



基于动态超图与多尺度特征融合的遥感图像目标检测

赵文清, 李溢晔

引用本文:

赵文清, 李溢晔. 基于动态超图与多尺度特征融合的遥感图像目标检测[J]. *智能系统学报*, 2026, 21(2): 399-409.
ZHAO Wenqing, LI Yiye. Remote sensing image object detection based on dynamic hypergraphs and multi-scale feature fusion[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(2): 399-409.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202508009>

您可能感兴趣的其他文章

融合视觉显著性再检测的孪生网络无人机目标跟踪算法

Siamese network combined with visual saliency re-detection for UAV object tracking
智能系统学报. 2021, 16(3): 584-594 <https://dx.doi.org/10.11992/tis.202101035>

结合模糊特征检测的鲁棒核相关滤波跟踪法

Robust KCF tracking algorithm combined with fuzzy feature detection
智能系统学报. 2021, 16(2): 323-329 <https://dx.doi.org/10.11992/tis.201912010>

多特征融合的异视角目标关联算法

Target association from different perspectives based on multi-feature fusion
智能系统学报. 2020, 15(5): 847-855 <https://dx.doi.org/10.11992/tis.202006037>

快速的圆投影图像匹配算法

Fast image matching algorithm based on circular projection
智能系统学报. 2020, 15(1): 84-91 <https://dx.doi.org/10.11992/tis.201903037>

多层卷积特征的真实场景下行人检测研究

Research on pedestrian detection based on multi-layer convolution feature in real scene
智能系统学报. 2019, 14(2): 306-315 <https://dx.doi.org/10.11992/tis.201710019>

基于Object Proposals并集的显著性检测模型

Saliency detection model based on the union of Object Proposals
智能系统学报. 2018, 13(6): 946-951 <https://dx.doi.org/10.11992/tis.201801009>

DOI: 10.11992/tis.202508009

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20251224.1701.009>

基于动态超图与多尺度特征融合的遥感图像目标检测

赵文清^{1,2}, 李溢晔¹

(1. 华北电力大学 控制与计算机工程学院, 河北 保定 071003; 2. 河北省能源电力知识计算重点实验室, 河北 保定 071003)

摘要: 遥感图像目标尺度差异大、背景复杂, 而目标检测模型存在多尺度感知能力不足、全局语义特征建模能力差的问题。本文提出了一种基于动态超图与多尺度特征融合的遥感图像目标检测模型。构造多尺度空洞卷积特征融合模块, 并设计对应的特征提取网络, 充分提取多尺度特征; 构造动态门控超图模块, 以此构建全局语义特征建模网络, 强化对目标特征区域的感知, 弱化复杂背景的干扰; 提出多通道坐标注意力模块, 结合坐标注意力机制与多尺度通道交互, 增强特征的表达。在 DIOR 和 RSOD 两个数据集上设计了多组消融实验, 结果表明, 本模型在 DIOR 数据集与 RSOD 数据集上的平均精度均值比 YOLO11 模型分别提升 2.5 和 2.3 个百分点, 显著提升了遥感图像目标检测的精度。为验证本模型的有效性, 设计对比实验与不同方法比较, 实验结果证明本模型的检测效果优于其他方法。

关键词: 遥感图像; 目标检测; 多尺度; 特征融合; 动态超图; 语义特征; 坐标注意力; 空洞卷积

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2026)02-0399-11

中文引用格式: 赵文清, 李溢晔. 基于动态超图与多尺度特征融合的遥感图像目标检测 [J]. 智能系统学报, 2026, 21(2): 399-409.

英文引用格式: ZHAO Wenqing, LI Yiye. Remote sensing image object detection based on dynamic hypergraphs and multi-scale feature fusion [J]. CAAI transactions on intelligent systems, 2026, 21(2): 399-409.

Remote sensing image object detection based on dynamic hypergraphs and multi-scale feature fusion

ZHAO Wenqing^{1,2}, LI Yiye¹

(1. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China; 2. Hebei Key Laboratory of Knowledge Computing for Energy & Power, Baoding 071003, China)

Abstract: Remote sensing images exhibit significant variations in target scales and complex backgrounds, while existing object detection models suffer from limited multi-scale perception and insufficient global semantic modeling capabilities. To address these challenges, a remote sensing object detection framework based on dynamic hypergraph and multi-scale feature fusion was proposed. First, a multi-scale dilated convolution feature fusion module was constructed, and a feature extraction network was designed to fully extract multi-scale features. Second, a dynamic gated hypergraph module was developed to establish a global semantic feature modeling network, which enhanced target feature perception while weakening complex background interference. Finally, a multi-channel coordinate attention module was presented, combining coordinate attention mechanisms with multi-scale channel interactions to strengthen feature representation. Ablation experiments are conducted on the DIOR and the RSOD datasets, demonstrating that the proposed model achieves 2.5 and 2.3 percent age point improvements in mean average precision over the YOLO11 baseline. Comparative experiments validate the superiority of the proposed model, showing enhanced detection performance against other methods.

Keywords: remote sensing image; object detection; multi-scale; feature fusion; dynamic hypergraph; semantic features; coordinate attention; dilated convolution

作为人工智能与遥感技术交叉渗透的重要研究领域, 遥感图像目标检测在城市规划、灾害监

测和军用侦察等关键场景中展现出广阔的应用潜力。然而, 遥感图像具有目标尺度差异大、背景复杂和小目标特征稀疏等特点^[1], 使得传统的目标检测算法面临显著挑战。

随着机器学习相关技术的发展, 一系列基于

收稿日期: 2025-08-08. 网络出版日期: 2025-12-25.

基金项目: 国家自然科学基金项目 (62371188).

通信作者: 赵文清. E-mail: zhaowenqing@ncepu.edu.cn.

深度学习的目标检测方法相继涌现。依据算法是否预先生成目标候选区域这一核心特征,目标检测算法可划分为单阶段目标检测算法与双阶段目标检测算法两大类。

YOLO(you only look once)^[2]系列是单阶段目标检测算法的代表。该系列算法能够基于输入图像的像素信息直接预测出目标边界框的坐标参数与对应类别概率。SSD(single shot multi-box detector)^[3]也是单阶段目标检测算法的典型代表。该算法通过多尺度特征图上开展边界框预测任务,具备对不同尺寸目标的检测能力。吴珺等^[4]在YOLOv5s目标检测网络基础上,提出了轻量网络模型,实现轻量化部署。唐嘉潞等^[5]基于CenterNet^[6]提出加入注意力机制和特征融合的瑕疵检测算法,提高了小目标检测精度。梁礼明等^[7]提出一种基于MobileViT(mobile vision Transformer)和多尺度特征聚合的遥感图像目标检测算法,提升了对遥感目标的检测性能。周莲等^[8]提出了一种小样本目标检测算法,提升了遥感图像小样本目标检测的精度。

R-CNN(region-based convolutional neural networks)^[9]系列是双阶段目标检测算法的主要代表。这类算法先生成候选区域,再对这些区域进行分类和定位。其检测精度较高,但训练繁琐且检测速度慢。Fast R-CNN^[10]将特征提取、分类和回归整合到一个网络,简化了训练过程并提升了效率。Faster R-CNN^[11]引入区域提议网络(region proposal network, RPN)实现端到端训练,适用于高精度场景。

随着Transformer的成功,一些基于Transformer的目标检测算法陆续被提出。DETR(detection Transformer)^[12]采用CNN提取特征,经编码器-解码器结构处理,预测目标类别和边界框位置。Huang等^[13]提出一个名为DEIM(DETR with improved matching)的高效训练框架,加速基于DETR的实时目标检测的收敛。Meng等^[14]提出的Conditional DETR通过改进解码器加速训练收敛。Ye等^[15]提出一个密度感知的动态查询DETR,实现了高效和准确的微小物体检测。

由于遥感图像自身的特点,传统的目标检测算法直接应用于遥感图像目标检测效果不尽人意。因此,一些针对于遥感图像的目标检测算法陆续被提出来。赵文清等^[16]提出了一种改进YOLOv5s的遥感图像目标检测算法,引入Swin Transformer网络结构与坐标注意力机制,提高了遥感图像目标检测的精度。文斌等^[17]提出了基

于FNM-Net的轻量级遥感目标检测网络,提高了模型的检测速度。Zhao等^[18]提出了一种自适应注意力特征融合的多尺度目标检测算法,提升了遥感图像目标检测的精度。单慧琳等^[19]提出了一种增强小目标特征的多尺度神经网络,改善了光学遥感图像小目标检测性能。Yue等^[20]提出了一种语义校正与聚焦网络,提高了对遥感目标的检测精度。Ren等^[21]提出了一个交叉阶段部分强化匹配检测器,改善了遥感图像目标检测效果。赵文清等^[22]提出了一种结合倒残差自注意力机制的目标检测方法,构造倒残差自注意力机制骨干网络,提升了遥感图像目标检测精度。虽然上述方法在遥感场景中取得了一定的进展,但是模型多尺度感知能力不足,全局语义建模薄弱的问题仍未被有效解决,在一定程度上限制了遥感图像目标检测的性能。

针对上述问题,并结合遥感图像目标尺度差异大,背景复杂等特点,本文提出一种基于动态超图与多尺度特征融合的遥感图像目标检测模型(dynamic hypergraphs and multi-scale feature fusion network, DHMF-Net)。本文的主要工作如下:

1)提出了一个多尺度空洞卷积特征融合模块(multi-scale dilated convolution feature fusion block, MSDCFB),利用不同大小的感受野,提取多尺度的目标特征。

2)提出了一个动态门控超图模块(dynamic gated hypergraph module, DGHGM),通过动态超图建立全局语义特征模型,抑制复杂背景的干扰。

3)提出了一个多通道坐标注意力模块(multi-channel and coordinate attention module, MCAM)结合坐标注意力机制与多尺度通道交互,增强特征的表达。

1 相关技术

1.1 YOLO11 网络

本文提出的DHMF-Net模型是基于YOLO11网络改进的。YOLO11的骨干网络以CSPDarknet(cross stage partial Darknet)为基础,主要用于提取图像的特征信息。相较于YOLOv8的骨干网络,YOLO11采用C3k2模块替换了C2f模块,同时在骨干网络的尾部添加了一个C2PSA(cross stage partial pyramid slice attention)模块。YOLO11的特征融合网络采用PANet(path aggregation network)结构分别从自顶而下和自底向上两个方向聚合不同尺度的特征信息。YOLO11检测头的分类损失分支使用两个深度可分离卷积(depthwise separable convolution, DSC)替代了标准卷积,减少了冗

余计算, 提高了效率。

与 YOLO 系列的其他模型相比, YOLO11 的性能更好, 参数量与计算量更低, 在计算资源受限的环境中更具效率。因此, 本文选择 YOLO11 作为基线模型, 并在此基础上提出了 DHMF-Net 模型。

1.2 超图神经网络

由于传统的图卷积神经网络仅限于处理数据间的成对连接关系, 难以对复杂的数据关系建模, 超图神经网络(hypergraph neural networks, HGNN)^[23]应运而生。HGNN 利用超图结构对复杂数据关系进行建模。超图中的超边可连接两个或更多节点, 能表示复杂数据之间的关系, 是一种更加灵活的数据建模工具。

HGNN 通过超边卷积操作, 在表征学习中处理数据之间的相关性。其将数据的潜在关系以超图结构来表示, 并通过超边卷积操作对高阶数据相关性进行表征学习。HGNN 核心思想是通过“节点-超边-节点”的两阶段消息传递机制, 实现节点特征的更新。通过超图卷积操作, HGNN 能够学习数据中的高阶相关性, 实现信息在超图结构上的高效传播和聚合, 使得节点能够综合考虑多种复杂关系, 获得更全面的特征信息。

本文提出的 DGHGM 模块借鉴了 HGNN 的思想, 通过动态超图构建全局语义特征模型, 抑制复杂背景的干扰。

1.3 空洞卷积

空洞卷积(dilated convolution)^[24]是一种在卷积核中插入空洞(零点)的卷积操作, 主要用于扩

大卷积的感受野。具体而言, 空洞卷积通过在卷积核中插入多个零点, 使得卷积操作能够跳跃式地覆盖更大的区域, 从而扩大感受野。这使得网络能够在不增加参数量的情况下捕获更长距离的依赖关系。通过使用不同空洞率(dilation rate)的空洞卷积, 可以在不同尺度上提取特征, 实现多尺度特征融合。这种特性使得空洞卷积可同时捕获局部细节和全局上下文信息, 对于目标检测任务尤为重要。

本文提出的 MSDCFB 模块利用不同空洞率的空洞卷积, 提升模型对多尺度目标的特征提取能力。

2 网络结构与方法

本文提出的 DHMF-Net 模型结构如图 1 所示。首先, 构建了一个基于多尺度空洞卷积的特征提取网络。利用 MSDCFB 模块将不同空洞率的空洞卷积与多分支结构创新性结合, 扩大卷积操作的感受野, 提升了网络对多尺度特征的提取能力。同时, 本文借鉴了 Hyper-YOLO^[25]模型中基于超图的跨层次跨位置表示网络(hypergraph-based cross-level and cross-position representation network)的思想, 并结合本文提出的 DGHGM 模块, 构建出一个全局语义特征建模网络, 提升了 DHMF-Net 模型抗复杂背景干扰的能力。此外, 本文构建了一个基于多通道坐标注意力的特征融合网络。利用 MCAM 模块的坐标注意力增强目标的特征信息, 利用多个深度可分离卷积组捕获丰富的上下文信息, 增强目标的特征。

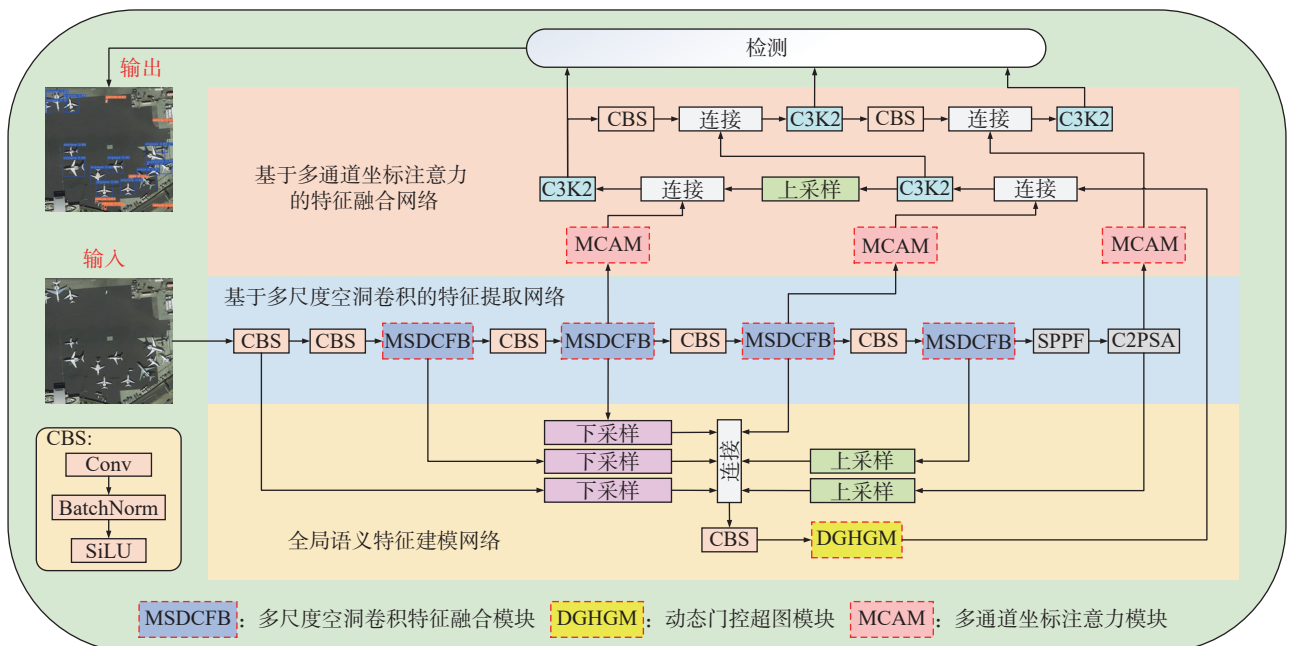


图 1 基于动态超图与多尺度特征融合的遥感图像目标检测模型

Fig. 1 DHMF-Net model

2.1 基于多尺度空洞卷积的特征提取网络

在基于多尺度空洞卷积的特征提取网络中,输入图像先后经过多个卷积层和 MSDCFB 模块,用于提取多尺度目标特征。然后再通过 SPPF (spatial pyramid pooling fast) 模块进行多尺度特征融合,最后通过 C2PSA 模块进行特征强化。

其中,MSDCFB 模块的结构如图 2 所示。本模块采用了多分支特征融合设计,以增强模型的多尺度特征感知能力;使用了一个轻量级通道注意力机制——SE_Block(squeeze-and-excitation Block)^[26],在仅增加少量参数的情况下,强化模型对关键特征的聚焦能力;使用多个 Bottleneck 模块堆叠,逐步提取深层次的语义特征,提升模型对复杂目标的识别能力,同时缓解梯度消失问题,确保深层网络的稳定;利用通道扩展与融合机制,综合不同层次和不同尺度的特征,提升了模型对目标类别的识别能力。

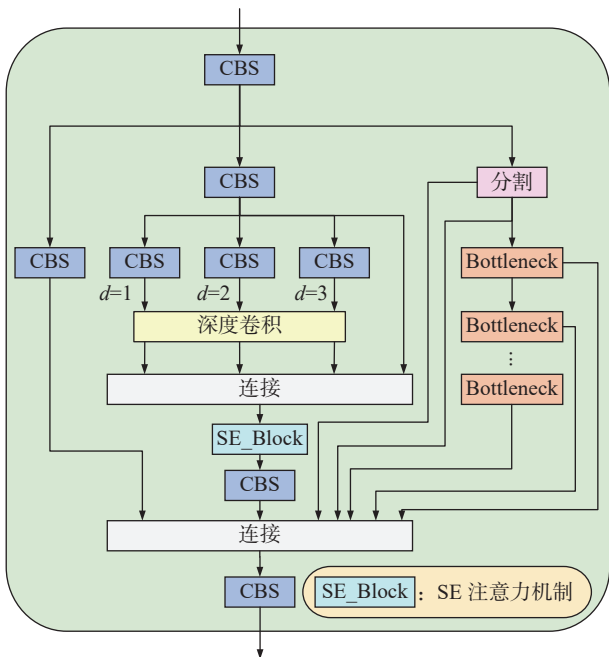


图 2 多尺度空洞卷积特征融合模块
Fig.2 MSDCFB module

与 YOLO11 中传统的 C3k2 模块相比,MSDCFB 模块结合了多分支特征融合与多尺度空洞卷积,扩大了模型的感受野,对不同尺度目标的特征提取能力更强。

在 MSDCFB 模块中,输入的特征图首先经过一个卷积层来扩展通道维度,为后续多分支处理提供更丰富的信息。然后,特征图经过 3 个分支的处理后再进行通道拼接,最终利用 1×1 卷积融合多分支特征,保留关键信息,减少冗余计算。其中,3 个处理特征的分支分别是特征降维分支、空洞卷积分支和特征分割处理分支。特征降维分

支直接通过 1×1 卷积对特征图降维,以保留原始特征的浅层信息,确保模型对简单目标的快速响应。空洞卷积分支利用 3 个空洞率分别为 1、2、3 的深度卷积(depth-wise convolution, DWConv)构成空洞卷积,在不增加参数量的情况下扩展感受野,捕捉不同尺度的上下文信息。然后,通过通道拼接与 SE_Block 注意力机制,动态调整各通道的重要性权重,抑制噪声信息的干扰,增强模型对关键特征(如目标轮廓、纹理)的响应。特征分割处理分支将初始特征进行拆分,并利用 Bottleneck 模块堆叠来挖掘深层特征。每个 Bottleneck 模块处理前一阶段的输出,并逐步细化多尺度特征,形成层次化的特征表示,提升目标检测的精度和泛化能力。此外,Bottleneck 模块通过使用残差连接来缓解模型在训练中的梯度消失问题,加强深层网络结构的稳定性。总之,MSDCFB 模块通过结合多分支特征融合、轻量级注意力机制、Bottleneck 模块堆叠和通道扩展融合机制,增强了特征提取网络对多尺度目标的感知能力。经过特征提取网络对输入图像的多尺度特征提取,为后续全局语义特征建模网络和基于多通道坐标注意力的特征融合网络提供了丰富的特征信息。

2.2 全局语义特征建模网络

在全局语义特征建模网络中,首先,将特征提取网络中 6 个不同尺度的特征图进行通道维度的拼接,然后再利用 1×1 的卷积层进行通道压缩。最后,利用 DGHGM 模块对压缩后的特征图进行语义特征建模,其结构如图 3 所示。

DGHGM 模块利用余弦相似度(cosine_similarity)来计算原始特征与投影特征的相似度,并结合可学习阈值动态生成超边掩码,隐式构建超图,以此筛选出高相关性的特征区域;利用双向门控机制提取局部特征,同时动态控制节点(特征点)与超边(高阶关系)之间的信息流动;利用深度卷积聚合局部特征,模拟超边内的特征交互,捕捉深层的语义特征,提升模型的语义建模能力。

对比传统的 HGNN 方法,DGHGM 模块利用可学习阈值参数与双向门控机制实现了超边的动态生成以及超边与节点之间的动态交互。这可以更加高效地筛选出特征图中的高阶语义特征,增强语义建模能力。

在 DGHGM 模块中,为降低计算的复杂度,特征图先经过 1×1 的卷积层来压缩通道数目,得到特征图 \mathbf{x} 。然后,构建一个节点到超边的门控

网络。使用 DWConv 和 Sigmoid 函数生成一个空间注意力权重 v_{2e} , 并与原始特征图 x 逐元素相乘, 得到局部门控特征 $x_{v_{2e}}$ 。此过程模拟超图中节点特征向超边聚合的过程, 主要作用是筛选出重要特征用于动态构建超边。然后, 再利用 DWConv 和 SiLU 函数实现对 $x_{v_{2e}}$ 的消息聚合。这一过程模拟了超边内的局部特征交互, 生成初始超边特征 E 。

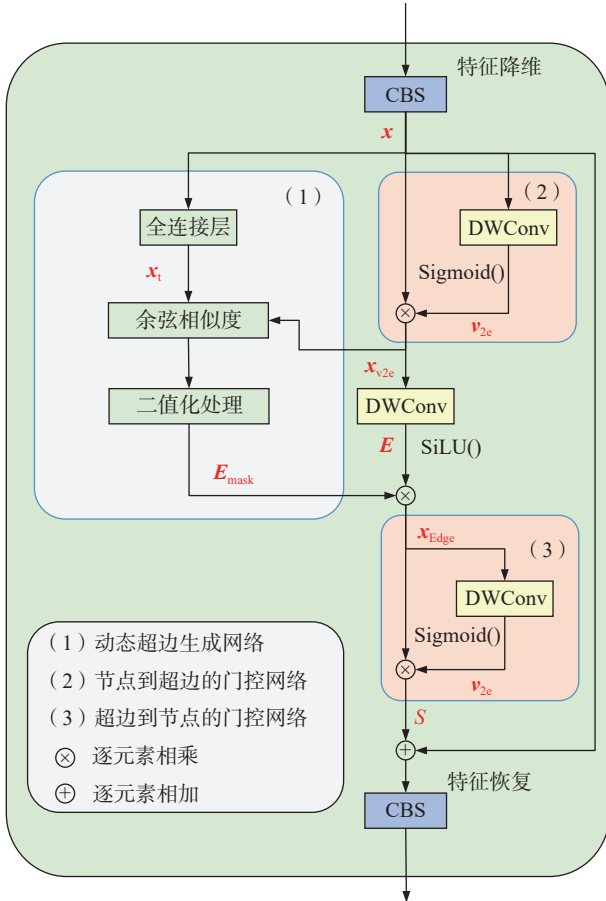


图 3 动态门控超图模块
Fig. 3 DGGM module

同时, 本文基于余弦相似度构建出动态超边生成网络。特征图 x 经过全连接层将原始特征投影到相似性空间, 并生成对应的投影特征 x_t 。然后, 利用余弦相似度来计算投影特征 x_t 与局部门控特征 $x_{v_{2e}}$ 之间的相似度。之后, 将相似度矩阵与预先设置好的可学习阈值进行二值化处理, 大于阈值的部分构成超边掩码 E_{mask} 。这一过程将超图结构隐式编码在特征相似度计算中, 通过超边掩码机制实现超边的按需生成。这种设计既提升了计算效率, 又保留了超图的拓扑特性。最后, 将初始超边特征 E 与超边掩码相乘得到超边特征 x_{Edge} 。这一过程将特征图中潜在的高阶语义特征聚合, 筛选出高相关性的特征区域, 增强了模型对上下文语义信息的建模能力。

接下来, 将超边特征 x_{Edge} 应用于超边到节点的门控网络中。再次使用 DWConv 和 Sigmoid 函数生成空间注意力权重 v_{2e} , 并与超边特征 x_{Edge} 逐元素相乘, 从而将超边捕获的高阶语义特征发散到各个节点中, 得到高阶特征 S 。最后, 将 S 与原始特征 x 相加以保留原始特征, 并使用卷积操作恢复通道数目, 为后续目标检测任务提供丰富的语义特征。

上述计算过程的公式为

$$x_{v_{2e}} = x \otimes v_{2e} \quad (1)$$

$$E = \text{SiLU}(\text{DWConv}(x_{v_{2e}})) \quad (2)$$

$$E_{mask} = \text{Bina}(\text{cosine_similarity}(x_t, x_{v_{2e}})) \quad (3)$$

$$x_{Edge} = E \otimes E_{mask} \quad (4)$$

$$S = x_{Edge} \otimes v_{2e} \quad (5)$$

式中: \otimes 表示逐元素相乘, DWConv() 表示深度卷积操作, SiLU() 表示使用 SiLU 作为损失函数, cosine_similarity() 表示使用余弦相似度比较, Bina() 表示使用阈值进行二值化处理。

这一过程充分利用了不同尺度的特征信息, 增强了模型对全局语义特征的建模能力。同时, 通过动态超边的生成与更新, 模型可以不断强化对目标特征区域的感知。

经过全局语义特征建模网络对多个层次特征的处理, 为后续基于多通道坐标注意力的特征融合网络提供了更加全面的语义特征。

2.3 基于多通道坐标注意力的特征融合网络

为了进一步增强模型对小目标的特征表达能力, 本文构建了一个基于多通道坐标注意力的特征融合网络。在整个特征融合网络部分, 分别将特征提取网络中 3 个不同尺度的特征图和经全局语义特征建模网络处理后的特征图作为输入, 经过多个 MCAM 模块与 C3k2 模块的处理, 实现特征增强与融合。其中, MCAM 模块的结构如图 4 所示。

本模块利用多尺度通道交互机制, 使用不同尺度的深度可分离卷积来增强多尺度目标感知能力; 利用坐标注意力机制, 生成归一化的坐标网格, 并将空间位置信息与通道信息融合, 强化了目标特征的表达; 同时, 本模块利用可学习权重参数, 对多尺度通道交互的输出进行加权融合, 并根据输入特征动态调整不同尺度分支的权重。

与 CBAM^[27] 注意力相比, MCAM 模块结合了多通道交互、可学习权重和坐标注意力机制, 让模型更聚焦关键区域, 对目标的感知与识别能力更强。

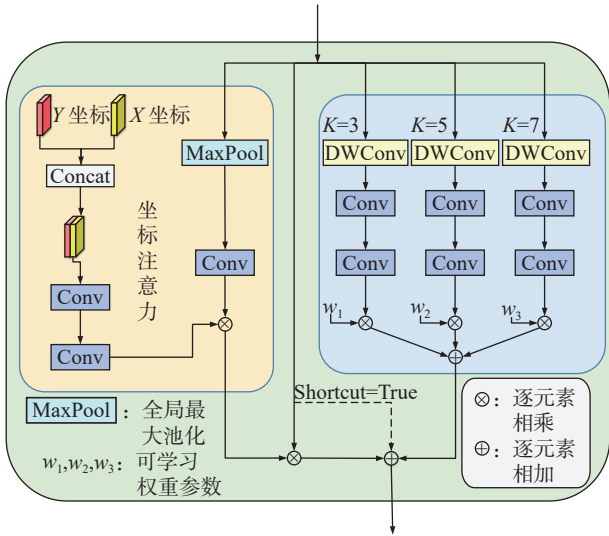


图 4 多通道坐标注意力模块
Fig. 4 MCAM module

当输入信息进入 MCAM 模块后, 会被 3 个不同的深度可分离卷积分支处理。其中, 3×3 的深度可分离卷积用于提取小尺度的特征, 5×5 的深度可分离卷积提取中等尺度的特征, 7×7 的深度可分离卷积提取大尺度的特征。这一过程可以提取到多尺度的上下文信息。深度可分离卷积的使用既保持了各通道特征的独立性, 又降低了模型的计算量。此外, 本模块预先设置了可学习的权重参数, 与多分支处理后的特征进行动态加权并融合。可学习的权重参数可根据任务需求自适应调整各分支特征的权重, 避免固定比例导致的信息丢失。

同时, 本模块利用了坐标注意力机制。首先, 根据特征图的空间尺寸创建归一化的 X 坐标和 Y 坐标网格, 并利用显式位置编码将空间坐标信息(目标的相对位置)融入注意力机制, 以增强对小目标的特征表达。然后, 使用 1×1 的卷积对坐标信息进行编码, 生成与输入特征同尺寸的注意力权重。此外, 输入信息会经过全局最大池化, 以增强模型对特征显著区域的关注度。接着利用 1×1 的卷积生成全局通道权重, 进一步筛选重要特征。然后, 将坐标注意力权重信息与全局通道权重信息逐元素相乘, 得到最终的坐标注意力权重。这一过程将目标的空间位置信息与显著特征信息相融合, 强化了模型对目标的定位和识别能力。最后, 将坐标注意力权重与原始特征相乘后与多尺度通道交互的加权特征相融合, 再通过残差连接防止原始特征丢失, 输出增强后的特征图。该特征图与全局语义特征建模网络的输出相结合后, 经过多层 C3k2 与卷积层的处理, 实现多尺度的特征融合。

经过特征融合网络对特征图进行特征增强与融合, 输出的不同尺寸特征图再通过检测头的处理, 实现对目标的分类与回归, 进而完成目标检测任务。

3 实验结果与分析

3.1 实验环境与参数设置

本实验使用的是 ubuntu20.04 LTS 操作系统, CPU 为 Intel Core i7 11700K, GPU 为 NVIDIA GeForce RTX 4070 SUPER。实验使用的相关技术框架是 PyTorch 2.2.0、CUDA 11.8、Python 3.8.20。实验训练过程中, epochs 设置为 300, batch 设置为 16, 初始学习率为 0.01, 最终学习率为 0.0001, 优化器采用了随机梯度下降算法(stochastic gradient descent, SGD)。在训练开始时, 首先进行 3 个 epoch 的 warm-up 训练。在 warm-up 训练阶段动量参数设置为 0.8, 训练结束后修改为 0.937。

3.2 数据集与评价指标

本研究选取 RSOD 与 DIOR 两个公开数据集开展实验验证。RSOD 数据集聚焦 4 类典型地物目标(操场、油桶、飞机、立交桥), 由 936 幅图像与 6950 个标注实例构成; DIOR 数据集涵盖 20 个目标类别, 总计包含 23463 幅遥感图像, 对应标注的目标实例数量达 192472 个。在数据集划分方面, 针对 RSOD 数据集, 本研究按照 8:2 的比例划分, 训练集 746 幅图像, 验证集 190 幅图像; 对于 DIOR 数据集, 本研究采用随机划分方式, 选取 11725 幅图像构建训练集, 余下 11738 幅图像作为验证集。

本研究采用多维度评估指标对模型性能进行分析: 针对检测精度的评估, 选取平均精度(average precision, AP)及平均精度均值(mean average precision, mAP)作为核心指标; 模型的大小通过参数量(Parameters)进行表征; 推理速度则以每秒处理帧数(frames per second, FPS)作为关键评估标准。

3.3 消融实验结果与分析

为验证 DHMF-Net 模型中 3 个创新模块的实际效果, 本研究在 DIOR 与 RSOD 两个数据集上设计并开展了针对性的消融实验。其中, 在 DIOR 数据集上进行消融实验的结果如表 1 所示。由表 1 可知, 基线模型 YOLO11 在 DIOR 数据集上的 mAP 为 78.0%。在基线模型的基础上分别加入 DGHGM 模块和 MSDCFB 模块后, mAP 先后达到 78.6% 和 79.5%。在基线模型的基础上同时加入 DGHGM 和 MCAM 模块 mAP 为 78.7%; 同时加入 DGHGM 和 MSDCFB 模块 mAP 为 80.0%。在基线模型的基础上, 同时加入 DGHGM、MSD-

CFB 和 MCAM 模块后, mAP 达到了 80.5%, 比基线模型提升了 2.5 个百分点。

表 1 DIOR 数据集消融实验结果
Table 1 Ablation results on the DIOR dataset %

方法	mAP
YOLO11	78.0
YOLO11+DGHGM	78.6
YOLO11+MSDCFB	79.5
YOLO11+DGHGM+MCAM	78.7
YOLO11+DGHGM+MSDCFB	80.0
YOLO11+DGHGM+MSDCFB+MCAM	80.5

注: 加粗为效果最好。

模型在 RSOD 数据集上的消融实验结果如表 2 所示。基线模型 YOLO11 在 RSOD 数据集上的 mAP 为 93.0%。在基线模型的基础上分别加入 DGHGM 模块和 MSDCFB 模块后, mAP 先后达到 93.6% 和 94.4%。在基线模型的基础上同时加入 DGHGM 和 MCAM 模块 mAP 为 94.9%; 同时加入 DGHGM 和 MSDCFB 模块后 mAP 为 95.1%。在基线模型的基础上, 同时加入 DGHGM、MSDCFB 和 MCAM 模块后, mAP 达到了 95.3%, 较基线模型提升了 2.3 个百分点。

表 2 RSOD 数据集消融实验结果
Table 2 Ablation results on the RSOD dataset %

方法	mAP
YOLO11	93.0
YOLO11+DGHGM	93.6
YOLO11+MSDCFB	94.4
YOLO11+DGHGM+MCAM	94.9
YOLO11+DGHGM+MSDCFB	95.1
YOLO11+DGHGM+MSDCFB+MCAM	95.3

注: 加粗为效果最好。

对比两个数据集上的消融实验结果, DHMF-Net 模型在 RSOD 数据集上的 mAP 达到了 95.3%, 在 DIOR 数据集上的 mAP 达到了 80.5%, 两者相差 14.8 个百分点。出现这种差异的原因是 RSOD 数据集的图片数目和目标类别都远少于 DIOR 数据集, 降低了目标检测的难度。

3.4 对比实验结果与分析

为了进一步证明 DHMF-Net 模型的有效性, 本文分别在 DIOR 和 RSOD 两个数据集上, 将所提出的 DHMF-Net 模型与其他的主流目标检测方法进行对比实验。其中, 不同算法在 DIOR 数据集上的对比实验结果如表 3 所示。

表 3 不同方法在 DIOR 数据集上的检测结果比较
Table 3 Comparison of detection results of different methods in DIOR dataset

方法	骨干网络	输入尺寸	mAP/%	参数量/ 10^6
SSD	VGGNet	300×300	55.7	26.3
Faster R-CNN	VGGNet	600×1000	52.1	136.7
CenterNet	Resnet50	640×640	55.2	32.7
RetinaNet	Resnet50	640×640	61.6	37.9
YOLOv3	Darknet53	640×640	58.0	61.5
YOLOv4	CSPDarknet53	640×640	65.0	52.5
YOLOv5	CSPDarknet	640×640	69.6	7.1
YOLOX	CSPDarknet	640×640	72.2	8.9
YOLOv8	CSPDarknet	640×640	77.2	3.0
AAFNet ^[18]	Modified CSPDarknet	640×640	75.3	14.1
SCFNet ^[20]	Modified CSPDarknet	640×640	69.9	32.1
StrMCsDet ^[21]	CSPDarknetC5	608×608	65.6	41.4
YOLO11	CSPDarknet	640×640	78.0	2.6
DHMF-Net	Modified CSPDarknet	640×640	80.5	3.8

注: 加粗表示效果最好。

由表 3 可知, 在目标检测精度方面, DHMF-Net 模型比 SSD、Faster R-CNN、CenterNet^[6] 和 RetinaNet 这些经典方法有显著提升。与 YOLOv3、

YOLOv4、YOLOv5、YOLOX、YOLOv8 和 YOLO11 这些主流的目标检测方法相比, DHMF-Net 模型仍有较大的提升。与 AAFNet^[18]、SCFNet^[20] 和

StrMCsDet^[21] 这些改进的遥感图像目标检测方法相比, DHMF-Net 模型的 mAP 分别提升了 5.2、10.6、14.9 百分点。通过与不同方法的对比, 验证了本文提出的 DHMF-Net 模型的有效性, 进一步说明了 DGHGM、MSDCFB 和 MCAM 这 3 个模块对提升模型的检测精度有重要作用。

从模型参数量上分析, DHMF-Net 模型在保持较高检测精度的同时, 仅有 3.8×10^6 的参数量, 显著低于大多数方法。这意味着本文提出的方法计算开销更小, 对硬件要求更低。虽然 YOLOv8 与 YOLO11 的参数量比 DHMF-Net 模型分别减少了 0.8×10^6 、 1.2×10^6 , 但是其检测精度也比 DHMF-

Net 模型降低了 3.3 和 2.5 百分点。这意味着尽管 DHMF-Net 模型在引入 DGHGM、MSDCFB 和 MCAM 这 3 个模块时增加了一定的参数量, 但检测精度有明显的提升。

各个对比方法在 DIOR 数据集上的检测结果详细比较如表 4 所示。表中 C1—C20 依次对应 DIOR 数据集中 20 个不同类别目标的检测精度。通过对比分析可得, 本研究提出的 DHMF-Net 模型在 17 类目标的检测精度上取得了最优结果。虽然在剩余 3 个类别的检测精度未达到最高值, 但是 DHMF-Net 模型的检测精度仍处于较高的水平。

表 4 不同方法在 DIOR 数据集的检测结果详细比较
Table 4 Comparison of detailed detection results of different methods in DIOR dataset

方法	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	mAP	%	
SSD	58.6	67.1	68.1	83.6	26.2	77.2	53.5	67.7	48.2	75.2	56.7	54.3	50.7	34.9	67.9	28.2	77.8	46.0	18.9	52.5	55.7		
Faster R-CNN	47.9	64.7	68.6	84.0	23.7	76.4	53.1	57.5	47.0	74.6	56.8	42.0	49.2	16.6	70.5	20.9	73.9	52.9	12.2	49.1	52.1		
CenterNet	65.7	64.7	69.2	84.8	25.8	73.8	46.8	54.0	48.0	69.5	56.9	39.7	48.2	45.1	47.6	39.6	79.8	50.3	30.2	65.1	55.2		
RetinaNet	71.8	65.7	71.1	87.9	30.9	79.2	57.3	69.9	54.8	79.4	74.1	55.8	53.3	50.1	70.7	40.2	83.8	45.9	21.4	68.4	61.6		
YOLOv3	74.7	54.6	69.4	83.8	27.2	73.5	47.7	50.2	46.9	57.7	44.2	57.8	47.3	88.6	29.3	72.3	85.8	27.3	47.4	73.4	58.0		
YOLOv4	84.8	65.5	74.7	85.1	36.3	78.6	52.3	57.2	54.9	71.3	69.2	58.2	56.2	88.0	38.7	67.8	85.8	49.0	49.9	75.9	65.0		
YOLOv5	85.9	76.1	72.3	89.4	43.6	80.8	61.5	59.9	58.0	75.5	73.8	62.1	57.6	89.1	55.7	72.7	86.9	55.5	53.8	82.7	69.6		
YOLOX	89.3	72.0	75.3	90.2	47.8	79.3	61.5	60.1	66.2	74.2	76.8	58.1	62.3	89.9	71.1	77.5	89.9	61.0	57.3	83.5	72.2		
YOLOv8	80.2	87.8	84.7	89.4	47.3	78.5	73.6	89.0	72.5	83.0	81.8	67.4	63.2	91.0	80.2	75.1	91.6	67.6	51.7	87.4	77.2		
AAFNet ^[18]	92.7	81.7	81.0	90.8	49.7	81.3	69.9	67.9	70.4	80.4	77.3	64.0	63.2	90.4	68.5	78.0	90.6	65.1	56.6	85.7	75.3		
StrMCsDet ^[21]	78.6	58.4	81.3	72.0	38.1	79.2	37.1	49.3	49.5	56.8	62.9	35.5	42.5	54.9	66.0	66.6	80.8	38.3	38.3	34.9	65.6		
YOLO11	79.6	90.7	86.1	89.5	46.8	78.5	77.4	90.3	70.5	85.4	81.8	68.1	63.7	90.9	81.6	75.1	91.5	74.7	51.8	86.5	78.0		
DHMF-Net	85.2	92.7	88.0	91.2	51.0	81.4	80.1	93.2	75.4	86.4	83.6	68.5	66.0	91.9	83.8	76.4	92.9	77.6	55.1	90.2	80.5		

注: 加粗表示效果最好。

不同方法在 RSOD 数据集上的检测结果比较如表 5 所示。在目标检测精度方面, 通过对比可知, DHMF-Net 模型的效果比 SSD、Faster R-CNN、CenterNet 和 RetinaNet 这些经典方法有显著提升。与 YOLO 系列的主流方法 YOLOv3、YOLOv4、

YOLOv5、YOLOX、YOLOv8 和 YOLO11 相比, DHMF-Net 模型的 mAP 分别提升了 13.7、7.5、11.7、5.9、5.9、2.3 百分点。此外, 与改进的遥感图像目标检测方法 IRSAM^[22] 相比, DHMF-Net 模型的 mAP 依然优于 IRSAM。

表 5 不同方法在 RSOD 数据集上的检测结果比较
Table 5 Comparison of detection results of different methods in RSOD dataset

方法	骨干网络	输入尺寸	mAP/%	参数量/ 10^6
SSD	VGGNet	300×300	76.4	26.3
Faster R-CNN	VGGNet	600×1000	80.5	136.7
CenterNet	Resnet50	640×640	77.3	32.7
RetinaNet	Resnet50	640×640	81.7	37.9
YOLOv3	Darknet53	640×640	81.6	61.5
YOLOv4	CSPDarknet53	640×640	87.8	52.5

续表 5

方法	骨干网络	输入尺寸	mAP/%	参数量/10 ⁶
YOLOv5	CSPDarknet	640×640	83.6	7.1
YOLOX	CSPDarknet	640×640	89.4	8.9
YOLOv8	CSPDarknet	640×640	89.4	3.0
IRSAM ^[22]	IRSAB	640×640	94.7	4.7
YOLO11	CSPDarknet	640×640	93.0	2.6
DHMF-Net	Modified CSPDarknet	640×640	95.3	3.8

注: 加粗表示效果最好。

从模型参数量上分析, DHMF-Net 模型的参数量显著低于大多数方法。虽然在引入 DGHGM、MSDCFB 和 MCAM 这 3 个模块后 DHMF-Net 模型参数量略高于 YOLO11 和 YOLOv8, 但是其检测精度优于 YOLO11 和 YOLOv8。通过比较不同方法在 RSOD 数据集上的检测结果, 进一步验证了 DHMF-Net 模型的有效性。

为更全面地评估 DHMF-Net 模型的性能, 本文在 RSOD 数据集上与 YOLO 系列模型进行了全方面对比, 对比结果如表 6 所示。结果表明, DHMF-Net 模型的检测精度最高; 在推理速度与参数量方面, DHMF-Net 模型的性能低于 YOLO11 和 YOLOv8 模型, 优于其他 YOLO 系列模型。

表 6 YOLO 系列与 DHMF-Net 对比结果
Table 6 Comparison results of YOLO and DHMF-Net

方法	mAP/%	推理速度/(帧/s)	参数量/10 ⁶
YOLOv3	81.6	30	61.5
YOLOv4	87.8	28	52.5
YOLOv5	83.6	48	7.1
YOLOX	89.4	49	8.9
YOLOv8	89.4	68	3.0
YOLO11	93.0	86	2.6
DHMF-Net	95.3	53	3.8

3.5 可视化结果分析

为更直观地展示 DHMF-Net 模型的效果, 本文将 DIOR 数据集和 RSOD 数据集上的目标检测结果进行可视化展示, 结果分别如图 5、6 所示。

由图 5 可知, 在 DIOR 数据集的可视化结果中, 第 1 张与第 2 张图片因为背景复杂, 干扰了目标的特征信息, 使用 YOLO11 检测时出现漏检与误检。而 DHMF-Net 模型利用全局语义特征建模, 抑制了复杂背景对目标检测的影响, 正确检测出目标。在第 3 张图片中, 由于目标尺度差异大, 导致 YOLO11 出现了漏检。而 DHMF-Net 模

型利用多尺度感知机制, 减少了目标的漏检。

由图 6 可知, 在 RSOD 数据集的可视化结果中, 通过对比 YOLO11 与 DHMF-Net 的检测结果可看出, 无论是背景复杂的目标, 还是尺度差异大的目标, DHMF-Net 的检测效果均优于 YOLO11。

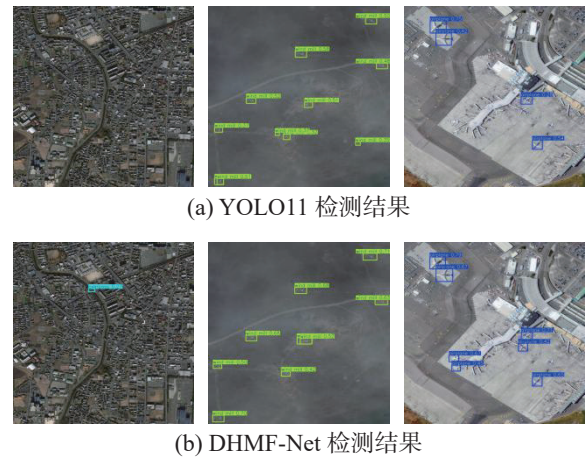


图 5 DIOR 数据集可视化结果
Fig. 5 Visualization detection results on the DIOR dataset

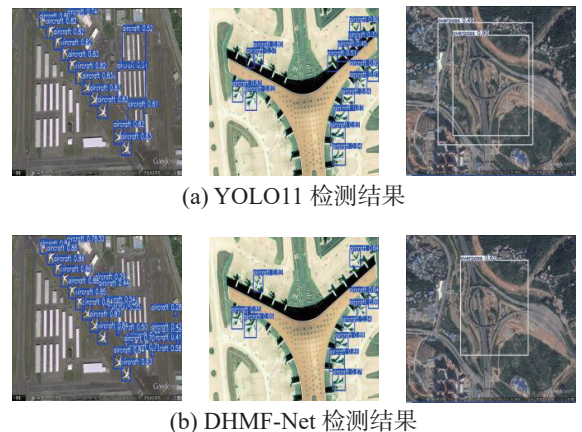


图 6 RSOD 数据集可视化结果
Fig. 6 Visualization detection results on the RSOD dataset

4 结束语

本文提出了一种基于动态超图与多尺度特征融合的遥感图像目标检测模型, 提升了模型的多

尺度感知能力和全局语义建模能力。通过在 DIOR 和 RSOD 两个遥感图像数据集上进行大量实验, DHMF-Net 模型取得了较好的效果。与基线模型相比, DHMF-Net 模型分别提升了 2.5 和 2.3 百分点, 克服了遥感图像目标检测中背景复杂, 目标尺度差异大的问题。

但是, DHMF-Net 模型的参数量仍有优化空间。未来研究将探索在保持较高检测精度的前提下, 更加轻量化的遥感图像目标检测模型。

参考文献:

- [1] 徐丹青, 吴一全. 光学遥感图像目标检测的深度学习算法研究进展[J]. *遥感学报*, 2024, 28(12): 3045–3073.
XU Danqing, WU Yiquan. Progress of research on deep learning algorithms for object detection in optical remote sensing images[J]. *National remote sensing bulletin*, 2024, 28(12): 3045–3073.
- [2] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779–788.
- [3] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Computer Vision–ECCV 2016. Cham: Springer, 2016: 21–37.
- [4] 吴珺, 董佳明, 刘欣, 等. 注意力优化的轻量目标检测网络及应用[J]. *智能系统学报*, 2023, 18(3): 506–516.
WU Jun, DONG Jiaming, LIU Xin, et al. Lightweight object detection network and its application based on the attention optimization[J]. *CAAI transactions on intelligent systems*, 2023, 18(3): 506–516.
- [5] 唐嘉璐, 杨钟亮, 张淞, 等. 结合显微视觉和注意力机制的毛羽检测方法[J]. *智能系统学报*, 2022, 17(6): 1209–1219.
TANG Jialu, YANG Zhongliang, ZHANG Song, et al. Detection of yarn hairiness combining microscopic vision and attention mechanism[J]. *CAAI transactions on intelligent systems*, 2022, 17(6): 1209–1219.
- [6] ZHOU Xingyi, WANG Dequan, KRAHENBUHL P. Objects as points[EB/OL]. (2019–04–16)[2025–08–03]. <https://arxiv.org/abs/1904.07850>.
- [7] 梁礼明, 冯耀, 龙鹏威, 等. 基于 MobileViT 和多尺度特征聚合的遥感图像目标检测[J]. *智能系统学报*, 2024, 19(5): 1168–1177.
LIANG Liming, FENG Yao, LONG Pengwei, et al. Remote sensing image object detection based on MobileViT and multiscale feature aggregation[J]. *CAAI transactions on intelligent systems*, 2024, 19(5): 1168–1177.
- [8] 周莲, 何楚, 汪鼎文, 等. 光学遥感图像的小样本目标检测[J]. *遥感学报*, 2024, 28(7): 1693–1701.
ZHOU Lian, HE Chu, WANG Dingwen, et al. Few-shot object detection in optical remote sensing images[J]. *National remote sensing bulletin*, 2024, 28(7): 1693–1701.
- [9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580–587.
- [10] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440–1448.
- [11] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 39(6): 1137–1149.
- [12] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Computer Vision–ECCV 2020. Cham: Springer, 2020: 213–229.
- [13] HUANG Shihua, LU Zhichao, CUN Xiaodong, et al. DEIM: DETR with improved matching for fast convergence[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2025: 15162–15171.
- [14] MENG Depu, CHEN Xiaokang, FAN Zejia, et al. Conditional DETR for fast training convergence[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2022: 3631–3640.
- [15] YE Xianhang, XU Chang, ZHU Haoran, et al. Density-aware DETR with dynamic query for end-to-end tiny object detection[J]. *IEEE journal of selected topics in applied earth observations and remote sensing*, 2025, 18: 13554–13569.
- [16] 赵文清, 康恽瑾, 赵振兵, 等. 改进 YOLOv5s 的遥感图像目标检测[J]. *智能系统学报*, 2023, 18(1): 86–95.
ZHAO Wenqing, KANG Yijin, ZHAO Zhenbing, et al. A remote sensing image object detection algorithm with improved YOLOv5s[J]. *CAAI transactions on intelligent systems*, 2023, 18(1): 86–95.
- [17] 文斌, 张俊, 王浚银, 等. 基于 FNM-Net 的轻量级遥感目标检测算法[J]. *现代电子技术*, 2025, 48(13): 1–10.
WEN Bin, ZHANG Jun, WANG Junyin, et al. Lightweight remote sensing object detection algorithm based on FNM-Net[J]. *Modern electronics technique*, 2025, 48(13): 1–10.
- [18] ZHAO Wenqing, KANG Yijin, CHEN Hao, et al. Adaptively attentional feature fusion oriented to multiscale ob-

- ject detection in remote sensing images[J]. *IEEE transactions on instrumentation and measurement*, 2023, 72: 5008111.
- [19] 单慧琳, 王硕洋, 童俊毅, 等. 增强小目标特征的多尺度光学遥感图像目标检测[J]. *光学学报*, 2024, 44(6): 374–386.
SHAN Huilin, WANG Shuoyang, TONG Junyi, et al. Multi-scale optical remote sensing image target detection based on enhanced small target features[J]. *Acta optica sinica*, 2024, 44(6): 374–386.
- [20] YUE Chenke, YAN Junhua, ZHANG Yin, et al. SCFNet: Semantic correction and focus network for remote sensing image object detection[J]. *Expert systems with applications*, 2023, 224: 119980.
- [21] REN Shougang, FANG Zhiruo, GU Xingjian. A cross stage partial network with strengthen matching detector for remote sensing object detection[J]. *Remote sensing*, 2023, 15(6): 1574.
- [22] 赵文清, 赵振寰, 巩佳潇. 结合倒残差自注意力机制的遥感图像目标检测[J]. *智能系统学报*, 2025, 20(1): 64–72.
ZHAO Wenqing, ZHAO Zhenhuan, GONG Jiexiao. Remote sensing image object detection based on inverted residual self-attention mechanism[J]. *CAAI transactions on intelligent systems*, 2025, 20(1): 64–72.
- [23] FENG Yifan, YOU Haoxuan, ZHANG Zizhao, et al. Hypergraph neural networks[J]. *Proceedings of the AAAI conference on artificial intelligence*, 2019, 33(1): 3558–3565.
- [24] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015–11–23)[2025–08–03]. <https://arxiv.org/abs/1511.07122>.
- [25] FENG Yifan, HUANG Jiangang, DU Shaoyi, et al. Hyper-YOLO: when visual object detection meets hypergraph computation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2025, 47(4): 2388–2401.
- [26] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132–7141.
- [27] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision–ECCV 2018. Cham: Springer International Publishing, 2018: 3–19.

作者简介:



赵文清, 教授, 博士, 主要研究方向为人工智能与图像处理。获河北省科技进步二等奖、三等奖各 1 项。发表学术论文 50 余篇。E-mail: zhaowenqing@ncepu.edu.cn。



李溢晔, 硕士研究生, 主要研究方向为深度学习与遥感图像处理。E-mail: liyipei2025@163.com。

[责任编辑: 刘冰洁]