

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

基于超图计算的高效传递多尺度特征小目标检测算法

孙叶美, 桑学婷, 张艳, 刘国瑞, 陈帅宇

引用本文:

孙叶美, 桑学婷, 张艳, 等. 基于超图计算的高效传递多尺度特征小目标检测算法[J]. 光电工程, 2025, 52(5): 250061.

Sun Y M, Sang X T, Zhang Y, et al. Hypergraph computed efficient transmission multi-scale feature small target detection algorithm[J]. *Opto-Electron Eng*, 2025, 52(5): 250061.

<https://doi.org/10.12086/oe.2025.250061>

收稿日期: 2025-02-28; 修改日期: 2025-04-08; 录用日期: 2025-04-08

相关论文

基于多尺度特征融合的遥感图像小目标检测

马梁, 苟于涛, 雷涛, 靳雷, 宋怡萱

光电工程 2022, 49(4): 210363 doi: 10.12086/oe.2022.210363

基于多尺度特征增强的高效Transformer语义分割网络

张艳, 马春明, 刘树东, 孙叶美

光电工程 2024, 51(12): 240237 doi: 10.12086/oe.2024.240237

特征协同与细粒度感知的遥感图像小目标检测

肖振久, 张杰浩, 林瀚翰

光电工程 2024, 51(6): 240066 doi: 10.12086/oe.2024.240066

多级特征筛选和任务动态对齐的声呐图像小目标检测

王燕, 王宏辉, 刘树东, 张艳, 郝泽玉

光电工程 2024, 51(10): 240196 doi: 10.12086/oe.2024.240196

更多相关论文见光电期刊集群网站 



光电工程
Opto-Electronic Engineering

<http://cn.ojournal.org/oe>



 OE_Journal



Website

基于超图计算的高效传递多尺度特征小目标检测算法

孙叶美, 桑学婷, 张 艳, 刘国瑞*, 陈帅宇*

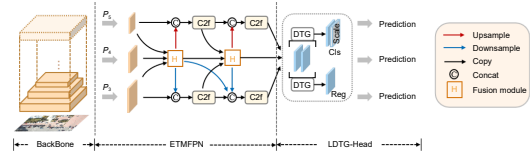
天津城建大学计算机与信息工程学院, 天津 300380

摘要: 无人机航拍图像具有背景复杂、目标小且密集的特点。在无人机航拍图像检测中, 存在小目标检测精度低和模型参数量大的问题。因此, 提出基于超图计算的多尺度特征高效传递小目标检测算法。首先, 设计高效传递多尺度特征金字塔网络作为颈部网络, 在中间层融合多层特征, 并将其向相邻各层直接传递, 有效缓解传递路径冗长导致的信息丢失问题。另外, 特征融合环节借助超图对高阶特征进行建模, 从而增强模型的非线性表达能力。其次, 设计轻量化动态任务引导检测头, 借助共享机制在参数量较少的情况下, 有效解决传统解耦头中分类与定位任务空间不一致导致检测目标不准确的问题。最后, 基于层自适应幅度的剪枝轻量化模型, 进一步减小模型体积。实验结果表明, 此算法在 VisDrone2019 数据集上表现出比其他架构更优越的性能, 精度 $mAP_{0.5}$ 和参数量分别达到了 42.4% 和 4.8 M, 与基准 YOLOv8 相比参数量降低了 54.7%, 该模型实现检测性能与资源耗费之间的良好平衡。

关键词: 小目标检测; 超图; 解耦头; 轻量化

中图分类号: TP391.4

文献标志码: A



孙叶美, 桑学婷, 张艳, 等. 基于超图计算的高效传递多尺度特征小目标检测算法 [J]. 光电工程, 2025, 52(5): 250061

Sun Y M, Sang X T, Zhang Y, et al. Hypergraph computed efficient transmission multi-scale feature small target detection algorithm[J]. *Opto-Electron Eng*, 2025, 52(5): 250061

Hypergraph computed efficient transmission multi-scale feature small target detection algorithm

Sun Yemei, Sang Xueting, Zhang Yan, Liu Guorui*, Chen Shuaiyu*

School of Computer and Information Engineering, Tianjin Chengjian University, Tianjin 300380, China

Abstract: UAV aerial images have the characteristics of complex background, small and dense targets. Aiming at the problems of low precision and a large number of model parameters in UAV aerial image detection, an efficient multi-scale feature transfer small target detection algorithm based on hypergraph computation is proposed. Firstly, a multi-scale feature pyramid network is designed as a neck network to effectively reduce the problem of information loss caused by lengthy transmission paths by fusing multi-layer features in the middle layer and transmitting them directly to adjacent layers. In addition, the feature fusion process uses hypergraphs to model higher-order features, improving the nonlinear representation ability of the model. Secondly, a lightweight dynamic task-guided detection head is designed to effectively solve the problem of inaccurate detection targets caused by

收稿日期: 2025-02-28; 修回日期: 2025-04-08; 录用日期: 2025-04-08

基金项目: 国家级大学生创新创业训练计划 (202410792012); 天津市哲学社会科学规划项目 (TJGL19XSX-045)

*通信作者: 刘国瑞, lgr@tcu.edu.cn; 陈帅宇, chenshuaiyu0906@163.com。

版权所有©2025 中国科学院光电技术研究所

inconsistent classification and positioning task space in the traditional decoupling head with a small number of parameters through sharing mechanism. Finally, the pruning lightweight model based on layer adaptive amplitude is used to further reduce the model volume. The experimental results show that this algorithm has better performance than other architectures on VisDrone2019 dataset, with the accuracy $mAP_{0.5}$ and parameter number reaching 42.4% and 4.8 M, respectively. Compared with the benchmark YOLOv8, the parameter number is reduced by 54.7%. The model achieves a good balance between detection performance and resource consumption.

Keywords: small target detection; hypergraph; decoupling head; light weight

1 引言

无人机具有体积小、隐蔽性好、操作简便等优势, 在多领域广泛应用, 然而, 无人机航拍图像普遍存在背景复杂、目标尺寸小、分布稀疏不均等特点, 导致目标检测精度低, 且需依赖复杂、庞大的模型才能提升精度。在无人机航拍任务中面临着检测环境复杂多变以及模型规模受限的问题, 研发高效、精准的目标检测方法至关重要。

目前, 主流目标检测方法大体可分为两阶段算法(如 R-CNN^[1]系列、Mask R-CNN^[2]、Cascade R-CNN^[3])和一阶段算法(如 YOLO^[4]系列、SSD^[5]、RetinaNet^[6])。两阶段算法先生成候选区域再进行目标定位和分类, 精度较高但计算成本大。相比之下, 一阶段算法直接通过单一网络进行定位和分类, 具有更高的实时性, 更适合无人机航拍图像的目标检测需求。选择 YOLOv8 作为基础网络, 并对其特征提取和表示能力进行优化, 以提升小目标检测精度并降低模型复杂度。

YOLO 框架的颈部模块^[7-10]是融合高层语义信息和低层空间细节的关键组件, 对增强模型在不同尺度下目标检测能力至关重要。然而, 传统的特征金字塔网络 FPN^[11]仅考虑单向信息流, 存在信息利用率不足的问题。尽管 BiFPN^[12]、PAFPN^[13]等方法设计多路径融合方式, 但计算成本较高, 且对特征间的复杂交互关系建模能力有限。AFPN^[14]、MAF-YOLO^[15]、DAMO-YOLO^[16]和 Gold-YOLO^[17]等方法虽然在一定程度上改进了特征融合策略, 但仍未充分挖掘不同尺度、位置和空间特征之间复杂的非线性高阶信息交互。增强对高阶特征的感知可以帮助模型过滤掉背景中的干扰信息。超图^[18-19]结构, 通过其独特的超边连接多个节点的方式, 能够有效地建模数据间的复杂交互关系, 为更深入地理解和利用特征间的复杂交互关系提供了新的途径, 且显著提升 YOLO 模型的目标检测

性能。

在传统的解耦检测头中, 目标检测任务被分解为两个独立的子任务: 目标定位和目标分类。分类分支主要关注图像的中心区域, 因为该区域通常包含丰富的语义信息, 有利于目标类别的识别; 回归分支则主要关注图像的边界区域, 因为该区域包含关于目标位置的关键信息, 有利于精确地定位目标边界框。然而, 在这种解耦设计中, 不同尺度的检测头分别进行卷积操作, 导致模型参数利用效率低, 且容易过拟合, 并增加模型计算负担。为了解决这个问题, Liu 等^[20]提出了一种轻量级的解耦头, 通过动态计算特定任务的重要区域来提高效率; StarDL-YOLO 则在检测头中引入了共享卷积层 (LSCD), 显著降低了计算量^[21]。这些方法在一定程度上实现了轻量化, 但由于两个分支是并行且独立训练的, 特征之间缺乏有效的交互, 且可能存在空间信息不对齐的问题, 最终导致预测结果中出现空间不一致性, 例如分类分支可能准确地识别目标类别, 但回归分支却由于空间信息错位而预测出不准确的边界框。

为进一步降低模型的复杂度, 许多学者针对模型轻量化进行了研究。典型的轻量级网络包括 SqueezeNet^[22]、ShuffleNet^[23]、MobileNet^[24]、EfficientNet^[25]和 GhostNet^[26]。但由于轻量级网络结构过于简单, 导致网络的特征提取与融合能力减弱, 无法充分提取待检测目标的特征信息和位置信息。王舒梦等^[27]提出了一种基于轻量化 PartialConv 卷积的模型轻量化方法。张佳承等^[28]在颈部网络中引入了 GSConv 和 slim-neck, 有效减少了模型的参数数量。此外, 也有研究^[29-30]表明, 通过合理的超参数调整和神经元剪枝技术, 可以有效地去除对神经网络性能影响较小的神经元, 从而降低模型参数数量和计算负担, 提升模型效率。

因此, 针对上述小目标检测中存在的问题, 提出

了一种基于超图计算的高效传递多尺度特征小目标检测算法 (hypergraph computing-based efficient transmission multi-scale feature network, HETMNet), 旨在使算法具有较高精准率的同时减少参数数量。主要贡献概述如下:

1) 设计了一种通用的高效传递多尺度特征金字塔网络 (efficient transmission multi-scale feature pyramid network, ETMFPN) 作为颈部网络, 实现了融合特征以最短路径向不同层级传递, 极大促进了高低层特征的充分交换, 有效减少了信息的损失。在融合阶段, 设计了高阶自适应融合模块 (high order adaptive fusion module, HOAF), 不仅能够实现多尺度特征的自适应融合, 而且通过引入超图计算, 显著增强了模型对高阶特征的感知能力, 有效抑制背景中的噪声, 增强复杂背景下小目标检测能力。

2) 设计了轻量化动态任务引导检测头 (lightweight dynamic task guide detection head, LDTG-Head)。基于共享机制将不同层级的 FPN 语义信息通过相同的 LDTG-Head, 促进分类和定位任务的动态对齐, 避免两个分支在预测中产生空间不一致的问题, 并显著降低了参数量。

3) 对改进模型采取基于层自适应幅度的剪枝 (LAMP)^[31] 方法进行剪枝, 使得模型在保持性能的前提下有效减少模型参数量, 便于模型进行轻量化部署。

2 检测算法

选用 YOLOv8 作为基础网络, 设计一种基于超图计算的高效传递多尺度特征小目标检测算法 (HETMNet)。HETMNet 的模型体系结构由主干、ETMFPN 颈部以及 LDTG-Head 三个部分构成。

HETMNet 网络整体的架构如图 1 所示。首先, 采用 CSPDarkNet53 作为主干进行特征提取。随后, 将获取的特征图 P_3 、 P_4 、 P_5 输入至 ETMFPN 中, 以高阶特征为基础实现深、浅层特征信息自适应融合, 整合跨层次和跨位置的信息并分散到其他各层中, 使每一层都可以获得跨三种分辨率的多样化特征。最后, 再经过 LDTG-Head 检测头, 进行多尺度预测。

2.1 高效传递多尺度特征金字塔网络

特征金字塔旨在从深度神经网络的主干中提取具有不同分辨率的特征, 进而实现特征融合以增强模型的多尺度检测能力。然而, 现有的特征金字塔网络由于特征传递路径冗长, 普遍存在传递过程信息丢失的问题, 模型对小目标的检测性能受到了限制。为了有效解决这些问题, 设计了 ETMFPN, 如图 1 所示。在融合阶段, ETMFPN 设计 HOAF 模块引入超图计算同时对浅层 P_{n-1} 和深层 P_{n+1} 进行多尺度特征融合, 为网络提供细粒度的边缘细节信息和有助于目标分类的粗粒度语义信息, 如图 2 所示。

首先, 深层特征 X_{n+1} 与浅层特征 X_{n-1} 分别通过双线性插值算法和小波池化与当前层特征 X 尺寸对齐, 并通过 1×1 卷积调整通道数。普通池化在进行下采样时, 往往会造成高频分量丢失, 导致特征丢失或特征平滑, 进而使图像细节信息受损。但小波池化能够把特征分解为高频子带与低频子带, 通过对高频和低频子带进行加和操作, 留存了低频分量里缺失的高频成分, 有效减少浅层纹理细节特征丢失。然后, 当前层特征 X 经过超图计算模块 (图 3), 将 P 层特征的尺寸从 (H, W, C) (H 和 W 分别表示图像的高度和宽度, C 表示通道数量) 变换到 $(H \times W, C)$ 得到特征 X_1 , 根据 X_1 计算特征与特征之间的距离, 并以每个特征点为中心将小于距离阈值的特征点构建成超边, 所有超边的

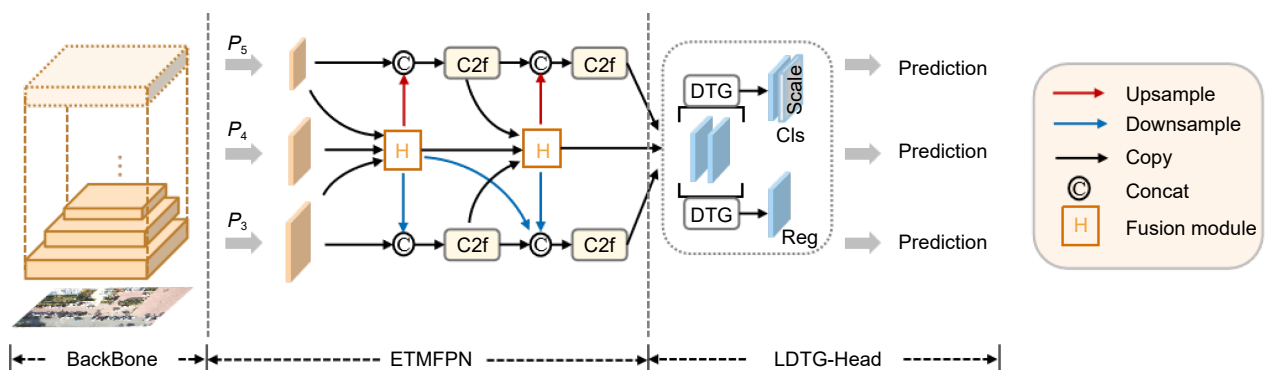


图 1 基于超图计算的高效传递多尺度特征小目标检测算法
Fig. 1 Efficient transfer multi-scale feature small target detection algorithm based on hypergraph computing

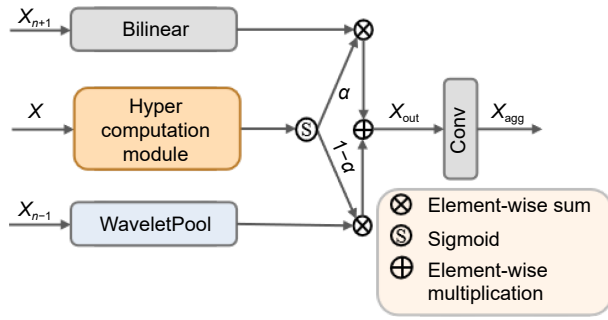


图 2 HOAF 模块

Fig. 2 HOAF module

集合则构成超图 X_d , 如图 4 所示, 其中 d 表示距中心特征点的距离, ε 表示距离阈值。

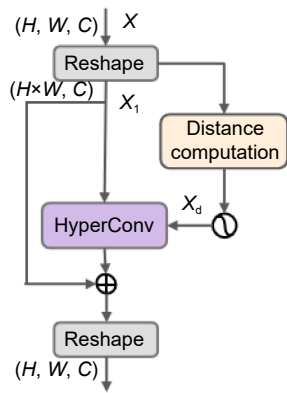


图 3 超图计算模块

Fig. 3 Hyper computation module

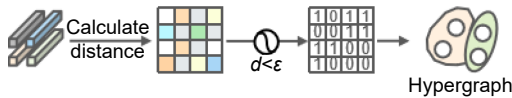


图 4 距离计算部分执行过程

Fig. 4 Distance computation part of the execution process

为了实现超图结构中高阶信息的有效传递, 引入超图卷积^[32]捕捉节点之间的高阶交互关系。将特征 X_1 和超图 X_d 作为超图卷积的输入, 利用均值聚合执行两阶段超图信息传递操作。第一阶段是从节点 v 到边 e 传递相邻节点与其边的关系; 第二阶段则是从边 e 到节点 v , 将超边上的信息重新分配给对应的节点。获得丰富高阶特征后, 为进一步增强原始特征的表达能力, 将高阶特征与原始特征进行残差连接, 对原始特征进行补充, 引入更多的复杂语义信息, 表达式为

$$HC(X, Y) = X + D_v^{-1} Y D_e^{-1} Y^T X \theta, \quad (1)$$

式中: $HC(\cdot)$ 表示超图卷积; X 表示节点特征矩阵; Y 表示超图的关联矩阵; D_v 和 D_e 分别表示顶点和超

边的对角矩阵; θ 为可训练参数。

随后, 将获得的高阶特征输入激活函数得到 α , 根据目标的尺寸自适应地选择合适的特征进行融合, 挖掘了高阶特征与不同分辨率特征之间的非线性关系, 从而促进了特征图内交叉位置的相互作用。如果 $\alpha > 0.5$, 则优先考虑细粒度局部特征, 与低分辨率特征进行矩阵相乘, 反之则强调上下文特征。之后, 将不同分辨率的特征相加, 经过 1×1 卷积进一步融合特征得到 X_{agg} , 构建更全面的特征表示。具体的表达式为

$$\alpha = \text{Sigmoid}(HC(X, X_d)), \quad (2)$$

$$X_{out} = \alpha(B(X_{n+1})) + (1 - \alpha)\text{WavePool}(X_{n-1}), \quad (3)$$

$$X_{agg} = \text{Conv}(X_{out}), \quad (4)$$

式中: $\text{Sigmoid}(\cdot)$ 为激活函数; B 为双线性插值; $\text{WavePool}(\cdot)$ 为小波池化; $\text{Conv}(\cdot)$ 为卷积操作。

最后, 在特征传递阶段, ETMFPN 将融合后的特征直接向不同尺度相邻层直线传递, 相比传统逐层传递特征的方式, 有效缩短了传递的路径, 减少了在传递过程中池化或卷积等操作导致的信息损失。通过 ETMFPN 的设计, 使得颈部每一层都丰富了多尺度特征信息, 有效提升了模型针对多尺度特征的检测效能。

2.2 轻量化动态任务引导检测头

目前的单阶段算法中, 每个检测头采用独立处理不同尺寸特征的方式, 这不仅增加了模型计算量, 而且检测头将任务分为分类和定位两个并行任务分别进行训练, 这还可能导致特征图在空间维度不对齐的问题。为了解决上述问题, 设计检测头 LDTG-Head, 其结构如图 5 所示。LDTG-Head 运用共享权重, 使颈部特征由同一个解耦头进行处理, 在最大程度上降低了复杂度, 也有效解决了检测头中分类与定位之间空间不对准的问题。LDTG-Head 的具体运行过程如下:

首先, 该模块采用多层 3×3 分组卷积作为经过颈部的输出 X_{FPN} 的特征提取器, 以此来强化分类与定位任务之间的信息交互, 从而得到特征 X_{inter} , 交互过程如式 (5) 和式 (6) 所示。分组卷积如图 6 所示, 其原理是依据分组数 (g) 对通道进行划分, 各分组分别执行卷积操作后再进行拼接, 其中 h_1 代表 C_2 个 filters 的高度、 w_1 代表 C_2 个 filters 的宽度。普通卷积与分组卷积的参数量计算公式分别如式 (7) 和式 (8) 所示, 从式中能够清晰地看出, 运用分组卷积后, 参数量仅为普通卷积参数量的 $1/g$, 这显著降低了计算复杂度

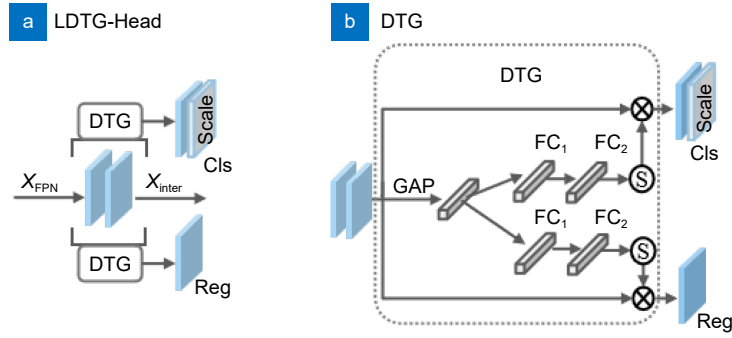


图 5 LDTG-Head 和 DTG 模块结构示意图。(a) LDTG-Head; (b) DTG 模块
Fig. 5 Structure diagram of LDTG-Head and DTG module. (a) LDTG-Head; (b) DTG module

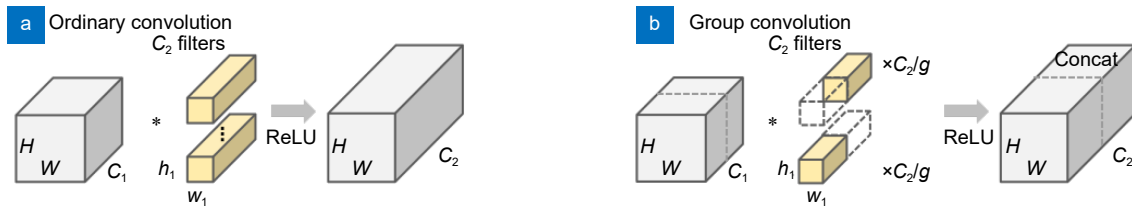


图 6 普通卷积与分组卷积对比。(a) 普通卷积; (b) 分组卷积

Fig. 6 Comparison between ordinary convolution and group convolution. (a) Ordinary convolution; (b) Group convolution

和模型参数量。

$$X_{\text{inter}_1} = \text{Group Conv}_{3 \times 3}(X_{\text{FPN}}), \quad (5)$$

$$X_{\text{inter}} = \text{Group Conv}_{3 \times 3}(X_{\text{inter}_1}), \quad (6)$$

式中: $\text{Group Conv}_{3 \times 3}$ 表示 3×3 的分组卷积; X_{FPN} 表示经过 FPN 生成的语义信息。

$$P_{\text{Conv}} = c_1 \times k \times k \times c_2, \quad (7)$$

$$P_{\text{Group Conv}} = \left(\frac{c_1}{g} \times k \times k \times \frac{c_2}{g} \right) \times g, \quad (8)$$

式中: c_1 表示输入通道数; k 表示卷积核的大小; c_2 表示输出通道数; g 表示分组数。

然后, 考虑到分类和定位任务在关注目标上存在差异, 分别聚焦于不同的层级以及感受野, 并且前期的交互任务不可避免地在这两个不同任务引入了一定程度的特征冲突。所以, 设计了任务分解模块 DTG, 通过通道注意力机制动态计算特定任务的特征, 促进任务分解。具体而言, 通过全局平均池化层获得通道维度的全局特征。接着, 经过由两个卷积的全连接层构成的激励阶段, 对通道注意力进行加权计算。之后, 利用权重对特征图进行动态计算, 生成分类任务和定位任务特定的特征。具体过程的表达式为

$$w = \delta(\text{FC}_2(\text{ReLU}(\text{FC}_1(\text{GAP}(X_{\text{inter}}))))), \quad (9)$$

$$X_{\text{task_cls}} = \text{Scale} \cdot \text{Conv}(w \cdot X_{\text{inter}}), \quad (10)$$

$$X_{\text{task_reg}} = \text{Conv}(w \cdot X_{\text{inter}}), \quad (11)$$

式中: δ 表示 sigmoid 函数; $\text{GAP}(\cdot)$ 表示全局平均池层; FC_1 和 FC_2 表示通道注意力的激励阶段由两个 1×1 卷积的全连接层; Scale 表示对不同尺度特征进行缩放; $X_{\text{task_cls}}$ 表示生成的分类任务特定的特征; $X_{\text{task_reg}}$ 表示生成的定位任务的特定特征。

2.3 模型剪枝

经过 ETMFPN 和 LDTG-Head 的轻量化设计后, 模型在参数量和模型大小方面有所改善, 但仍然存在参数冗余和模型过大的问题。为了在保证模型对小目标检测效果的同时, 减少参数冗余和计算量, 进一步采用基于幅度的层自适应稀疏度的剪枝算法 (layer-adaptive sparsity for the magnitude-based pruning, LAMP) 对改进后的模型进行剪枝, 最大限度降低其对算力资源的消耗。

在深度学习网络中, 网络连接的权重代表了输入信号的重要性。LAMP 方法的核心思想是根据每一层参数的重要性来自适应进行剪枝, 修剪权重较小、对模型贡献不大的节点。与传统的统一全局剪枝不同, LAMP 是根据各层实际情况进行针对性处理, 有效解决了传统全局剪枝导致的层失效问题。

LAMP 具体剪枝过程如下:

首先, 为每层计算权重 $|W|$ 并按照升序排列, 根

据式 (12) 计算 LAMP 分数 (*score*)。

$$score(u, W) = \frac{(W[u])^2}{\sum_{p \geq u} (W[p])^2}, \quad (12)$$

式中: $W(u)$ 、 $W(p)$ 分别表示索引 u 、 p 映射的权重项。由于按照 $u < p$ 的顺序排列, 此时 $W(u) < W(p)$ 。LAMP 分数保证了每一层至少有一个评分为 1 的最优通道, 有效避免了层失效问题。

然后, 预先设置一个修剪比例, LAMP 将 *score* 小于或等于阈值的权重对应的掩码矩阵元素置为 0, 标记为需要剪枝的权重。接着更新权重, 重复以上步骤, 自动确定层级稀疏性, 直到满足预先设定的稀疏性要求。这样的设计能有效平衡通道重要性, 实现自适应的逐层级稀疏。最后, 剪枝后的模型继续进行微调训练, 从而恢复或提升模型性能。

3 实验及结果分析

3.1 实验数据集

本次实验选用 VisDrone2019 数据集, 该数据集内的图像采集于不同场景以及各类天气、光照条件下。其中, 涵盖了 6471 张训练图像、548 张验证图像以及 1610 张测试图像, 共有 10 类非均匀分布的对象。这些图像涵盖了各种角度、高度、光照条件和分辨率, 其中物体尺寸较小, 数量较大, 可以有效验证网络模型用于小目标检测的性能。图 7 展示了该数据集的部分图像。

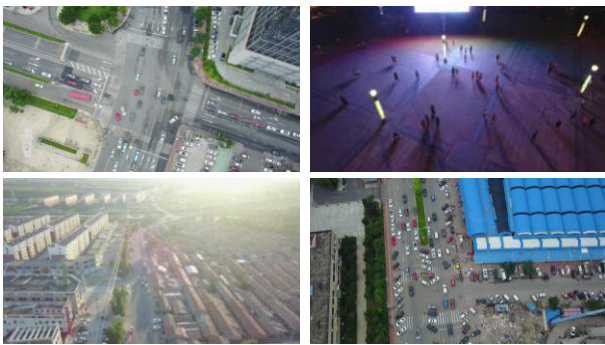


图 7 VisDrone2019 数据集部分图像
Fig. 7 Partial images of the VisDrone2019 dataset

3.2 实验环境及评价指标

在本实验中, CPU 为 12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz, 显卡为 RTX 3080, 显存为 10 GB。输入图像统一调整大小为 640 pixel×640 pixel。在训练设置方面, 选用随机梯度下

降法 (stochastic gradient descent, SGD) 优化器, 动量设定为 0.9, 权重衰减值设为 0.0005, 初始学习率为 0.01。训练时的批大小设定为 4, 对于非最大抑制 (non-maximum suppression, NMS), 其交并比 (intersection over union, IoU) 阈值被设置成 0.7, 迭代次数为 300 次。

为了评估本算法的有效性, 采用在 IoU 阈值为 0.5 时, 所有目标类别平均均值 (mean average precision, mAP) 精度和 IoU 阈值处于 0.50 至 0.95 (步长设定为 0.05) 的 10 个阈值下的检测精度的平均值 $mAP_{0.5:0.95}$ 作为综合衡量模型性能的相关评价指标。采用模型参数量、计算量、模型大小等用来衡量模型的复杂程度, 复杂程度越小代表模型更加轻量化, 更适合边缘设备的部署。同时, 采用平均精度 (average precision, AP) 来评估模型针对单个目标类别的检测性能。AP 和 mAP 计算过程为

$$Precision = \frac{TP}{TP + FP}, \quad (13)$$

$$Recall = \frac{TP}{TP + FN}, \quad (14)$$

$$AP = \int_0^1 Precision(Recall) d(Recall), \quad (15)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N}, \quad (16)$$

式中: TP 表示真阳性, 即实际为正样本且被准确地标记为正样本的情况; FP 表示假阳性, 即原本属于负样本, 却被错误地标识成了正样本; FN 表示正样本被错误地判定为负样本; N 表示类别总数。

3.3 对比实验

为了检验所提模型的有效性, 在 VisDrone 数据集中, 将其与其他先进的检测算法进行对比。在表 1 中, 呈现出了数据集内 10 类目标在不同算法下的 AP 、 $mAP_{0.5}$ 结果, 以及每种算法所的参数量。可以看出, 对比其他算法, HETMNet 在行人、人、汽车、面包车、卡车、三轮车、遮阳棚三轮车、巴士和摩托等尺寸较小和较大的目标上 AP 表现出了较好的检测性能。这表明 HETMNet 在各种场景下都有较好的检测性能, 能在使用较少模型参数量, 在不同类型目标的检测性能上超越其他算法, 尤其是 YOLOv9、YOLOv10、Mamba-YOLO 等目前最新的算法。值得一提的是, HETMNet 的参数量只有 4.8 M, 在 $mAP_{0.5}$ 检测精度高于参数量达 54.2 M 的 YOLOX。进一步验

表 1 不同算法在 VisDrone 数据集上的平均精度和参数量对比结果

Table 1 Comparison results of AP and params of different algorithms on VisDrone dataset

Method	AP/%										$mAP_{0.5} / \%$	Params/M
	Pedestrian	People	Bicycle	Car	Van	Truck	Tricycle	Awning-Tricycle	Bus	Motor		
Faster R-CNN ^[33]	20.9	14.8	7.3	51.0	29.7	19.5	14.0	8.8	30.5	21.2	21.8	—
Cascade R-CNN	22.2	14.8	7.6	54.6	31.5	21.6	14.8	8.6	34.9	21.4	23.2	—
YOLOv5	39.0	31.3	11.2	73.5	35.4	29.5	20.5	11.1	43.1	37.0	33.2	7.0
YOLOX ^[34]	34.8	24.5	16.9	72.4	34.4	40.5	23.1	17.8	53.1	36.0	35.3	54.2
YOLOv7 ^[35]	37.9	34.6	9.4	76.1	36.3	29.8	20.1	10.6	43.2	41.8	34.0	6.0
YOLOv8	40.3	30.1	12.5	77.7	45.8	36.0	27.0	14.9	54.5	42.7	38.2	10.6
YOLOv9 ^[36]	42.4	33.4	14.2	79.5	45.8	39.4	29.5	16.9	57.5	44.5	40.3	6.8
YOLOv10 ^[37]	43.4	34.7	14.5	80.5	46.5	37.3	28.0	15.8	55.2	45.7	40.2	6.9
Mamba-YOLO ^[38]	41.0	31.7	11.5	79.3	44.2	33.2	26.1	13.8	56.0	43.0	38.0	5.7
BDAD-YOLO ^[39]	37.8	30.0	10.5	77.0	42.4	32.5	24.8	13.6	53.0	39.6	36.1	3.2
HETMNet(ours)	45.2	35.8	14.2	81.6	49.7	41.1	30.7	18.0	60.5	51.1	42.4	4.8

证 HETMNet 在处理无人机图像目标检测算法时具有更大的优势, 也更适合边缘设备的部署。

3.4 消融实验结果与分析

为验证多尺度特征高效传递金字塔网络 (ETMFPN)、轻量化动态任务引导检测头 (LDTG-Head)、LAMP 剪枝对 HETMNet 的贡献, 设计了 5 组消融实验, 实验所得结果如表 2 所示。表 2 中模型 1 为原始 YOLOv8 模型, 模型 2 表示使用 ETMFPN 替换模型 1 颈部的 PAFPN 网络, ETMFPN 通过缩短传递路径实现了模型的轻量化, 模型参数量降低 0.9 M、模型大小也减少了 1.9 MB。另外, 在模型性能上也没有下降, 通过 HOAF 模块捕捉高阶特征自适应融合各层特征, 丰富了特征的非线性表达能力, 使当前目标与上下文信息结合, $mAP_{0.5:0.95}$ 指标提升 2.3%。模型 3 表示在模型 1 的基础上将原始检测头替换成 LDTG-Head, 从实验结果可以看出, LDTG-Head 在检测精度上大幅提升, $mAP_{0.5}$ 提升了 2.6%, $mAP_{0.5:0.95}$ 上提升了 2.0%, 同时设计的轻量化共享检测头也使得模型在复杂度上变得更加轻量化, 参数量

减少了 23.6%, 模型大小减少了 23.7%。模型 4 表示在模型 1 的基础上同时使用 ETMFPN 和 LDTG-Head, 从实验结果可以看出模型的参数量减少 1.4 M, 模型大小也减少 2.7 MB。同时, 检测精度也相比单独使用时有了相互促进的作用。模型 5 表示在模型 4 的改进模型上使用 LAMP 剪枝, 在保持原始精度的同时 $mAP_{0.5}$ 提升了 0.4%, $mAP_{0.5:0.95}$ 提升 0.3%, 参数量降低了 47.8%, 计算量降低了 28.7%, 模型大小减少了 8.9 MB。通过减少冗余的通道, 实现了模型的轻量化, 同时也实现了较优的检测精度。

为了验证所提 ETMFPN 具有更好的效果, 在不同的金字塔网络上进行了消融实验, 如表 3 所示。可以看出 ETMFPN 相较于双路径金字塔网络 PAFPN 更适合小目标检测且更加轻量化, 与先进的解决跨层特征丢失问题的金字塔网络 GoldYOLO 和 BiFPN 相比, 其 $mAP_{0.5}$ 与 $mAP_{0.5:0.95}$ 更具有优势。证明了 HOAF 通过获取高阶特征并在不同层之间直接传递的设计, 具有更好的效果。

为了实现模型检测精度与参数量之间的平衡, 选

表 2 HETMNet 在 VisDrone 数据集上的消融实验结果

Table 2 Results of HETMNet's ablation experiment on the VisDrone dataset

Model	ETMFPN	LDTG-Head	LAMP	$mAP_{0.5} / \%$	$mAP_{0.5:0.95} / \%$	Params/M	GFLOPs	Size/MB
Model 1	—	—	—	38.2	22.8	10.6	28.5	21.5
Model 2	√	—	—	41.5	25.1	9.7	29.5	19.6
Model 3	—	√	—	40.8	24.8	8.1	24.3	16.4
Model 4	√	√	—	42.0	25.7	9.2	28.6	18.8
Model 5	√	√	√	42.4	26.0	4.8	20.4	9.9

择了不同的 LAMP 剪枝比例进行实验, 如表 4 所示。剪枝比例等于剪枝前模型计算量与剪枝后模型计算量之比。剪枝比例越大, 模型的参数量越小, 为了在保持原有检测精度的同时实现了模型的轻量化, 选择的剪枝比例为 1.40, 检测精度比未剪枝的模型 $mAP_{0.5}$ 增长 0.4%, 参数量降低 47.8%。

3.5 可视化分析

为了充分验证 HETMNet 在不同场景里的适用性, 从 VisDrone 数据集中选取了部分涵盖夜间环境、高空视野以及密集遮挡等复杂状况的图像进行测试, 检测结果如图 8 所示。对比图 8(a, b) 可知, 在光线昏暗并且小目标存在密集遮挡的场景下, 与基准算法相比,

HETMNet 对小目标展现出更高的检测准确率。对比图 8(c, d) 可知, 在高空条件下, HETMNet 的总体检测精度优于基准算法, 能够更加聚焦目标的中心位置, 有效抑制背景噪声, 同时保留对目标判定更为关键的特征信息, 在复杂背景中展现出更为出色的检测性能。对比图 8(e, f) 可知, HETMNet 在密集遮挡状况下, 可以精准检测出远视距位置的汽车小目标。整体来讲, HETMNet 模型在复杂场景, 尤其是远视距环境中的检测性能得到了显著提升。

4 结 论

针对无人机航拍图像检测存在的小目标检测精度

表 3 ETMFPN 消融实验结果

Table 3 ETMFPN ablation experiment results

Model	$mAP_{0.5}/\%$	$mAP_{0.5:0.95}/\%$	Params/M
+PAFPN	38.2	22.8	10.6
+GoldYOLO	40.5	24.4	13.0
+BiFPN	40.9	24.9	7.0
+ETMFPN	41.5	25.1	9.7

表 4 LAMP 消融实验结果

Table 4 LAMP ablation experiment results

Speed_up	$mAP_{0.5}/\%$	Params/M
—	42.0	9.2
1.25	42.6	5.6
1.35	42.4	5.0
1.40	42.4	4.8
1.45	42.3	4.6

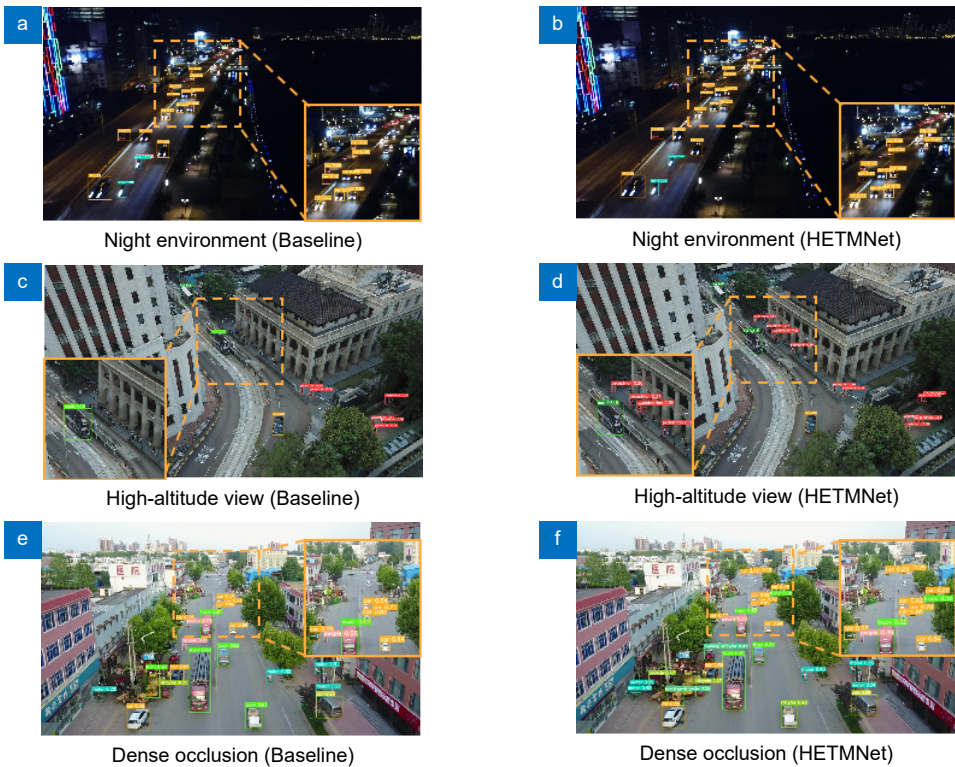


图 8 VisDrone 数据集上的复杂场景目标检测效果对比。(a)(b) 夜间环境; (c)(d) 高空视角; (e)(f) 密集遮挡

Fig. 8 Comparison of target detection effectiveness in complex scenes on VisDrone dataset. (a)(b) Night environment; (c)(d) High-altitude view; (e)(f) dense occlusion scenes

低和模型参数量大的问题, 提出基于超图计算的多尺度特征高效传递小目标检测算法 HETMNet。在颈部设计 ETMFPN 替换原始的 PAFPN, 缩短特征传递路径, 减少信息损失。利用超图构建高阶特征丰富模型的上下文特征。在目标预测方面, 设计 LDTG-Head, 利用共享机制大大减少了参数量实现了轻量化, 同时使用动态任务引导对齐分类和定位任务, 再逐一分解, 提高任务之间的相关性。引入 LAMP 剪枝, 自适应选择对模型贡献小的冗余通道进行修建, 使模型能够在保持原始性能的基础上进一步减少复杂度。在 VisDrone 数据集上进行实验, 实验结果显示, 在模型参数量仅为 4.8 M 时, $mAP_{0.5}$ 达到了 42.4%。相较于基准算法, 该模型参数量降低 54.7%, 而精度提升了 4.2%。验证了 HETMNet 模型的有效性, 能较好地应对无人机图像目标检测任务。下一步将继续研究高效目标检测算法, 保证精准度和轻量化的同时, 进一步提升检测的实时性。

利益冲突: 所有作者声明无利益冲突

参考文献

- [1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 580–587. <https://doi.org/10.1109/CVPR.2014.81>.
- [2] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, 2017: 2980–2988. <https://doi.org/10.1109/ICCV.2017.322>.
- [3] Cai Z W, Vasconcelos N. Cascade R-CNN: delving into high quality object detection[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 6154–6162. <https://doi.org/10.1109/CVPR.2018.00644>.
- [4] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779–788. <https://doi.org/10.1109/CVPR.2016.91>.
- [5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
- [6] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, 2017: 2999–3007. <https://doi.org/10.1109/ICCV.2017.324>.
- [7] Ghiasi G, Lin T Y, Le Q V. NAS-FPN: learning scalable feature pyramid architecture for object detection[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 7029–7038. <https://doi.org/10.1109/CVPR.2019.00720>.
- [8] Jiang Y Q, Tan Z Y, Wang J Y, et al. GiraffeDet: a heavy-neck paradigm for object detection[C]//*Proceedings of the 10th International Conference on Learning Representations*, 2022.
- [9] Ma M, Pang H. SP-YOLOv8s: an improved YOLOv8s model for remote sensing image tiny object detection[J]. *applied sciences*, 2023, 13(14): 8161.
- [10] Wang K X, Liew J H, Zou Y T, et al. PANet: few-shot image semantic segmentation with prototype alignment[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*, 2019: 9196–9205. <https://doi.org/10.1109/ICCV.2019.00929>.
- [11] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 936–944. <https://doi.org/10.1109/CVPR.2017.106>.
- [12] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10778–10787. <https://doi.org/10.1109/CVPR42600.2020.01079>.
- [13] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 8759–8768. <https://doi.org/10.1109/CVPR.2018.00913>.
- [14] Yang G Y, Lei J, Zhu Z K, et al. AFPN: asymptotic feature pyramid network for object detection[C]//*Proceedings of 2023 IEEE International Conference on Systems, Man, and Cybernetics*, 2023: 2184–2189. <https://doi.org/10.1109/SMC53992.2023.10394415>.
- [15] Xue Y J, Ju Z Y, Li Y M, et al. MAF-YOLO: multi-modal attention fusion based YOLO for pedestrian detection[J]. *Infrared Phys Technol*, 2021, 118: 103906.
- [16] Xu X Z, Jiang Y Q, Chen W H, et al. DAMO-YOLO: a report on real-time object detection design[Z]. arXiv: 2211.15444, 2023. <https://doi.org/10.48550/arXiv.2211.15444>.
- [17] Wang C C, He W, Nie Y, et al. Gold-YOLO: efficient object detector via gather-and-distribute mechanism[C]//*Proceedings of the 37th International Conference on Neural Information Processing Systems*, 2023: 2224.
- [18] Gao Y, Zhang Z Z, Lin H J, et al. Hypergraph learning: methods and practices[J]. *IEEE Trans Pattern Anal Mach Intell*, 2020, 44(5): 2548–2566.
- [19] Gao Y, Feng Y F, Ji S Y, et al. HGNN*: general hypergraph neural networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2023, 45(3): 3181–3199.
- [20] Liu Y Y, Yu Z Y, Zong D L, et al. Attention to task-aligned object detection for end-edge-cloud video surveillance[J]. *IEEE Internet Things J*, 2024, 11(8): 13781–13792.
- [21] Shen Q, Zhang L, Zhang Y X, et al. Distracted driving behavior detection algorithm based on lightweight StarDL-YOLO[J]. *Electronics*, 2024, 13(16): 3216.
- [22] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5MB model size[Z]. arXiv: 1602.07360, 2016. <https://doi.org/10.48550/arXiv.1602.07360>.
- [23] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 6848–6856. <https://doi.org/10.1109/CVPR.2018.00716>.
- [24] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[Z].

- arXiv: 1704.04861, 2017.
<https://doi.org/10.48550/arXiv.1704.04861>.
- [25] Tan M X, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks[C]//*Proceedings of the 36th International Conference on Machine Learning*, 2019: 6105–6114.
- [26] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 1577–1586. <https://doi.org/10.1109/CVPR42600.2020.00165>.
- [27] Wang S M, Xu H Y, Zhu X Z, et al. Lightweight small object detection algorithm based on improved YOLOv8n aerial photography: PECS-YOLO[J]. *Comput Eng*, 2024.
王舒梦, 徐慧英, 朱信忠, 等. 基于改进 YOLOv8n 航拍轻量化小目标检测算法: PECS-YOLO[J]. *计算机工程*, 2024.
- [28] Zhang J C, Wei J, Chen Y S. Improved YOLOv8 real-time lightweight robust hedge detection algorithm[J]. *Comput Eng*, 2024.
张佳承, 韦锦, 陈义时. 改进 YOLOv8 的实时轻量化鲁棒绿篱检测算法[J]. *计算机工程*, 2024.
- [29] Gale T, Elsen E, Hooker S. The state of sparsity in deep neural networks[Z]. arxiv: 1902.09574, 2019.
<https://doi.org/10.48550/arXiv.1902.09574>.
- [30] Evci U, Gale T, Menick J, et al. Rigging the lottery: making all tickets winners[C]//*Proceedings of the 37th International Conference on Machine Learning*, 2020: 276.
- [31] Lee J, Park S, Mo S, et al. Layer-adaptive sparsity for the magnitude-based pruning[C]//*Proceedings of the 9th International Conference on Learning Representations*, 2021.
- [32] Feng Y F, Huang J G, Du S Y, et al. Hyper-YOLO: when visual object detection meets hypergraph computation[J]. *IEEE Trans Pattern Anal Mach Intell*, 2025, 47(4): 2388–2401.
- [33] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(6): 1137–1149.
- [34] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[Z]. arXiv: 2107.08430, 2021.
<https://doi.org/10.48550/arXiv.2107.08430>.
- [35] 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 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 7464–7475.
<https://doi.org/10.1109/CVPR52729.2023.00721>.
- [36] Wang C Y, Yeh I H, Liao H Y M. YOLOv9: learning what you want to learn using programmable gradient information[C]//*Proceedings of the 18th European Conference on Computer Vision*, 2025: 1–21.
https://doi.org/10.1007/978-3-031-72751-1_1.
- [37] Wang A, Chen H, Liu L H, et al. YOLOv10: real-time end-to-end object detection[C]//*Proceedings of the 38th Conference on Neural Information Processing Systems*, 2024.
- [38] Wang Z Y, Li C, Xu H Y, et al. Mamba YOLO: SSMs-based YOLO For object detection[Z]. arXiv: 2406.05835v1, 2024.
<https://doi.org/10.48550/arXiv.2406.05835>.
- [39] Sun J Y, Xu M J, Zhang J P, et al. Optimized and improved YOLOv8 target detection algorithm from UAV perspective[J]. *Comput Eng Appl*, 2025, 61(1): 109–120.
孙佳宇, 徐民俊, 张俊鹏, 等. 优化改进 YOLOv8 无人机视角下目标检测算法[J]. *计算机工程与应用*, 2025, 61(1): 109–120.

作者简介



孙叶美 (1990-), 女, 山东临沂人, 硕士, 实验师, 研究方向为基于深度学习的小目标检测、语义分割、图像超分辨率重建等以及人工智能其他延伸应用方向。

E-mail: sunyemei1216@163.com



桑学婷 (2000-), 女, 硕士研究生, 主要研究方向为小目标检测、机器视觉。

E-mail: xt_sang@126.com



张艳 (1982-), 女, 副教授, 博士, 主要研究方向为图像处理、机器视觉。

E-mail: yanzhang0910@163.com



【通信作者】刘国瑞 (1981-), 男, 河北邢台人, 博士, 副教授, 主要从事数据库技术、数据挖掘、机器学习、并行计算、智能算法、微电子技术等方面的研究。

E-mail: lgr@tcu.edu.cn



【通信作者】陈帅宇 (2004-), 男, 本科, 主要从事图像处理方面的研究。

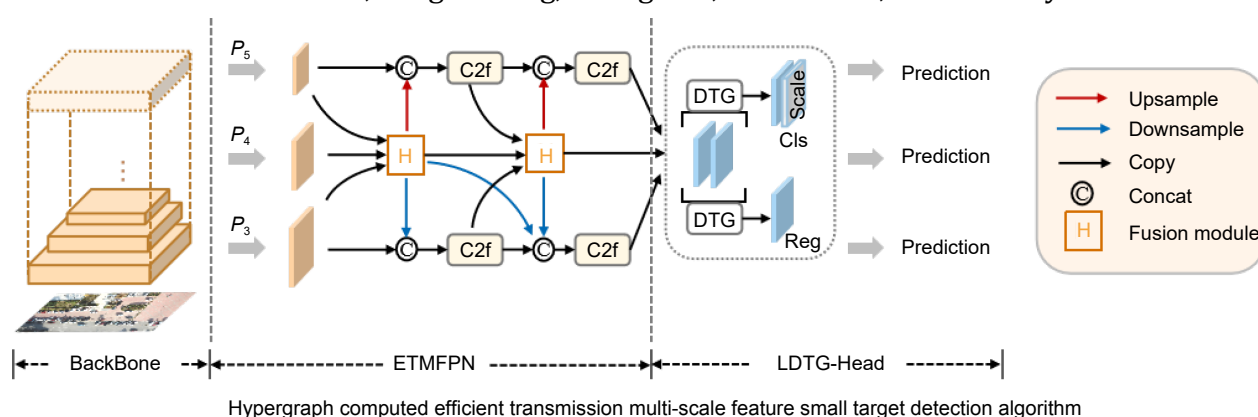
E-mail: chenshuaiyu0906@163.com



扫描二维码, 获取PDF全文

Hypergraph computed efficient transmission multi-scale feature small target detection algorithm

Sun Yemei, Sang Xueting, Zhang Yan, Liu Guorui*, Chen Shuaiyu*



Overview: Aiming at the characteristics of UAV aerial images such as complex background, small target size and dense distribution due to high-angle shooting, as well as the common problems of insufficient accuracy and parameter redundancy in existing detection models, this paper proposes an efficient multi-scale feature transfer small target detection algorithm based on hypergraph computation. By systematically improving network architecture, feature fusion mechanism and model compression strategy, the algorithm achieves an effective balance between detection performance and computational efficiency. In terms of network architecture design, this study innovatively constructs a multi-scale feature pyramid network as a neck structure. Different from the traditional feature pyramid layer-by-layer transmission, this network transmits the features of the middle layer directly to the adjacent layers through the cross-layer feature aggregation mechanism, which significantly shortens the feature transmission path. Specifically, by integrating shallow high-resolution features and deep semantic features, the spatial information loss caused by long-distance transmission is effectively alleviated, so that the location information and texture features of small targets can be completely preserved. In the feature fusion stage, hypergraph is introduced to break through the limitation of binary relation of traditional graph neural networks. By connecting multiple feature nodes with hyperedge and establishing a high-order feature interaction model, the nonlinear correlation between the object and the complex background in UAV images can be accurately described. This hypergraph structure can not only capture the geometric correlation between objects but also model the potential relationship between the interference factors such as illumination change and occlusion and the object features. Secondly, a lightweight dynamic task-guided detection head is designed to effectively solve the problem of inaccurate detection targets caused by inconsistent classification and positioning task space in the traditional decoupling head with a small number of parameters by sharing mechanism. Finally, a layer adaptive pruning amplitude strategy is used to break through the limitation of the traditional global pruning threshold. By analyzing the weight distribution characteristics of each convolution layer, the calculation model of the pruning coefficient based on layer sensitivity is established. Experimental results show that the proposed algorithm performs better than other architectures on VisDrone2019 dataset, with an accuracy of 42.4% and many parameters of 4.8 M. Compared to the benchmark YOLOv8, the number of parameters has been reduced by 54.7%. This model achieves a good balance between detection performance and resource consumption.

Sun Y M, Sang X T, Zhang Y, et al. Hypergraph computed efficient transmission multi-scale feature small target detection algorithm[J]. *Opto-Electron Eng*, 2025, 52(5): 250061; DOI: [10.12086/oee.2025.250061](https://doi.org/10.12086/oee.2025.250061)

Foundation item: National College Students Innovation and Entrepreneurship Training Program (202410792012), Tianjin Philosophy and Social Science Planning Project (TJGL19XSX-045)

School of Computer and Information Engineering, Tianjin Chengjian University, Tianjin 300380, China

* E-mail: lgr@tcu.edu.cn; chenshuaiyu0906@163.com