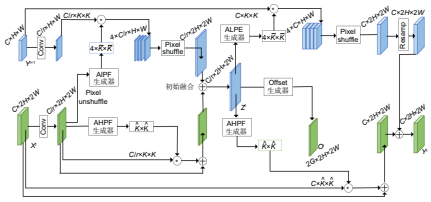
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 2 月, 第 53 卷, 第 2 期

基于频域感知特征融合的无人机航拍图像小目标检测算法



贾小军[1,2*], 唐　彪[1,2], 骆淑云[2], 梁诺雷[1,2], 张春花[1]

[1]嘉兴大学人工智能学院, 浙江 嘉兴 314001;

[2]浙江理工大学信息科学与工程学院, 浙江 杭州 310018

摘要: 小目标检测一直是计算机视觉处理中一项极具挑战性的任务, 针对传统无人机航拍图像小目标多且密集, 而传统检测算法存在漏检、误检的问题, 提出一种基于频域感知特征融合的 FFC-YOLO (FASFF-FreqFusion-CAA YOLO) 目标检测算法。首先, 通过增加 P2 检测层, 重新设计了针对小目标检测的 Detect-FASFF (four adaptively spatial feature fusion) 检测头结构。然后, 将传统的特征融合采样方法替换为频率感知特征融合 FreqFusion (frequency-aware feature fusion) 方法并结合 BiFPN (bidirectional feature pyramid network) 结构提升了算法处理密集小目标图像的能力。最后, 在 C2PSA 模块中加入 CAA (context anchor attention) 注意力机制, 增强了目标上下文特征信息的关联。在 VisDrone2019 数据集中, 基于 FFC-YOLO 目标检测算法的 mAP@0.5 为 40%, 较传统算法如 Fast-RCNN、RetinaNet 分别提升了 28.4%、18.6%, 较 YOLO 算法如 YOLO v8n、v10n、v11n 分别提升了 8.0%、7.6%、7.8%。在自建小目标数据集 tiny-data 中, 基于 FFC-YOLO 的 3 类目标 sperson、lperson 和 wperson 的检测结果较 YOLOv11n, mAP@0.5、*P*、*R* 分别提升了 9.2%、8.7%、5.9%。实验结果表明, FFC-YOLO 小目标检测算法能够更好地应用于无人机航拍图像检测领域。

关键词: FFC-YOLO; 小目标检测; 无人机航拍图像; 频域感知特征融合; YOLOv11n

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250208 | CSTR: 32245.14.oee.2026.250208 | 中图分类号: TP391.4 | 文献标识码: A

贾小军, 唐彪, 骆淑云等. 基于频域感知特征融合的无人机航拍图像小目标检测算法 [J]. 光电工程, 2026, **53**(2): 250208

Jia X J, Tang B, Luo S Y et al. A small target detection algorithm for UAV aerial images based on frequency domain perception feature fusion[J]. *Opto-Electron Eng*, 2026, **53**(2): 250208

1 引　言

近年来, 随着无人机 (unmanned aerial vehicle, UAV) 技术的快速发展, 航拍图像在灾害救援、农业监测、交通管理、军事侦察等领域得到了广泛应用。然而, 航拍图像中的小目标检测 (如车辆、行人、小型设施等) 仍面临严峻挑战。这类目标通常仅占图像中极少的像素 (如小于 32 pixel×32 pixel), 且受限于无人机飞行高度、复杂背景干扰、多尺度变化及光照条件等因素, 导致目标特征模糊、信噪比低, 传统检测方法难以实现精准识别。

传统基于手工特征的检测算法如 HOG[1]、SIFT[2] 等, 因依赖人工设计特征, 泛化能力不足, 难以应对小目标的多样性与复杂性。随着深度学习技术的兴起, 基于卷积神经网络 CNN 的方法显著提升了小目标检测性能。研究者通过多尺度特征融合 FPN、注意力机制 CBAM[3]、

收稿日期: 2025-07-15; 修回日期: 2025-09-24; 录用日期: 2025-09-29; 网络出版日期: 2026-02-13

基金项目: 嘉兴市科技局公益类项目 (2024AY40009)

*通信作者: 贾小军, xjjiad@sina.com

版权所有©2026 中国科学院光电技术研究所

超分辨率重建等技术增强小目标特征表达，并优化 Faster R-CNN[4]、YOLO 系列[5-7] 网络的结构设计和训练策略以提升检测精度。然而，如何用无人机有限的算力平衡检测精度与实时性，仍是实际部署的核心难题。

优化提高无人机航拍图像的小目标检测精度已成为计算机视觉的主流研究方向，张杨等[8] 提出使用卷积层更少的 ResNet-18 作为骨干网络改进 Faster R-CNN，优化特征融合方法增强了目标特征与背景特征的对比度。李钟华等[9] 在 YOLOv5 框架的基础上设计了多尺度结构感知模块，并借助 Transformer 设计了全局上下文信息模块，提高了平均精度，但针对小目标检测的改进不足。李利霞等[10] 为了增加小目标特征信息，提出浅层特征增强模块来学习底层特征信息，在 VisDrone2019 数据集上提升了 3.5% 的精度，但并未对模型做轻量化处理，精度还有提升空间。韩俊等[11] 以 YOLOv5 为基础，构造了多尺度特征提取模块 LM-fem 增强特征提取能力，设计了新的混合域注意力模块 S-ECA 和自适应权重动态融合结构 AFF，在 VisDrone2021 数据集上提升了 6.1% 的检测精度，但因未对网络模型轻量化处理，所需算力过大。牛为华等[12] 针对小目标多且小目标易被噪声淹没等问题，基于 YOLOv7，提出增加一个检测小目标的头部，改进特征融合方式来优化特征图的表达能力，在 VisDrone2019 数据集上精度提升了 2.8%。常凯旋等[13] 提出了一种基于双模态图像融合的算法，通过双路特征融合模块融合红外和可见光的特征和在颈部网络层引入 BRA (bi-level routing attention) 模块，增强了小目标的检测能力，在 DroneVehicle 数据集上平均精度均值提升了 14.78%。郑华伟等[14] 针对图像检测精度低的问题，以 YOLOv5 为基准网络，通过在骨干网络部分引入可变形注意力模块，网络头部加入浅层特征图等方式，在 NWPU VHR-10 数据集上精度达到 94.3%。张润梅等[15] 针对无人机航拍过程受到恶劣环境影响等问题，提出一种改进算法 SSG-YOLOv7。通过引入 SimAM 注意力机制和 GhostConv 卷积模块，然后使用 Soft NMS 优化锚框置信度，在 VisDrone2019数据集上较原算法精度提升了 10.45%。唐彪等[16] 以 YOLOv10n 为基准网络，加入一个 160 pixel×160 pixel 尺寸的小目标检测头并在每个检测头前加入大尺寸可分离注意力机制 LSKA (large separable kernel attention)，最后使用 BiFPN 融合更多层级特征并优化网络结构，在 VisDrone2019 数据集上较基准网络精度提升了 4.6%。马滔等[17] 将 YOLOv11 的骨干网络替换为 VanillaNet，加入了自适应空间融合检测头，在 HIT-UAV 红外小目标数据集上精度达到了 81.3%，但并未验证模型泛化能力，并且参数量还是较大。

综上所述，现有的无人机航拍图像小目标检测算法虽然都提高了检测精度，但都存在一定的问题。小目标的特征信息很难提取且易被背景遮挡，因此在网络中多次使用采样方法易造成特征信息丢失过多，这种情况在低分辨率图像和小目标密集图像中更为严重。为了解决以上的检测问题，本文提出一种基于频域感知特征融合的检测算法 FFC-YOLO (FASFF-FreqFusion-CAA-YOLO)，该算法能够有效地提高小目标检测的精度，在自建小目标数据集上的泛化实验也验证了模型的泛化性，能够较好地适用于无人机航拍图像的小目标检测。

2 FFC-YOLO 目标检测算法

2.1 FFC-YOLO 网络结构

YOLOv11 是 Ultralytics 团队于 2024 年 9 月发布的 YOLO 系列，包括 YOLOv11n、s、m、l、x 等。YOLOv11 通过模块架构的迭代 (如 C3k2、C2PSA、SPFF 模块) 与任务导向优化，在目标检测领域实现了精度、效率与轻量化的三位一体突破。本文选用精度和体积更合适的 YOLOv11n 网络作为基准，优化网络结构来提高模型对于小目标的检测能力，提出一种优化算法：FFC-YOLO，旨在提升小目标的检测能力。本文的改进包括：

1) 设计自适应空间特征融合的四检测头结构 FASFF (four adaptively spatial feature fusion)，替换掉原 YOLOv11 的三检测头结构，使得模型检测小目标的能力更加优秀。

2) 用频域感知特征融合 (frequency-aware feature fusion, FreqFusion) 方法替换掉采样方法并与 BiFPN 结构结合，使模型可以实现多层级特征融合的同时还可以解决密集图像预测任务中的类别不一致和边界位移问题。

3) 在 C2PSA 中采用上下文锚点注意力 CAA (context anchor attention)，构建成 C2PSA-CAA 模块，以并行方式捕捉远距离的上下文信息。改进后的网络模型结构如图 1 所示。

2.2 自适应融合 FASFF 四检测头结构

在无人机航拍图像目标检测任务中，小目标具有特征不明显、尺度小等特点。YOLOv11 模型采用的三检测头 (1/8、1/16、1/32) 虽能满足大部分目标的尺度检测，

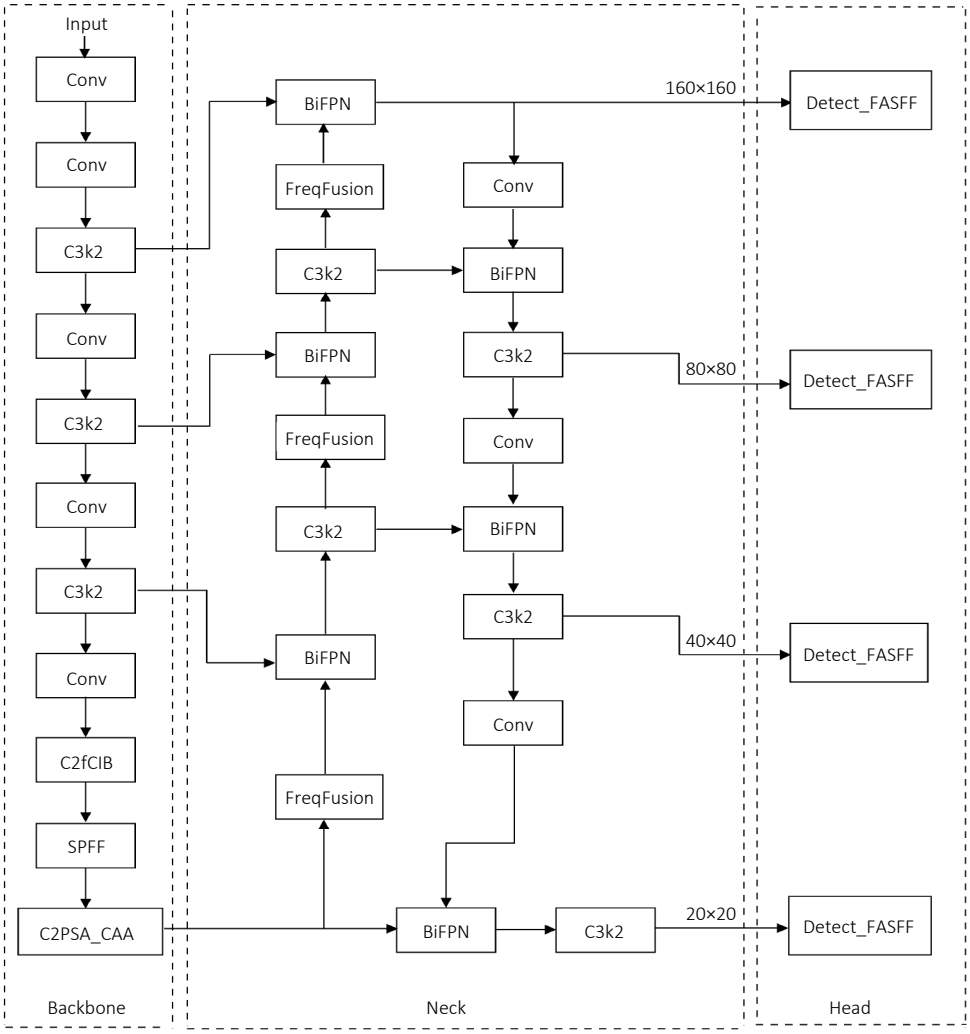


图 1 FFC-YOLO 网络结构

Fig. 1 FFC-YOLO network structure

但在处理像素点小于 32 pixel×32 pixel 的目标时检测精度较低，因此如何提高小目标的检测精度是检测模型的改进重点。小目标检测精度低的原因在于以下两点：

1) 没有针对超小目标的特征分辨率。1/8 尺度特征图的分辨率为 80 pixel×80 pixel 的网格，难以精确地捕捉到超小目标的形态特征。

2) 浅层的 (1/8 尺度) 高分辨率特征无法有效地传递信息至检测头，且各尺度的特征信息交互很有限。原始 YOLOv11 的 FPN-PAN 结构无法充分交互浅层与深层之间的特征信息。

为此，将 YOLOv11 模型的 Head 部分检测头替换为自适应空间特征融合检测头 ASFF[18] 并新增了一个 1/4 尺度的检测头，该检测头生成的 160 pixel×160 pixel 的高分辨率特征图能够有效的提升模型对超小目标的检测能力，形成新的检测头称为 FASFF 检测头结构，如图 2 所示。

FASFF 结构首先从骨干网络中提取底层特征 C2 (高分辨率，小目标细节丰富) 和高层特征 C3、C4、C5 (大感受野，语义信息强)。然后通过加权双向金字塔特征网络结构与频域感知特征融合方法提高特征融合能力。通过不同层级的特征融合生成 P2、P3、P4、P5，各层计算如下：

$$\begin{cases} P_5^{\text{out}} = \text{Conv}(P_5^{\text{in}}) \\ P_4^{\text{out}} = \text{Conv}(P_4^{\text{in}} + FreqFusion(P_4^{\text{out}})) \\ P_3^{\text{out}} = \text{Conv}(P_3^{\text{in}} + FreqFusion(P_3^{\text{out}})) \\ P_2^{\text{out}} = \text{Conv}(P_2^{\text{in}} + FreqFusion(P_2^{\text{out}})) \end{cases}. \tag{1}$$

然后，FASFF 结构为每个层级 P2-P5 分配空间自适

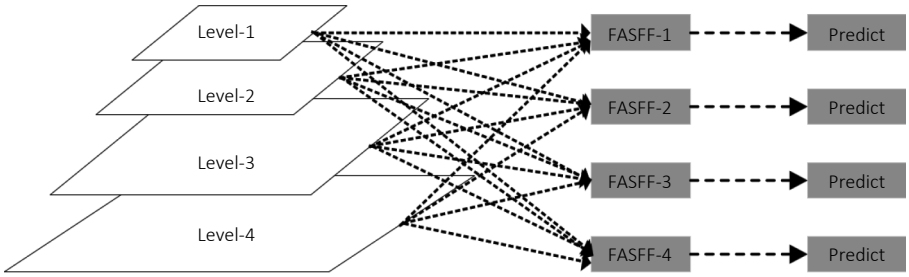


图 2 FASFF 结构

Fig. 2 FASFF structure

应权重，再将各层特征与其对应的空间权重逐元素相乘后求和，生成融合特征。具体计算如下：

$$\boldsymbol{y}_{ij}^{l}=\alpha_{ij}^{l}\cdot\boldsymbol{x}_{ij}^{1\to l}+\beta_{ij}^{l}\cdot\boldsymbol{x}_{ij}^{2\to l}+\gamma_{ij}^{l}\cdot\boldsymbol{x}_{ij}^{3\to l}+\delta_{ij}^{l}\cdot\boldsymbol{x}_{ij}^{4\to l}, \tag{2}$$

式中：$\boldsymbol{y}_{ij}^{l}$表示生成的融合特征；$\boldsymbol{x}_{ij}^{n\to l}$表示从 n 层逐层映射到 l 层(i,j)位置上的特征向量；α_{ij}^{l}、β_{ij}^{l}、γ_{ij}^{l}、δ_{ij}^{l}为各层级对应的特征参数分别映射到 l 层的特征权重，并且满足以下关系：

$$\alpha_{ij}^{l}=\frac{\mathrm{e}^{\lambda_{\alpha_{ij}}^{l}}}{\mathrm{e}^{\lambda_{\alpha_{ij}}^{l}}+\mathrm{e}^{\lambda_{\beta_{ij}}^{l}}+\mathrm{e}^{\lambda_{\gamma_{ij}}^{l}}}, \tag{3}$$

$$\alpha_{ij}^{l}+\beta_{ij}^{l}+\gamma_{ij}^{l}+\delta_{ij}^{l}=1. \tag{4}$$

2.3 频域感知特征融合方法

标准特征融合方法是当前计算机视觉领域应用最广泛的方式，其将低级的高分辨率特征和高级的粗略特征融合，以获得细节和语义信息。标准的特征融合方法可以表示为

$$Y^{l}=\mathcal{F}^{\mathrm{UP}}(Y^{l+1})+X^{l}, \tag{5}$$

式中：$X^{l}\in\mathbb{R}^{C\times 2H\times 2W}$；$Y^{l+1}\in\mathbb{R}^{C\times H\times W}$分别代表骨架生成的第 l 层特征和上采样后的第 l+1 层特征；$\mathcal{F}^{\mathrm{UP}}$表示上采样。但是这种方式存在两个显著影响密集预测的缺陷：类别内不一致性和边界位移。因此，通过引入频域感知特征融合 FreqFusion[19] 方法来解决传统特征融合方法存在的问题。主要组件结构如图 3 所示，包括：

1) 自适应低通滤波器生成器 (ALPF)，如图 3(a) 所示。预测空间变化的低通滤波器，在上采样过程中减少对象内部的高频成分，降低类别内不一致性。

2) 偏移生成器 (Offset)，如图 3(b) 所示。通过重采样，用更一致的特征替换大的不一致特征，使得同一类目标特征更稳定一致。

3) 自适应高通滤波器生成器 (AHPF)，如图 3(c) 所示，增强在下采样过程中丢失的高频细节边界信息。

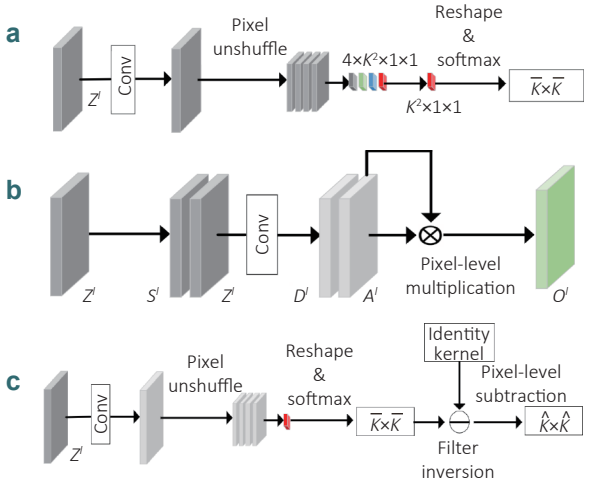


图 3 FreqFusion 主要组件结构：(a) ALPF 生成器；(b) Offset 生成器；(c) AHPF 生成器

Fig. 3 The structural composition of the key components of FreqFusion; (a) ALPF generator; (b) Offset generator; (c) AHPF generator

这些组件共同组成了 FreqFusion 结构，如图 4 所示。其中，偏移量生成器Z^{l}和S^{l}作为输入并预测偏移量。它们由两个 3×3 卷积层组成，用于预测偏移量方向和偏移量尺度，D^{l}表示偏移的方向，A^{l}用于控制偏移的大小，O^{l}是每个高级特征像素的最终预测偏移量，G 表示偏移组的数量。r 表示用于减少三个生成器后续计算成本的通道缩减率，$\overline{K}$和$\widehat{K}$分别表示低通滤波器和高通滤波器，Resamp 代表重采样过程，⊕和⊛分别表示像素级相加和卷积操作。FreqFusion 通过两个主要阶段运行，即初始融合和最终融合。初始融合首先通过像素反洗牌操作将特征的空间维度缩小一半，并利用 ALPF 生成器将压缩的低级特征作为输入，并生成一个初始的低通滤波器来对压缩的高级特征进行上采样，使得低电平中存在的高分辨率结构可以有利于上采样粗略的高电平特征。然后通过与像素反洗牌操作相反的像素洗牌操作后与低级特

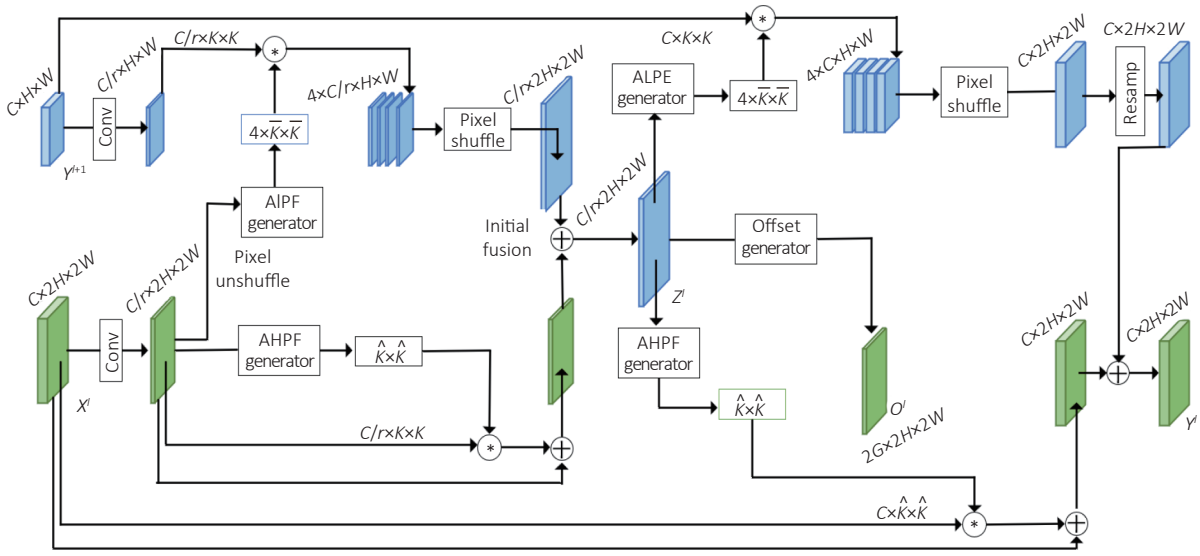


图 4 FreqFusion 结构

Fig. 4 FreqFusion structure

征融合，一部分输入 ALPF 生成器，另一部分输入 AHPF 生成器，AHPF 生成器生成的精细调谐的高频细节丰富了特征表示，促进了更有效的下游处理。增强后的初始融合使后续生成器能够更好地适应特征内容，从而产生更精细的最终融合结果。

2.4 上下文锚点注意力 (CAA) 模块

C2PSA (Cross stage partial with pyramid squeeze attention) 是 YOLOv11 新增的核心模块，是 CSP 结构与 PSA 注意力机制的结合体。通过 CSP 分段处理把特征图分为两部分，一部分直接传递，另一部分通过 PSA 注意力模块融合，最终拼接融合形成 C2PSA 模块。PSA 注意力机制通过动态调整特征来调整不同位置的关注度，增强模型对目标细节的感知，提升复杂场景下的检测精度。但是上下文特征的信息和关键区域的特征融合不足。为了充分把握检测过程中不同尺度目标的上下文相互依赖性，提高检测精度。本文引入了上下文锚点注意力 CAA[20] 模块，结合 C2PSA 模块形成全新的 C2PSA-CAA 模块。CAA 注意力模块的核心在于充分组合和利用其独特的全局平均池化和一维带状卷积，这种策略能够使网络通过精准捕捉图像中近距离和远距离像素的潜在关系，增强中心区域特征学习能力，其结构如图 5 所示。首先对输入特征进行全局平均池化，提取全局上下文信息。然后通过水平和垂直方向的 1D 卷积，近似模拟大卷积核的效果，捕获长距离的像素关系。最后通过 Sigmoid 函数生成注意力权重，用于增强模型输出特征，从而提升模型对长距离上下文信息的感知能力。

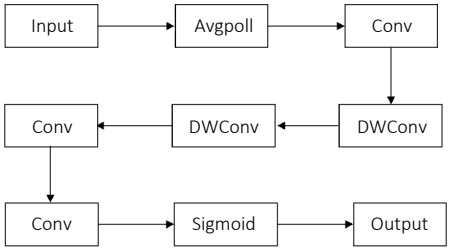


图 5 CAA 结构

Fig. 5 CAA structure

3 实验结果与分析

3.1 数据集

实验使用 VisDrone2019 数据集[21]，该数据集是由天津大学制作的开源无人机目标检测数据集，取自中国 14 个不同城市。该数据集具有环境、光照和天气复杂且多样的特点。待检测的稀疏和密集目标样本数足够多且小目标占比大。默认的训练集、验证集、测试集分别为 6471 张图片、548 张图片和 1610 张图片。数据集共包含 10 类检测对象，分别为行人、人、自行车、汽车、面包车、卡车、三轮车、遮阳篷三轮车、公交车和摩托

车，每个类别都有若干个样本，各类别的目标检测数量总和如表 1 所示。

表 1 VisDrone2019 数据集
Table 1 VisDrone2019 dataset

Class	Number
Pedestrian	79337
People	27059
Bicycle	10477
Car	144865
Van	24950
Truck	12871
Tricycle	4803
Awning-tricycle	3243
Bus	5926
Motor	29642

3.2 实验环境

实验平台的操作系统为 Windows 10 专业版，Intel (R) Core (TM) i9-10900K CPU@3.70 GHz，64 G，GeForce RTX 3090 24 G，编码平台为 PyCharm，Python 3.9，深度学习的主要框架是 Pytorch，CUDA 12.0。

网络训练共进行 200 轮 (epoch)，初始学习率 0.01，目标学习率 0.1。训练策略上，前 3 个 epoch 采用 Warmup 预热，后续则应用余弦退火算法以实现学习率的平稳衰减。

3.3 评价指标

为建立标准化性能基准并实现跨模型可比性研究，本研究采用多维量化指标体系，包含以下核心评价要素：准确率 (precision, *P*)、召回率 (recall, *R*)、平均精度 mAP@0.5、帧率 (FPS)、参数量 (Params)。其中 *P* 定义模型预测结果的可信度；*R* 表征模型覆盖真实目标的能力；mAP@0.5 是基于交并比阈值 (*IoU*=0.5) 的综合评估，全面反映模型定位与分类性能的协同表现；FPS 是单位时间内可处理的图像帧数；Params 是模型可训练参数总量，决定边缘部署可行性。计算公式如下：

$$P = \frac{TP}{TP+FP}, \tag{6}$$

$$R = \frac{TP}{TP+FN}, \tag{7}$$

$$AP = \int_0^1 P(R)\mathrm{d}R, \tag{8}$$

$$mAP = \frac{1}{N}\sum_{i=0}^{n} AP_i, \tag{9}$$

式中：*TP* (true positive) 真阳性表示模型正确地检测出了真实目标；*FP* (false positive) 假阳性表示模型错误地将背景或非目标区域识别为目标；*FN* (false negative) 假阴性表示真实存在的目标没有被模型检测出来。

3.4 实验结果与数据分析

为了从数据上评估模型的性能，选取了多种主流的模型进行比较。表 2 展示了不同算法在相同的实验条件下在 VisDrone2019 数据集上的对比实验数据。由实验数据可知，FFC-YOLO 相比 YOLOv11n 基准模型 mAP@0.5 提升了 7.8%。相比于二阶段算法 Fast-RCNN，mAP@0.5 提升了 28.4%，同时模型的参数量明显减少。与其他一阶段算法如 RetinaNet 相比，精度更优越。与 YOLO 算法，如 YOLOv3-tiny、YOLOv5n、YOLOv8n、YOLOv10n、YOLOv11n、YOLOv11s 相比，mAP@0.5 分别提升了 16.9%、8.8%、8.0%、7.6%、7.8%、0.7%。FFC-YOLO 在参数量低于 YOLOv11s 模型 15.78 M 的

表 2 不同网络在 VisDrone2019 数据集上的检测结果
Table 2 Detection results of different networks on VisDrone2019 dataset

Networks	mAP@0.5/%	Params/M	*P*/%	*R*/%	FPS
Fast-RCNN	11.6	28.4	12.2	10.3	6
RetinaNet	21.4	19.8	26.9	34.2	22
YOLOv3-tiny	23.1	9.52	39.5	29.4	124
YOLOv5n	31.2	2.18	42.1	31.3	138
YOLOv8n	32.0	2.69	42.3	32.6	147
YOLOv10n	32.4	2.69	42.2	32.9	136
YOLOv11n	32.2	2.58	41.5	32.9	126
YOLOv11s	39.3	19.8	49.3	37.9	132
FFC-YOLO	**40.0**	**4.02**	**49.0**	**38.8**	**106**

情况下，精度却提升了 0.7%。为了更好地比较改进模型和原始模型对每种类别的识别能力，统计 mAP@0.5 值，可视化结果如图 6 所示。图 6(a) 表示原始模型 YOLOV11n 的实验结果，对于遮挡严重的密集小目标，由于其尺寸小，容易被环境和其它目标遮挡、重叠，所以检测效果并不理想。由图 6(b) 可知，FFC-YOLO 在所有类别的检测精度都有显著提升。更直观的展示效果如表 3 所示。从表中可以看出，FFC-YOLO 相较于 YOLOv11n，对于行人、人、自行车、汽车、面包车、卡车、三轮车、遮阳篷三轮车、公交车和摩托车这 10 类检测目标，mAP@0.5 值分别提升了 12.3%、11.0%、6.1%、7.3%、7.0%、4.8%、4.8%、2.8%、8.9% 和 13.0%。这充分表明，在小目标类别检测任务中，FFC-YOLO 具有显著的优势。

为了验证在 C2PSA 模块中加入 CAA 注意力机制的必要性以及优越性，选取了一些主流注意力机制如 CBAM 和 ECA，以及模型优化模块如 iRMB 和 EMA，进行对比实验，实验结果如表 4 所示。其中 Base 为 YOLOv11n 中加入 FASFF 检测头和 FreqFusion 机制后的基线网络，在基线网络中每个检测头前加入并行且不融入 C2PSA 模块的 CAA 注意力机制与融入 C2PSA 模块后的 C2PSA-CAA 注意力机制作对比。可以看到，CAA 注意力机制与 C2PSA 融合后的 C2PSA-CAA 模块检测精度达到了最高 (40.0%)，比单独加入 CAA 注意力

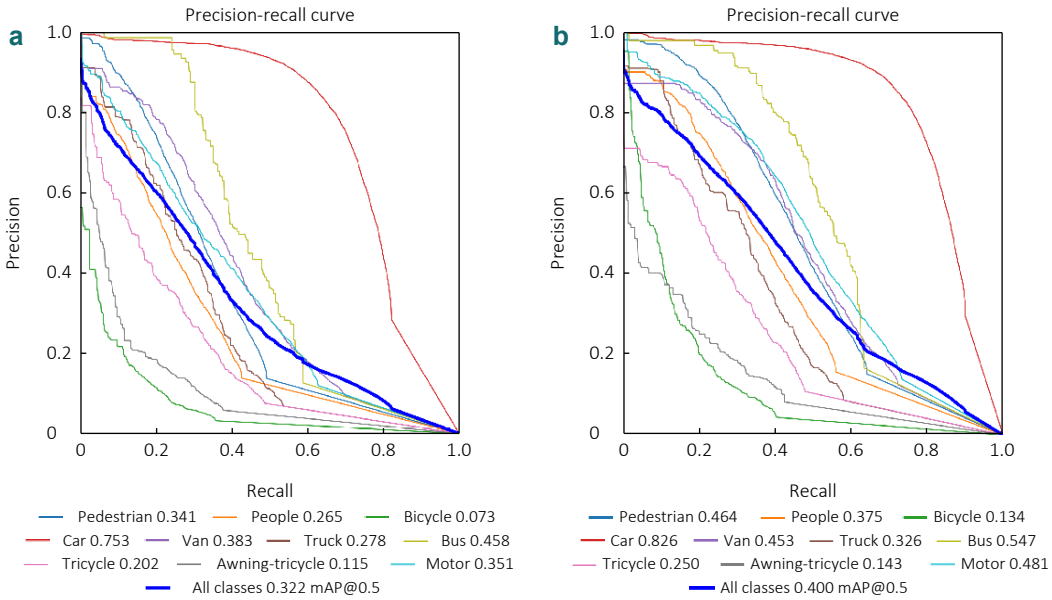


图 6 十种识别目标 mAP@0.5 值对比。(a) YOLOv11n 检测结果；(b) FFC-YOLO 检测结果

Fig. 6 Comparison of mAP@0.5 values for ten recognition targets; (a) Detection results of YOLO v11n; (b) Detection results of FFC-YOLO

表 3 密集小目标检测 mAP@0.5 值结果对比

Table 3 Dense small object detection mAP@0.5 results comparison

Class	mAP@0.5/%		Improved value
	YOLOv11n	FFC-YOLO	
Pedestrian	34.1	46.4	12.3
People	26.5	37.5	11.0
Bicycle	7.3	13.4	6.1
Car	75.3	82.6	7.3
Van	38.3	45.3	7.0
Truck	27.8	32.6	4.8
Tricycle	20.2	25.0	4.8
Awning-tricycle	11.5	14.3	2.8
Bus	45.8	54.7	8.9
Motor	35.1	48.1	13.0

表 4 添加不同注意力机制及模型优化模块的对比实验结果

Table 4 Comparative experiment results with adding different attention mechanisms or model optimization modules

Type	mAP@0.5/%	Params/M	*P*/%	*R*/%
Base	39.4	4.04	48.8	39.0
Base+caa	39.7	4.22	43.1	32.9
Base+C2PSA-CAA	40.0	4.02	49.0	38.8
Base+C2PSA-CBAM	39.6	4.04	48.2	37.8
Base+C2PSA-ECA	39.8	4.02	49.1	38.2
Base+C2PSA-iRMB	39.8	4.06	50.3	38.2
Base+C2PSA-EMA	39.7	4.00	48.8	38.8

机制提升了 0.3%，且参数量较单独加入 CAA 注意力机制下降了 0.2 M。在与主流注意力机制的对比中，C2PSA-CAA 检测精度比 CBAM 和 ECA 分别提升了 0.4% 和 0.2%，参数量和 C2PSA-ECA 相等，比 C2PSA-CBAM 减少了 0.02 M，且 *R* 值达到了最高 (38.8%)。而且，在与模型优化模块 iRMB 和 EMA 的对比中，C2PSA-CAA 检测精度比 iRMB 和 EMA 分别提升了 0.2% 和 0.3%，虽然参数量比 C2PSA-EMA 增加了 0.02 M，但 *P* 值提高了 0.2%。综上可知，C2PSA-CAA 模块更加适用于航拍图像小目标检测任务。

为了验证 FFC-YOLO 网络模型对像素在 32 pixel×32 pixel 及以下小目标的检测能力，另外选择了数据集 tiny-data 进行模型性能测试。tiny-data 数据集为无人机航拍图像下的自建自标注小目标图像数据集，数据集内标注的目标像素尺寸均小于 32 pixel×32 pixel。数据集内图片通过翻转、缩放、色域调整等方式进行数据增强后，按照 8:1:1 分为训练集、测试集和验证集，包括训练集图片 1152 张，测试集图片 145 张，验证集图片 145 张。目标类别为 sperson、lperson、wperson，分别代表站立状态的人，坐卧状态的人和水中状态的人；目标数量分别为 35240 个、26930 个、18110 个。YOLO v11n 和 FFC-YOLO 在此数据集上的检测结果如表 5 所示。结果表明，FFC-YOLO 在识别所有类别的目标都比 YOLOv11n 出色。FFC-YOLO 在识别 sperson 类时表现出色，精度较高，mAP@0.5 和 Precision 达到了最高的 75.4%、73.3%。FFC-YOLO 在识别 wperson 类精度较低，但 mAP@0.5 值也达到了 57.8%。从实验数据来看，在面对小目标数据集 tiny-data，FFC-YOLO 的小目标检测效果令人满意。

表 5　FFC-YOLO 在 tiny-data 数据集上的检测结果

Table 5　Detection results of FFC-YOLO on tiny-data dataset

Networks	Class	mAP@0.5/%	*P*/%	*R*/%
YOLO v11n	Sperson	67.5	64.4	65.8
	Lperson	62.0	61.1	60.8
	Wperson	49.1	57.6	45.6
	All	59.5	61.0	57.4
FFC-YOLO	Sperson	75.4	73.3	69.9
	Lperson	73.0	67.3	73.2
	Wperson	57.8	68.4	46.9
	All	68.7	69.7	63.3

3.5 消融实验

为了验证 FFC-YOLO 网络模型中各改进模块的实际效果，对 FFC-YOLO 网络模型进行消融实验。实验结果如表 6 所示，其中：Baseline 为未作任何修改的 YOLOv11n 的数据；模型 A 为替换 FASFF 检测头之后的数据，加入 FASFF 检测头后的 mAP@0.5 提升 5.0%，*P* 值也提升了 5.8%；模型 B 为单独加入频域感知特征融合机制 FreqFusion，可以看出加入 FreqFusion 机制后，参数量不仅下降了 0.14 M，而且检测精度提升了 0.5%，*P* 值达到 43.1%；模型 C 为在 A 的基础上加入频域感知特征融合机制 FreqFusion，可以看出加入 FreqFusion 机制后，参数量不仅较 A 下降了 0.01 M，而且检测精度提升了 2.2%，*R* 值达到 39%；模型 D 为在 B 的基础上加入 C2PSA-CAA 模块，可以看出参数量较 B 下降了 0.16 M，但检测精度下降了 0.1%；模型 E 为在 A 的基础上加入 CAA 注意力机制，可以看出参数量基本不变，但检测精度有 0.2% 的下降，*P* 值提升了 0.2%，*R* 值下降了 0.3%；模型 F 为在 C 的基础上加入了 C2PSA-CAA 模块，降低了 0.02 M 参数量的同时检测精度提升了 0.6%。从实验结果来看，FFC-YOLO 满足了检测精度、准确率和召回率较高的特点，可以更加有效的适用于无人机航拍图像小目标检测领域。

3.6 实验检测结果

在数据集 VisDrone2019 中，FFC-YOLO 的小目标检测能力在图 7 所示的实验结果中得到了凸显。相比 YOLOv11n (如图 7(a-c) 所示)，FFC-YOLO (如图 7(d-f) 所示) 显著提高了对远处小目标 (如行人、自行车、摩托车) 的检测准确率。YOLOv11n 在图 7(a, b) 中均存在较多小尺度车辆和行人的漏检问题，行人、摩托车和远处

表 6　消融实验结果

Table 6　Ablation experiment results

Network	FASFF	FreqFusion	C2PSA-CAA	mAP@0.5/%	Params/M	*P*/%	*R*/%
Baseline				32.2	2.58	41.5	32.9
A	√			37.2	4.05	47.3	36.5
B		√		32.7	2.44	43.1	32.9
C	√	√		39.4	4.04	48.8	39.0
D		√	√	32.6	2.42	43.7	32.7
E	√		√	37.0	4.04	47.5	36.2
F	√	√	√	40.0	4.02	49.0	38.8

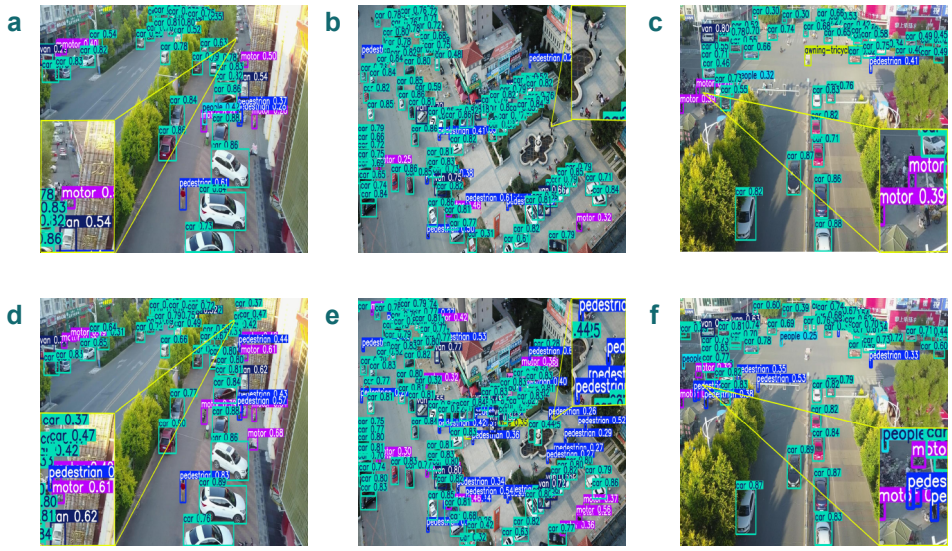


图 7 YOLOv11n 与 FFC-YOLO 检测结果。(a-c) YOLOv11n 检测结果；(d-f) FFC-YOLO 检测结果

Fig. 7 Detection results between YOLOv11n and FFC-YOLO. (a-c) Detection results of the YOLOv10n; (d-f) Detection results of the FFC-YOLO

汽车漏检数多达 80 个，整体对小尺度目标敏感度不足。然而，FFC-YOLO 不仅有效减少了各类尺度目标的漏检，而且检测置信度更高。综上，FFC-YOLO 在处理像素小于 32×32 的小目标上比 YOLO v11n 更精准，更适用于无人机航拍图像的小目标检测。

为了验证 FFC-YOLO 检测小目标的优越性，将 YOLOv11n 和改进后的模型用于检测自建数据集 tiny-data。YOLOv11n 检测结果如图 8(a, b) 所示，FFC-YOLO 检测结果如图 8(c, d) 所示。对于 3 类小目标 sperson、lperson 和 wperson，FFC-YOLO 检测出的目标数量均高于 YOLOv11n。对比图 8(a) 和图 8(c)，FFC-YOLO 检测出 3 类目标的个数比 YOLOv11n 分别高出了 25 个，1 个和 7 个。对比图 8(b) 和图 8(d)，FFC-YOLO 检测出 3 类目标的个数比 YOLOv11n 分别多出了 10 个，3 个和 14 个。结果表明，对于远处或密集条件下的小目标，FFC-YOLO 的检测能力更加出色。

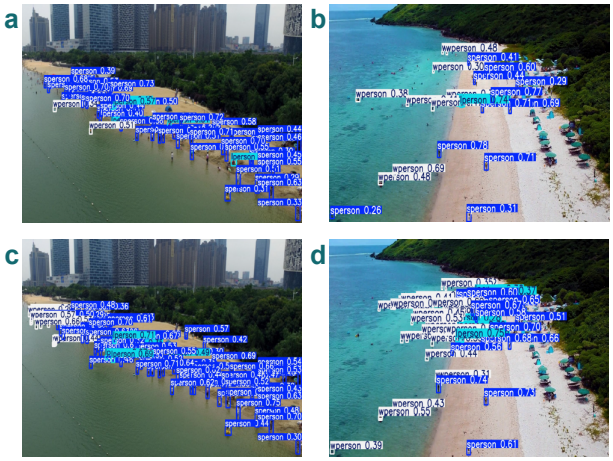


图 8 YOLOv11n 与 FFC-YOLO 对于 tiny-data 的检测结果。(a, b) YOLOv11n 检测结果；(c, d) FFC-YOLO 检测结果

Fig. 8 Detection results of YOLOv11n and FFC-YOLO on tiny-data. (a, b) The detection results of YOLOv11n; (c, d) The detection results of FFC-YOLO

4 结 论

为应对无人机航拍图像小目标检测中特征模糊、识别难度高的挑战，本文提出了新的小目标检测算法 FFC-YOLO。该算法从三个层面实现了创新：1) 增强特征提取：将头部检测头替换为 FASFF 结构，利用其四个分支强化对小目标的关注及多尺度检测能力；2) 优化特征融合：设计频域感知特征融合策略改进传统采样过程，并将此融合模块与 BiFPN 结构耦合，有效提升了融合效率，解决了密集小目标场景下的特征丢失、类别误判和边界偏移问题；3) 引入上下文感知：在 C2PSA 模块中加入上下文锚点注意力机制，使特征提取过程能更好地利用目标周围的语义信息。在 VisDrone2019 数据集中，基于 FFC-YOLO 较传统算法如 Fast-RCNN、RetinaNet 分别提升 28.4%、18.6%，较 YOLO 算法如 YOLOv8n、YOLOv10n、YOLOv11n 分别提升 8.0%、7.6%、7.8%。对于行人、人、自行车、汽车、面包车、卡车、三轮车、

遮阳篷三轮车、公交车和摩托车这 10 类检测目标，FFC-YOLO 相较于 YOLOv11n，mAP@0.5 值分别提升了 12.3%、11.0%、6.1%、7.3%、7.0%、4.8%、4.8%、2.8%、8.9% 和 13.0%。在自建小目标数据集 tiny-data 中，基于 FFC-YOLO 的 3 类目标 sperson、lperson 和 wperson 的检测结果较 YOLOv11n，mAP@0.5、*P*、*R* 分别提升了 9.2%、8.7%、5.9%。在小目标类别检测任务中，FFC-YOLO 检测优势较为突出。后续工作旨在利用模型剪枝和蒸馏技术精简模型，实现高效检测算法的实际部署。

作者贡献

唐彪、梁诺雷：研究资料收集、分析、论文写作；贾小军、洛淑云、张春花：指导论文选题、论文修改、研究资料收集。

利益冲突

所有作者声明无利益冲突

参考文献

1. Li X Y, Wang Z F, Wu H. Research on UAV target tracking algorithm for simulation scenarios[J]. *J Syst Simul*, 2025.
   李心仪, 王振飞, 吴晗. 面向仿真场景的无人机目标跟踪算法研究[J]. 系统仿真学报, 2025.
2. Cao K, Li N, Wang L Z, et al. Research on obstacle avoidance strategy of patrol UAV based on monocular vision[J]. *Chin J Electron Science Devices*, 2025, **48**(1): 86−90.
   曹凯, 李宁, 王立志, 等. 基于单眼视觉的巡检无人机避障策略研究[J]. 电子器件, 2025, **48**(1): 86−90.
3. Zhuang Y, Fu X J, Li S, et al. UAV small target detection based on XMB-YOLOv5s[J]. *Comput Mod*, 2025, (4): 29−35,41.
   庄瑜, 傅晓锦, 李莎, 等. 基于 XMB-YOLOv5s 的无人机小目标检测[J]. 计算机与现代化, 2025, (4): 29−35,41.
4. Ghen M Q, Yu L J, Zhi C, et al. Improved faster R-CNN for fabric defect detection based on Gabor filter with genetic algorithm optimization[J]. *Comput Ind*, 2022, **134**: 103551.
5. Terven J, Córdova-Esparza D M, Romero-González J A. A comprehensive review of YOLO architectures incomputer vision: from YOLOv1 to YOLOv8 and YOLO-NAS[J]. *Mach Learn Knowl Extr*, 2023, **5**(4): 1680−1716.
6. Wang A, Chen H, Liu L H, et al. YOLOv10: real-time end-to-end object detection[C]//*Proceedings of the 38th International Conference on Neural Information Processing Systems*, 2025: 3429.
7. Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements[Z]. arXiv: 2410.17725, 2024. https://arxiv.org/abs/2410.17725.
8. Zhang Y, Cheng Z D, Zhu B. Small target detection algorithm of drones based on improved faster R-CNN[J]. *Chin J Quantum Electron*, 2022, **39**(3): 354−363.
   张杨, 程正东, 朱斌. 基于改进 Faster R-CNN 的无人机小目标检测算法[J]. 量子电子学报, 2022, **39**(3): 354−363.
9. Li Z H, Lin C J, Zhu H L, et al. Small object detection based on structure perception and global context information[J]. *Comput Eng Appl*, 2024, **60**(9): 292−298.
   李钟华, 林初俊, 朱恒亮, 等. 基于结构感知和全局上下文信息的小目标检测[J]. 计算机工程与应用, 2024, **60**(9): 292−298.
10. Li L X, Wang X, Wang J, et al. Small object detection algorithm in UAV image based on feature fusion and attention mechanism[J]. *J Graphics*, 2023, **44**(4): 658−666.
    李利霞, 王鑫, 王军, 等. 基于特征融合与注意力机制的无人机图像小目标检测算法[J]. 图学学报, 2023, **44**(4): 658−666.
11. Han J, Yuan X P, Wang Z, et al. UAV dense small target detection algorithm based on YOLOv5s[J]. *J Zhejiang Univ (Eng Sci)*, 2023, **57**(6): 1224−1233.
    韩俊, 袁小平, 王准, 等. 基于 YOLOv5s 的无人机密集小目标检测算法[J]. 浙江大学学报 (工学版), 2023, **57**(6): 1224−1233.
12. Niu W H, Wei Y L. Small target detection in aerial photography images based on improved YOLOv7 algorithm[J]. *Electron Opt Control*, 2024, **31**(1): 117−122.
    牛为华, 魏雅丽. 基于改进 YOLOv7 的航拍小目标检测算法[J]. 电光与控制, 2024, **31**(1): 117−122.
13. Chang K X, Huang J H, Sun X Y, et al. Optical small target detection method by drone based on dual-modal image fusion[J]. *Laser Optoelectron Prog*, 2025, **62**(4): 0428001.
    常凯旋, 黄建华, 孙希延, 等. 基于双模态图像融合的无人机光学小目标检测算法[J]. 激光与光电子学进展, 2025, **62**(4): 0428001.
14. Zheng H W, Wang F, Gao J B. DES-YOLO: a more accurate object detection method[J]. *Opto-Electron Eng*, 2024, **51**(11): 240212.
    郑华伟, 王飞, 高建邦. DES-YOLO: 一种更精确的目标检测方法[J]. 光电工程, 2024, **51**(11): 240212.
15. Zhang R M, Xiao Y F, Jia Z N, et al. Improved YOLOv7 algorithm for target detection in complex environments from UAV perspective[J]. *Opto-Electron Eng*, 2024, **51**(5): 240051.
    张润梅, 肖钰霏, 贾振楠, 等. 改进 YOLOv7 的无人机视角下复杂环境目标检测算法[J]. 光电工程, 2024, **51**(5): 240051.
16. Tang B, Jia X J, Luo S Y, et al. BL-YOLO: a target detection algorithm for aerial images captured by drones[J]. *J Optoelectron·Laser*, 2025. http://kns.cnki.net/kcms/detail/12.1182.o4.20250423.1727.002.html.
    唐彪, 贾小军, 骆淑云, 等. BL-YOLO: 无人机航拍图像目标检测算法[J]. 光电子·激光, 2025. http://kns.cnki.net/kcms/detail/12.1182.o4.20250423.1727.002.html.
17. Ma T, Xiong Y Z, Pan Q N, et al. Infrared small object detection for UAV imagery based on VSA-YOLO11n[J]. *Infrared Technol*, 2025. http://kns.cnki.net/kcms/detail/53.1053.tn.20250509.1602.002.html.
    马滔, 熊英灼, 潘庆娜, 等. 基于 VSA-YOLO 11n 的无人机红外小目标检测[J]. 红外技术, 2025. http://kns.cnki.net/kcms/detail/53.1053.tn.20250509.1602.002.html.
18. Ye R, Ma Z F, Gao Q, et al. Target detection of tea disease based on improved YOLOv5s-ECA-ASFF algorithm[J]. *J Chin Agric Mech*, 2024, **45**(1): 244−251.
    叶荣, 马自飞, 高泉, 等. 基于改进 YOLOv5s-ECA-ASFF 算法的茶叶病害目标检测[J]. 中国农机化学报, 2024, **45**(1): 244−251.
19. Chen L W, Fu Y, Gu L, Yan C, et al. Frequency-aware feature fusion for dense image prediction[J]. *IEEE Trans Pattern Anal Mach Intell*, 2024, **46**(12): 10763−10780.
20. Cai X H, Lai Q X, Wang Y W, et al. Poly kernel inception network for remote sensing detection[C]//*Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 27706–27716. https://doi.org/10.1109/CVPR52733.2024.02617.
21. Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10778–10787. https://doi.org/10.1109/CVPR42600.2020.01079.

作者简介



【通信作者】贾小军 (1974-)，四川邻水人，嘉兴大学教授，研究方向为计算机视觉与模式识别，图像处理。

E-mail：xjjiad@sina.com

A small target detection algorithm for UAV aerial images based on frequency domain perception feature fusion

Jia Xiaojun[1,2*], Tang Biao[1,2], Luo Shuyun[2], Liang Nuolei[1,2], Zhang Chunhua[1]

[1] College of Artificial Intelligence, Jiaxing University, Jiaxing, Zhejiang 314001, China;

[2] School of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou, Zhejiang 310018, China

Abstract

Objective Small object detection remains a significant and persistent challenge within the computer vision domain. This challenge is particularly acute in the context of drone aerial imagery, where objects captured from high altitudes appear extremely small, densely packed, and often exhibit blurred features due to low resolution and environmental factors. Traditional object detection algorithms frequently suffer from performance degradation in these scenarios, manifesting as high rates of missed detections and false positives. To address these critical limitations, this paper proposes a novel object detection algorithm named FFC-YOLO (FASFF-FreqFusion-CAA YOLO), which is built upon a frequency-aware feature fusion framework.

Methods The core of the FFC-YOLO algorithm lies in three systematic and synergistic innovations designed to enhance feature representation, fusion, and contextual understanding specifically for small and dense objects. First, to bolster multi-scale feature extraction, we redesign the detection head by incorporating a Four Adaptively Spatial Feature Fusion (Detect-FASFF) structure. This is augmented with an additional P2 detection layer dedicated to capturing finer-grained spatial information crucial for identifying minuscule targets. The FASFF mechanism leverages four adaptive branches to dynamically aggregate features across different scales, thereby significantly improving the model's sensitivity and discriminative power for small objects across varying sizes.Second, we comprehensively optimize the feature fusion pathway to counteract the information loss common in conventional sampling methods (e.g., up-sampling and down-sampling). We introduce a Frequency-aware Feature Fusion (FreqFusion) module. This module operates by decomposing features into different frequency components within the frequency domain, allowing for more deliberate and effective integration of high-frequency details (essential for edge and texture information of small objects) and low-frequency semantics. This FreqFusion module is then elegantly coupled with a Bidirectional Feature Pyramid Network (BiFPN) structure. The resulting fusion architecture enables efficient, multi-level bidirectional cross-scale information flow, effectively mitigating the problems of feature obliteration, category misclassification, and bounding box localization drift in scenes crowded with small objects.

Third, to enhance the model's ability to leverage surrounding semantic context—a vital clue when object appearance is ambiguous—we integrate a Context Anchor Attention (CAA) mechanism into the C2PSA module. This attention mechanism functions by establishing dynamic associations between anchor points and their contextual surroundings, enabling the feature extraction process to selectively focus on and amplify informative regions around potential targets. This contextual awareness strengthens feature representation and aids in distinguishing small objects from complex backgrounds.

Results and Discussions The efficacy of the proposed FFC-YOLO algorithm is rigorously validated through extensive experiments on two benchmark datasets. On the public VisDrone2019 dataset, a standard benchmark for drone vision, FFC-YOLO achieves a mean Average Precision (mAP@0.5) of 40.0%. This represents a substantial performance gain of 8.0%, 7.6%, and 7.8% over the widely-

Foundation item: Public Welfare Projects of Jiaxing Science and Technology Bureau(2024AY40009)

*Correspondence: Jia X J, E-mail: xjjiad@sina.com

recognized YOLOv8n, YOLOv10n, and YOLOv11n baselines, respectively. To further verify its generalization capability and prowess in detecting extremely small targets, we conducted additional experiments on a custom-built, challenging dataset named "tiny-data", which contains three categories of progressively smaller persons ("sperson", "lperson", "wperson"). On this dataset, FFC-YOLO demonstrates superior results compared to the strong YOLOv11n baseline, with improvements of 9.2% in mAP@0.5, 8.7% in Precision (*P*), and 5.9% in Recall (*R*). These consistent gains across both public and private datasets robustly confirm that the FFC-YOLO algorithm possesses strong generalization ability and excels in the demanding task of drone-based aerial image analysis, particularly for long-range and dense small object detection.

Conclusions In conclusion, this paper presents a comprehensive and effective solution to the small object detection problem in aerial imagery. By innovating at the levels of feature extraction (FASFF head), feature fusion (FreqFusion-BiFPN), and contextual modeling (CAA attention), the FFC-YOLO framework successfully addresses key shortcomings of prior methods. The experimental evidence underscores its potential for practical application in UAV-based surveillance, traffic monitoring, and search-and-rescue operations. For future work, we plan to explore model compression techniques such as pruning and knowledge distillation to streamline the model for efficient deployment on resource-constrained edge devices, thereby bridging the gap between high accuracy and real-time performance.

Keywords: FFC-YOLO; small target detection; UAV aerial image; FreqFusion; YOLOv11n

Jia X J, Tang B, Luo S Y et al. A small target detection algorithm for UAV aerial images based on frequency domain perception feature fusion[J]. *Opto-Electron Eng*, 2026, **53**(2): 250208; DOI: 10.12086/oee.2026.250208

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文