

基于 SSE-YOLO 的红外小目标检测算法

大妹, 姜麟, 陶友凤, 胡淼

(昆明理工大学 理学院, 云南 昆明 650500)

摘要: 针对红外成像面积小、分辨率低、易被遮挡导致漏检、检测精度低等问题, 本文提出了一种基于 SSE-YOLO 的红外小目标检测算法。首先在 YOLOv8s 的基础上引入深度非跨步卷积模块, 避免检测过程中细粒度信息的丢失并提高特征学习的效率; 其次在特征提取阶段增加专门针对小目标的检测层, 以提升模型对红外小目标的提取能力; 此外设计了一种高效的双注意力机制 (efficient dual-attention mechanism, EDAM), 自适应地学习每个通道和空间位置的重要性, 从而更好地捕捉图像中的关键信息; 然后使用 Shape_IoU 损失函数来聚焦边框自身形状与自身尺度, 进一步提高边框回归的精确度; 最后在 FLIR 数据集和艾睿光电公司拍摄的数据集上进行了一系列实验。结果表明: 本文所提方法在两种数据集上的平均精度分别达到了 89.8% 与 92.1%, 相比于原始的模型分别提高了 3.3% 与 2.9%。

关键词: YOLOv8s; 红外小目标检测; 深度非跨步卷积; Shape_IoU 损失函数; 双注意力机制

中图分类号: TP391 文献标识码: A 文章编号: 1001-8891(2025)04-0475-09

Improved Infrared Small Target Detection Algorithm Based on SSE-YOLO

DA Mei, JIANG Lin, TAO Youfeng, HU Miao

(Faculty of Science, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: To address the problems of a small infrared imaging area, low resolution, and ease of occlusion—resulting in incorrect detection, missed detection, and low detection accuracy—this paper proposes an infrared small-target detection algorithm based on SSE-YOLO. Firstly, a depth non-stepwise convolution module is introduced on the basis of YOLOv8s to avoid the loss of fine-grained information during the detection process and to improve the efficiency of feature learning. Then, a detection layer specialized for small targets is added in the feature extraction stage to improve the model's ability to extract infrared small targets. In addition, an efficient dual attention mechanism (EDAM) is designed to adaptively learn the importance of each channel and spatial location to better capture key information in the image. Secondly, the Shape_IoU loss function is used to focus on the shape of the boundary itself and its scale, which further improves the accuracy of boundary regression. Finally, a series of experiments were conducted on the FLIR dataset and a dataset captured by IRay. The results show that the average accuracies of the proposed method on the two datasets reach 89.8% and 92.1%, which are 3.3% and 2.9% higher than those of the original model, respectively.

Key words: YOLOv8s, infrared small target detection, space-to-depth convolution, shape_IoU Loss, dual-attention mechanism

0 引言

红外成像技术作为一项重要的无源探测技术, 已被应用于自动驾驶、军事侦察等诸多领域^[1], 是当前

重点研究和发展的方向之一^[2]。

近年来, 人们开始将深度学习算法引入红外目标检测领域, 并取得了不少成果。当前目标检测算法包括两类, 分别为双阶段 (two-stage) 和单阶段 (one-

收稿日期: 2024-03-28; 修订日期: 2024-05-10。

作者简介: 大妹 (1998-), 女, 云南西双版纳人, 硕士研究生, 主要研究方向为计算机视觉。E-mail: damei107698@163.com。

通信作者: 姜麟 (1969-), 男, 湖南益阳人, 教授, 硕士生导师, 主要从事网络计算等方面的教学与科研工作。E-mail: tojianglin@163.com。

基金项目: 国家自然科学基金项目 (11761042)。

stage) 目标检测方法。两阶段目标检测的代表算法主要有 R-CNN^[3] (region-based convolutional neural network) 系列, 包括: R-CNN^[3]、Fast R-CNN^[4]、Faster R-CNN^[5]、FPN^[6]等。这类方法的准确度相对较高, 但速度却不太理想。为了提高检测速度, 出现了以 YOLO^[7] 系列和 SSD^[8] 系列为代表的单阶段目标检测算法, 这类算法具备良好的实时检测效果以及泛化能力。

杨等人^[9]提出一种基于 YOLOv5 的红外车辆目标检测算法, 在浅层特征层引入 RFBs 模块, 以提高小目标的感受野及检测效果。这一方法显著提高了小目标的精度, 同时降低了漏检率。李等人^[10]基于 YOLOv5s 模型使用跨域迁移学习策略、引入通道注意力机制 Squeeze-and-Excitation 来对红外数据进行检测, 改进后的算法有效提高了精度。鞠等人^[11]为解决图像中小目标检测率低和虚警率高的问题, 提出了基于 YOLOv3 的改进方法。作者通过降采样和上采样技术, 构建了一个输出为 4 倍降采样的特征融合目标检测层, 显著提高了 YOLOv3 的精度和召回率。

近年来, 一些研究人员在红外小目标检测中, 结果取得了不同的提升, 但仍有一些不足之处, 如多数注意力机制 SE (squeeze-and-excitation) 仅关注了通道注意力, 没考虑空间方面的注意力, 而空间特征信息对于构建空间注意图至关重要; 当使用传统的卷积结构时, 由于池化层的存在, 会导致许多重要的信息丢失, 同时也会忽略掉整体与局部之间的关联。

本文针对上述问题提出了一种新的红外小目标检测算法 SSE-YOLO, 本文的主要贡献如下:

① 在解决传统卷积神经网络性能下降的问题时, 引入新的非跨步卷积来提高特征学习效率。

② 本文通过增加小目标检测层, 以便更好地保留小目标的特征信息, 提高检测精度。

③ 设计了一种轻量化的注意机制模块, 该模块能够自适应地学习每个通道和空间位置的重要性, 从而更好地捕捉图像中的关键信息, 进一步提高了检测精度。

④ 为了提升小目标检测的准确性, 本研究采用 Shape_IoU_Loss 优化损失函数, 以便有效处理边界框微小偏差, 并显著降低漏检率。

1 YOLOv8s 模型

YOLOv8^[12] 是 Ultralytics 公司研发的 YOLO 系列模型之一, 结构如图 1 所示, 包括 Backbone、Neck 和 Head 部分。

YOLOv8 通过采用梯度流更为丰富的 C2f 模块, 并对不同尺度的模型进行通道数的调整, 成功降低了模型的计算量, 从而提高了收敛速度和检测效果。

特征增强模块采用了 PANet (path aggregation network)^[13] 结构, 通过上采样和通道融合将 PANet 的 3 个输出分支引入检测模块。检测模块采用解耦头结构将回归分支和预测分支分离, 进一步加快了模型的收敛速度。

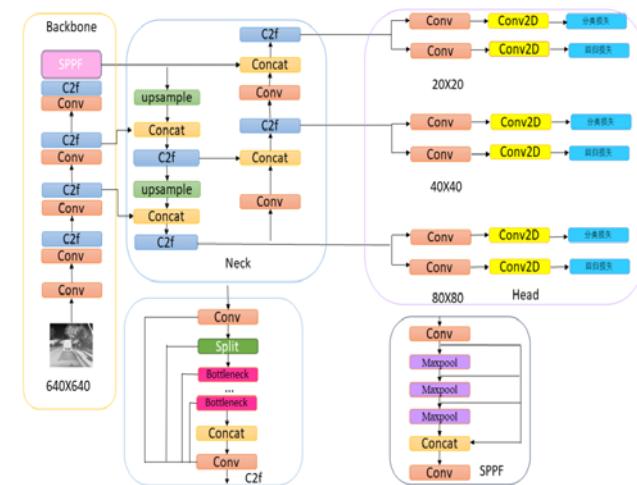


图 1 YOLOv8 网络结构
Fig.1 YOLOv8 network structure

2 改进 YOLOv8s 模型

2.1 SSE-YOLO 模型

红外小目标检测在诸多领域应用广泛, 如自动驾驶、安防监控、军事侦察, 对高精确率与召回率的模型需求尤为庞大。

SSE-YOLO 在 YOLOv8s 的基础上, 设计 EDAM 注意力机制并额外添加一个 160×160 的小目标检测层, 通过 SPD_Conv 模块代替 YOLOv8s 中的卷积层, 使用 Shape_IoU_Loss 聚焦边框自身形状与自身尺度计算损失, 进一步加快收敛速度。SSE-YOLO 网络结构如图 2 所示。

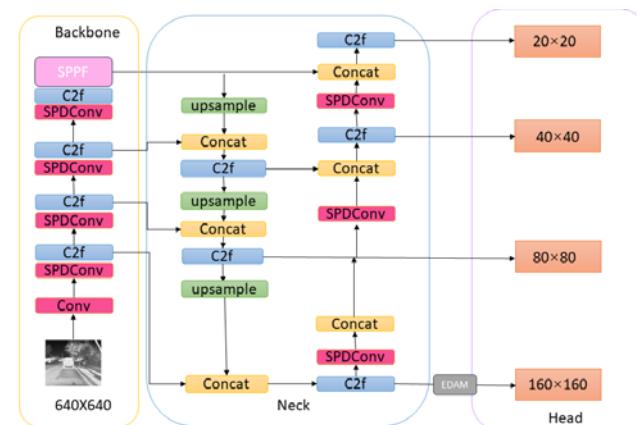


图 2 SSE-YOLO 网络结构
Fig.2 SSE-YOLO network structure

2.2 小目标检测层

在 FLIR 数据集中, 红外小目标遮挡十分严重,

不易被检测到,因此为了提高检测精度,本文在原有的YOLOv8s模型的基础上,额外添加了一个 160×160 小目标检测头,以增强对小目标的敏感度。具体而言,SSE-YOLO从主干网络的第3层开始提取特征,并使用Concat拼接将颈部网络结构中提取的浅层特征与EDAM注意力机制提取的上下文信息进行融合。

2.3 空间-深度非跨步卷积

在处理低分辨率图像和微小目标检测场景时,传统卷积神经网络的性能显著下降。这一问题主要源于现有网络结构中跨卷积层和池化层的使用,导致细粒度信息丧失以及特征表示的学习效率降低。为了应对这一挑战,我们引入了一种新的卷积神经网络模块——SPD-Conv^[14]。由该模块取代传统的跨卷积层和池化层。结构如图3所示。

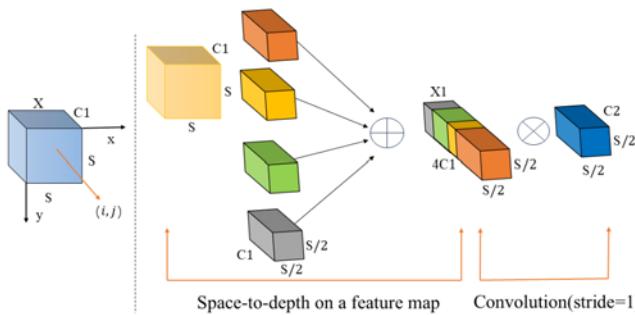


图3 SPD_Conv网络结构

Fig.3 SPD_Conv network structure

一般地,对于任意大小为 $S \times S \times C_1$ 的特征图 X ,可以切出一系列子特征图,如公式(1)所示:

$$\begin{aligned} f_{0,0} &= X[0:S:scale, 0:S:scale], \\ f_{1,0} &= X[1:S:scale, 0:S:scale], \dots, \\ f_{scale-1,0} &= X[scale-1:S:scale, 0:S:scale]; \\ f_{0,1} &= X[0:S:scale, 1:S:scale], f_{1,1}, \dots, \\ f_{scale-1,1} &= X[scale-1:S:scale, 1:S:scale]; \\ &\vdots \\ f_{0,scale-1} &= X[0:S:scale, scale-1:S:scale], f_{1,scale-1}, \dots, \\ f_{scale-1,scale-1} &= X[scale-1:S:scale-1:S:scale-1] \end{aligned} \quad (1)$$

在SPD层中, X 为初始特征图,该特征图可由 $i+x$ 和 $i+y$ 的scale切出子图 $f_{i,j}$,其中scale为比例因子,每个子图都按比例对 X 进行下采样。图3中的scale为2,每个子图尺寸为 $S/2 \times S/2 \times C_1$ 。特征图 X_1 由4张子图在通道维度上拼接所得,其尺寸为 $S/2 \times S/2 \times 4C_1$ 。在进行SPD特征变换后,添加一层步长为1的卷积层,将特征图 X_1 的尺寸变为 $S/2 \times S/2 \times C_2$,其中 $C_2 < 4C_1$ 。

2.4 EDAM注意力机制

EDAM是一种高效的双注意力机制,旨在提升数据中关键区域特征表达的同时,减少网络的参数量和计算量。如图4所示,该注意力机制不仅考虑了数据在通道和空间维度上的重要性,使模型能够集中注意力于目标检测有决定性作用的区域上。同时,采用组卷积和通道打乱技术^[15],分别对其进行加权求和得到全局上下文信息,以此来保证模型在具有优秀建模能力的同时,减少模型的参数量和计算量。EDAM的整体计算过程可以表示为:

$$F_c = A_c(F) \otimes F \quad (2)$$

$$F_{cs} = A_s(F_c) \otimes F_c \quad (3)$$

式中: $F \in R^{P \times S}$ 表示输入特征; P 和 S 分别为传感器数量和序列长度; \otimes 为 element-wise product 表示对应元素相乘; A_c 和 A_s 分别是通道注意力模块和空间注意力模块。

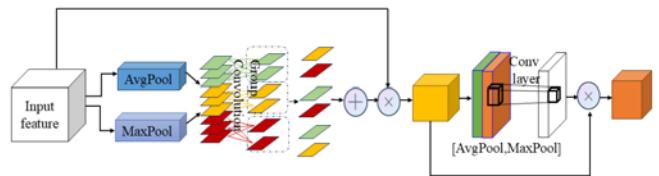


图4 EDAM注意力机制

Fig.4 EDAM attention mechanism

通道注意力模块主要致力于调整不同通道之间的权重,以便网络能够更有效地关注对小目标检测任务有用的特征。通道注意力机制能够动态地调整各个通道的权重,使模型能够集中处理关键区域的特征,进而增强对目标特征的感知能力。其具体的计算步骤如下所述:

$$F_c = \sigma[S_c(G(\text{AvgPool}(F))) + S_c(G(\text{MaxPool}(F)))] \quad (4)$$

式中: F_c 为通道注意力模块输出的特征图; σ 为 Sigmoid 激活函数; S_c 和 G 分别表示通道打乱和组卷积操作; $\text{AvgPool}(\cdot)$ 是全局平均池化操作; $\text{MaxPool}(\cdot)$ 是全局最大值池化操作。

空间注意力模块主要专注于调整不同空间位置上的特征响应,以便更好地捕捉目标的空间分布信息。该模块通过自适应地学习每个空间位置的权重,使得模型能够更有针对性地关注对目标检测任务重要的区域。在复杂环境下的目标检测任务中,对检测过程中误检、漏检现象具有不同的贡献。具体计算过程如下:

$$F_s = \sigma[\text{Conv2D}(\text{Con}(\text{AvgPool}(F)), (\text{MaxPool}(F)))] \quad (5)$$

式中: F_s 表示空间注意力模块输出的特征图; Conv2D 表示二维卷积运算。 $\text{Con}(\cdot)$ 表示 concatenation 组合操作,其他符号的含义同通道注意力模块。

相比于传统CBAM^[16]与SE^[17]注意机制,EDAM引入了组卷积和通道打乱技术,替代传统的卷积和全连

接层操作。这种方法可以使模型在保持有效性能的同时，大幅度降低模型的参数数量，从而提高计算效率。

2.5 Shape_IoU_loss 函数

边界框回归损失函数有助于模型从训练数据中学习如何准确地预测目标的位置，从而提高目标检测的定位精度和区域感知能力。这对于实现高质量的目标检测结果至关重要。现有的 YOLOv8 采用 CIoU^[18] 损失函数进行边界框的回归计算，如式(6)~(9)所示：

$$\text{CIoU} = \text{IoU} - \frac{\rho^2(b, b^{\text{gt}})}{c^2} - \alpha v \quad (6)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2 \quad (7)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (8)$$

则损失函数可定义为：

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha v \quad (9)$$

式中：IoU 衡量的是预测边界框与真实边界框之间的重叠度^[19]，而 $\rho^2(b, b^{\text{gt}})$ 表示预测边界框与真实边界框中心点之间的欧氏距离。此外 α 是用于平衡不同组成部分重要性的正权重参数， v 表示预测框与真实框的长宽比， w 、 h 分别为预测框的宽、高。现有的边界回归方法通常考虑目标边界框与预测边界框之间的几何关系来计算损失，而忽略了边界框本身的形状和尺度对回归结果的影响。本文引入了 Shape-IoU^[20] 方法，其中 Shape-IoU 损失函数的结构如图 5 所示，这一方法可以更精确地计算损失，因为它专注于边界框本身形状和尺度，同时该方法考虑到了边界框形状和尺度因素对回归结果的影响，并在计算损失时将这些因素纳入考量。

Shape_IoU 损失函数如公式(10)~(16)：

$$\text{IoU} = \frac{|B \cap B^{\text{gt}}|}{|B \cup B^{\text{gt}}|} \quad (10)$$

$$ww = \frac{2 \times (w^{\text{gt}})^{\text{scale}}}{(w^{\text{gt}})^{\text{scale}} + (h^{\text{gt}})^{\text{scale}}} \quad (11)$$

$$hh = \frac{2 \times (h^{\text{gt}})^{\text{scale}}}{(w^{\text{gt}})^{\text{scale}} + (h^{\text{gt}})^{\text{scale}}} \quad (12)$$

$$\text{distance}^{\text{shape}} = \frac{hh \times (x_c - x_c^{\text{gt}})^2}{c^2} + ww \times \frac{(y_c - y_c^{\text{gt}})^2}{c^2} \quad (13)$$

Shape_IoU 损失的形状代价 $\mathcal{Q}^{\text{shape}}$ 的计算为公式(14)所示， θ 控制着形状代价的程度。

$$\mathcal{Q}^{\text{shape}} = \sum_{t=w,h} (1 - e^{-w_t})^\theta, \theta = 4 \quad (14)$$

$$w_w = hh \times \frac{|w - w^{\text{gt}}|}{\max(w, w^{\text{gt}})}, w_h = ww \times \frac{|h - h^{\text{gt}}|}{\max(h, h^{\text{gt}})} \quad (15)$$

式中： w^{gt} 和 h^{gt} 分别代表真实框的宽度和高度； scale 是缩放因子，与数据集中目标的大小相关； ww 和 hh 分别表示水平方向和垂直方向上的权重系数^[21]，其值与真实框的形状有关。相应的边界框回归损失如下：

$$L_{\text{shape_IoU}} = 1 - \text{IoU} + \text{distance}^{\text{shape}} + 0.5 \times \mathcal{Q}^{\text{shape}} \quad (16)$$

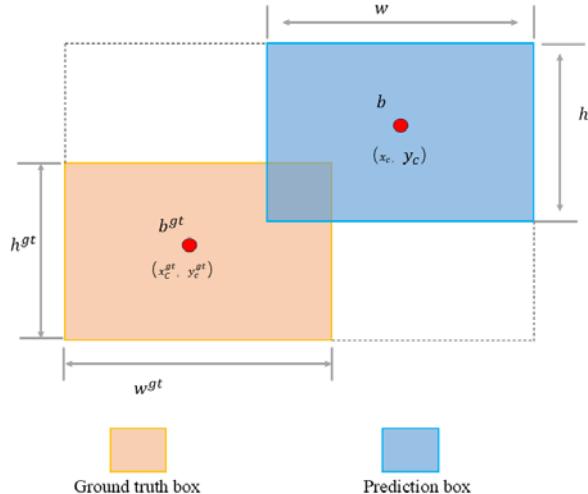


图 5 Shape_IoU 损失

Fig.5 Shape_IoU loss

3 实验结果与分析

3.1 实验平台与参数

该模型是在 NVIDIA GeForce RTX 3090 GPU 上使用 Pytorch 1.8 框架进行搭建，本文实验参数设置如表 1 所示。

表 1 实验参数

Table 1 Experimental parameters	
Parameter	Value
Learning rate	0.01
Weight decay factor	0.0005
Batch size	32
Epoch	300
Image size	512×640

3.2 数据集

本文采用的数据集由 FLIR 公司采集，该数据集中的图像都经过了降分辨率处理，使得其分辨率较低。低分辨率图像通常具有模糊、失真等问题，对算法的性能提出了更高的要求。该数据集是计算机视觉领域广泛使用的红外图像库。原始的 FLIR 数据集包含的标签种类有人、汽车、自行车、摩托车、公交车等 15 类目标，本文挑选了目标数目最多的 3 类：汽车、人和自行车，其具体数量如表 2 所示。

表 2 数据类别及数量
Table 2 Data categories and quantities

Data categories	Quantities
Cars	29691
People	16356
Bicycle	3618

图 6 显示, 本文数据集中红外小目标的图像宽高比例集中在 0.4% 以内, 其占图像面积比例集中在 0.2%, 符合文献[20]对小目标的定义。

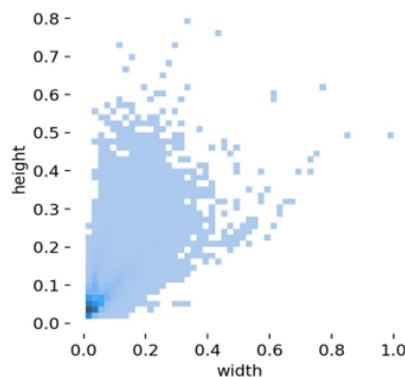


图 6 数据集大小占比
Fig.6 Data set size percentage

3.3 评估指标

为了评估本研究提出的改进算法的效果, 实验使用平均精度 (AP)、召回率 (R)、平均精度的均值 (mAP) 和红外目标图的检测帧数 (FPS) 作为性能评估指标。

精度反映模型检测出来的目标是真实目标的比例, TP 表示正确认别红外数据的目标个数, FP 表示错误识别红外数据的目标个数。精度的表达式如下:

$$p = \frac{TP}{TP + FP} \quad (17)$$

召回率反映总样本中被模型预测正确的比例, FN 表示漏检红外数据集的目标个数。召回率的表达式如下:

$$R = \frac{TP}{TP + FN} \quad (18)$$

检测精度代表模型对于这类目标的检测准确度, 通常以 $P-R$ 图中曲线求积分来计算, 平均精度的表达式如下:

$$AP = \int_0^1 p(r)dr \quad (19)$$

平均精度的均值就是所有类别的平均 AP 值, 在多类别目标检测中用于衡量识别精度。平均精度均值的表达式如下:

$$mAP = \frac{1}{k} \sum_{i=1}^k AP_i \quad (20)$$

目标检测的 FPS 实时性要求是指在单位时间内能够处理的图像帧数, 通常用帧率 (frames per second, FPS) 来衡量。FPS 的计算公式如下:

$$FPS = \frac{\text{Numbers}}{\text{Spend Time}} \quad (21)$$

4 消融实验

4.1 注意力机制消融实验

由于小目标特征不明显, 常出现误检和漏检现象。为了突显小目标特征, 本文在 YOLOv8s_SPD_Small (v8s_S_S) 基础上添加了 EDAM 注意力机制, 该注意力机制采用了组卷积和通道打乱技术, 并与其他主流注意力机制进行了对比, 以验证其在小目标检测中的有效性。

如表 3 所示, 实验结果表明, 加入 SimAM^[22]、CPCA^[23]、SegNeXt^[24]、MLCA^[25]、EDAM 等注意机制相比于原始的 YOLOv8s 模型 mAP 值分别提高了 2.3%、1.8%、2.4%、1.7%、3.06%; 召回率分别提高了 1.32%、2.37%、3.19%、3.29%、5.99%。实验结果表明本文设计的 EDAM 注意力机制优于其他注意力机制。

表 3 注意力机制对比

Table 3 Comparison of attention mechanisms

Model	mAP@0.5	mAP@0.5-0.95		P	R
		P	R		
v8s	86.50%	46.50%	91.08%	78.93%	
v8s_S_S	88.60%	51.30%	92.51%	79.78%	
+SimAM	88.80%	52.60%	91.20%	80.25%	
+CPCA	88.30%	52.80%	92.10%	81.30%	
+SegNeXt	88.90%	52.50%	90.75%	82.12%	
+MLCA	88.20%	51.20%	93.70%	82.22%	
+EDAM	89.56%	52.30%	93.67%	84.92%	

SSE-YOLO 算法的注意力机制改进思路是将 C2f 的输出作为 EDAM 模块的输入。表 4 中的 $Feat_i$ ($i=0,1,2,3$) 分别为颈部 C2f 输出的特征层。

表 4 不同特征层的效果

Table 4 Effects of different feature layers

	mAP@0.5	mAP@0.5-0.95		P	R
		P	R		
v8s	86.50%	46.50%	91.08%	78.93%	
+Feat0	89.56%	52.30%	93.67%	84.92%	
+Feat1	89.39%	50.30%	93.24%	80.47%	
+Feat2	89.19%	47.00%	92.89%	81.23%	
+Feat3	89.35%	50.10%	92.48%	83.29%	

如表 4 实验结果所示, 针对本文实验场景, 将 EDAM 模块嵌入颈部输出的特征中有不同的效果,

本文 YOLOv8s_S_S_EDAM 将模块加入 Feat0 中与小目标检测层进行融合研究。

4.2 模块消融实验

为了验证本文提出的算法对红外小目标的有效性, 进行了一系列的消融实验, 本文实验主要由 mAP@0.5% 及 mAP@0.5%~0.95% 两个评价指标来评估实验效果, 实验结果如表 5 所示, 实验一为原始 YOLOv8s 模型实验结果, 在表中“√”表示针对原始算法的改进点, 在本文中分别进行了实验二、三、四、五组消融实验。

如表 5 所示, 额外添加了一个专门用于小目标检测的层, 将 mAP 值提升了 1.64%, 增强了对小目标的检测能力。在实验三中, 基于实验二的结果, 我们替换了原始 YOLOv8s 算法的卷积层, 引入了 SPD_Conv。这一改动使模型的平均精度比实验二提高了 0.46%, 更有效地解决了细粒度信息丢失和特征学习效率低的问题。实验四在前两组实验的基础上设计了一种高效的双注意力机制 (EDAM), mAP 相比实验三的平均精度提升了 0.96%, 实验五在实验四的基础上更改了损失函数, 实验结果表明该改进算法比

实验四的平均精度提升了 0.25%。

根据实验结果, 每项算法的优化对最终结果都产生了正向促进作用。此外, 通过对 4 项算法的组合进行改动应用, 使得该算法的平均精度达到了最优水平。

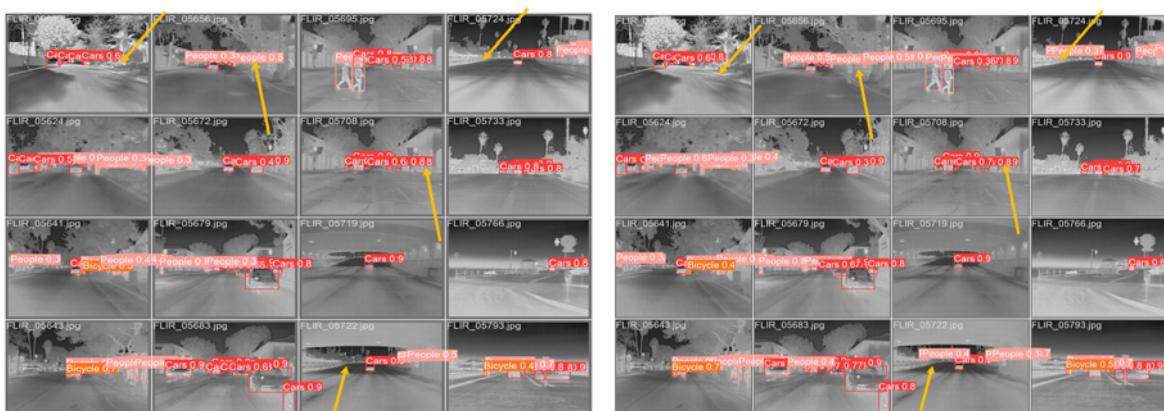
图 7 表示选取若干张红外小目标图像进行对比, 图中可以看出红外小目标在不同场景下的检测效果, 在红外数据集中, 通过对比可以看出, YOLOv8s 算法漏检严重, 很多微小的目标未被检测出来, 而本文算法能够非常精确地将微小目标全部检测出来, 说明改进后的算法对小目标检测效果显著, 漏检率低, 泛化能力强, 对小目标检测具有通用性。

图 8 为 P_R 曲线对比图, 曲线与坐标轴围成的面积越大, 表明算法在精确率和召回率方面达成了更好的平衡, 从而反映了改进算法的优越性。从图中可以看出, 提出的 SSE-YOLO 模型对每个目标类别都有不同程度的提升, Cars 类 AP 值从 88.1% 到 90.1%, 提升 2 个百分点; People 类 AP 值从 88.3% 到 91.4%, 提升 3.1 个百分点; Bicycle 类 AP 值从 83.1% 到 87.9%, 提升 4.8 个百分点; 3 种红外目标检测的 AP 值均有提升, 体现出了模型良好的泛化能力。

表 5 改进 YOLOv8s 的消融实验

Table 5 Experimental on ablation of improved YOLOv8s

Model	+head	SPD	EDAM	Shape_IoU	mAP@0.5	mAP@0.5-0.95
Experiment 1	×	×	×	×	86.50%	46.50%
Experiment 2	√	×	×	×	88.14%	47.40%
Experiment 3	√	√	×	×	88.60%	51.30%
Experiment 4	√	√	√	×	89.56%	52.30%
Experiment 5	√	√	√	√	89.80%	53.00%



(a) YOLOv8s

(b) Improved YOLOv8s

图 7 YOLOv8s 和改进后的 YOLOv8s 在 FLIR 数据集上的可视化结果对比

Fig. 7 Comparison of visualization results between YOLOv8s and improved YOLOv8s on FLIR dataset

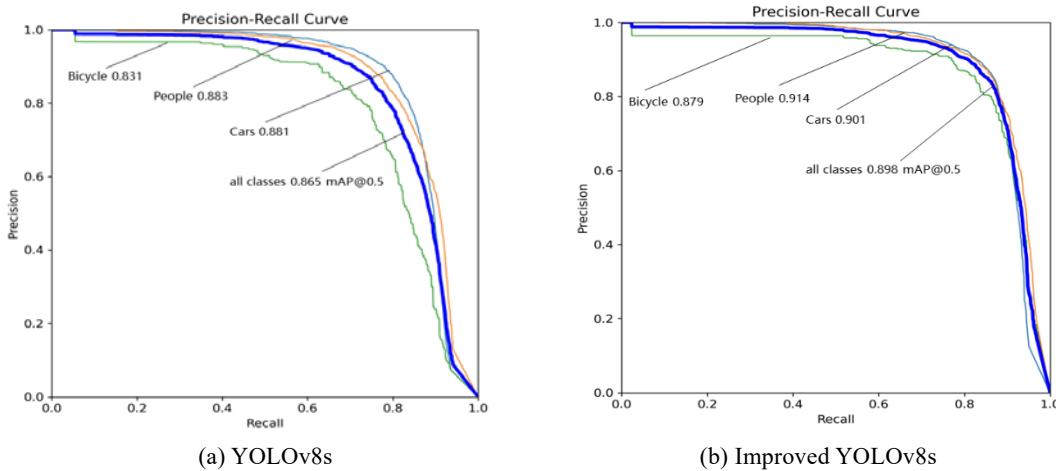


图 8 初始的 YOLOv8s 和改进后 YOLOv8s 在 FLIR 数据集上的 P_R 结果对比

Fig.8 Comparison of P_R results between initial YOLOv8s and improved YOLOv8s on FLIR dataset

4.3 艾睿光电红外数据集实验

为了验证改进模型在不同数据集上的优秀性能, 本研究基于艾睿光电公司拍摄的红外图像数据集进行验证, 该数据图像为 640 像素 \times 418 像素, 共计 11952 张图像, 并进行了详细的实验分析, 详情如表 6 所示。

表 6 艾睿光电数据集标签信息

Table 6 IRay dataset labeling information

Data categories	Quantities
Car	92979
Truck	10618
Person	6170
Bus	3968
Cyclist	4626
Bike	47478

实验结果如表 7 所示, 可以看出, 将新的数据引

用到本文所提网络中达到了不错的效果, 提出的 SSE-YOLO 算法的平均精度的均值达到了 92.1%, 与原始的 YOLOv8s 相比, 平均精度提升了 2.9%, 召回率达到了 88.3%, 相比于原始模型提高了 4.7%, 实验选取若干张红外小目标图像进行对比, 检测性能对比图如图 9 所示。

表 7 艾睿光电数据集检测性能对比

Table 7 IRay dataset detection performance comparison

Model	mAP@0.5	mAP@0.5-	
		0.95	P R
YOLOv8s	89.2%	58.8%	86.1% 83.6%
SSE-YOLO	92.1%	62.4%	86.8% 88.3%

从实验结果可以看出, 在复杂背景下原始的算法不仅精度低而且存在严重的漏检情况, 经过改进后, 有效提高精度同时降低了漏检率。

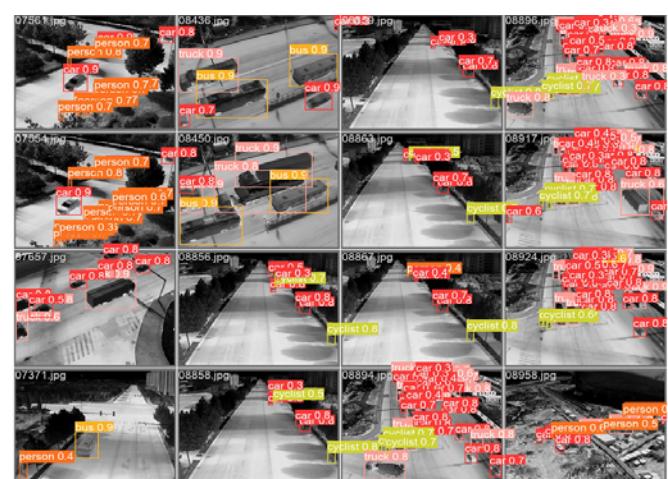
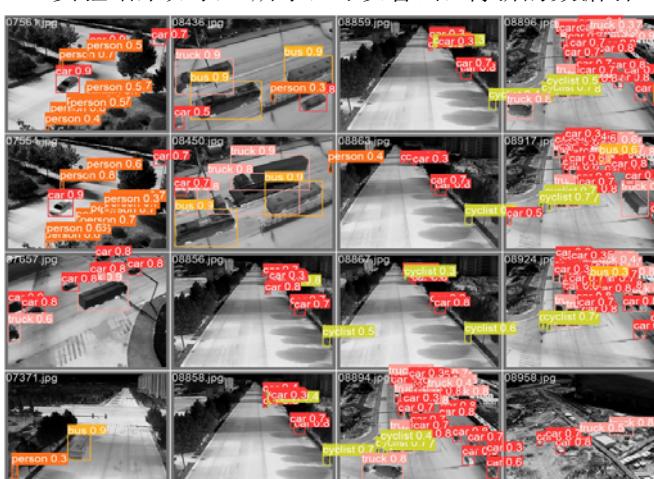


图 9 YOLOv8s 和改进后的 YOLOv8s 在艾瑞光电数据集上的可视化结果对比

Fig. 9 Comparison of visualization results between YOLOv8s and improved YOLOv8s on IRay dataset

4.4 不同目标检测算法的结果比较

SSE-YOLO 算法是基于 YOLOv8s 模型基础之上的改进。为验证其在 FLIR 数据集上出色的检测性能, 将 SSE-YOLO 与现阶段主流的一阶段目标检测算法进行了对比, 包括 YOLOv3-tiny^[26]、YOLOv5s、在 YOLOv5s 的基础上添加一个小目标检测层的 YOLOv5s_S 以及 YOLOv6s^[27], 实验结果如表 8 所示。

表 8 不同目标检测算法的结果对比

Table 8 Comparison of results of different target detection algorithms

Model	mAP@0.5	mAP@0.5-0.95	P	R	FPS	Detection time/ms
v3-tiny	60.10%	40.13%	72.92%	53.31%	218.8	1.1
v5s	81.20%	41.80%	88.50%	74.70%	83.1	2.4
v5s_S	82.6%	45.90%	90.10%	78.81%	80.1	2.6
v6s	84.0%	47.8%	87.10%	77.60%	100.5	1.4
v8s	86.50%	46.50%	93.08%	78.93%	91.4	1.6
v8s_S	88.14%	48.56%	92.76%	81.81%	84.2	2.3
SSE-YOLO	89.80%	53.00%	92.67%	84.73%	82.8	2.2

5 结论

本文提出了一种 SSE-YOLO 的模型对红外图像中的目标进行识别, 优化了红外目标检测精度低的问题。首先, 引入了一种专为低分辨率图像和小目标细化检测设计的深度非跨步卷积模块。该模块旨在防止在检测过程中丢失细微信息, 同时提高了特征学习的效率; 此外, 增加了用于小目标的额外检测层, 旨在提高模型对红外小目标的检测能力; 其次, 在颈部 C2f 输出的特征层上设计了一种高效的双注意力机制, 使模型更好地关注重要的特征, 抑制一些无关紧要的通道; 最后使用了 Shape_IOU 损失函数来聚焦边框自身形状与自身尺度, 进一步提高边框回归的精确度。分别在不同数据集上进行验证, 均达到了良好的效果。

参考文献:

- [1] 张睿, 李允臣, 王家宝, 等. 基于深度学习的红外目标检测综述[J]. 计算机技术与发展, 2023, 33(11): 1-8.
ZHANG Rui, LI Yunchen, WANG Jiaobao, et al. A review of infrared target detection based on deep learning[J]. *Computer Technology and Development*, 2023, 33(11): 1-8.
- [2] 薛如翔, 卫俊杰, 周华伟, 等. 基于可见光-红外跨域迁移的红外弱小目标检测[J]. 计算机科学, 2024, 51(10): 287-294.
XUE Ruxiang, WEI Junjie, ZHOU Huawei, et al. Infrared weak target detection based on visible-infrared cross-domain migration[J]. *Computer Science*, 2024, 51(10): 287-294.
- [3] 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.
- [4] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [5] REN S, HE K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(6): 1137-1149.
- [6] LIN T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [7] 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.
- [8] LIU W, Anguelov D, Erhan D, et al. Ssd: single shot multibox detector[C]// Proceedings of Computer Vision-ECCV, 14th European Conference, Part I, 2016: 21-37.
- [9] 杨雪, 修吉宏, 刘小嘉, 等. 航空机载红外图像的车辆目标自主检测识别[J]. 激光与红外, 2023, 53(12): 1877-1884.
YANG Xue, XU Jihong, LIU Xiaojia, et al. Autonomous detection and recognition of vehicle targets from airborne infrared images[J]. *Laser and Infrared*, 2023, 53(12): 1877-1884.
- [10] 李晓佩, 张寅宝, 李严培, 等. 一种基于 YOLOv5s 的红外图像目标检测改进算法[J]. 激光与红外, 2023, 53(7): 1043-1051.
LI Xiaopei, ZHANG Yinbao, LI Yanpei, et al. An improved algorithm for infrared image target detection based on YOLOv5s[J]. *Laser and*

- Infrared*, 2023, **53**(7): 1043-1051.
- [11] 鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, **39**(7): 253-260.
- JU Meran, LUO Haibo, WANG Zhongbo, et al. Improved YOLO V3 algorithm and its application to small target detection[J]. *Journal of Optics*, 2019, **39**(7): 253-260.
- [12] Rejin Varghese, Sambath M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness[C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), 2024: DOI: 10.1109/ADICS58448.2024.10533619.
- [13] 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.
- [14] Sunkara R, LUO T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2022: 443-459.
- [15] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//Proc. of IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2018: 6848-6856.
- [16] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2018, **11211**: 3-19.
- [17] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, DOI: 10.1109/CVPR.2018.00745.
- [18] DU S, ZHANG B, ZHANG P, et al. An improved bounding box regression loss function based on CIOU loss for multi-scale object detection[C]//International Conference on Pattern Recognition and Machine Learning of IEEE, 2021: DOI:10.1109/PRML52754.2021.9520717.
- [19] YU J, JIANG Y, WANG Z, et al. Unitbox: an advanced object detection network[C]//Proceedings of the 24th ACM International Conference on Multimedia, 2016: 516-520.
- [20] ZHANG H, ZHANG S. Shape-IoU: more accurate metric considering bounding box shape and scale[J]. arXiv preprint arXiv:2312.17663, 2023.
- [21] CHEN C Y, LIU M Y, TUZEL O, et al. R-CNN for small object detection[C]//Proceeding of Asian Conference on Computer Vision, 2016, **10115**: 214-230.
- [22] YANG L, ZHANG R Y, LI L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine Learning, PMLR, 2021: 11863-11874.
- [23] HUANG H, CHEN Z, ZOU Y, et al. Channel prior convolutional attention for medical image segmentation[J]. arXiv preprint arXiv:2306.05196, 2023.
- [24] GUO M H, LU C Z, HOU Q, et al. Segnext: rethinking convolutional attention design for semantic segmentation. arXiv 2022[J]. arXiv preprint arXiv:2209.08575, 2022.
- [25] WAN D, LU R, SHEN S, et al. Mixed local channel attention for object detection[J]. *Engineering Applications of Artificial Intelligence*, 2023, **123**: 106442.
- [26] Adarsh P, Rathi P, Kumar M. YOLO v3-tiny: object detection and recognition using one stage improved model[C]//2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020: 687-694.
- [27] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.