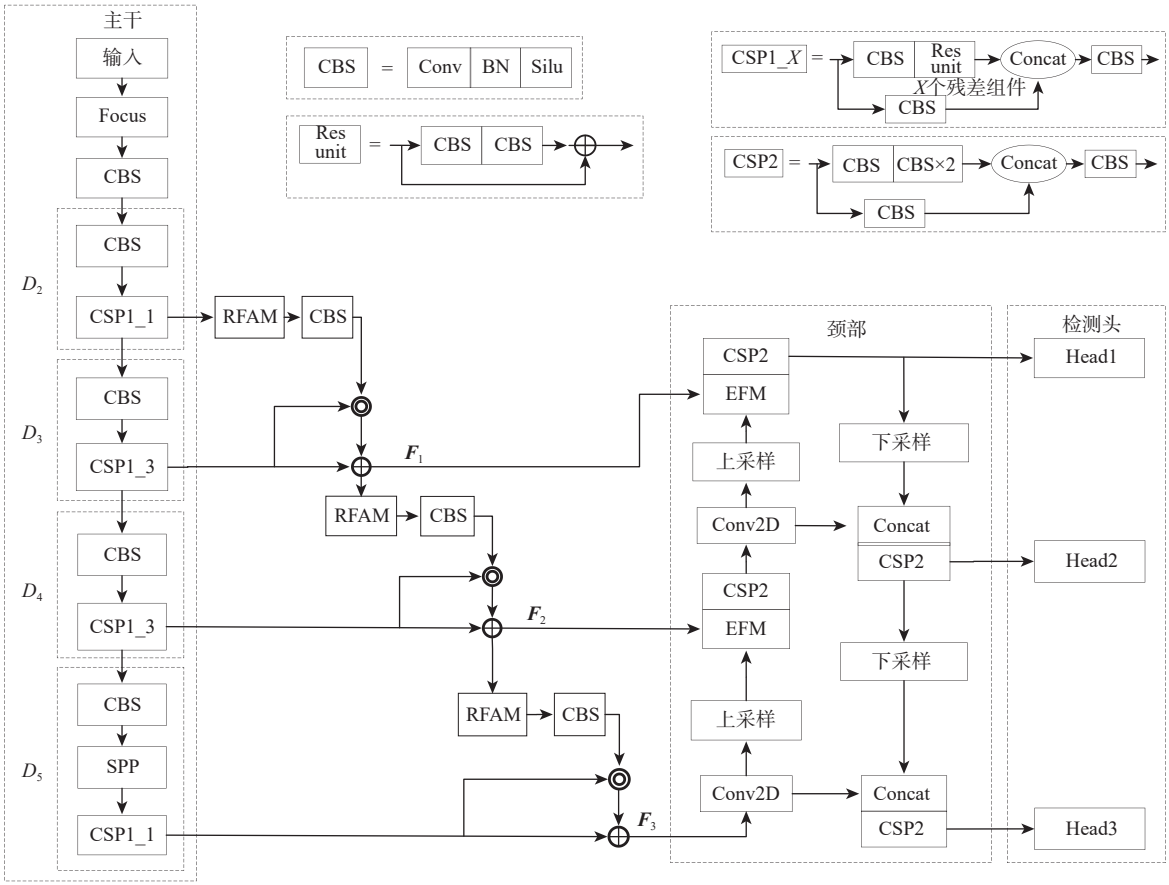


图 2 交通场景小目标检测网络

Fig. 2 Small target detection networks in traffic scenario

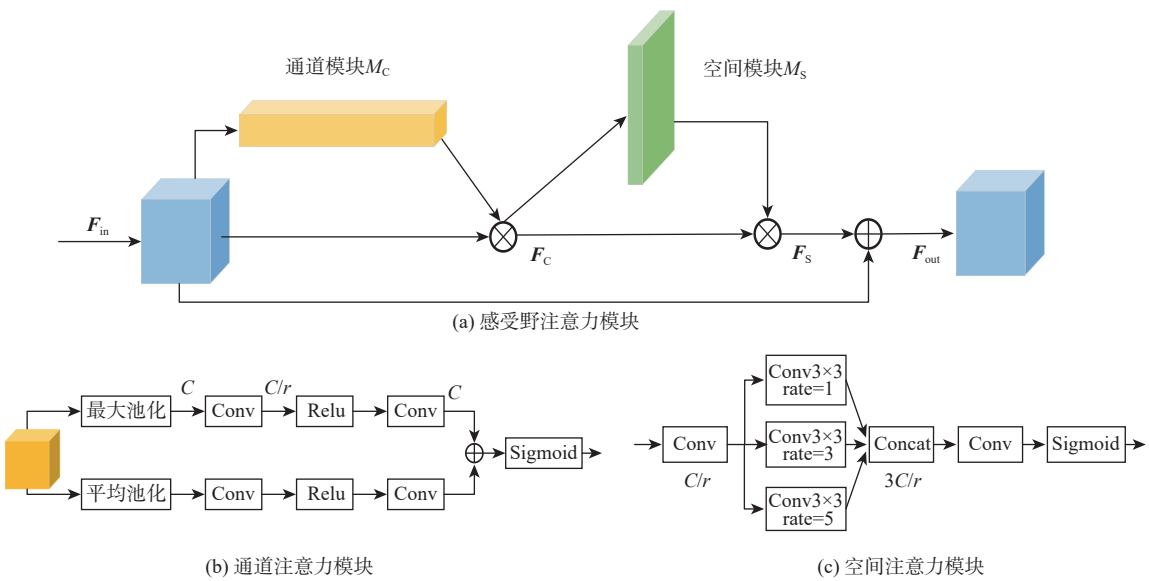


图 3 感受野注意力模块

Fig. 3 Receptive field attention module

式中：⊗表示基于元素的乘法；$\boldsymbol{F}_C \in \mathbf{R}^{C\times H\times W}$和$\boldsymbol{F}_S \in \mathbf{R}^{C\times H\times W}$分别为通道注意力和空间注意力的输出；⊕表示基于元素的加法。

图 3 中，M_C首先通过自适应平均池化和自适应最大池化聚合特征信息，沿着空间维度压缩特征，生成$1\times1\times C$的特征图$\boldsymbol{F}^C_{\text{avg}}$和$\boldsymbol{F}^C_{\text{max}}$，将二维特征通道变成具有全局感受野的实数；其次，经卷积操作先对通道降维到C/r，其中，r为简约比；接着，再经卷积扩展回C通道，降低网络计算量的同时增加了非线性能力；最后，通过逐元素求和合并两特征图，再经 Sigmoid 函数得到 0～1 之间的权重系数，即每个特征通道的重要性，通过乘法逐通道加权到之前的

特征上完成通道维度上对原始特征的重标定。M_C部分计算式为

$$M_C(\boldsymbol{F}_{in}) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(\boldsymbol{F}_{in})) \oplus \mathrm{MLP}(\mathrm{MaxPool}(\boldsymbol{F}_{in}))) \tag{4}$$

式中: σ为 Sigmoid 函数; $\mathrm{AvgPool}(\boldsymbol{F}_{in})$和$\mathrm{MaxPool}(\boldsymbol{F}_{in})$得到平均池化和最大池化的特征$\boldsymbol{F}^C_{avg}$和$\boldsymbol{F}^C_{max}$; MLP 表示通道降维并恢复的过程。

M_S部分为进一步保留图像特征, 首先, 对输入特征图$\boldsymbol{F}_{in} \in \mathbf{R}^{C\times H\times W}$采用简约比$r$进行通道降维; 然后, 通过空洞卷积模块对特征图的空间权重进行学习; 最后, 经1×1卷积恢复通道数。空洞卷积有助于构建更有效的空间特征图, 更好地利用上下文信息, 在浅层特征中通过并行不同膨胀系数且公约数不大于 1 的空洞卷积扩大浅层小目标感受野, 学习注意分数来决定具有多个感受野的通道重要性, 保证在接下来的运算中以更少的噪声将浅层细节特征传递给相邻层, 以提高小目标的检测效果。M_S部分计算式为

$$M_S(\boldsymbol{F}_C) = \sigma(f^{1\times1}(f^{d}(f^{1\times1}(\boldsymbol{F}_C)))) \tag{5}$$

式中: $f^{1\times1}$表示1×1卷积运算; f^{d}表示空洞卷积模块。

在常规的认知里, 特征融合能增强决策的预期目标特征, 但融合后的特征不仅包括期望, 还包括复杂背景噪声。浅层特征与深层特征融合时需要的仅仅是底层细粒度信息对相邻深层特征进行细节补充, 为避免融合带来的特征冗余, 将经 RFAM 处理过的浅层特征经卷积下采样后与相邻层进行矩阵乘法计算, 抑制浅层带来的背景噪声, 突出浅层关键信息对深层进行特征补充, 计算过程为

$$\boldsymbol{f}_i = \boldsymbol{F}_i \odot \boldsymbol{F}_{i-1} \tag{6}$$

式中: $\boldsymbol{F}_i$表示从主干网络中得到的第 i 个有效特征层, i=1,2,3; ⊙为哈达码积, 将计算后的结果$\boldsymbol{f}_i$赋给相邻深层通道$\boldsymbol{F}_i$, 如下:

$$\boldsymbol{F}_i = \boldsymbol{F}_i \oplus \boldsymbol{f}_i \tag{7}$$

对主干网络引出的有效特征层进行该处理, 依次得到 3 个特征输出$\boldsymbol{F}_1$、$\boldsymbol{F}_2$、$\boldsymbol{F}_3$作为新的有效特征层传入 FPN 进行后续特征融合操作。处理后的特征层具有更丰富的细节特征和上下文信息, 使得小目标检测性能得到有效提升。

2.2　有效融合模块

考虑到深层特征表示的更多是目标的整体特征和语义信息, 而浅层特征代表了目标的细节信息, 随着网络层的不断加深, 一方面网络获取更好的场景特征, 有助于区分目标与背景噪音, 另一方面, 在下采样的过程中会不断丢失细节信息, 不利于小目标的检测。因此, 在金字塔特征融合结构中, 将深层特征上采样后与横向浅层特征进行 Concat 操作, 逐层融合底层细节和深度语义信息, 实现不同层间的特征集成。然而考虑到不同检测层对应不同尺度的目标, 不难理解层间的特征融合会存在一定的信息冲突, 因此, 引入 EFM 模块对深层特征上采样过程中的 Concat 过程进行替换, 通过对上采样特征和横向浅层特征进一步处理, 合理权衡特征之间的关系以实现更有效融合的目的, 整体结构如图 4 所示。

在金字塔上采样带来深层语义信息的同时, 为避免深层的整体特征冗余对来自主干的横向浅层细节特征造成抑制, 采用图 4 中的第 2 部分, 在上采样时对深层特征进行处理。若将目标特征视为局部稀疏矩阵, 则特征在局部图像中具有奇异值, 对应着矩阵中隐含的重要信息, 且重要性和奇异值大小正相关。一般认为小奇异值由图像中的噪声引起, 并对其取 0 来去除噪声。因此, 采用全局最大池化汇聚特征, 在每个通道内对所有元素保留局

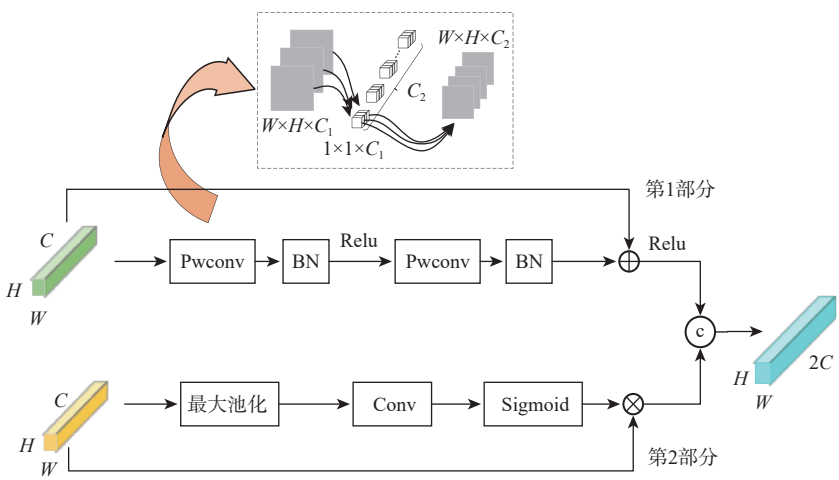


图 4　有效融合模块

Fig. 4　Effective fusion module

部最大值来代表目标整体特征，并能有效去除冗余噪声。池化后的特征经1D卷积进行学习，其中，卷积核的大小通过函数自适应变化，在网络的深层可以更多地进行跨通道交互。最后，经Sigmoid进行权值归一化处理，将归一化权重和原输入特征图逐通道相乘得到加权特征图，用作与横向特征Concat的基准。计算过程如下：

$$\boldsymbol{P}_1=\sigma(f(\tau(\boldsymbol{X})))\otimes\boldsymbol{X} \tag{8}$$

式中：f和τ分别表示1D卷积和全局最大池化。

在实际的检测过程中，目标在整个图像中占较小区域，浅层细节特征显得至关重要，因此，采用图4中的第1部分对融合中的横向有效特征层进行处理，增强目标细节特征。选择点态卷积(pointwise convolution, Pwconv)来聚合每个空间位置的局部上下文，使其与局部通道的空间信息相互作用，使网络更加关注具有局部高对比度的信息特征，以突出目标细微特征，在像素级对小目标特征进行增强，有效地解决了小目标像素和细节特征缺失的问题。计算过程如下：

$$\boldsymbol{P}_2=\beta(\mathrm{BN}(f_{\mathrm{PWC2}}(\beta(\mathrm{BN}(f_{\mathrm{PWC1}}(\boldsymbol{X})))))\oplus\boldsymbol{X}) \tag{9}$$

式中：f_{PWC}、BN和β分别表示点态卷积、批量归一化层和Relu激活函数。

对经EFM模块处理过的$\boldsymbol{P}_1$与$\boldsymbol{P}_2$进行Concat操作，将网络深层整体特征信息与浅层细节信息进行有效融合，进一步完善目标特征。

2.3 损失函数

损失函数用来度量模型预测值与真实值之间的差异程度。YOLOv5的损失计算由分类损失，定位损失及置信度损失组成，置信度损失和分类损失采用二元交叉熵计算，而定位损失采用的是CIOU损失函数，计算过程如下：

$$L_{\mathrm{CIOU}}=1-\Delta_2=1-\left(\Delta_1-\frac{\rho^2(b,b^{\mathrm{gt}})}{c^2}-\alpha v\right) \tag{10}$$

式中：Δ_1和Δ_2分别为IOU和CIOU；b和b^{gt}分别为预测框和真实框的中心点；c为两框最小外接矩形的对角线长度。v为长宽比的一致性；α表示权衡参数，二者定义如下：

$$v=\frac{4}{\pi^2}\left(\arctan\frac{W^{\mathrm{gt}}}{H^{\mathrm{gt}}}-\arctan\frac{W}{H}\right)^2 \tag{11}$$

$$\alpha=\frac{v}{(1-\Delta_1)+v} \tag{12}$$

CIOU虽然考虑了边界框回归的重叠面积、中心点距离、长宽比等因素，但式(11)中的v反映的是长宽比的差异，而不是真实框和预测框间实际的长宽差值，存在长宽比值相等时度量失效的问题。本文提出一种新的定位损失计算方法，与CIOU相比，采用将长宽差异分开度量的方式，同时增加两框主对角之间距离的几何度量，弥补现有的几何度量在特殊情况下失效的情况，更好地促进定位损失的收敛和优化。计算方法如下：

$$\Delta_3=\Delta_1-\left(\frac{\rho^2(b,b^{\mathrm{gt}})}{c^2}+\frac{\rho^2(W,W^{\mathrm{gt}})}{{C_{\mathrm{w}}}^2}+\frac{\rho^2(H,H^{\mathrm{gt}})}{{C_{\mathrm{h}}}^2}+\frac{\rho^2(A,A^{\mathrm{gt}})+\rho^2(B,B^{\mathrm{gt}})}{c^2}\right) \tag{13}$$

$$L_{\mathrm{loc}}=1-\Delta_3 \tag{14}$$

式中：Δ_3为SEIOU；C_{w}和C_{h}分别为预测框与真实框最小外接矩形的长和宽；$\rho(b,b^{\mathrm{gt}})$为两框中心点间距；$\rho(W,W^{\mathrm{gt}})$为两框宽度的差值；$\rho(H,H^{\mathrm{gt}})$为两框长度的差值；$\rho(A,A^{\mathrm{gt}})$为两框左上角间距，$\rho(B,B^{\mathrm{gt}})$为两框右下角间距。

图5为SEIOU损失的边界框回归图，其中，黄色和蓝色虚线框分别为预测框和真实框，黑色实线框为两框最小外接矩形。由图5可知，本文提出的计算方法能兼顾重叠面积、中心点、长宽比等一系列度量，并通过分开计算长宽差异值弥补了CIOU有关长宽比的模糊定义，此外当预测框与真实框在没有完全重合的情况下，$\rho(A,A^{\mathrm{gt}})$与$\rho(B,B^{\mathrm{gt}})$便不为0，这大大降低了度量失效的概率，保证了惩罚项的有效性，获取更为精准的回归结果。

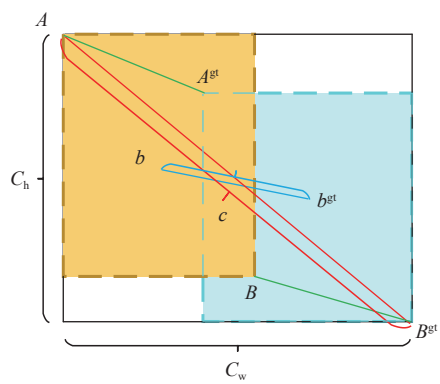


图5 预测框回归

Fig. 5 Prediction box regression

网络整体的损失函数计算式为

$$L=\lambda_1L_{\mathrm{loc}}+\lambda_2L_{\mathrm{cls}}+\lambda_3L_{\mathrm{obj}} \tag{15}$$

式中：λ_1、λ_2、λ_3为平衡系数；L_{loc}、L_{cls}和L_{obj}分别为定位损失、分类损失和置信度损失。只针对正样本进行分类损失和定位损失的计算，置信度损失针对所有样本进行。

3 实验与结果分析

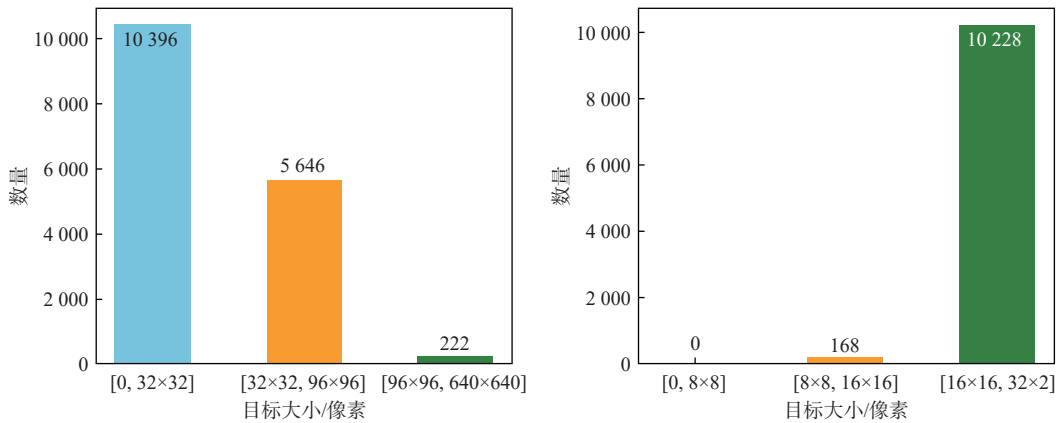
3.1 实验环境与参数设置

本文采用Pytorch深度学习框架搭建网络，操

作系统为 Window11, CPU 为 i5-12600KF, GPU 为 NVIDIA GeForce RTX4070Ti, 显存大小为 12 GB。CUDA 版本为 12.0, Pytorch 版本为 1.9.1, Python 版本为 3.7.11。实验共训练 100 个 epoch, Batchsize 设置为 16, IOU 阈值设置为 0.5, 初始学习率为 1×10^{-2}, 学习率下降方式为 step, 采用随机梯度下降(stochastic gradient descent, SGD)法优化学习率, 权值衰减为 5×10^{-4}, 动量为 0.937。

3.2　数据集

本文主要基于 CCTSDB[19] 和 S2TLD[20] 数据集开展实验。其中, CCTSDB 数据集分为指示、禁止、警告 3 个类别。根据目标检测领域 COCO 数据集的评价指标, 小目标是指目标的真实框像素面积小于32像素×32像素的目标, 中目标是指像素面积在[32像素×32像素,96像素×96像素]范围的目标, 大目标是指像素面积大于96像素×96像素的目标。图 6 为 CCTSDB 训练集的目标大小分布情况, 可以看出, 真实框主要集中在[0像素,32像素×32像素]大小, 又以[16像素×16像素,32像素×32像素]为主, 属于小目标范畴。此外, 采用 S2TLD 数据集中 1 280 像素×720 像素大小的原始图片开展实验, 分为红、黄、绿、关 4 类。为进一步验证本文网络对交通场景小目标的检测效果, 评估了在 TLD 数据集上的检测效果, TLD 数据集是在真实交通环境中采集图像自制的交通灯数据集, 包括在郊区和城市道路的白天和黑夜的数据, 分为红、黄、绿、关 4 类。



(a) 道路交通标识训练集小、中、大目标分布情况　　(b) 交通标识小目标分布情况

图 6　CCTSDB 训练集目标大小分布

Fig. 6　Target size distribution of CCTSDB training set

3.3　评价指标

本文采用准确率 P、召回率 R、均值平均精度 (mean average precision, mAP)、每秒帧数 (frames per second, FPS) 等指标来评估模型的性能。根据 IOU 将预测框分为正确正样本 (true positives, TP)、正确负样本 (true negatives, TN)、错误正样本 (false positives, FP) 和错误负样本 (false negatives, FN)。P 和 R 计算式分别为

$$P=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}} \tag{16}$$

$$R=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}} \tag{17}$$

$$A_{\mathrm{P}}=\int_0^1 P(R)\mathrm{d}R \tag{18}$$

$$m_{\mathrm{AP}}=\frac{1}{N}\sum_{i=1}^{N}A_{\mathrm{P}i} \tag{19}$$

式中: N 为目标类别数; $P(R)$ 指的是基于 P 和 R 的曲线; A_{P} 为曲线下的面积, 对所有类别的 A_{P} 值求平均得到 m_{AP}。

3.4　CCTSDB 数据集检测效果

3.4.1　CCTSDB 数据集消融实验

在 CCTSDB 数据集上对交通标志进行检测, 将改进模块依次加到网络中, 训练结果如表 1 所示。可以看出, 单个模块加入网络中时, mAP 相比 YOLOv5-s 分别提高了 1.52%、1.26%、0.31%, 说明了部署单个模块的有效性, 将所有改进都加入到网络中时, mAP 达到 93.04%, 比原网络提高了 2.54%, 说明了本文算法的整体有效性。

本文通过引入 FSM 等有针对性的改进措施解

表 1　CCTSDB 数据集消融实验

Table 1　Ablation experiments of CCTSDB dataset

YOLOv5-s	FSM	EFM	SEIOU	mAP/%
√				90.50
√	√			92.02
√		√		91.76
√			√	90.81
√	√	√		92.56
√	√	√	√	93.04

决小目标检测困难的问题，训练结果如表2所示，在COCO的评价指标下将图像分为小、中、大3类进行检测效果展示，其中，AP和AR分别表示平均准确率和平均召回率。可以看出，在IOU=0.5:0.95的阈值条件下，改进前后小目标的检测效果有了明显提升，AP和AR分别提升了3.4%和2%，基本满足小目标检测的要求。

表2 CCTSDB数据集上不同尺度目标检测结果

Table 2 Detection results of different scale targets on CCTSDB dataset

算法	AP/%			AR/%		
	小目标	中目标	大目标	小目标	中目标	大目标
YOLOv5-s	38.5	67.4	83.5	47.5	74.4	89.5
本文算法	41.9	68.4	84.9	49.5	75.0	89.3

图7对YOLOv5-s算法和本文算法的检测效果进行展示对比。

从第1行对比图可以看出，本文算法通过获取更多的上下文信息辅助判断可以有效地避免误检，从第2行、第3行对比图可以看出，补充更浅层的细节特征能缓解漏检问题，对小目标的检测有较好的改善。虽然本文算法相比原网络有一定改善，但在第3行仍漏检一个目标，这是因为恶劣天气会影响图像采集质量，当目标又排列密集时会对检测精度产生一定影响。

3.4.2 CCTSDB数据集不同算法对比实验

在CCTSDB数据集上将本文算法与Faster R-CNN、Efficientdet[21]、YOLOv8-n和YOLO-FR[16]等算法的检测结果进行对比分析，验证本文算法对交通标志检测的有效性。实验在相同的参数设置和操作系统下展开，检测结果如图8所示。

由图8可知，本文算法与Faster R-CNN及一系列轻量化算法相比，有着很好的精度优势。YOLO-

(a) YOLOv5-s (b) 本文算法

图7 CCTSDB数据集上检测效果对比图

Fig. 7 Comparision of detection results on CCTSDB dataset

FR算法通过特征重组下采样，同时利用内容感知特征重组(content-aware reassembly of features, CARAFE)算子完成特征融合上采样来减少特征损失，通过实验结果可以看出，在CCTSDB数据集上效果并不理想。

表3给出了不同算法的参数对比，结合图8进行分析。本文算法与Faster R-CNN算法、Efficientdet算法、YOLOx-s算法[22]及YOLO-FR算法相比，在检测速度为113帧/s占优的前提下，检测精度也取得不错的效果，与YOLOv4-tiny算法、YOLOv7-tiny算法、YOLOv8-n算法等轻量化网络算法，虽然检

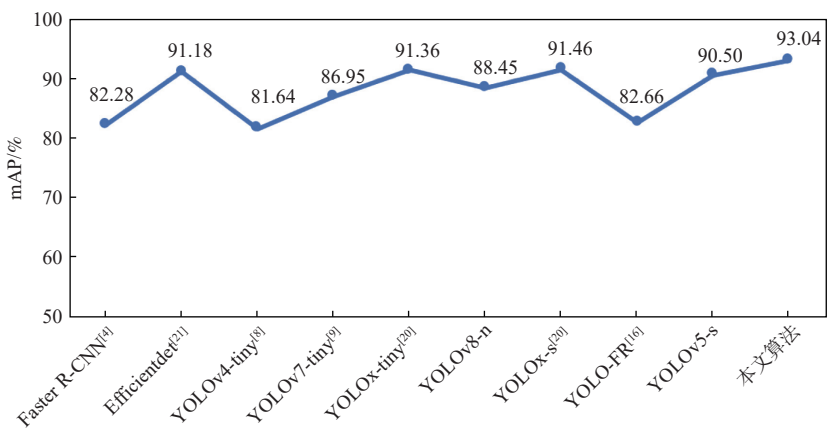


图8 不同算法在CCTSDB数据集上检测结果

Fig. 8 Detection results of different algorithms on CCTSDB dataset

表 3　不同算法参数对比

Table 3　Comparison of different algorithm parameters

算法	mAP/%	参数量	检测速度/(帧·s^{-1})
Faster R-CNN[4]	82.28	28.4×10^6	29
Efficientdet[21]	91.18	6.5×10^6	34
YOLOv4-tiny[8]	81.64	5.9×10^6	271
YOLOv7-tiny[9]	86.95	6.0×10^6	140
YOLOv8-n	91.36	3.0×10^6	146
YOLOx-tiny[20]	88.45	5.0×10^6	103
YOLOx-s[20]	91.46	8.9×10^6	96
YOLO-FR[16]	82.66	6.5×10^6	94
YOLOv5-s	90.50	7.1×10^6	118
本文算法	93.04	8.8×10^6	113

测速度稍慢, 但在检测精度上分别有 11.4%、6.09%、1.68% 的优势。从参数量的角度分析, 本文算法相比原 YOLOv5-s 算法以轻微的代价取得了与 YOLOx 算法、Faster R-CNN 算法、Efficientdet 算法相比时的检测速度与精度优势, 与 YOLOv4-tiny 算法、YOLOv7-tiny 算法、YOLOv5-s 算法相比时的精度优势。

从推理速度的角度分析, 该参数受模型并行化程度、内存访问率、模型分支数量等多个因素影响。通过表 3 可以看出, 虽然 Efficientdet 算法参数量少, 但运行速度较慢。分析网络结构可知, Efficientdet 算法以 Efficientnet 为主干网络、BiFPN 为特征融合部分, 与常规的 FPN 相比, BiFPN 的构建更加复杂, 不仅通过 5 个特征层进行多次采样和堆叠操作, 而且在特征融合部分对 BiFPN 进行重复操作。

对 Efficientdet 算法和本文算法的 memory 和 MemR+MemW 进行计算, 结果如表 4 所示。其中, memory 是节点推理时所需要的内存大小, 直接影响模型可以处理的数据量和速度, 如果内存不足, 模型需要频繁进行内存交换, 较大的内存交换量可能会成为性能瓶颈, 导致模型推理速度下降。MemR+MemW 表示 MemRead 和 MemWrite, 是网络运行时内存读写大小, 如果读写较大导致速度缓慢, 节点计算将受到限制, 也会使得模型推理速度下降, 影响模型性能。

表 4　内存相关参数量对比

Table 4　Comparison of memory-related parameters

算法	memory/MB	(MemR+MemW)/MB
Efficientdet[21]	673.26	1 320.00
本文算法	333.75	558.43

由表 4 可以看出, 相比于本文算法, Efficientdet 算法的模型结构较为复杂, 需要占用更多的内存, 因此, 虽然其参数量很小, 但检测速度仍会受影响。而本文算法经 FSM、EFM 改进小目标检测性能, 同时在推理阶段将 Conv 与 BN 层相融合, 在不损失精度的情况下, 保证了模型的推理速度, 二者之间得到了很好的平衡, 极大地丰富了网络的应用范围。

3.5　S2TLD 数据集检测效果

3.5.1　S2TLD 数据集消融实验

在 S2TLD 数据集上验证本文算法对交通灯的检测效果, 训练结果如表 5 所示。随着改进模块加入到网络中, 精度不断提升, 最终达到 86.79%, 相比原网络提高了 3.62%。

表 5　S2TLD 数据集上的消融实验

Table 5　Ablation experiments on S2TLD dataset

YOLOv5-s	FSM	EFM	SEIOU	mAP/%
√				83.17
√	√			83.35
√	√	√		86.37
√	√	√	√	86.79

表 6 为 IOU=0.5:0.95的阈值条件下算法改进前后不同尺度目标的效果对比, 可以看出小目标的 AP 和 AR 分别提升了 4% 和 3.1%。

表 6　S2TLD 数据集上不同尺度目标检测结果

Table 6　Detection results of different scale targets on S2TLD dataset

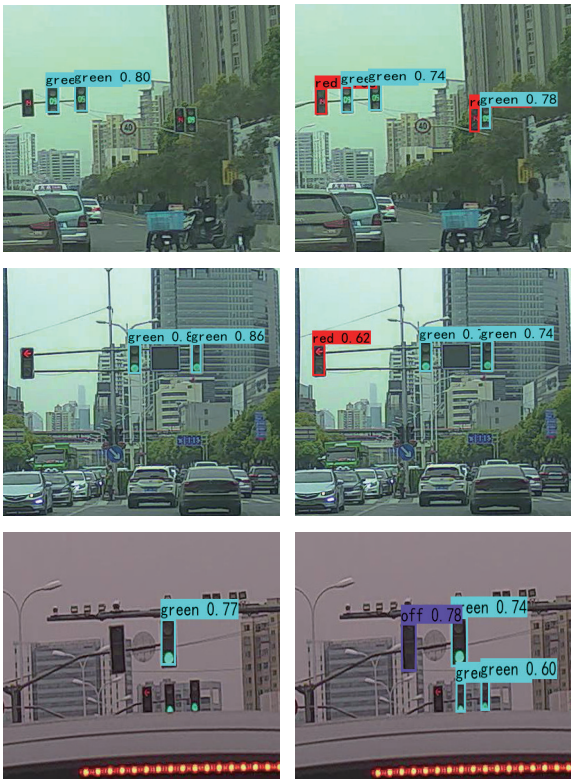
算法	AP/%			AR/%		
	小目标	中目标	大目标	小目标	中目标	大目标
YOLOv5-s	30.4	53.2	76.7	42.6	63.8	78.1
本文算法	34.4	54.4	76.3	45.7	64.6	81.2

图 9 展示了算法改进前后对交通灯的检测效果。可以看出, 针对原网络检测过程中小目标分辨率低, 缺乏像素信息的问题, 本文算法通过补充目标细节特征、获取上下文信息辅助检测、采用有效特征融合策略等措施, 有效避免漏检, 提升了检测效果。但在第 3 行低能见环境会进一步减弱小目标的特征信息, 对检测效果造成影响, 导致本文算法没能完全避免漏检。

3.5.2　S2TLD 数据集不同网络对比实验

将本文算法在 S2TLD 数据集上与 YOLOv4-tiny, YOLOv8-n 等轻量网络进行对比, 分析各类别及整体的检测精度的变化趋势, 检测效果如图 10 所示。

由图 10 中折线图的整体趋势可以发现, 与 YOLOv4-tiny 算法和 YOLOv7-tiny 算法相比, 本文



(a) YOLOv5-s (b) 本文算法

图 9 S2TLD 数据集上检测效果对比

Fig. 9 Comparision of detection results on S2TLD dataset

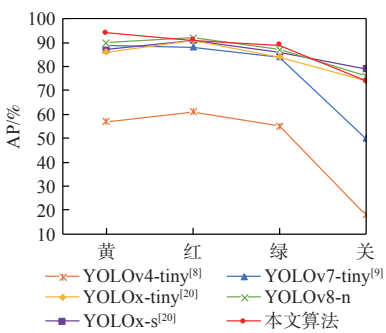


图 10 不同算法在 S2TLD 数据集上检测结果

Fig. 10 Detection results of different algorithms on S2TLD dataset

算法在各类别上检测效果差别不大，解决了“关”类的目标特征不明显导致的精度过低的问题，体现了本文算法的稳定性，能更好地应对各种复杂的路况。与包括 YOLOv8-n 算法在内的其他算法相比，本文算法在黄、绿这 2 类的检测精度上都有一定优势。

3.6 TLD 数据集检测效果

为进一步验证算法在复杂交通环境及不同光照条件下的稳定性，在 TLD 数据集上对交通灯进行检测。训练结果如表 7 所示，表 7 在不同类别及整体 mAP 上对算法改进前后的检测效果进行对比，可以看出，各类别的 AP 值分别提升了 2.42%、5.27%、3.86%、5.77%。mAP 达到 82.34%，提升了 4.33%，检测效果有了显著的提升。

表 7 算法改进前后 TLD 数据集上的检测结果

Table 7 Detection results on TLD dataset before and after algorithm improvement

算法	AP/%				mAP/%
	红	绿	黄	关	
YOLOv5-s	87.56	80.21	74.71	69.55	78.01
本文算法	89.98	85.48	78.57	75.32	82.34

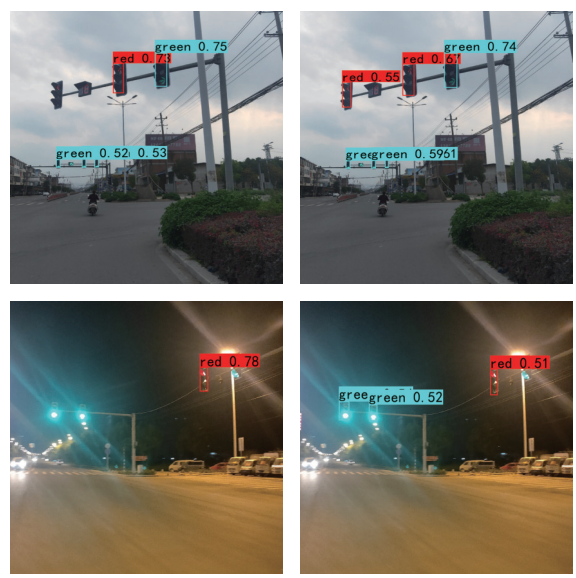
在 TLD 数据集上对改进前后不同尺度目标的检测效果进行对比。如表 8 所示，在 IOU=0.5:0.95 的阈值条件下，可以看出，虽然大目标上的 AP、AR 略有降低，但在小目标上 AP 和 AR 分别提升了 2.2% 和 1.5%，在中等目标上 AP 和 AR 分别提升了 2.6% 和 1%。考虑到交通灯数据以小目标居多，且整体 mAP 有较大提升，因此，仍能很好地满足检测需求。

表 8 TLD 数据集上不同尺度目标检测结果

Table 8 Detection results of different scale targets on TLD dataset

算法	AP/%			AR/%		
	小目标	中目标	大目标	小目标	中目标	大目标
YOLOv5-s	29.3	50.0	62.3	42.9	62.5	72.3
本文算法	31.5	52.6	58.5	44.4	63.5	65.3

图 11 展示了本文算法的检测效果对比，可以看到，在郊区及城市道路的白天、黑夜环境下，都能取得很好的检测效果，有效避免漏检情况。



(a) YOLOv5-s (b) 本文算法

图 11 TLD 数据集上检测效果对比

Fig. 11 Comparision of detection results on TLD dataset

3.7 PASCAL VOC 数据集实验结果

为进一步验证本文算法的有效性，在通用数据

集 PASCAL VOC 上进行试验验证, 实验结果如表 9 所示。

由表 9 可知, 在 PASCAL VOC 数据集的 20 个类别中, 本文算法在自行车、鸟、船等 15 个类别上有较好的精度提升, 虽然在飞机等 5 个类别上与原算法存在一些微小差距, 但可以看出, 对最终精度影响很小, mAP 提升了 2.01%, 证明了本文算法的整体有效性。

表 9　算法改进前后 PASCAL VOC 数据集上的检测结果

Table 9　Detection results on PASCAL VOC dataset before and after algorithm improvement

算法	AP/%										
	飞机	自行车	鸟	船	瓶	巴士	汽车	猫	椅子	牛	
YOLOv5-s	92.31	87.58	85.12	70.27	68.97	92.96	91.26	84.16	67.57	85.74	
本文算法	91.95	90.13	92.32	74.15	69.61	95.42	91.02	86.89	69.72	89.46	
算法	AP/%										mAP/%
	餐桌	狗	马	摩托车	人	盆栽	羊	沙发	火车	电视监视器	
YOLOv5-s	60.81	82.28	92.67	83.67	92.60	59.70	87.68	75.89	86.50	89.42	81.84
本文算法	62.11	85.22	95.06	85.08	92.52	63.05	90.84	75.80	88.02	88.75	83.85

3.8　将改进模块部署到 YOLOv7-tiny

为进一步验证本文提出的改进模块的有效性和普适性, 将 FSM、EFM 等模块依次部署到 YOLOv7-tiny 算法中。首先, 在 YOLOv7-tiny 的主干网络和原 PANet 融合网络之间, 经 FSM 从主干网络更浅层引出特征层, 对相邻深层进行细节补充后得到 3 个新的有效特征层, 传入 PANet 进行后续特征融合的操作; 此外, 采用 EFM 模块对原网络融合主干特征和上采样信息的 concat 部分进行替换, 最后, 用 SEIOU 替换 YOLOv7-tiny 的原损失函数。在 CCTSDB 数据集上检测效果如表 10 所示。

由表 10 可知, 相比 YOLOv7-tiny 原网络, 部署单个模块检测精度分别提升了 2.98%, 1.56% 和 0.58%, 整体提高了 3.98%, 体现了本文算法的有效性和普适性。表 11 给出不同尺度目标检测结果对比, 可以看出, 在IOU=0.5:0.95的阈值条件下, 小目标的 AP 和 AR 分别提升了 6.1% 和 5.2%, 改善了交通场景小目标的检测效果, 证明了本文算法具有较好的泛化性。

表 10　YOLOv7-tiny 网络消融实验

Table 10　Ablation experiment of YOLOv7-tiny network

算法	AP/%			mAP/%
	禁止	指示	警告	
YOLOv7-tiny	88.41	88.00	84.45	86.95
+ FSM	90.42	89.09	90.28	89.93
+ EFM	89.80	89.07	86.66	88.51
+ SEIOU	88.64	88.88	85.07	87.53
+ FSM+ EFM	91.77	90.18	90.37	90.77
+FSM+EFM+SEIOU	90.80	90.62	91.38	90.93

表 11　网络改进前后不同尺度目标检测结果

Table 11　Detection results of different scale targets before and after network improvement

算法	AP/%			AR/%		
	小目标	中目标	大目标	小目标	中目标	大目标
YOLOv7-tiny	34.9	67.1	83.3	44.4	74.8	89.7
+FSM+EFM+SEIOU	41.0	67.6	82.1	49.6	74.8	88.5

4　结　论

1) 分析实验结果可看出, 本文算法在 CCTSDB、S2TLD 和 TLD 数据集上的 mAP 分别达到了 93.04%、86.79%、82.34%, 相比原 YOLOv5 算法分别提高了 2.54%、3.62%、4.33%, 检测速度达到了 113 帧/s, 在检测速度和精度上都有很好的表现。

2) 本文算法在 CCTSDB、S2TLD 和 TLD 数据集上相比原算法, 小目标的 AP 值分别提升了 3.4%、4%、2.2%, 小目标检测效果有明显提升。

3) 本文算法在 PASCAL VOC 通用数据集上相比原算法在 15 个类别上都有较好的精度优势, mAP 提高了 2.01%, 证明了本文算法的有效性。

4) 将本文的改进模块添加到 YOLOv7-tiny 算法, mAP 提高了 3.98%, 小目标的 AP 提升了 6.1%, 证明了改进的泛化性和可靠性。

考虑到交通场景中的遮挡目标存在像素重叠问题, 会进一步加大检测难度, 后续研究会针对该类问题做进一步改进。

参考文献（References）

[1] WANG Q Y, LI X Y, LU M. An improved traffic sign detection and recognition deep model based on YOLOv5[J]. IEEE Access, 2023, 11: 54679-54691.

[2] LIU C S, LI S, CHANG F L, et al. Machine vision based traffic sign detection methods: review, analyses and perspectives[J]. IEEE Access, 2019, 7: 86578-86596.

[3] 孙迎春, 潘树国, 赵涛, 等. 基于优化 YOLOv3 算法的交通灯检测[J]. 光学学报, 2020, 40(12): 143-151.

SUN Y C, PAN S G, ZHAO T, et al. Traffic light detection based on optimized YOLOv3 algorithm[J]. Acta Optica Sinica, 2020, 40(12): 143-151(in Chinese).

[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] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2980-2988.

[6] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]//Proceedings of the Computer Vision-ECCV. Beilin: Springer, 2014: 346-361.

[7] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL].(2018-04-08)[2024-01-02].https://arxiv.org/abs/1804.02767.

[8] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2024-01-02]. http://arxiv.org/abs/2004.10934.

[9] 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464-7475.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.

[11] 井方科, 任红格, 李松. 基于多尺度特征融合的小目标交通标志检测[J]. 激光与光电子学进展, 2024, 61(12): 372-380.
JING F K, REN H G, LI S. Small object traffic sign detection based on multi-scale feature fusion[J]. Laser and Optoelectronics Progress, 2024, 61(12): 372-380(in Chinese).

[12] 钱伍, 王国中, 李国平. 改进 YOLOv5 的交通灯实时检测鲁棒算法[J]. 计算机科学与探索, 2022, 16(1): 231-241.
QIAN W, WANG G Z, LI G P. Improved YOLOv5 traffic light real-time detection robust algorithm[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(1): 231-241(in Chinese).

[13] YAO Y B, HAN L, DU C J, et al. Traffic sign detection algorithm based on improved YOLOv4-Tiny[J]. Signal Processing: Image Communication, 2022, 107: 116783.

[14] SHI Y L, LI X D, CHEN M M. SC-YOLO: a object detection model for small traffic signs[J]. IEEE Access, 2023, 11: 11500-11510.

[15] LI J Y, LIU C N, LU X C, et al. CME-YOLOv5: an efficient object detection network for densely spaced fish and small targets[J]. Water, 2022, 14(15): 2412.

[16] MOU X G, LEI S, ZHOU X. YOLO-FR: a YOLOv5 infrared small target detection algorithm based on feature reassembly sampling method[J]. Sensors, 2023, 23(5): 2710.

[17] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2020: 1571-1580.

[18] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759-8768.

[19] ZHANG J M J, ZOU X, KUANG L D, et al. CCTSDB 2021: a more comprehensive traffic sign detection benchmark[J]. Human-centric Computing and Information Sciences, 2022, 12: 23.

[20] YANG X, YAN J C, LIAO W L, et al. Scrdet++: detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 2384-2399.

[21] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 10778-10787.

[22] CE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL].(2021-07-18)[2024-01-02].https://arxiv.org/abs/2107.084.

Small target detection algorithm for traffic scenes based on improved YOLOv5-s

WANG Kun*, FENG Kangwei

(College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

Abstract: A traffic scene tiny target detection method based on enhanced YOLOv5-s was presented to address the issue that the properties of small targets in traffic scenes, such as traffic signs and traffic lights, are not readily apparent. Firstly, a feature supplement module (FSM) was designed to supplement the features of the adjacent deep detection layers by further obtaining the shallow details, which effectively improved the detection effect of small targets, and avoided feature redundancy by matrix operation between adjacent layers. Second, in order to reduce feature conflict and improve the effectiveness of the pyramid feature fusion, an effective fusion module (EFM) was created to handle the horizontal shallow feature and the upsampled feature, respectively. Then, the super enhanced intersection over union (SEIOU) loss calculation method was proposed to improve the regression effect and detection accuracy by adding the distance measurement between the main diagonal of the ground truth box and the prediction box. Finally, experiments were carried out on CCTSDB, S2TLD, the Traffic lights dataset and the PASCAL VOC dataset. According to the results, the proposed algorithm's accuracy has increased by 2.54%, 3.62%, 4.33%, and 2.01%, respectively, and its detection speed has reached 113 frames per second, making it appropriate for detecting jobs in real-world traffic situations.

Keywords: YOLOv5-s algorithm; small target detection; feature supplement; feature fusion; loss function

Received: 2024-01-03; **Accepted**: 2024-02-25; **Published Online**: 2024-03-12 20:21

URL: link.cnki.net/urlid/11.2625.V.20240312.1540.004

Foundation item: National Natural Science Foundation of China (62173331)

* **Corresponding author**. E-mail: yogo_w@163.com