计算机科学
COMPUTER SCIENCE

基于深度特征强化与路径聚合优化的目标检测

王晓峰，黄俊俊，谭文雅，沈紫璇

引用本文

王晓峰，黄俊俊，谭文雅，沈紫璇. 基于深度特征强化与路径聚合优化的目标检测[J]. 计算机科学，2025，52(11)：184-195.

WANG Xiaofeng, HUANG Junjun, TAN Wenya, SHEN Zixuan. Object Detection Based on Deep Feature Enhancement and Path Aggregation Optimization [J]. Computer Science, 2025, 52(11): 184-195.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.241100107

基于深度特征强化与路径聚合优化的目标检测

王晓峰　黄俊俊　谭文雅　沈紫璇
武汉科技大学计算机科学与技术学院　武汉 430070
武汉科技大学智能信息处理与实时工业系统湖北省重点实验室　武汉 430070
(wangxiaofeng@wust.edu.cn)

摘　要　在深度网络的前馈过程中,输入数据的特征信息会被抽象和压缩,导致部分对于目标检测关键的特征信息被弱化。基于YOLOv11n,提出了深度特征强化与路径聚合优化的目标检测方法。首先,设计全局-局部特征增强模块GLFEM(Global-Local Feature Enhancement Module),结合特征图局部特征与全局特征,强化深层网络特征的表达能力。然后,设计自适应特征增强模块AFEM(Adaptive Feature Enhancement Module),根据特征的可靠性动态增强深层网络的特征提取能力。最后,对路径聚合特征金字塔网络进行优化,融合了不同层次之间的特征信息,减少了层次之间的语义信息差。在VisDrone,NWPU VHR-10和TinyPerson这3个公共数据集上的实验结果表明,该方法的平均检测精度相较于当前先进的目标检测器均有所提升。在自建数据集AirportTiny上进行实验,该方法同样取得了不错的效果,具有良好的泛化能力。

关键词:目标检测;深层网络;路径聚合;特征信息;特征强化
中图分类号　TP391.4

Object Detection Based on Deep Feature Enhancement and Path Aggregation Optimization

WANG Xiaofeng,HUANG Junjun,TAN Wenya and SHEN Zixuan
School of Computer Science and Technology,Wuhan University of Science and Technology,Wuhan 430070,China
Hubei Provincial Key Laboratory of Intelligent Information Processing and Real-time Industrial System,Wuhan University of Science and Technology,Wuhan 430070,China

Abstract　In deep networks,the feature information of the input data is gradually abstracted and compressed during the feed-forward process,resulting in some of the feature information that is crucial for object detection being diluted or lost. Based on YOLOv11n,an object detection method with deep feature enhancement and path aggregation optimization is proposed. Firstly,GLFEM is designed to combine the local features of the feature map with the global features to strengthen the expression ability of the deep network features. Then,AFEM is designed to dynamically enhance the feature extraction ability of the deep network according to the reliability of the features. Finally,the path aggregation feature pyramid network is optimized to fuse the feature information between different levels and reduce the semantic information difference between levels. Experimental results on three public datasets,VisDrone,NWPU VHR-10,and TinyPerson,show that the average detection accuracy of the proposed method is improved compared to current state-of-the-art object detectors. Experiments on the self-built dataset AirportTiny also show the proposed method achieves good performance,it has good generalisation ability.

Keywords　Object detection,Deep network,Path aggregation,Feature information,Feature enhancement

1　引言

目标检测是计算机视觉中的一项基本任务,旨在通过定位边界框并预测相应的分类分数来检测图像或视频中感兴趣的目标[1],其在医学图像诊断[2-3]、智能交通系统[4-5]和农业自动化[6-8]等领域都有着广泛的应用。

现代目标检测器分为基于两阶段和基于单阶段两大类[9-11]。两阶段检测器以RCNN系列[12-14]为例,首先需要生成候选区域,然后在候选区域中进一步分类并回归边界框。这类方法具有较高的检测精度,但计算复杂度高、速度较慢。单阶段检测器主要以YOLO系列[15-21]和SSD[22]为例,此类方法在单一网络结构中直接完成目标定位和分类,无需生成候选区域,处理速度快,实时性较强。尽管这些方法取得了一定成功,但仍然存在诸多挑战。当输入数据经过逐层特征提

到稿日期:2024-11-18　返修日期:2025-02-16
基金项目:国家自然科学基金(62302351);湖北省自然科学基金(2022CFB018)
This work was supported by the National Natural Science Foundation of China(62302351) and Natural Science Foundation of Hubei Province(2022CFB018).
通信作者:黄俊俊(3514469387@qq.com)

取和空间变换后，大量特征信息会被削弱甚至丢失。信息退化或损失会引发梯度流出现偏差，导致深度网络在目标和输入之间建立错误关联，最终导致训练后的模型产生错误的预测。

最近，基于 YOLO 系列端到端的检测器[23-26]通过增强感受野来缓解深层网络目标特征信息减弱的问题。然而，这些方法普遍缺乏全局上下文整合机制，对长距离的全局依赖建模不足，难以有效捕获图像中全局的关系，导致目标检测性能不佳。此外，它们未能有效处理路径聚合特征金字塔网络中不同层次间潜在的信息冲突，而这种冲突会削弱跨层特征的融合效果，降低特征的判别能力，最终可能导致误检和漏检的情况增加。

针对上述问题，本文基于 YOLOv11n，提出了 YOLO-FE-PA(Feature Enhancement and Path Aggregation)目标检测算法，通过增强深层网络特征并优化路径聚合特征金字塔网络来提高目标检测精度。具体来说，首先设计了全局-局部特征增强模块 GLFEM，其中包含局部特征提取分支和全局上下文增强分支。全局分支旨在提取全局边缘或纹理信息，以弥补局部分支有限的感受野。然后，设计了自适应特征增强模块 AFEM，通过稠密连接实现特征复用与传递，增强网络的特征提取能力，并引入提前停止策略，动态终止可靠性高的特征图的特征增强过程，减少计算冗余。最后，提出双向特征融合网络 BFFN(Bidirectional Feature Fusion Network)，用于融合路径聚合特征金字塔网络中不同层次的特征，使不同层次之间的特征信息可以直接交互，避免了多阶段传输中的信息丢失或退化。

本文的创新点如下：

1)提出了全局-局部特征增强模块 GLFEM，融合特征图的局部特征和全局上下文特征，减少卷积操作对局部感受野的依赖，避免上下文语义信息丢失，增强特征图的表达能力；

2)设计了自适应特征增强模块 AFEM，通过稠密连接机制与提前停止策略动态增强网络的特征提取能力，还设计了置信度估计器来评估特征图的可靠性；

3)构建了双向特征融合网络 BFFN，该网络通过融合路径聚合特征金字塔网络中不同层级的特征，有效缩小了层次间的语义差距。

2 相关工作

2.1 YOLOv11

YOLOv11 作为现阶段 YOLO 系列算法的最新版本，由 Ultralytics 推出，属于单阶段目标检测模型[27]，相较于 YOLOv8，YOLOv11 在速度和精度方面均有提升。其主要改进包括：将主干与颈部网络中的 C2f 模块升级为 C3K2 模块，在 SPPF[28]模块之后引入新的 C2PSA[1)](Cross Stage Partial with Spatial Attention)块，以及将头部网络的卷积层用深度卷积层进行替换。

图 1 为 YOLOv11 的网络结构图。YOLOv11 整体由 3 部分组成：主干(Backbone)特征提取网络、颈部(Neck)路径聚合特征金字塔网络和头部(Head)检测网络。主干网络主要由 CBS 层、C3K2 块、SPPF 层以及 C2PSA 块组成。CBS 层对输入图像进行下采样，在逐步减少空间维度的同时增加特征图的深度，是主干特征提取网络的基础。YOLOv11 不再使用 YOLOv8 中使用的 C2f 块，而是引入了更加高效的 C3K2 块，该块是 Cross Stage Partial Bottleneck 的优化版本，由两个较小的卷积(内核大小为 2)组成，以降低计算成本，同时保持性能。SPPF 层利用最大池化层跨多个尺度对特征图进行空间池化，从而进一步提取多尺度特征。C2PSA 增强了特征图中的空间注意力，这有助于模型专注特定感兴趣的区域，从而提高对不同大小和位置物体的检测准确率。颈部路径聚合特征金字塔网络采用上采样和拼接层来聚合不同分辨率的特征图，使模型能够有效捕获多尺度信息。头部检测网络通过结合深度卷积和核大小为 1 的点卷积，构成深度可分离卷积。深度可分离卷积替代了原 YOLOv8 检测头中的 3×3 卷积，极大地降低了模型的计算复杂度和减少了参数量。头部检测网络的主要功能是定位目标的边界框坐标并预测类别信息，其结构包括 3 个检测层(P3，P4 和 P5)，分别处理来自颈部网络的不同尺度特征图，以实现对不同大小目标的检测。

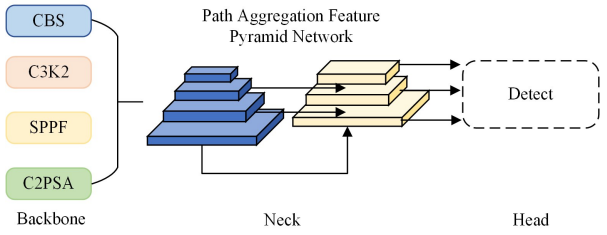


图 1 YOLOv11 总体架构

Fig. 1 Overall architecture of YOLOv11

虽然 YOLOv11 在 YOLOv8 的基础上进行了大量轻量化改进，在不损失检测精度的同时有效减少了计算量和参数量，但其主干网络的 C3K2 块以及 C2PSA 块仍然是由卷积层搭建的，卷积层在提取图像局部特征方面具有优势，但对全局特征的提取却有一定的局限性[29]，从而影响对复杂场景或多目标的理解，导致目标漏检和误检。而本文方法通过设计全局上下文增强分支，在局部特征的基础上结合全局特征，为特征图提供了更加全面的语义信息。

2.2 深度网络特征

深度网络通过增加隐藏层，可以学习到更加抽象和复杂的特征，提升了模型的表达能力[30-32]。更深的网络，意味着更好的非线性表达能力，可以学习更加复杂的变换，从而可以拟合更加复杂的特征输入。

在深度网络中，输入数据在前馈过程中容易丢失信息，这种现象通常被称为信息瓶颈[21]。当输入数据经过深度网络的逐层特征提取和空间变换后，会不可避免地丢失一些细节信息。在训练开始前，网络中的权重和偏置是随机初始化的，在训练过程中逐渐调整这些参数，但在初始阶段，它们可能导致信息在传递过程中被扭曲或丢失，这些信息对于目标检测

1) https://learnopencv.com/yolo11/

任务来说是至关重要的。输入数据中的信息丢失可能导致模型无法准确捕捉数据的真实分布和特征，从而降低模型的预测性能和分类准确性。信息丢失也可能导致梯度消失或爆炸等问题，使得模型在训练过程中难以收敛到最优解。

Que 等[23]使用深度可分离卷积和空洞卷积设计双残差感受野增强模块，用于扩大主干网络的感受野特征，缓解了下采样过程中目标信息丢失的问题。Li 等[24]提出多层次特征融合模块，通过多感受野卷积和注意力机制融合目标的多尺度特征信息，以增强目标的特征表示。Ni 等[26]提出并行多尺度特征提取模块，将输入特征馈入两个并行的扩张卷积和可变形卷积中，生成具有不同感受野的自适应权重，并将生成的权重信息融合到浅层特征图中，增强了目标的特征提取能力。这些方法大多采用卷积变体和注意力机制设计对应的模块，虽然表现出更好的特征学习能力，但这些卷积变体对全局特征提取不足，并且注意力机制的引入对硬件资源要求高，参数较多，也容易导致梯度消失或梯度爆炸。

与以上方法相比，本文方法在深层网络设计全局上下文分支，以增强网络提取全局特征的能力。同时，引入稠密连接机制，有效缓解了梯度消失或梯度爆炸问题。此外，所提出的提前停止策略也可以有效降低推理成本。

2.3 路径聚合特征金字塔网络

由于图像中的物体可能具有不同尺度，因此构建能够融合高级和低级特征的多尺度特征图变得尤为重要[33]。鉴于图像中目标尺度的多样性，路径聚合特征金字塔网络 PAFPN[34](Path Aggregation Feature Pyramid Network)通过横向连接和多尺度特征融合机制，确保了不同尺度的高效利用。随后，还激发了其一系列改进版本的发展，如双向特征金字塔网络 BiFPN[35](Bidirectional Feature Pyramid Network)、全局特征金字塔网络 GFPN[36](Global Feature Pyramid Network)和跨层级强化特征金字塔网络 CFEPN[37](Cross-level Feature-Fusion Enhanced Pyramid Network)，以及其他的一些变体[38]。

BiFPN 通过重复双向跨尺度连接和带权重的特征融合机制，不仅增强了特征间的信息流动，还确保了关键特征在融合过程中的有效保留，进一步提高了目标检测的准确率。GFPN 利用自注意力机制级联多个全局特征融合模块，并充分利用双路径主干相互补充的空间及语义特征，促进各个尺度分支融合全局多尺度信息，从而增强了顶部特征空间几何表征能力。CFEPN 以跨层级的方式融合 CBAM 注意力机制来改进 PAFPN，减弱了融合时噪声信息对深层特征的干扰，增强了模型对目标的检测能力。

这些金字塔网络通过构建复杂的连接路径增强了特征之间的交流，并融合注意力机制强化了重要特征的表示。然而，它们在融合不同层次之间的特征方面仍然存在语义信息差这一局限性。尽管 PAFPN 已经尝试通过引入自下而上的路径来突破这一局限性，但不同层级特征之间仍然存在特征矛盾。因此，本文在 PAFPN 的基础上，提出双向特征融合网络，通过对不同层级特征进行融合来减小层级间的特征矛盾。

3 YOLO-FEPA 方法

3.1 网络整体架构

YOLO-FEPA 由主干(Backbone)、颈部(Neck)和头部(Head)3 个部分组成，模型的整体结构如图 2 所示。

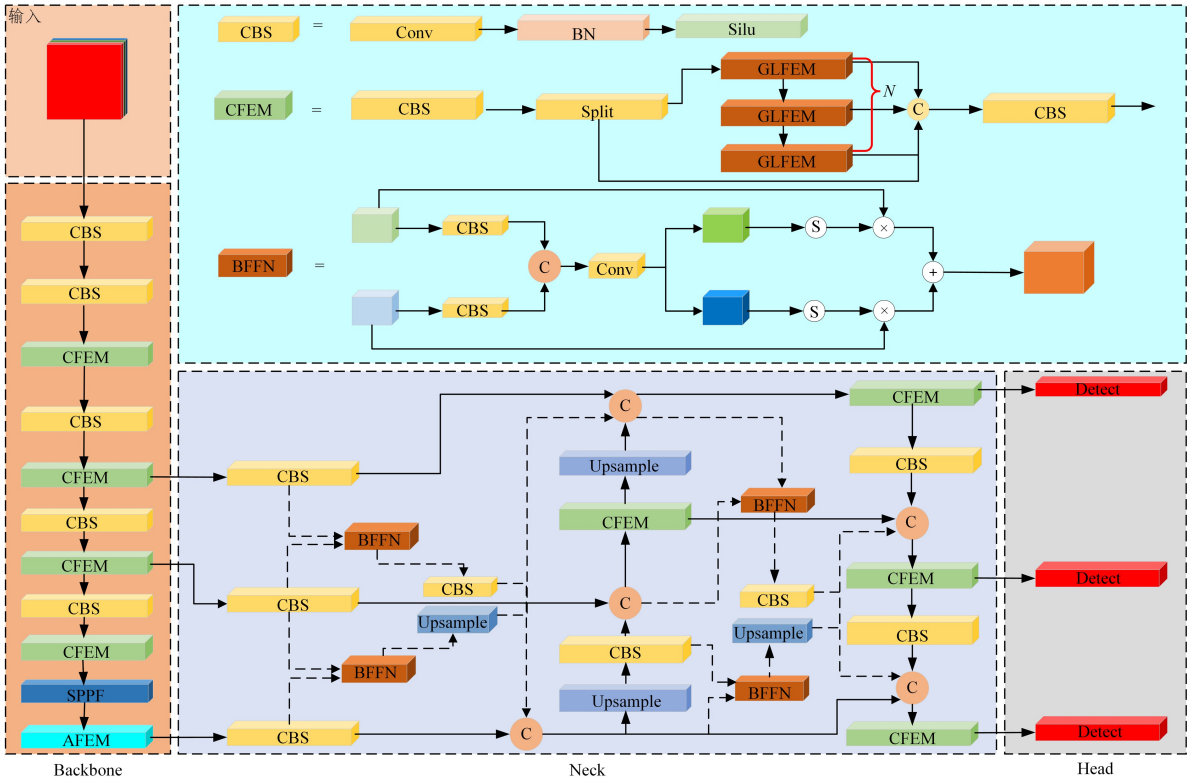


图 2 YOLO-FEPA 总体架构

Fig. 2 Overall architecture of YOLO-FEPA

首先，输入图像通过主干网络进行深度特征提取。主干网络包括 CBS（Convolution-BN-SiLU）模块和 CFEM 模块（Contextual Feature Enhancement Module）。CBS 模块对图像特征进行特征提取，CFEM 模块对图像特征进行特征增强。然后，主干网络输出的特征图再输入 AFEM 模块进一步对特征图进行自适应特征增强。接着，颈部网络 OPAFPN（Optimized Path Aggregation Feature Pyramid Network）将主干网络输出的不同尺度的特征图进行特征融合，将多尺度特征转换为一系列图像特征。最后，将经过 OPAFPN 处理后的不同尺度的图像特征传递至头部网络，用于判断物体类别、生成预测框和置信度分数。

3.2 全局-局部特征增强模块

卷积层在特征提取阶段通常提取的是图像局部细节特征，而忽略了图像的全局特征[39]。此外，神经网络在前向传播过程中，随着网络层数加深，特征图在传递过程中可能会遭遇信息瓶颈，导致部分特征信息逐渐减弱或丢失[40]。

基于以上考虑，本文提出了全局-局部特征增强模块 GLFEM，如图 3 所示。它引入了一个局部特征提取分支并行一个全局上下文增强分支，用于在深层网络中增强特征图的特征。具体来说，假设输入特征图为 $\boldsymbol{X}\in\mathbb{R}^{C\times H\times W}$，其中 C 为通道数，H 和 W 为特征图的高和宽。

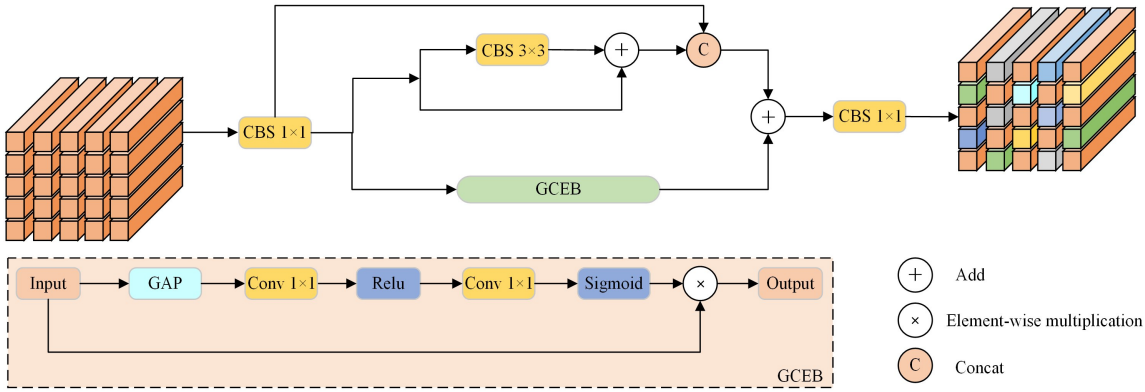


图 3 全局-局部特征增强模块

Fig. 3 Global-local feature enhancement module

1）局部特征提取分支

首先，输入特征图 $\boldsymbol{X}$，先用一层 1×1 的卷积层调整特征图通道数。

$$F(\boldsymbol{X})=Conv_{1\times1}*\boldsymbol{X} \tag{1}$$

其中，$F(\boldsymbol{X})\in\mathbb{R}^{C'\times C\times1\times1}$，$C'$ 为调整后的通道数，$*$ 表示卷积运算。然后使用 3×3 的卷积层对特征图 $F(\boldsymbol{X})$ 进行局部特征提取，以提取局部上下文特征 $\boldsymbol{X}_{\text{local}}$。

$$\boldsymbol{X}_{\text{local}}=Conv_{3\times3}*F(\boldsymbol{X}) \tag{2}$$

其中，$\boldsymbol{X}_{\text{local}}\in\mathbb{R}^{C'\times C'\times3\times3}$。接下来使用残差连接防止梯度消失或梯度爆炸，最后对输入特征图和输出特征图进行拼接，以丰富局部上下文特征表示。

$$\boldsymbol{F}_{\text{local}}=Concat(F(\boldsymbol{X})+\boldsymbol{X}_{\text{local}},F(\boldsymbol{X})) \tag{3}$$

2）全局上下文增强分支

为了增强局部上下文有限的感受野，并行引入了全局上下文增强分支 GCEB（Global Context Enhanced Branch），通过全局平均池化操作对特征图进行全局空间聚合，得到全局特征向量 $G(F(\boldsymbol{X}))\in\mathbb{R}^{C}$。

$$G(F(\boldsymbol{X}))=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F(\boldsymbol{X})(i,j) \tag{4}$$

即每个通道的平均值。全局平均池化后的特征通过 1×1 卷积进行通道压缩，再经过 ReLU 激活函数提升网络的非线性能力。

$$\boldsymbol{Y}_1=\text{ReLU}(Conv_{1\times1}*G(F(\boldsymbol{X}))) \tag{5}$$

其中，$Conv_{1\times1}$ 是 1×1 卷积的权重矩阵；ReLU 是修正线性单元激活函数，用于引入非线性特征表达。接下来，$\boldsymbol{Y}_1$ 经过另一个 1×1 卷积生成通道注意力权重，之后通过 Sigmoid 函数，将结果限制在[0，1]。

$$\boldsymbol{Y}_2=\sigma(Conv_{1\times1}*\boldsymbol{Y}_1) \tag{6}$$

其中，σ 是 Sigmoid 激活函数；$\boldsymbol{Y}_2$ 是每个通道的权重，表示不同通道的重要程度。再将输入特征 $F(\boldsymbol{X})$ 与生成的通道注意力权重 $\boldsymbol{Y}_2$ 进行逐元素相乘，对原始特征进行重新加权，以增强对有用特征的表达，同时抑制不重要特征，从而实现特征增强。

$$\boldsymbol{Y}_{\text{global}}=F(\boldsymbol{X})\odot\boldsymbol{Y}_2 \tag{7}$$

其中，$\odot$ 表示逐元素相乘。接着，将局部分支与全局分支进行融合，实现特征的累加和互补，提升特征图的特征表达能力。

$$\boldsymbol{X}_{\text{fused}}=\boldsymbol{F}_{\text{local}}+\boldsymbol{Y}_{\text{global}} \tag{8}$$

最后，经过 1×1 卷积操作将特征重新映射，以实现通道扩展，并增强特征的非线性表达。

3.3 自适应特征增强模块

在 YOLOv11n 中，输入特征图首先经过主干网络完成深层次特征提取，随后被传递至 C2PSA 模块。C2PSA 模块通过多头注意力机制[41]，增强了模型的特征表达能力。该模块通过级联多个 PSABlock 块对特征进行提取。然而，随着网络深度的增加，在反向传播过程中，梯度信息可能会逐渐衰减或爆炸，使得深层网络的训练难度加大。同时，多层次的特征传递可能导致特征退化问题，削弱模型的检测性能。此外，YOLOv11n 主干网络输出的特征图通道数高达 1 024 通道，对于已提取到良好特征的特征图来说，将这些特征图逐层传递至多个 PSABlock 块中进行重复特征提取，会造成计算冗余的问题。

为了解决上述问题，本文提出了自适应特征增强模块 AFEM，该模块能够增强主干网络输出的特征，并减少计算冗

余，如图 4 所示。AFEM 由自适应密集网络增强模块 ADEM（Adaptive DenseNet Enhancement Module）和卷积层组成。假设由主干网络输出的特征为 $\boldsymbol{F}_b \in \mathbb{R}^{C\times H\times W}$，其中 C 为通道数，H 和 W 为特征图的高和宽。首先通过 1×1 卷积进行标记：

$$\boldsymbol{F}_0 = Conv_{1\times1} * \boldsymbol{F}_b \tag{9}$$

其中，$*$ 表示卷积运算。然后，标记特征通过 L 层 PSABlock 块进行特征增强，并利用置信度估计器对其进行特征筛选。

$$\boldsymbol{F}_{L+1} = H_{\text{PSABlock}}([E(\boldsymbol{F}_0), E(\boldsymbol{F}_1), \cdots, E(\boldsymbol{F}_L)]) \tag{10}$$

其中，$[\boldsymbol{F}_L]$ 表示所有之前层的特征拼接，E 表示置信度估计器。最后，将 $\boldsymbol{F}_{L+1}$ 与 $\boldsymbol{F}_b$ 进行拼接，以丰富 $\boldsymbol{F}_{L+1}$ 的特征表示，再采用 1×1 卷积层对特征维度进行对齐，得到输出特征 $\boldsymbol{F}_g$。

$$\boldsymbol{F}_g = Conv_{1\times1}(Concat(\boldsymbol{F}_b, \boldsymbol{F}_{L+1})) \tag{11}$$

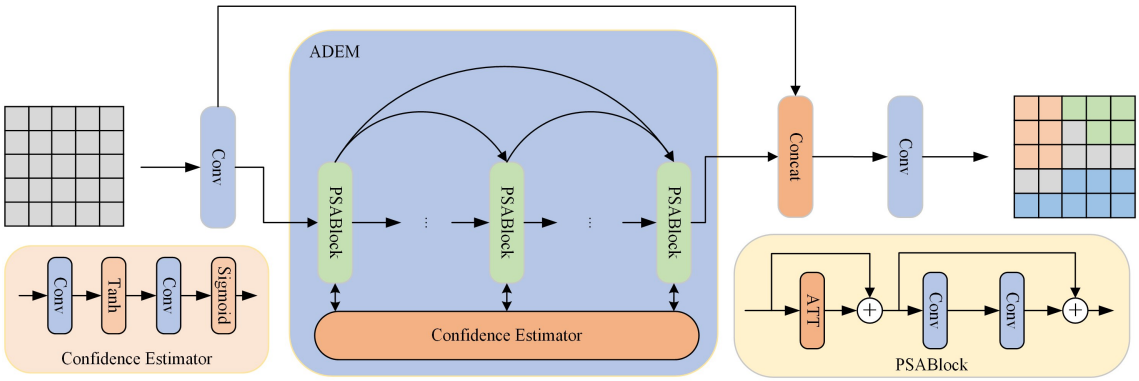


图 4 自适应特征增强模块

Fig. 4 Adaptive feature enhancement module

ADEM 模块包括 PSABlock 块以及用于特征筛选的置信度估计器。ADEM 模块在所有的 PSABlock 之间引入了稠密连接，以增强网络的特征提取能力，并缓解网络的梯度消失或梯度爆炸问题。置信度估计器通过积累置信度分数来生成二进制停止掩码，自适应地停止特征图进行进一步处理，从而减少计算冗余。

1）稠密连接机制

稠密连接[42]的核心特点是每一层的输出都会作为下一层的输入传递给后续的所有层。这意味着每一层都可以直接访问前面所有层的特征。这种连接方式使得模型能够复用前面提取的特征，而不用重复学习相同的信息。对于深层神经网络，复用前面层次的特征能够有效避免信息丢失，保持特征的完整性，进而增强整体的特征表达，并且每一层的输出都直接传递到后续的所有层，可以缓解梯度消失或梯度爆炸问题。

2）提前停止策略

为了避免已提取到良好特征的特征图继续被输入 PSABlock 块中造成计算冗余，提出了一种提前停止策略。其核心思想是：随着网络的加深，对于质量较好的特征图，可以自适应地停止它们继续进入 PSABlock 模块进行进一步计算。那么，如何准确地获取停止信号呢？

具体而言，采用置信度估计器来评估特征图的质量。置信度越高，表明特征图的质量越好。为了在不同网络深度下停止特征图继续进行计算，定义了一个置信度阈值 K，置信度估计器根据设定的阈值判断是否应该停止对当前特征图的进一步处理。

置信度估计器：假设第 n 个 PSABlock 的特征图为 $\boldsymbol{F}^n \in \mathbb{R}^{H\times W\times C}$，$\boldsymbol{F}^n$ 的置信度为 $\boldsymbol{C}^n \in \mathbb{R}^{H\times W\times 1}$，它是由置信度估计器测量而来。该估计器由 Conv-Tanh-Conv-Sigmoid 层组成：

$$\boldsymbol{C}^n = Sigmoid(Conv(Tanh(Conv(\boldsymbol{F}^n)))) \tag{12}$$

其中，$Conv$ 是 3×3 卷积层。

受不确定性驱动损失[43]的启发，采用任意不确定性进行置信度估计，将具有高不确定性的特征转换为低置信度表示，将具有低不确定性的特征转换为高置信度表示。具体而言，假设待预测特征图为 $\boldsymbol{I}$，已标注的真实图像为 $\boldsymbol{I}_{\text{GT}}$，使用参数项 θ 对任意不确定性进行建模。为了准确估计 θ，使用拉普拉斯分布对似然函数进行建模，该函数可以表示为：

$$\ln p(\boldsymbol{I}_{\text{GT}}, \theta | \boldsymbol{I}) = -\frac{\| \boldsymbol{I}_{\text{GT}} - \boldsymbol{I} \|_1}{\theta} - \ln\theta - \ln 2 \tag{13}$$

建模 $\theta = \frac{1}{(\boldsymbol{C}^n)\uparrow}$，其中 $\uparrow$ 表示对 $\boldsymbol{C}^n$ 进行上采样，使其尺寸与 $\boldsymbol{I}_{\text{GT}}$ 对齐。然后，式(13)可以重新表述为：

$$L = \sum_{n=1}^{N}\left(\boldsymbol{C}^n \| \boldsymbol{I}_{\text{GT}} - \boldsymbol{I} \|_1 + \log\left(\frac{1}{\boldsymbol{C}^n + \epsilon}\right)\right) \tag{14}$$

其中，$\epsilon = 1\times10^{-8}$ 是用于稳定训练的常数。通过对特征图进行置信度估计，可以将高置信度的特征图停止在当前层，避免其进行进一步的计算。

3.4 优化的路径聚合特征金字塔网络

PAFPN 在进行特征融合和传播的过程中，顶部的高级特征需要跨越多个中间层级才能实现与底部低级特征的融合。这一过程虽促进了特征的跨层级交互，但也伴随着潜在风险：在自上而下的传播路径中，高层特征的丰富语义信息可能在逐层传递中发生丢失或弱化；与此同时，自下而上的路径也可能导致底层信息在传播和交互过程中丢失或退化，进而加剧了不同层次特征间语义信息的鸿沟。

为了解决上述问题，对 PAFPN 进行了优化，以减少不同层次之间的信息差，如图 5 所示。由于非相邻分层特征之间的语义差距大于相邻分层特征之间的语义差距，直接对非相邻分层特征进行融合可能导致信息传递不合理。因此，本文设计了双向特征融合网络 BFFN，它允许来自不同层级的特征信息进行逐步融合。通过双向流动的特征传递机制，深层次的语义特征和浅层次的细节特征可以得到充分融合，从而

减少不同层次间的语义信息差。例如，在融合底层特征层 P3 与顶层特征层 P5 时，BFFN 首先对 P3 和 P4 层进行特征融合，再将融合后的特征引入 P5 层。P3 层与 P4 层之间的特征融合缩小了它们之间的语义差距，由于 P4 和 P5 是相邻的分层特征，因此 P3 和 P5 之间的语义差距也就间接缩小。

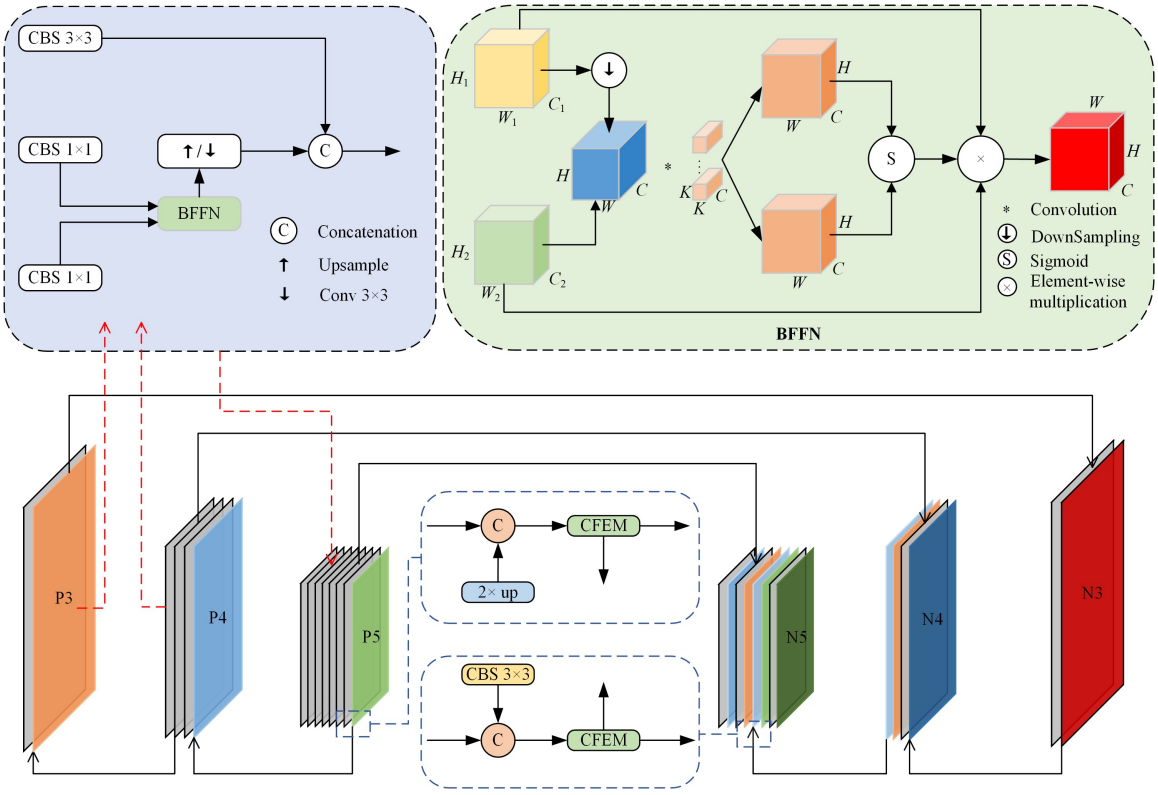


图 5 优化的路径聚合特征金字塔网络

Fig. 5 Optimized path aggregation feature pyramid network

双向特征融合网络：假设输入的低层特征为 $\boldsymbol{F}_{\text{low}} \in \mathbb{R}^{H_1 \times W_1 \times C_1}$，它主要包含丰富的细节信息，但语义信息较少；输入的高层特征为 $\boldsymbol{F}_{\text{high}} \in \mathbb{R}^{H_2 \times W_2 \times C_2}$，它主要包含全局语义信息，但细节信息较少。将低层特征下采样到高层特征，同时对高层特征用 1×1 卷积对齐通道数，从而得到 $\boldsymbol{F}'_{\text{high}}$ 和 $\boldsymbol{F}'_{\text{low}}$。

$$\begin{aligned}\boldsymbol{F}'_{\text{high}} &= Conv_{1\times1}(\boldsymbol{F}_{\text{high}})\\ \boldsymbol{F}'_{\text{low}} &= Conv_{3\times3}(\boldsymbol{F}_{\text{low}})\end{aligned} \tag{15}$$

这样就得到了尺寸一致且通道数对齐的高、低层特征，记作 $\boldsymbol{F}'_{\text{high}}, \boldsymbol{F}'_{\text{low}} \in \mathbb{R}^{H' \times W' \times C'}$。对齐后的特征 $\boldsymbol{F}'_{\text{high}}$ 和 $\boldsymbol{F}'_{\text{low}}$ 通过拼接操作进行融合：

$$\boldsymbol{F}_{\text{concat}} = Concat(\boldsymbol{F}'_{\text{low}} + \boldsymbol{F}'_{\text{high}}) \tag{16}$$

其中，$\boldsymbol{F}_{\text{concat}} \in \mathbb{R}^{H' \times W' \times 2C'}$ 包含了高层与低层的特征信息。接着，$\boldsymbol{F}_{\text{concat}}$ 通过一个 1×1 卷积层进行进一步的通道整合，输出融合特征 $\boldsymbol{F}_{\text{fused}}$。对融合特征 $\boldsymbol{F}_{\text{fused}}$ 应用 Sigmoid 激活函数，生成自适应权重矩阵 $\boldsymbol{W}_{\text{high}}$ 和 $\boldsymbol{W}_{\text{low}}$：

$$\begin{aligned}\boldsymbol{W}_{\text{high}} &= \sigma(\boldsymbol{F}_{\text{fused}})\\ \boldsymbol{W}_{\text{low}} &= \sigma(\boldsymbol{F}_{\text{fused}})\end{aligned} \tag{17}$$

然后，将这两个权重矩阵分别与高层特征和低层特征进行逐元素乘积，得到加权后的特征：

$$\begin{aligned}\boldsymbol{F}^{\text{weighted}}_{\text{high}} &= \boldsymbol{F}_{\text{high}} \odot \boldsymbol{W}_{\text{high}}\\ \boldsymbol{F}^{\text{weighted}}_{\text{low}} &= \boldsymbol{F}_{\text{low}} \odot \boldsymbol{W}_{\text{low}}\end{aligned} \tag{18}$$

加权操作可以视为一种自适应特征选择机制，可以自动调整高、低层特征在输出中的占比，适应不同层级的语义表达需求。再将加权后的特征进行相加得到聚合特征 $\boldsymbol{F}_{\text{aggregated}}$：

$$\boldsymbol{F}_{\text{aggregated}} = \boldsymbol{F}^{\text{weighted}}_{\text{high}} + \boldsymbol{F}^{\text{weighted}}_{\text{low}} \tag{19}$$

聚合操作不仅保留了高层特征的语义信息，还兼顾了低层特征的空间细节，从而实现了不同层次特征的平衡表达，减少了高低层之间的语义鸿沟。

4 实验

4.1 实验设置

本文实验在 VisDrone，NWPU VHR-10，TinyPerson 公共数据集上进行评估，并选取了平均均值精度（mAP）、参数数量（Params）、浮点操作数（FLOPs）、每秒处理图像帧数（FPS）和模型大小（Size）作为评估模型性能的指标。

在训练过程中，使用 Adam 优化器从头开始训练，初始学习率设置为 0.01，权重衰减设置为 0.000 5，批量大小设置为 8，patience 设置为 50，总训练周期数为 300 个 epochs。训练过程在 NVIDIA GeForce RTX 3090 服务器上完成。

4.2 对比实验

本文对比了 YOLO-FEPA 与 RT-DETR，YOLOv11n，Faster-RCNN 等先进目标检测器在 VisDrone，NWPU VHR-10 和 TinyPerson 数据集上的性能表现，实验结果如表 1 所列。

具体而言，在 VisDrone 数据集上，YOLO-FEPA 不仅保持了 181 FPS 的推理速度，还提升了目标检测的准确性，mAP50 达到 36.7%，比 YOLOv11n 提升了 3 个百分点。在精度和速度等指标上超越了 Faster-RCNN，SSD 和 RT-DE-

TR 等主流算法，且模型大小仅为 5.9×10^6，与其他 YOLO 系列算法差距很小。

在 NWPU VHR-10 数据集上，YOLO-FEPA 的 mAP50 达到 88.7%，分别比 Faster-RCNN，YOLOv11n，SSD 和 RT-DETR 提升了 11.2 个百分点、7.6 个百分点、17.9 个百分点和 12.5 个百分点，推理速度也能达到 111 FPS。

在 TinyPerson 数据集上，YOLO-FEPA 同样表现优异，mAP50 相较于 YOLOv11n 提升了 4.4 个百分点，且在精度上同样优于其他主流算法，证明了其在不同复杂度和规模的数据集上均能保持稳定的性能。

表 1 对比实验结果

Table 1 Results of comparative experiments

数据集	模型	mAP50/%	mAP50-95/%	Params	FLOPs	FPS	Size
VisDrone	YOLOv11n	33.7	19.6	**2.6×10^6**	**6.5×10^9**	**221**	**5.20×10^6**
	YOLOv10n	33.3	19.5	2.7×10^6	8.1×10^9	162	5.50×10^6
	YOLOv8n	33.2	19.5	3.1×10^6	8.2×10^9	153	6.20×10^6
	SSD-300	23.5	11.8	24.3×10^6	88.1×10^9	41	4.62×10^7
	Faster-RCNN	36.1	21.5	40.8×10^6	205.7×10^9	19	1.08×10^8
	RT-DETR-R18	36.5	21.8	20.1×10^6	58.3×10^9	106	3.98×10^7
	本文算法	**36.7**	**21.9**	2.8×10^6	7.5×10^9	181	5.90×10^6
NWPU VHR-10	Faster-RCNN	77.5	55.7	40.8×10^6	205.7×10^9	15	1.08×10^8
	SSD-300	70.8	50.1	24.3×10^6	88.1×10^9	33	4.62×10^7
	RT-DETR-R18	76.2	54.6	20.1×10^6	58.3×10^9	85	3.98×10^7
	YOLOv11n	81.1	58.5	**2.6×10^6**	**6.5×10^9**	**179**	**5.20×10^6**
	本文算法	**88.7**	**60.1**	2.8×10^6	7.5×10^9	111	5.90×10^6
TinyPerson	Faster-RCNN	23.2	6.9	40.8×10^6	205.7×10^9	25	1.08×10^8
	SSD-300	11.4	2.6	24.3×10^6	88.1×10^9	39	4.62×10^7
	RT-DETR-R18	23.8	7.1	20.1×10^6	58.3×10^9	93	3.98×10^7
	YOLOv11n	24.1	7.6	**2.6×10^6**	**6.5×10^9**	**183**	**5.20×10^6**
	本文算法	**28.5**	**9.2**	2.8×10^6	7.5×10^9	125	5.90×10^6

4.3 消融实验

4.3.1 全局-局部特征增强模块的消融实验

为了探究 GLFEM 模块对基准模型的影响，分别在 VisDrone 数据集上对其局部分支和全局分支进行了消融实验，结果如表 2 所列。

表 2 GLFEM 的消融实验结果

Table 2 Results of the GLFEM ablation experiments

方法	mAP50/%	mAP50-95/%	Params	FLOPs	FPS
YOLOv11n	33.7	19.6	2.59×10^6	6.5×10^9	221
局部分支	33.4(↓0.3)	19.3	**2.45×10^6**	**6.3×10^9**	**239**
全局分支	33.6(↓0.1)	19.5	2.57×10^6	6.5×10^9	223
局部分支+全局分支	**34.3**(↑0.6)	**20.0**	2.63×10^6	6.7×10^9	211

实验结果表明，单独使用局部分支或全局分支时，模型的检测精度均有所下降。其中，仅使用局部分支时，mAP50 下降 0.3 个百分点，仅使用全局分支时，mAP50 下降 0.1 个百分点，说明全局特征在目标检测中具有重要作用。当在局部分支的基础上添加全局分支时，模型的 mAP50 值相较于基准模型有所提升，mAP50 提高了 0.6 个百分点，但参数量和计算量略有增加。

4.3.2 自适应特征增强模块的消融实验

为了分析 AFEM 模块对基准模型的影响，在 VisDrone 数据集上对 PSABlock 块是否使用稠密连接机制进行了消融实验，结果如表 3 所列。为进一步验证提前停止策略的有效性，在不启用稠密连接机制的前提下，将置信度估计器的置信度阈值 K 分别设为 0.0，0.3，0.5，0.7，0.9 和 1.0，阈值 K 越大，表示有更多的特征图输入 PSABlock 块进行特征提取。对不同的 K 值进行了消融实验，结果如表 4 所列。

表 3 稠密连接机制的消融实验结果

Table 3 Results of ablation experiments with dense connections mechanisms

方法	mAP50/%	mAP50-95/%	Params	FLOPs	FPS
Without DenseNet	33.7	19.6	2.59×10^6	6.5×10^9	221
With DenseNet	34.1(↑0.4)	19.9	2.55×10^6	6.4×10^9	228

表 4 提前停止策略的消融实验结果

Table 4 Results of ablation experiments with early stopping strategy

阈值	mAP50/%	mAP50-95/%	Params	FLOPs	FPS
$K=0.0$	32.7 (↓1.0)	19.0	**2.34×10^6**	**6.1×10^9** (↓**0.4×10^9**)	**259**
$K=0.3$	32.9 (↓0.8)	19.1	2.41×10^6	6.2×10^9 (↓0.3×10^9)	248
$K=0.5$	33.2 (↓0.5)	19.3	2.46×10^6	6.3×10^9 (↓0.2×10^9)	241
$K=0.7$	33.5 (↓0.2)	19.5	2.52×10^6	6.3×10^9 (↓0.2×10^9)	238
$K=0.9$	33.7 (↓0.0)	19.6	2.57×10^6	6.4×10^9 (↓0.1×10^9)	229
$K=1.0$	**33.7**	**19.6**	2.59×10^6	6.5×10^9	221

表 3 的实验数据表明，使用稠密连接机制后，模型的 mAP50 值提升 0.4 个百分点，说明稠密连接机制可以增强网络的特征提取能力。表 4 的实验结果显示，不使用 AFEM 模块时($K=0.0$)，模型的 mAP50 值较基准模型($K=1.0$)低 1 个百分点，同时，计算量随着 K 值的降低而减少。当 $K=0.9$ 时，模型的 mAP50 值与基准模型相同，并且计算量下降了 0.1×10^9 FLOPs。这表明，对于一部分已经提取到良好特征的特征图，适当停止这些特征图的计算，可以在一定程度上减少计算冗余且不影响精度。

4.3.3 路径聚合特征金字塔网络的消融实验

路径聚合特征金字塔网络通过传递并融合主干网络的

多尺度特征，提升了模型处理多尺度目标的检测能力。为了评估 BFFN 对路径聚合特征金字塔网络性能的具体影响，在 VisDrone 数据集上分别对 P3，P4 和 P5 层使用 BFFN 进行消融实验，结果如表 5 所列。

实验数据显示，使用 BFFN 对相邻层进行融合后，模型的平均检测精度均有所提高，具体表现为：P3 和 P4 融合后，mAP50 提升了 0.8 个百分点；P4 和 P5 融合后，mAP50 提升了 0.6 个百分点。而在跨层次融合中，效果则更为显著，对 P3 和 P5 进行融合后，mAP50 提升 1.4 个百分点，并且在更加严格的 mAP50-95 指标上也有提升，但模型的推理速度均有所下降。以上结果表明，BFFN 能够缩小层次间的语义信息差异，从而提升目标检测精度。

表 5 路径聚合特征金字塔网络的消融实验结果

Table 5 Results of ablation experiments on path aggregation feature pyramid networks

方法	mAP50/%	mAP50-95/%	Params	FLOPs	FPS
Without BFFN	33.7	19.6	**2.59×10^6**	**6.5×10^9**	**221**
P3+P4	34.5(↑0.8)	20.1	2.71×10^6	6.6×10^9	215
P4+P5	34.3(↑0.6)	20.0	2.73×10^6	6.6×10^9	210
P3+P5	**35.1(↑1.4)**	**20.5**	2.79×10^6	6.9×10^9	201

4.3.4 所提模块的消融实验

为了评估 GLFEM，AFEM 以及 OPAFPN 这 3 个模块对模型整体性能的独立及综合影响，分别对这些模块在 VisDrone 数据集上进行了消融实验，实验结果如表 6 所列。

表 6 所提模块的消融实验结果

Table 6 Results of ablation experiments with the proposed modules

实验组	CFEM	AFEM	OPAFPN	mAP50/%	mAP50-95/%	Params	FLOPs	FPS
1	×	×	×	33.7	19.6	2.59×10^6	6.5×10^9	221
2	√	×	×	34.3	20.0	2.63×10^6	6.7×10^9	211
3	×	√	×	34.1	19.9	**2.53×10^6**	**6.4×10^9**	**236**
4	×	×	√	36.1	21.5	2.85×10^6	7.5×10^9	179
5	√	√	×	34.5	20.2	2.57×10^6	6.5×10^9	219
6	√	×	√	36.5	21.8	2.89×10^6	7.9×10^9	168
7	×	√	√	36.4	21.6	2.79×10^6	7.3×10^9	192
8	√	√	√	**36.7**	**21.9**	2.83×10^6	7.5×10^9	181

在基准模型上分别使用 GLFEM，AFEM($K=0.9$)和 OPAFPN 后，模型的 mAP50 值和 mAP50-95 值都有所提升。具体来说：添加 OPAFPN 后，模型的 mAP50 值提升幅度最大，达到 36.1%；添加 GLFEM 后，检测精度有一定的提升；引入 AFEM 后，模型精度提升且推理速度也有所提高。接着，将这些模块两两组合进行实验，结果显示，组合后模型的 mAP50 较单独使用模块时进一步提升，表明三者之间兼容性良好。最后，将 3 个模块全部引入基准模型，实验显示，模型的 mAP50 达到 36.7%，但参数量和计算量也有所增加。

4.4 自建数据集实验

为了验证 YOLO-FEPA 模型的泛化能力，选取了小型机场作为测试环境，并采用 Intel D455 相机进行实地拍摄。该相机配备了全局快门的 RGB 传感器，确保了在高动态场景下的成像质量，其在 4 米距离内的深度测量精度优于 2%，RGB 传感器的视野范围(FOV)达到 90°×60°，能够满足多样化场景拍摄需求。相机的核心组件包括 Intel 实感视觉处理器 D4 及 Intel 实感深度模块 D455，它们共同提供了高精度的深度与色彩信息。

本文自建的 AirportTiny 数据集设计了卡车、公共汽车、汽车、飞机和人 5 类目标标签，涵盖了机场区域常见的移动与静态物体。整个数据集包含 837 张图片，每张图片都经过仔细筛选并使用 Labelimg 标注工具进行了精确的手动标注，确保了数据的质量与准确性。为了合理进行模型训练与评估，将数据集分为训练集(600 张)、测试集(150 张)和验证集(87 张)。

本文选取 RT-DETR 和 YOLOv11n 等先进算法在 AirportTiny 数据集上进行了对比实验。为了确保实验结果的公正性与可比性，所有算法的实验参数均与另外 3 个数据集保持一致，遵循了严格的实验设置规范，实验结果如表 7 所列。

表 7 在自建数据集上的对比实验结果

Table 7 Results of comparison experiments with self-built dataset

模型	mAP50/%	mAP50-95/%	Params	FLOPs	FPS
YOLOv11n	57.4	41.5	**2.6×10^6**	**6.500×10^9**	**208**
RT-DETR-R18	53.3	38.6	2.01×10^7	5.830×10^{10}	158
Faster-RCNN	35.1	20.9	4.08×10^7	2.057×10^{11}	39
本文算法	**65.6**	**46.6**	2.8×10^6	7.500×10^9	171

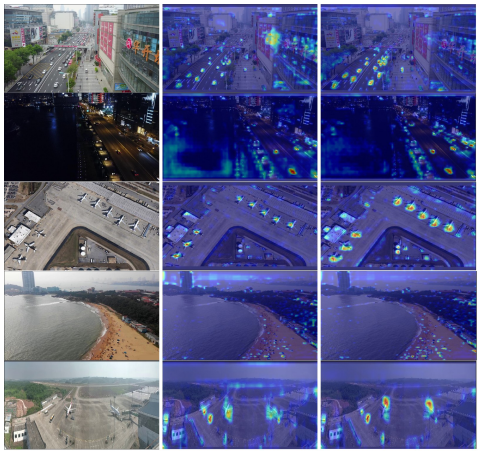
从表 7 中可以看出，YOLO-FEPA 模型在关键评估指标上展现出了一定优势。YOLO-FEPA 在 mAP50 指标上相较于 YOLOv11n 高出 8.2 个百分点，与其他优秀算法相比，同样表现出较大的优势，表明了该模型具有不错的泛化能力。在机场这样复杂的场景中，误报和漏报可能导致严重后果，更高的 mAP 值意味着对目标的识别更精准，能够有效减少误报和漏报情况的发生。

4.5 可视化效果分析

4.5.1 热力图

为了更好地验证 YOLO-FEPA 的特征增强效果，将特征图以热力图的形式进行可视化，如图 6 所示。可视化热力图描绘了哪些区域具有较高的激活值，这种方法可以更直观地理解模型对不同区域的关注程度。

图 6(a)是输入图像，图 6(b)则是基准模型生成的热力图，从中可以观察到，基准模型对特征图的关注度较为分散，未能有效聚焦于目标特征。相比之下，图 6(c)呈现了 YOLO-FEPA 生成的热力图。该图表明，所提算法提升了模型对特征图中待检测目标特征的专注度。此外，与基准模型相比，YOLO-FEPA 在一定程度上降低了图像背景特征的干扰。



(a)输入图像 (b)YOLOv11n (c)YOLO-FEPA

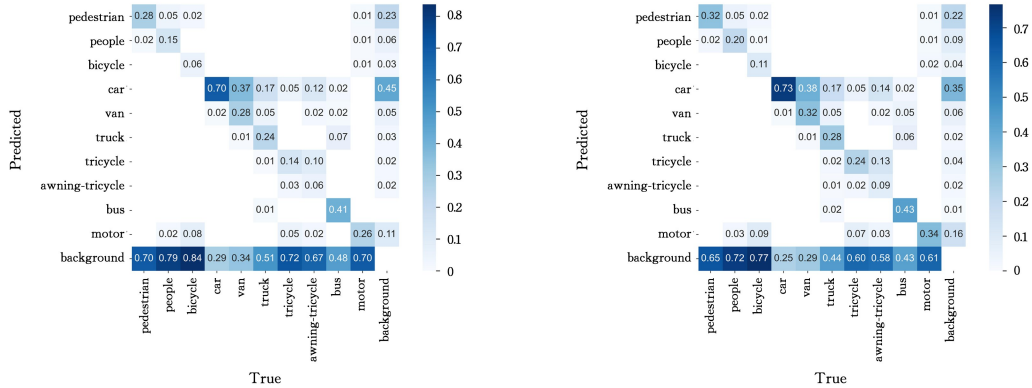
图 6 可视化热力图对比

Fig. 6 Comparison of visual heat maps

4.5.2 混淆矩阵

混淆矩阵是一种用于评估模型分类性能的表格，矩阵的每一行代表一个真实的类别，每一列代表预测的类别。它以实际类别和模型预测类别为基础，将样本分类结果进行统计，可以直观地比较类别间的分类准确性，并识别出模型在哪些类别表现好或表现差。矩阵的主对角线数值越大，表示模型正确检测出的样本数量越多；左下三角数值越大，表明模型存在较多的漏检；右上三角数值越大，表示模型存在较多的误检。图 7 分别为 YOLOv11n 与 YOLO-FEPA 的混淆矩阵。

相较于 YOLOv11n，本文方法在识别各个类别时展现出了更高的准确性，具体表现为，每个类别的检测正确率均有提升(YOLO-FEPA 的主对角线数值比 YOLOv11n 大)，同时漏检率降低(YOLO-FEPA 的左下三角数值比 YOLOv11n 小)。然而，值得注意的是，对于“bicycle”“tricycle”和“motor”这 3 个类别，YOLO-FEPA 的误检率略高于 YOLOv11n。



(a)YOLOv11n (b)YOLO-FEPA

图 7 混淆矩阵

Fig. 7 Confusion matrix

4.6 视觉效果对比分析

为了比较本文算法与基准算法在实际场景中的检测效果，分别在 TinyPerson，NWPU VHR-10，VisDrone 和自建数据集上进行可视化实验。将 YOLOv11n 与 YOLO-FEPA 进行视觉效果对比，并定量分析它们之间的漏检率(MR)和误检率(FDR)。

$$MR=\frac{FN}{TP+FN} \tag{20}$$

$$FDR=\frac{FP}{TP+FP} \tag{21}$$

其中，FN 是漏检的目标数，TP 是正确检测出的目标数，FP 是误检的目标数。

4.6.1 TinyPerson 数据集上的检测效果

TinyPerson 数据集专注于模拟复杂海洋环境下的人群目标检测任务，有 1610 张标记图像与 759 张未标记图像。数据集中显著特征为尺度较小的目标占据多数，且标注了两种关键类别，即海岸边的人员以及海水中的个体。为了验证所提算法的有效性，将其与 YOLOv11n 进行了可视化检测实验对比，实验结果如图 8 所示。

图 8 直观地展示了 YOLO-FEPA 模型在 TinyPerson 数据集上的目标检测任务所表现出的显著优势(红圈代表漏检，黑圈代表误检)。具体来说，根据式(20)和式(21)，在场景一中，YOLOv11n 的漏检率为 66.7%，误检率是 20.0%；YOLO-FEPA 的漏检率为 23.1%，误检率为 0.0%。在场景二中，YOLOv11n 的漏检率为 50.0%，误检率为 8.3%；YOLO-FEPA 的漏检率为 33.3%，误检率为 5.8%。对比分析两种算法的检测置信度分数后发现，YOLO-FEPA 在场景一和场景二中的检测置信度分数几乎均高于 YOLOv11n。

(a)输入图像 (b)YOLOv11n (c)YOLO-FEPA

图 8 TinyPerson 数据集上的检测结果(电子版为彩图)

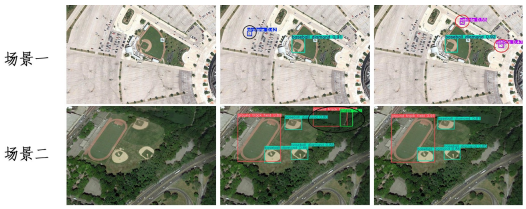
Fig. 8 Detection results on TinyPerson dataset

4.6.2 NWPU VHR-10 数据集上的检测效果

NWPU VHR-10 数据集由西北工业大学于 2014 年构建

并公开发布，是遥感图像领域内专注于高分辨率（VHR）目标检测任务的一项宝贵资源。该数据集广泛覆盖了 10 个不同的目标类别，共包含 3651 个精心标注的目标实例，旨在促进遥感图像分析技术的发展。数据集由 650 张正例图像组成，每张图像中均含有至少一个属于预定义类别的目标对象。此外，还包含 150 张反例图像，这些图像中未包含任何预设的目标类别对象。

实验结果如图 9 所示。在场景一中，YOLOv11n 的漏检率为 92.3%，误检率为 50.0%；YOLO-FEPA 的漏检率为 77%，误检率为 0.0%。在场景二中，YOLOv11n 的漏检率为 60.0%，误检率为 33.3%；YOLO-FEPA 的漏检率为 60.0%，误检率为 0.0%。进一步对比两种算法的检测置信度分数后发现：在场景一中，YOLO-FEPA 在检测棒球场时的置信度分数高于 YOLOv11n；在场景二中，两种算法在检测 3 个棒球场地时的置信度分数相近，但 YOLO-FEPA 在检测田径场时的置信度分数高于 YOLOv11n。



（a）输入图像 （b）YOLOv11n （c）YOLO-FEPA

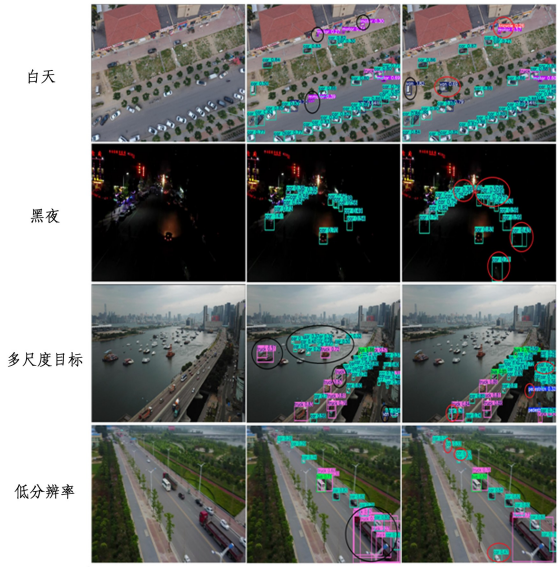
图 9 NWPU VHR-10 数据集上的检测结果

Fig. 9 Detection results on NWPU VHR-10 dataset

4.6.3 VisDrone 数据集上的检测效果

VisDrone 数据集由天津大学机器学习和数据挖掘实验室 AISKYEYE 团队收集，包括训练集、验证集和测试集，分别包含 6471 张、548 张和 3190 张图像。该数据集是在不同的场景、不同的天气和光照条件下通过不同的无人机平台收集的，覆盖范围广泛，包括位置、环境、物体和密度，旨在全面评估无人机视觉系统在不同情境下的性能，为智能交通、灾害响应、城市规划和农业监测等领域的研究和发展提供重要的基准和参考依据。数据集包含人、行人、面包车、汽车、卡车、巴士、摩托车、三轮车、遮阳敞篷三轮车和自行车 10 类目标。本文从 VisDrone 数据集中挑选了几类具有代表性的场景进行测试，覆盖了白天、黑夜、多尺度目标以及低分辨率等条件，以全面检验模型在不同场景下的性能表现，检测结果如图 10 所示。

在白天场景中，YOLOv11n 的漏检率为 10.2%，误检率为 10.2%；YOLO-FEPA 的漏检率为 7.5%，误检率为 2.6%。在黑夜场景中，YOLOv11n 的漏检率为 56.4%，YOLO-FEPA 的漏检率为 26.7%。黑夜场景下，光照条件不足，图像中物体特征很大程度上受到背景的干扰，导致基准模型产生较大的漏检率。在多尺度目标场景中，YOLOv11n 的漏检率为 47.4%，误检率为 30.2%；YOLO-FEPA 的漏检率为 25.4%，误检率为 3.8%。YOLOv11n 将水面上大部分的船舶识别为车辆，造成了大量的误检，而 YOLO-FEPA 在这个场景下的误检率较低。在低分辨率场景下，YOLOv11n 的漏检率是 35.3%，YOLO-FEPA 的漏检率是 23.5%。值得注意的是，YOLOv11n 在检测卡车这一类别时出现了大量的冗余框。进一步分析这两种算法的检测置信度分数后，在 4 个场景下，YOLO-FEPA 对所有检测目标的检测置信度分数几乎均高于 YOLOv11n。



（a）输入图像 （b）YOLOv11n （c）YOLO-FEPA

图 10 VisDrone 据集上的检测结果

Fig. 10 Detection results on VisDrone dataset

4.6.4 自建数据集上的检测效果

AirportTiny 数据集不仅涵盖了广泛尺寸范围的目标，从显著的大尺度物体到细微的小尺度目标一应俱全，还覆盖了 5 种常见的机场场景类别，为目标检测任务提供了全面而具有挑战性的测试平台。为了深入验证本文方法的泛化能力，选择了两个具有代表性的场景进行实验，检测结果如图 11 所示。

（a）输入图像 （b）YOLOv11n （c）YOLO-FEPA

图 11 自建数据集上的检测结果

Fig. 11 Detection results on self-built dataset

从图 11 所展示的检测结果中可以观察到，YOLO-FEPA 在降低漏检率方面展现出了显著的优势（黑圈代表漏检）。具体表现为：在场景一中，YOLOv11n 的漏检率为 33.3%，YOLO-FEPA 的漏检率为 22.2%；在场景二中，YOLOv11n 的漏检率为 66.7%，YOLO-FEPA 的漏检率为 25.0%。

机场环境中的漏检可能导致安全问题，如人员误入禁区或车辆冲突。YOLO-FEPA 显著降低了漏检率，有效减少了这些隐患，并在已检测目标的置信度上全面优于 YOLOv11n，表现出更高的可靠性和准确性。精准的目标检测为滑行道、停机位和地面车辆调度提供了可靠支持，减少了调度和设备维护错误的风险，提升了机场运营效率，这进一步证明了该模型在复杂场景下的泛化能力和实际应用价值。

结束语 本文提出了深度特征强化与路径聚合优化的目标检测方法，旨在解决深度神经网络在特征提取过程中特征

信息容易被弱化的问题。首先，提出了全局-局部特征增强模块 GLFEM，通过融合局部特征和全局特征，强化主干网络和颈部网络的特征表达能力。然后，设计了自适应特征增强模块 AFEM，其关键思想是使用稠密机制增强网络的特征提取能力，以缓解特征图在 PSABlock 块中因梯度消失或梯度爆炸而引起的特征退化问题，并设计置信度估计器以筛选特征质量良好的特征图，停止其继续输入 PSABlock 中，减少了计算冗余。之后，针对路径聚合特征金字塔网络中不同层次的语义信息不一致的问题，设计了双向特征融合网络 BFFN，融合不同层级的特征信息，有效降低了 PAFPN 不同层次间语义信息差。最后，在 VisDrone，NWPU VHR-10 和 TinyPerson 数据集上的实验证明，YOLO-FEPA 降低了模型的漏检率和误检率，并提高了检测置信度分数。同时，在自建的 AirportTiny 数据集上的测试也验证了该算法的优良泛化能力。

尽管 YOLO-FEPA 在检测精度上有所提升，但其增加的参数量和计算开销可能导致推理速度下降，尤其在边缘设备或移动设备上，这是难以接受的。然而，在云端部署或特定任务（如医学图像分析、工业检测）中，适当增加参数和计算开销以换取更高精度是可行的。未来的工作将探索在提高精度的同时，通过算法优化保持或减少参数量和降低计算复杂度，追求精度、参数和计算复杂度的平衡。

参 考 文 献

[1] CHEN C, QI J, LIU X, et al. Weakly Misalignment-free Adaptive Feature Alignment for UAVs-based Multimodal Object Detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 26836-26845.

[2] SOBEK J, MEDINA INOJOSA J R, MEDINA INOJOSA B J, et al. MedYOLO: A Medical Image Object Detection Framework [J]. Journal of Imaging Informatics in Medicine, 2024, 37: 3208-3216.

[3] WANG Q, LIU F, ZOU R, et al. Enhancing medical image object detection with collaborative multi-agent deep Q-networks and multi-scale representation [J]. EURASIP Journal on Advances in Signal Processing, 2023, 2023(1): 132.

[4] XU Q, LIN X, CAI M, et al. End-to-End Joint Multi-Object Detection and Tracking for Intelligent Transportation Systems [J]. Chinese Journal of Mechanical Engineering, 2023, 36(1): 138.

[5] ZHAO R, TANG S, SUPENI E E B, et al. A Review of Object Detection in Traffic Scenes Based on Deep Learning [J]. Applied Mathematics and Nonlinear Sciences, 2023, 9(1): 1-25.

[6] SARACENI L, MOTOI I M, NARDI D, et al. AgriSORT: A Simple Online Real-time Tracking-by-Detection framework for robotics in precision agriculture[C]// Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2024: 2675-2682.

[7] ZHAO P, ZHOU W, NA L. High-precision object detection network for automate pear picking [J]. Scientific Reports, 2024, 14(1): 14965.

[8] MUJKIC E, CHRISTIANSEN M P, RAVN O. Object Detection for Agricultural Vehicles: Ensemble Method Based on Hierarchy of Classes [J]. Sensors, 2023, 23(16): 7285.

[9] ZHAO Y, LYU W, XU S, et al. Detrs beat yolos on real-time object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 16965-16974.

[10] XIAO J S, ZHAO T, ZHOU J, et al. Small Target Detection Network Based on Context Augmentation and Feature Refinement [J]. Journal of Computer Research and Development, 2023, 60(2): 465-474.

[11] JIANG Z T, ZHAI F S, QIAN Y, et al. Low Illumination Object Detection Combined with Feature Enhancement and Multi-Scale Receptive Field[J]. Journal of Computer Research and Development, 2023, 60(4): 903-915.

[12] ZHANG K H, SHEN H K. Solder joint defect detection in the connectors using improved Faster- RCNN algorithm[J]. Applied Sciences, 2021, 11(2): 576.

[13] YANG A M, JIANG T Y, HAN Y, et al. Research on application of on-line melting in-SITU visual inspection of iron ore powder based on Faster R-CNN [J]. Alexandria Engineering Journal, 2022, 61(11): 8963-8971.

[14] KUMAR A, MANIKANDAN R. Brain tumor detection using deep neural network- based classifier[C]// Proceedings of the 2022 International Conference on Innovative Computing and Communications. Singapore: Springer, 2022: 173-181.

[15] TERVEN J, CÓRDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas [J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.

[16] SAPKOTA R, MENG Z C, CHURUVIJA M, et al. Comprehensive Performance Evaluation of YOLO11, YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environments[J]. arXiv: 2407. 12040, 2024.

[17] WANG C Y, BOCHKOVSKIY A, LIAO H Y 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. 2023: 7464-7475.

[18] HE Z, WANG K, FANG T, et al. Comprehensive Performance Evaluation of YOLOv11, YOLOv10, YOLOv9, YOLOv8 and YOLOv5 on Object Detection of Power Equipment [J]. arXiv: 2411. 18871, 2024.

[19] ZHANG Y, XIA Y. Object Detection Method with Multi-scale Feature Fusion for R-emote Sensing Images [J]. Computer Science, 2024, 51(3): 165-173.

[20] TERVEN J, CÓRDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of yolo architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-nas [J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.

[21] WANG C Y, YEH I H, LIAO H Y. Yolov9: Learning what you want to learn using programmable gradient information [C]// Proceedings of the European Conference on Computer Vision. Cham: Springer. 2025: 1-21.

[22] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot

multibox detector[C]// Proceedings of the Computer Vision-ECCV 2016:14th European Conference. Springer. 2016:21-37.

[23] QUE Y,GAN M H,LIU Z W. Object Detection with Receptive Field Expansion and Multi-branch Aggregation[EB/OL]. https://doi.org/10.11896/jsjkx.230600151.

[24] LI Y C,ZHANG R,WANG J B,et al. Re-parameter-ization Enhanced Dual-modal Realtime Object Detection Model[J]. Computer Science,2024,51(9):162-172.

[25] WANG J,CHEN Y,DONG Z,et al. Improved YOLOv5 network for real-time multi-scale traffic sign detection [J]. Neural Computing and Applications,2023,35(10):7853-7865.

[26] NI J,ZHU S,TANG G,et al. A small-object detection model based on improved YOLOv8s for UAV image scenarios [J]. Remote Sensing,2024,16(13):2465.

[27] WEI J,NI L,LUO L,et al. GFS-YOLO11:A Maturity Detection Model for Multi-Variety Tomato [J]. Agronomy,2024,14(11):2644.

[28] JOOSHIN H K,NANGIR M,SEYEDARABI H. Inception-YOLO:Computational cost and accuracy improvement of the YOLOv5 model based on employing modified CSP,SPPF,and inception modules [J]. IET Image Processing,2024,18(8):1985-1999.

[29] 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.

[30] CAGNETTA F,PETRINI L,TOMASINI U M,et al. How deep neural networks learn compositional data:The random hierarchy model [J]. Physical Review X,2024,14(3):031001.

[31] LIU M,WANG H,DU L,et al. Bearing-detr:A lightweight deep learning model for bearing defect detection based on RT-DETR [J]. Sensors,2024,24(13):4262.

[32] LA MALFA E,LA MALFA G,NICOSIA G,et al. Characterizing learning dynamics of deep neural networks via complex networks[C]// Proceedings of the 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence(ICTAI). IEEE,2021:344-351.

[33] 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.

[34] 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.

[35] 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.

[36] QIU Y F,XIN H. Target Detection Algorithm Based on Global Feature Fusion in Parallel Dual Path Backbone[J]. Journal of Frontiers of Computer Science and Technology,2024,18(12):3247-3259.

[37] SHI Y,WANG L,YAO Y P,et al. Small Object Detection Based on Enhanced Feature Pyramid and Focal-AIoU Loss[J]. Journal of Frontiers of Computer Science and Technology,2025,19(3):693-702.

[38] HAN B,HE L,KE J,et al. Weighted parallel decoupled feature pyramid network for object detection [J]. Neurocomputing,2024,593:127809.

[39] PENG Z,HUANG W,GU S,et al. Conformer:Local features coupling global representations for visual recognition[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:367-376.

[40] GEIGER B C,KUBIN G. Information bottleneck:Theory and applications in deep learning [J]. Entropy,2020,22(12):1408.

[41] LIU Z,WANG B,LI Y,et al. UnitModule:A lightweight joint image enhancement module for underwater object detection [J]. Pattern Recognition,2024,151:110435.

[42] HUANG G,LIU Z,VAN DER MAATEN L,et al. Densely connected convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:4700-4708.

[43] NING Q,DONG W,LI X,et al. Uncertainty-driven loss for single image super-resolution [J]. Advances in Neural Information Processing Systems,2021,34:16398-16409.

WANG Xiaofeng, born in 1978, Ph.D, professor, is a member of CCF (No. A8319M). His main research interests include object detection and image super resolution.

HUANG Junjun, born in 1998, postgraduate. His main research interests include object detection and image super resolution.

(责任编辑:何杨)