

频域增强及多维度特征扩散的遥感小目标检测算法*

周冬梅 王飞成 吴小所

(兰州交通大学电子与信息工程学院 兰州 730070)

摘要:针对遥感图像目标检测中背景复杂、目标尺度多变及对任意方向旋转敏感等核心难题,提出一种基于双域特征提取与多维度聚焦扩散机制的新型检测算法(DMADet),旨在提升模型在复杂遥感场景下的感知能力与鲁棒性。首先,设计双域协同特征提取网络(DDCNet),通过Scharr算子强化空间域边缘细节,并结合快速傅里叶变换提取频域全局表征,实现空间、频域特征的双向交互与互补融合;其次,为缓解特征金字塔跨层传递中的语义衰减问题,构建多维度聚焦扩散金字塔(MFDPN),采用通道分块加权策略自适应融合高低层特征,并引入轻量化双流协同注意力模块(LDC-Attention),增强多尺度上下文感知能力;最后,面向旋转目标检测,提出自适应旋转感知检测头,利用动态路由机制生成方向感知卷积核,有效提升模型对任意朝向目标的检测精度与旋转不变性。实验结果表明,DMADet在DOTA-v1.0、NWPU VHR-10和RSOD 3个主流遥感目标检测数据集上的mAP@0.5分别达到75.9%、96.8%和85.7%,显著优于当前先进的对比方法,充分验证了所提算法的有效性与优越性,不仅有效缓解了因背景干扰、尺度变化和旋转多样性导致的检测精度不足问题,还显著提升了目标定位的准确性与鲁棒性。

关键词: 频域增强;多维度特征;Scharr算子;遥感图像;ARC卷积;目标检测

中图分类号: TP391.4;TN98 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Remote sensing small target detection algorithm with frequency domain enhancement and multi-dimensional feature diffusion

Zhou Dongmei Wang Feicheng Wu Xiaosuo

(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: To detect objects in remote sensing imagery remains highly challenging due to complex background clutter, extreme variations in object scale, and high sensitivity to arbitrary orientations. To address these issues, this paper proposes DMADet, a novel detection framework based on dual-domain feature extraction and multi-dimensional focused diffusion, aiming to enhance perception capability and robustness in complex remote sensing scenarios. The method first introduces a dual-domain collaborative network (DDCNet) that jointly exploits spatial and frequency domain information: edge-aware features in the spatial domain are strengthened using Scharr operators, while global contextual representations are captured in the frequency domain via Fast Fourier transform, enabling bidirectional interaction and complementary fusion between the two domains. Second, to alleviate semantic degradation across layers in conventional feature pyramids, a multi-dimensional focused diffusion pyramid (MFDPN) is developed, which employs a channel-block weighting strategy to adaptively integrate high- and low-level features and incorporates a lightweight dual-stream collaborative attention module (LDC-Attention) to enhance multi-scale contextual awareness. Finally, an adaptive rotation-aware detection head is designed, leveraging a dynamic routing mechanism to generate orientation-sensitive convolution kernels, thereby significantly improving rotational invariance and detection accuracy for arbitrarily oriented objects. Extensive experiments demonstrate that DMADet achieves mAP@0.5 scores of 75.9%, 96.8%, and 85.7% on the DOTA-v1.0, NWPU VHR-10, and RSOD benchmark datasets, respectively—consistently outperforming current state-of-the-art methods. These results validate the effectiveness and superiority of the proposed approach in mitigating performance degradation caused by background interference, scale diversity, and rotational variance, while substantially

收稿日期: 2025-06-20 Received Date: 2025-06-20

* 基金项目: 国家自然科学基金(62161016, 61661025, 62161016)、国家重点研发计划(YFB3903604)、甘肃省科技计划项目(24JRZA104)、科研培育计划项目(2023011wys021)、“昆仑英才”人才引进科研项目(W2023 QLGLYCZX 034)资助

improving object localization accuracy and robustness in real-world remote sensing applications.

Keywords: frequency domain enhancement; multi-dimensional features; scharr; remote sensing images; arc convolution; object detection

0 引言

目标检测是计算机视觉的核心,支撑语义理解。无人机和遥感技术推动其在交通、农业、救灾等领域的应用。基于卷积神经网络(convolutional neural networks, CNN)的深度学习方法通过端到端特征学习显著提升性能,主流算法分为高效单阶段模型,如你只看一次(you only look once, YOLO),和高精度双阶段。然而,遥感图像背景复杂、小目标多、尺度变化大,使现有方法在精度与效率上面临挑战。

针对遥感图像背景复杂、小目标易受干扰等问题,近年研究提出了多种高效检测方法:Fang 等^[1]设计双向注意力融合模块,结合空间-通道协同注意力与跨层特征金字塔,增强复杂背景下的特征判别力;Gao 等^[2]引入全局语义交互与局部注意力机制,抑制背景噪声并强化小目标特征;Ji 等^[3]通过多分支空洞卷积扩展感受野,融合上下文信息以提升对象与背景的区别能力。

相较于两阶段模型,一阶段检测器易受背景干扰与目标密集影响,导致定位偏差和误判。对此,Zhao 等^[4]提出 PSO-YOLO,融合 Swin Transformer 与 STENek 模块以增强长程上下文建模;Wang 等^[5]构建 SR-YOLO,引入递归瓶颈 Transformer 与 Efficient Neck 提升密集场景性能。针对尺度变化大问题,Gao 等^[6]设计局部-全局融合模块捕获多尺度特征,Jiang 等^[7]通过双向密集传统特征金字塔网络(feature pyramid network, FPN)增强跨尺度表达。此外,王迎龙等^[8]提出 BG-YOLO,结合高分辨率检测头、Biformer 注意力、NWD 损失与 GhostNetV2 轻量化,在低慢小无人机检测中兼顾精度与速度;冉宁等^[9]提出 SBC-YOLOv8,集成 SAHI 切片、BiFPN + P2 及 CSP-OmniFusion,有效应对遥感小目标密集、微小与特征难提取等挑战。

现有方法多聚焦空间域浅层语义挖掘^[10],忽视频域特征的旋转不变性与全局表征能力^[11],且缺乏对边缘信息的有效利用,难以刻画复杂背景下小目标的模糊边界。FPN 虽支持多尺度提取,但其单向融合机制导致低层细节在上采样中衰减,且跨尺度关联受限,影响检测性能。为此,提出一种融合双域特征提取与多维度聚焦扩散金字塔的遥感目标检测算法。

本文设计了一种基于空频域协同的特征提取骨干网络(dual domain collaborative feature extraction backbone network, DDCNet)。该网络通过融合空间结构细节和频

域全局表征既聚焦于空间域特征的局部细节刻画,又强化了频域特征的全局语义理解。提出了一种多维度特征聚焦扩散金字塔网络(multi dimensional feature focused diffusion pyramid network, MFDPN)。该网络通过多维度特征筛选聚焦模块(multi dimensional feature filtering and focusing module, MDFM)将不同维度的特征进行多尺度的聚焦和扩散,提升了不同尺度特征的交互能力,减少了信息传递时的衰减和丢失。设计了轻量化双流协同注意力(lightweight dual-stream collaborative attention, LDC-Attention)自顶向下注入 MFDPN 金字塔网络中,保持网络效率的同时,充分利用了遥感场景中多层次的局部信息和全局视觉特征。构建了自适应旋转感知增强检测头,其中的自适应旋转卷积模块(adaptive rotation convolution, ARC)可根据输入特征图的几何特性自适应调整卷积核的旋转角度,从而捕获目标的方位特征,并通过角度感知权重分配策略,显著提升了模型对旋转目标的特征表征能力。

1 DMADet 目标检测算法

由于遥感图像存在复杂的几何场景、目标密集分布、目标的有向性以及目标尺度变化较大等问题,普通场景下的检测模型难以满足遥感图像中目标检测的需求。为了解决上述问题,提出了双域多维注意力检测算法网络模型(dual-domain multi-dimensional attention-based detector, DMADet),其整体架构如图 1 所示。

1.1 DDCNet 骨干网络

传统目标检测骨干网络主要聚焦于空间域特征学习,虽能提取部分有效语义信息,但忽略了频域中隐含的一些重要特征信息。为此,提出了一种双域(空间域与频域)协同的特征提取骨干网络,结构如图 2 所示。

骨干网络的核心在于构建跨尺度混合域模块(cross stage partial frequency-spatial module, CSP_FreqSpatial),该模块采用空间-频域双分支架构。在空间域分支,引入 Scharr 算子从输入中获取基于空间结构的细节和边缘特征;在频域分支,通过快速傅里叶变换与逆变换从频域捕捉到图像的低频和高频成分,从全局和局部的尺度上提取信息;最后通过加权的方式,将空域和频域特征相融合,使其能够同时考虑空间结构信息和频率信息的同时,增强了模型在多种场景下的表现能力。同时,为了减轻其计算负担,采用跨阶段部分连接(cross stage partial, CSP)策略通过分割和重组的方式降低计算量。

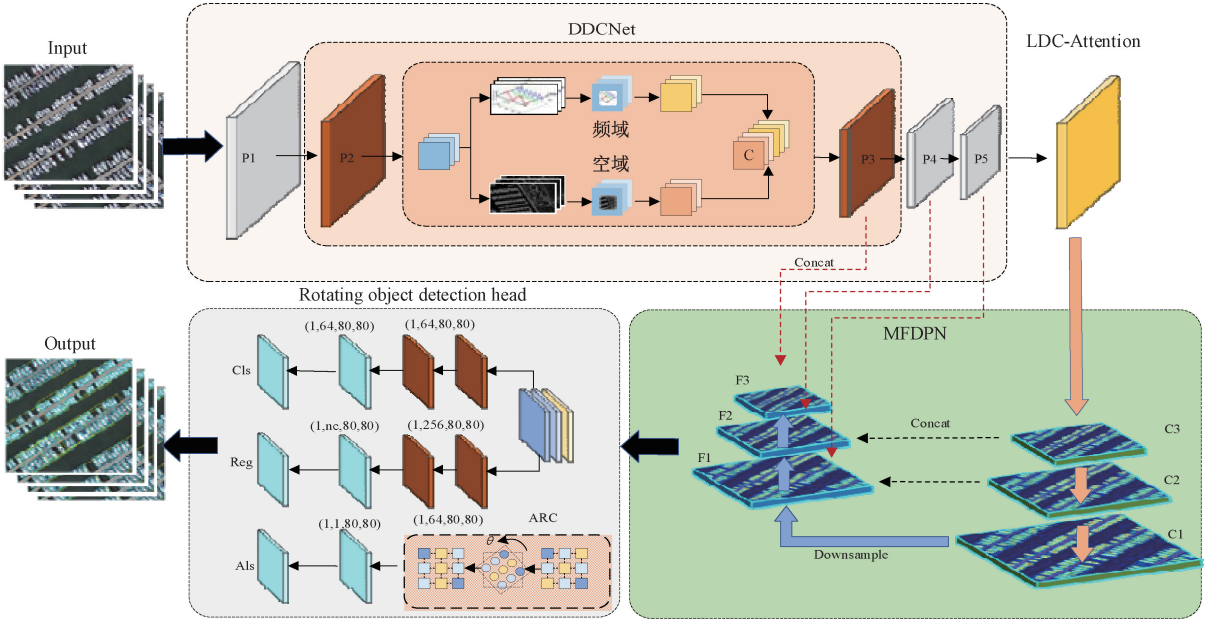


图1 DMADet 网络结构

Fig. 1 DMADet network structure diagram

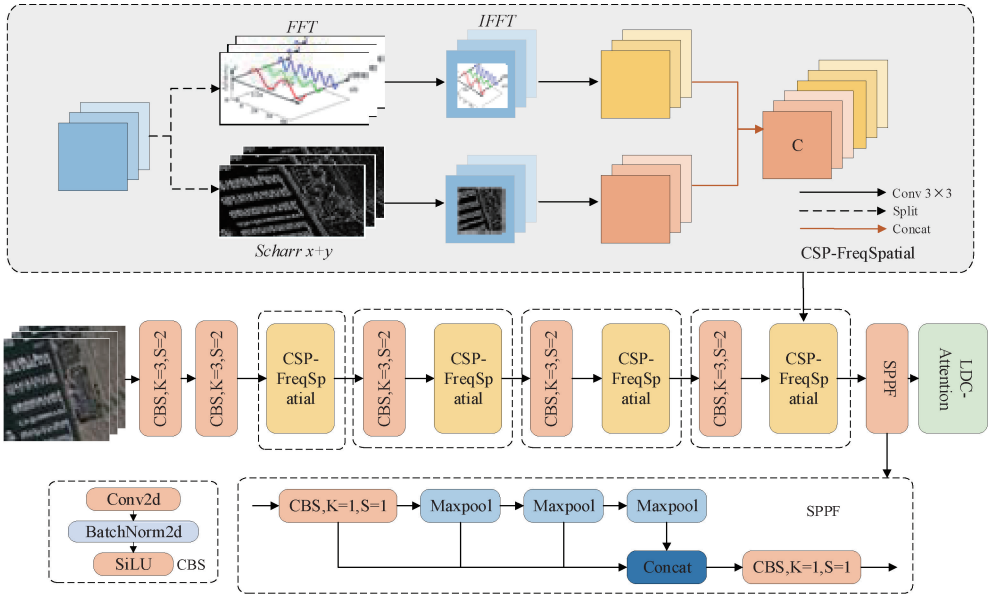


图2 DDCNet 结构

Fig. 2 DDCNet structure diagram

具体来说,设输入特征为 $X \in \mathbf{R}^{B \times H \times W \times C}$,先用二维实数快速傅里叶变换将输入从空间域转到频域,如式(1)所示。

$$\mathcal{F}(X) = RFFT2(X) \in \mathbf{C}^{B \times C \times H \times [W/2+1]} \quad (1)$$

其次,分离实部和虚部并拼接实虚分量,调整频域特征张量使其适应标准卷积,如式(2)所示。

$$r = \text{Re}(\mathcal{F}(X)) \quad (2)$$

$$i = \text{Im}(\mathcal{F}(X)) \quad (3)$$

$$F_{cat} = \text{Concat}(r, i) \quad (4)$$

$$F_{conv} = \text{Reshape}(F_{cat}) \quad (5)$$

式中: r 为实部分量; i 为虚部分量; F_{cat} 表示拼接操作, Reshape 操作将通道 C 与 F_{cat} 进行通道合并为 $2C$ 。利用卷积 $W_1 \in \mathbf{R}^{K \times K \times 2C \times C'}$ 提取频域特征并通过快速傅里叶逆变换恢复时域特征。

$$Z_1 = W_1 * F_{conv} \quad (6)$$

$$Z_{split} = \text{Reshape}(Z_1) \quad (7)$$

$$X_{ifft} = iRFFT2(Z_c(Z_{split}), s = (H, W)) \quad (8)$$

式中: $*$ 表示卷积操作; Z_{split} 表示恢复其复数结构; Z_c 表示重构复数张量; X_{fft} 表示频域特征输出, 通过进一步卷积 W_2 处理输出频域特征 $F_{freq} \in \mathbf{R}^{B \times C'' \times H \times W}$ 。

$$F_{freq} = W_2 * X_{fft} \quad (9)$$

同时, 在空域分支引入 Scharr 算子, 通过计算图像的梯度提取更具有针对性的边缘特征, 从而有效增强模型在复杂背景下对小目标的检测能力。Scharr 算子由 2 个 3×3 的卷积核组成, 分别用于检测图像中的水平和垂直边缘特征, 随后将水平和垂直方向的梯度加权求和, 结合这 2 个方向的梯度信息, Scharr 算子能有效地捕捉图像中局部区域内像素强度的突然变化, 从而识别出图像中的边缘, 生成包含显著边缘特征的特征图。利用 Scharr 算子进行卷积运算如下:

$$G_x(x, y) = \begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix} * X \in \mathbf{R}^{H \times W \times C} \quad (10)$$

$$G_y(x, y) = \begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix} * X \in \mathbf{R}^{H \times W \times C} \quad (11)$$

$$X_{scharr} = G_x(x, y) + G_y(x, y) \quad (12)$$

式中: $G_x(x, y)$ 和 $G_y(x, y)$ 分别表示 X 经水平及垂直边缘检测的输出特征图; X_{scharr} 表示 G_x 和 G_y 相加后生成的特征图; $*$ 表示卷积运算。

最后将含有 Scharr 算子提取的细节空间特征 $F_{spatial}$ 与频域特征进行融合。

$$F_{fusion} = F_{spatial} + F_{freq} \quad (13)$$

频域处理弥补了空间域在特征提取时遗漏的局部和全局语义信息, 空间域的处理则细化了边缘特征和细节

信息, 通过同时利用空域和频域的特征信息, 从而提高了对含有复杂背景和特定类型特征的遥感图像目标的检测能力。

1.2 多维度特征聚焦扩散金字塔网络

在光学遥感图像目标检测算法中, 当前主流算法普遍采用路径聚合金字塔网络 (path aggregation feature pyramid network, PAFPN) 等特征金字塔结构进行多尺度特征整合, 如图 3(a) 所示。

然而, PAFPN 仅支持相邻层融合, 缺乏远层特征交互, 限制了浅层细节与深层语义间的全局信息共享; 同时, 跨尺度特征在经多次卷积与下采样后易衰减丢失, 难以有效融合多层次优势。为此, 本文提出多维度聚焦扩散金字塔网络, 如图 3(b) 所示, 通过多尺度聚焦与扩散机制增强跨尺度特征交互, 显著减少信息衰减, 克服现有结构局限。

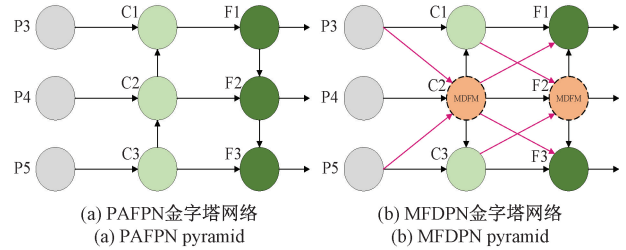


图3 金字塔网络

Fig. 3 Pyramid network

在 MFDPN 金字塔网络中, MDFM 模块能够整合不同层级的特征图, 并根据目标的大小和特性自适应地选择合适的特征进行特征融合, 如图 4 所示。

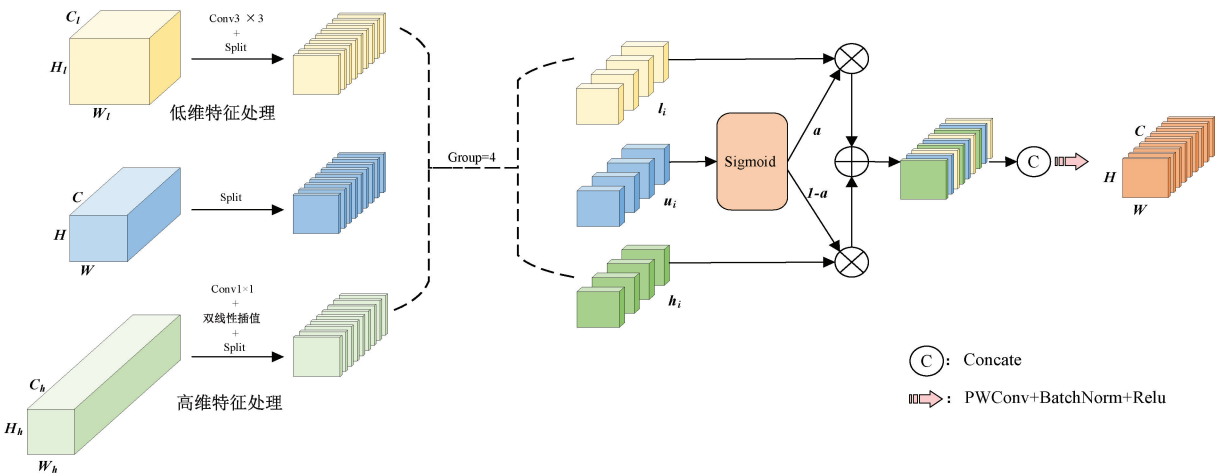


图4 MDFM 结构

Fig. 4 MDFM structure diagram

MDFM 首先通过双线性插值和 1×1 、 3×3 卷积操作,

将高维特征映射 $F_h \in \mathbf{R}^{H_h \times W_h \times C_h}$ 和低维特征映射

$F_l \in \mathbf{R}^{H_l \times W_l \times C_l}$ 与当前层的特征 $F_u \in \mathbf{R}^{H \times W \times C}$ 进行空间维度对齐,确保多尺度特征融合的一致性。随后为了解决传统特征融合方法中的通道信息耦合问题,在通道维度上将这些特征分为4个相等的特征分块,得到:

$$(h_i)_{i=1}^4 \in \mathbf{R}^{H \times W \times \frac{C}{4}} \quad (14)$$

$$(l_i)_{i=1}^4 \in \mathbf{R}^{H \times W \times \frac{C}{4}} \quad (15)$$

$$(u_i)_{i=1}^4 \in \mathbf{R}^{H \times W \times \frac{C}{4}} \quad (16)$$

其中, h_i 、 l_i 和 u_i 分别表示高维、低维和当前特征的第 i 个特征分块。各特征子块的权重分配通过以下方式实现:

$$\alpha = \text{sigmoid}(u_i) \quad (17)$$

$$u'_i = \alpha l_i + (1 - \alpha) h_i \quad (18)$$

式中: u_i 通过激活函数得到 $\alpha \in \mathbf{R}^{H \times W \times \frac{C}{4}}$, 而 $u'_i \in \mathbf{R}^{H \times W \times \frac{C}{4}}$

表示每个分块的选择性聚合结果。在通道维度上合并 $(u'_i)_{i=1}^4$ 后,得到 $F'_u \in \mathbf{R}^{H \times W \times \frac{C}{4}}$ 。最终输出 $\hat{F}_u \in \mathbf{R}^{H \times W \times \frac{C}{4}}$ 。

$$F'_u = [u'_1, u'_2, u'_3, u'_4] \quad (19)$$

$$\hat{F}_u = \delta(B(\text{Conv}(F'_u))) \quad (20)$$

其中,操作 Conv 、 B 和 δ 分别表示卷积、批量归一化和修正线性单元(rectified linear unit, ReLU)。当 $\alpha > 0.5$, 模型优先筛选细粒度特征;而 $\alpha < 0.5$, 则强调上下文特征,从而提高遥感图像中小目标的检测能力。

1.3 LDC-Attention

传统自注意力机制能够显著提升网络性能,但其在处理特征图时,需要对每个特征位置之间的关系进行建模,导致计算复杂度增加,并且卷积操作后的特征图之间存在高度的语义冗余^[12]。基于此,本文提出一种融合卷积和自注意力互补优势的混合建模范式,结构如图5所示。

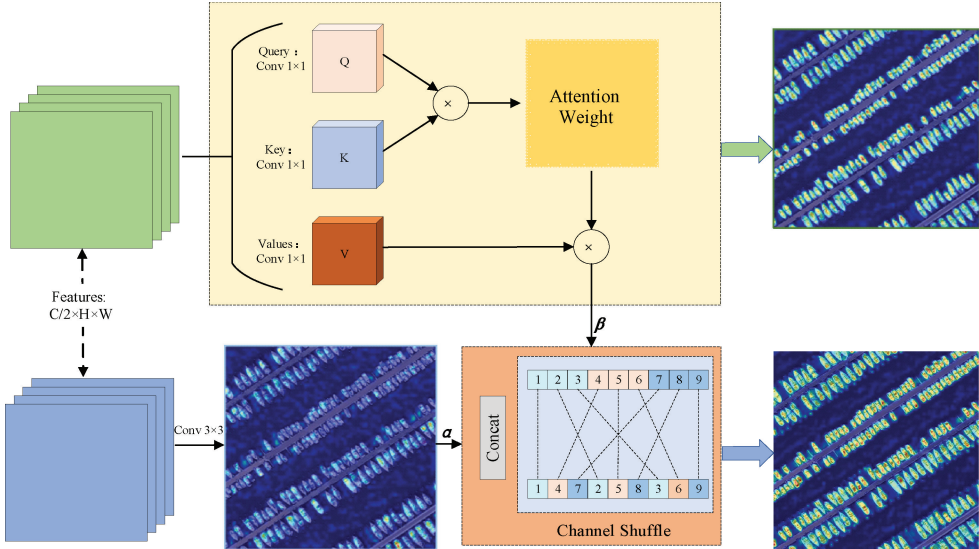


图5 LDC-Attention 结构

Fig.5 The LDC-attention structure diagram

常规卷积通过滑动窗口方式处理输入特征图,使得每个卷积核聚焦输入的局部区域,而自注意力可对特征图内的任意空间点建立长程依赖关系。因此, LDC-Attention 采用双分支并行处理结构,卷积分支利用 3×3 卷积聚焦邻近区域的特征细节,保留对微小目标的敏感性。自注意力分支通过压缩投影实现轻量化全局建模,先用 1×1 卷积将输入特征映射成查询(Q)、键(K)、值(V)卷积核矩阵,然后计算注意力权重并对值进行聚合,自注意力的具体输出过程如下:

设 $F \in \mathbf{R}^{H \times W \times C_{in}}$ 表示输入特征, $T \in \mathbf{R}^{H \times W \times C_{out}}$ 表示输出特征, H 和 W 分别表示特征空间维度高和宽。 $f_{ij} \in \mathbf{R}^{C_{in}}$ 和 $t_{ij} \in \mathbf{R}^{C_{out}}$ 为 F 和 T 中对应像素点坐标的特征向量, t_{ij}

的输出如式(21)所示。

$$t_{ij} = \sum_{a,b \in N_k(i,j)} A(\mathbf{W}_q^{(l)} f_{ij}, \mathbf{W}_k^{(l)} f_{ab}) \mathbf{W}_v^{(l)} f_{ab} \quad (21)$$

式中: $\mathbf{W}_q^{(l)}$ 、 $\mathbf{W}_k^{(l)}$ 、 $\mathbf{W}_v^{(l)}$ 分别表示用 1×1 卷积将输入特征映射成查询、键、值矩阵; $N_k(i,j)$ 表示以 (i,j) 为中心,卷积步长为 k 的局部空间像素特征; $A(\mathbf{W}_q^{(l)} f_{ij}, \mathbf{W}_k^{(l)} f_{ab})$ 则表示 $N_k(i,j)$ 位置的特征注意力权重 $A(q,k)$, 计算过程如式(22)所示。

$$A(q,k) = \text{softmax}_{N_k(i,j)} \left(\frac{(\mathbf{W}_q^{(l)} f_{ij})^G (\mathbf{W}_k^{(l)} f_{ab})}{\sqrt{d}} \right) \quad (22)$$

式中: d 是 $\mathbf{W}_q^{(l)} f_{ij}$ 的特征维度。自注意力机制的输出 $F_{att} \in \mathbf{R}^{H \times W \times C_{in}}$, 由所有位置的 t_{ij} 聚合得到:

$$F_{att} = [t_{ij}]_{i,j} \quad (23)$$

LDC-Attention 的输出 $F_{\text{out}} \in \mathbf{R}^{H \times W \times C_m}$ 如式(24)所示。

$$F_{\text{out}} = \alpha F_{\text{conv}} + \beta F_{\text{att}} \quad (24)$$

式中:参数 α 和 β 设置为 0.4 和 0.6,分配不同的权重给卷积和自注意力路径。最后,在输出端引入 Channel Shuffle 层,通过跨组信息交换促进通道间的信息交互,抑制冗余特征并选择重要的特征通道。

为系统评估 α 与 β 不同组合对模型性能的影响,开展了超参数敏感性实验,如表 1 所示。

表 1 注意力机制超参数实验分析

Table 1 Attention mechanism hyperparameters

参数	A	B	C	D	E	F
α	0	0.2	0.4	0.6	0.8	1
β	1	0.8	0.6	0.4	0.2	0
mAP@0.5/%	74.5	75	75.9	75.5	74.8	74
mAP@0.5 : 0.95/%	53	53.5	54.7	54	53.3	52.5

实验以 0.2 为步长调整参数取值,结果表明,当 α 为 0.4 β 为 0.6 时,模型在 mAP@0.5 和 mAP@0.5 : 0.95 指标上分别达到最优值 75.9% 和 54.7%。随着 α 增大至 1 β 减小至 0,模型性能呈现下降趋势,说明单纯依赖卷积特征或自注意力机制均会限制模型的特征表达能力。该结果进一步验证了卷积与自注意力路径协同融合对于提升目标检测性能的重要性。

1.4 自适应旋转感知增强检测头

遥感影像在垂直成像模式下待检测目标通常呈现显

著方向性分布,而传统目标检测器框架因标准卷积的平移不变性制约,难以有效捕捉旋转敏感特征,导致小目标定位偏移及密集目标误匹配等问题,如图 6 所示。

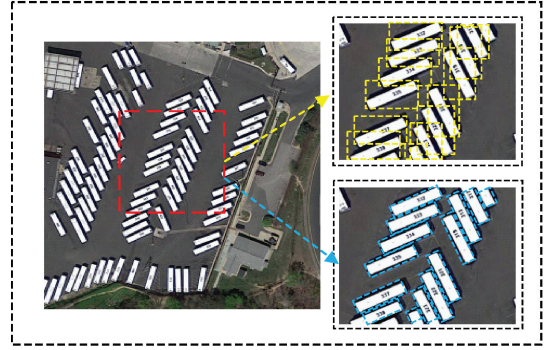


图 6 旋转框和水平框示意图

Fig. 6 Rotating and horizontal boxes

为了有效解决上述问题,设计了旋转感知增强检测器分支,在角度损失 (angle loss) 输入端引入卷积核尺寸为 3×3 的自适应旋转卷积模块 ARC^[13],如图 7 所示,ARC 卷积核可以根据不同的输入特征图进行自适应旋转,从而更好地捕捉有向目标的方向特征。具体而言,首先输入特征图 $x \in \mathbf{R}^{H \times W \times C_m}$ 通过路由函数框架进行自适应旋转角度和组合权重预测,路由函数初始化采用零均值截断正态分布,标准差为 0.2,有助于在网络学习初期产生较小的值,增强训练的稳定性。

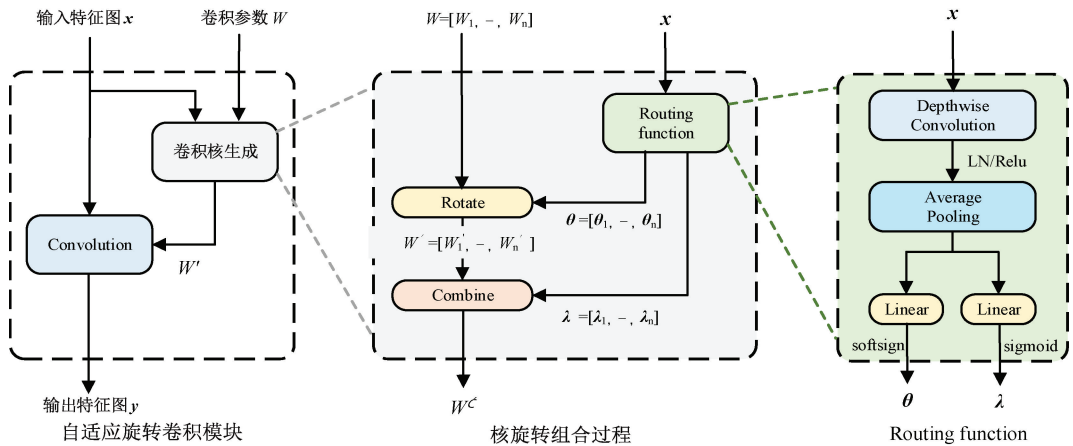


图 7 ARC 结构

Fig. 7 The ARC structure diagram

特征图 x 经过 3×3 的深度卷积进行特征提取,然后进行 ReLU,并在其后跟随一个平均池化层压缩空间维度,随后由激活函数 Softsign 和 Sigmoid 分别预测旋转角度 θ 和组合权重 λ ;在核旋转组合过程阶段,基于预测参数(旋转角度 $\theta = [\theta_1, \dots, \theta_n]$ 和组合权重 $\lambda = [\lambda_1, \dots,$

$\lambda_n]$) 对基础组卷积核 ($W = [W_1, \dots, W_n]$) 进行旋转变换与加权组合,生成具有方向感知能力的动态组卷积核 W' ,最后将 W' 与输入特征图进行多向卷积运算,并以逐元素的方式将输出特征图相加,最终生成旋转增强特征 $y \in \mathbf{R}^{H \times W \times C_{out}}$ 。

2 实验结果与分析

2.1 实验细节

所有实验基于 Linux 系统,使用 NVIDIA GeForce RTX 3090 GPU 进行模型演算实验,使用 Pytorch 2.5.1 深度学习框架,CUDA 版本为 12.4。输入图片的尺寸统一为 640×640 pixels,确保不同尺度目标在特征图中的清晰度。训练时,采用小批量梯度下降(stochastic gradient descent,SGD),模型的初始学习率为 0.01,学习率动量为 0.937,训练权重衰减系数为 0.000 5,训练轮次为 200 轮。为了保证模型在各类对比实验和消融实验中都具有公平的起始条件,所有实验均未使用任何形式的预训练权重,保证结果的可对比性。

2.2 数据集介绍

DOTAv1.0^[14]数据集是由武汉大学的计算机视觉实验室创建的,用于航空遥感影像的目标检测。共 15 个类别,包括 plan(C1)、ship(C2)、storage tank(C3)、baseball diamond(C4)、tennis court(C5)、baseball court(C6)、ground track field(C7)、harbor(C8)、bridge(C9)、large vehicle(C10)、small vehicle(C11)、helicopter(C12)、roundabout(C13)、soccer ball field(C14)、swimming pool(C15)。2 806 张尺寸在 800 到 20 000 的图片,共 188 282 个实例数。由于 DOTA 数据集图片尺寸大小不一,故对数据集进行分割扩充处理,以相邻图像间隔 500 pixel 将其分割成 1 024×1 024 pixels 的子图对其进行扩充,效果如图 8 所示。

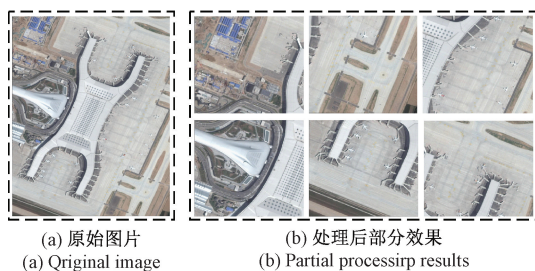


图 8 数据集分割示意图

Fig. 8 Schematic representation of the dataset segmentation

分割扩充后的数据集共 14 745 张,将其划分为 8 326 张训练集,3 345 张验证集和 3 074 张测试集。

NWPU VHR-10^[15]数据集是由西北工业大学发布的遥感图像目标检测公开数据集,共 800 张图片,10 个标签类别,分别为 airplane(C1)、ship(C2)、storage tank(C3)、baseball diamond(C4)、tennis court(C5)、basketball court(C6)、ground track field(C7)、harbor(C8)、bridge(C9)、vehicle(C10),其中拥有 650 张包含目标的图像和

150 张背景图片,共 3 775 个实例数。

RSOD^[16]数据集是一个专门用于遥感图像目标检测研究的公开数据集。共有 976 个图像数量,6 950 个实例数,4 个标签类别,分别为 aircraft、oil tank、overpass、playground。

2.3 评价指标

实验选取了 5 个主要指标来衡量模型的检测性能,包括精度(precision,P)、召回率(recall,R)、平均精确度均值(mean average precision,mAP)、模型参数量(Parameters)和浮点运算次数(giga floating-point operations per second,GFLOPs)。其中 P、R、mAP 的计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (25)$$

$$R = \frac{TP}{TP + FN} \quad (26)$$

$$mAP = \frac{1}{K} \sum_{i=1}^K AP_i \quad (27)$$

式中:TP、FP 和 FN 分别表示真阳性、假阳性和假阴性预测的数量;K 代表类别数。mAP 通过计算每个类别的 Precision-Recall 曲线下方的面积(average precision,AP),并取平均值。mAP@0.5 表示交并比(IoU)阈值为 0.5 时的 AP;mAP@0.5:0.95 则通过计算 IoU 在 0.5~0.95 范围内(步长 0.05)的 AP 均值。

2.4 对比实验与分析

1) 主干网络对比实验

为了评估双域协同的结构,作为特征提取网络的有效性,在 DOTA 数据集上,将其与 5 种经典的骨干网络做了对比实验:VanillaNet^[17]、EfficientViT^[18]、FasterNet^[19]、GhostNetV2^[20]、MobileNetV3^[21],所有对比实验是在最终模型上进行的,结果如图 9 所示。

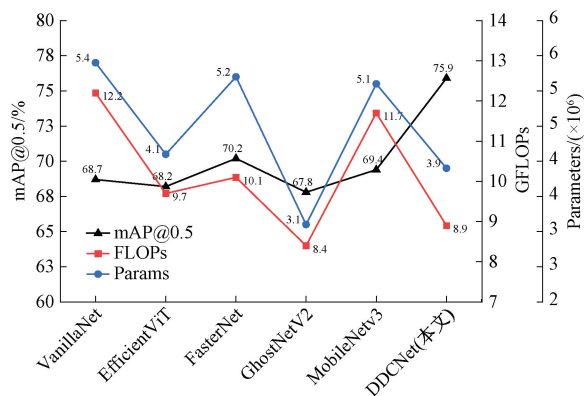


图 9 骨干网络对比实验

Fig. 9 The backbone network comparison experiment

实验结果显示,DDCNet 在目标检测基准测试中

mAP@ 0.5 达到 75.9%，相较于 VanillaNet (68.7%)、EfficientViT (68.2%) 及 FasterNet (70.2%) 分别实现了 7.2%、7.7% 与 5.7% 的性能提升，DDCNet 通过融合空间-频域特征提取机制获得了显著性能突破。其次，DDCNet 采用 CSPDarkNet53 的跨阶段部分连接结构，能在不同的网络阶段之间传递信息的同时，减少了重复梯度信息的流动，从而降低了计算冗余，提高了特征利用率，较 MobileNetv3 降低了 31.04% (Parameters) 和 31.46% (GFLOPs) 的计算需求，GhostNet 基于其轻量级的 Ghost 模块具有较低的计算成本和模型大小，但在表征小目标的细节方面存在一定的局限性，mAP@ 0.5 较低于 DDCNet。综合表明 DDCNet 不仅能够提供高精度的目标检测，还具有较高的计算效率和模型轻量化的特点。

2) 金字塔网络对比实验

为了验证多维度特征聚焦扩散金字塔网络的合理性，采用了几种典型的特征金字塔结构依次移植到本文的 DMADet 模型中进行比较，并保持相同的输入输出维度。如图 10 所示。

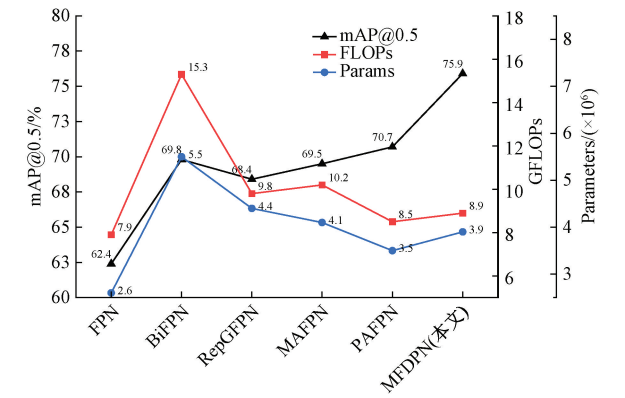


图 10 金字塔网络对比实验

Fig. 10 Pyramid network comparison experiment

MFDPN 在检测精度与计算效率的平衡性上展现出显著优势，其 mAP@ 0.5 达到 75.9%，优于其他方法。FPN^[22] 模型的计算量最小，其浮点数和参数量分别为 7.89 GFLOPs 和 2.64×10⁶，而它单向的特征融合不能有效地捕捉多尺度的遥感图像目标特征 mAP@ 0.5 为 62.4%。

尽管 PAFPN 较 FPN 增加了自底向上的特征传播路径，来融合底层的细节信息，使 mAP@ 0.5 达到了 70.7%，但其较长的传播路径容易造成信息的衰减，尤其在处理小目标时效果有限，而所提出的 MFDPN 在缩短了传播路径的同时，实现了跨维度特征的高效聚合。BiFPN^[23] 和 RepGFPN^[24] 通过更复杂的双向融合提高了网络的检测性能，但仍存在不同层次信息不能直接交互的问题，这可能导致其检测性能较为不足。与

MAFPN^[25] 相比，其通过多分支辅助融合网络来提升目标检测的性能，该方法在特征聚合后将结果分配到不同层，而所提的 MFDPN 则将多维度特征聚焦保留到中间层再分配，尤其针对微小目标检测中的低层细粒度特征进行加强聚焦，从而在 mAP@ 0.5 上提高了 6.4%。

3) 不同模型对比分析

为了验证 DMADet 在遥感图像目标检测任务中的先进性，实验选取了以轻量化为代表的 YOLO 系列算法，并在 DOTA v1.0 数据集上做对比实验分析，结果如表 2 所示。

表 2 YOLO 系列算法对比实验 (DOTA v1.0)

Table 2 Comparison experiment of YOLO series algorithms

Method	mAP@0.5/%	mAP@0.5 : 0.95/%	R/%	Params(×10 ⁶)
YOLOv3-Tiny	50.2	41.2	48.2	12.1
YOLOv5n	62.4	44.1	57.4	1.8
YOLOv6n	64.2	46.3	60.1	4.6
YOLOv7-Tiny	70.8	47.8	65.1	6.0
YOLOv8n	68.0	50.4	65.5	3.0
YOLOv10n	68.6	51.7	65.3	2.7
YOLO11n	70.3	52.3	66.4	2.59
YOLOv12n	71.4	52.8	68.3	2.6
DMADet	75.9	54.7	71.3	3.93

从实验结果中，可以看出 DMADet 相较于其他先进的 YOLO 系列算法在 mAP@ 0.5、mAP@ 0.5 : 0.95 和 R 指标上都优于其他算法模型。从 R 指标具体来看，DMADet 达到了 71.3%，相较于最新版本的 YOLO11n 和 YOLO12n 分别高了 4.9% 和 3%，这表明 DMADet 具有有效的多尺度特征表达能力，以及检测头的调整和注意力机制的引入，使其在遥感图像的有向性目标检测方面具有显著优势。相较而言，召回率最低的是 YOLOv3-Tiny，为 48.2%，其只具有两个尺寸的检测头，限制了在不同尺度目标检测上的表现，导致了严重的漏检问题。在 mAP@ 0.5 和 mAP@ 0.5 : 0.95 指标上，DMADet 保持了同样的优势，分别达到了 75.9% 和 54.7%，相较于紧随其后的 YOLOv12 模型分别提升了 4.5% 和 1.9%，由于 YOLOv3-Tiny 网络结构的简化，在两指标上表现较差仅为 50.2% 和 41.2%。YOLOv5n 牺牲了一定程度的精度以换取更快的速度和更低的计算复杂度，参数量仅为 1.8×10⁶，但 mAP@ 0.5 和 mAP@ 0.5 : 0.95 较 DMADet 低了 13.5% 和 10.6%。

此外，在与其他针对遥感图像目标检测的算法模型对比时，DMADet 在多个指标上依然保持着一定的优势，如表 3 所示。

具体来说，YOLO-CLSA、TPH-YOLO 和 PARDet 在对目标框的定位和尺寸检测上有所改进，但其在 mAP@ 0.5 : 0.95 指标上较低，分别为 40.2%、50.5% 和 43.05%，较 DMADet 分别降低了 14.0%、3.7% 和

表 3 其他改进算法在 DOTAv1.0 上的对比实验

Table 3 Comparative experiments of other improved algorithms on DOTAv1.0 (%)				
Method	mAP@0.5	mAP@0.5 : 0.95	P	R
YOLO-CLSA ^[26]	69.0	40.2	72.9	63.3
FR+CGL ^[27]	73.2	—	—	—
RA2DC-Net ^[28]	74.5	—	—	—
TPH-YOLO ^[29]	73.1	50.5	—	—
PARDet ^[30]	73.08	43.05	—	—
DMADet	75.9	54.7	80.4	71.3

11.15%,MFDPN 通过特征聚焦与扩散的方式,增加了多维度特征间的直接交互,因此在面对多尺度目标时,DMADet 具有更强的多尺度特征表达能力。RA2DC-Net 使用残差结构和注意力机制来增强特征焦点信息,尽管 mAP@0.5 达到了 74.5%,但仍然不及 DMADet 的 75.9%,这主要得益于其对于边缘细节特征的信息挖掘,使得在小目标的信息处理上优于前者。FR+CGL 基于其

耦合全局-局部(coupled global-local,CGL)网络来捕获更多的有效信息,mAP@0.5 达到了 73.2%,但其忽视了对频域的细节信息处理,较本文的 DMADet 降低了 2.7%。

为了验证所提出的算法模型在不同目标物体上的识别与分类效果,实验列出了不同先进算法模型在 DOTAv1.0 数据集上的各类别检测结果,如表 4 所示,DMADet 的平均 mAP@0.5 均优于其他算法模型,针对 plan(C1)、ship(C2)、storage tank(C3)等此类小目标的检测效果上,本文提出的 DMADet 算法模型均展现了最优的结果。值得注意的是,DMADet 在 bridge(C9)、large vehicle(C10)、roundabout(C13)、swimming pool(C15)这 4 类目标的检测效果上依然保持着显著的优势,而其目标存在复杂的背景以及多尺度的特点,尤为考验算法模型在局部、全局语义信息以及边缘细节信息的处理能力,这得益于 DMADet 算法模型的双域协同特征提取网络与多维度聚焦扩散金字塔在此方面具有的突出性能表现。

表 4 不同算法模型在 DOTAv1.0 上的各类别检测对比

Table 4 Comparison of different algorithm models in DOTAv1.0 for various categories detection (%)																
Method	mAP@0.5															mAP@0.5
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	(all)
YOLOv5	83.3	87.4	73.2	74.3	91.0	48.6	53.2	81.4	51.3	80.7	54.6	32.4	58.8	38.7	78.9	62.4
YOLOv7-Tiny	88.4	84.2	75.3	75.1	90.0	48.8	54.1	83.4	52.8	83.7	55.6	32.6	62.7	39.6	79.9	70.8
YOLOv8n	93.0	89.4	75.9	80.2	91.7	49.6	55.7	81.4	52.5	85.8	57.6	32.6	65.4	39.8	81.1	68.8
FDLR-Net ^[31]	89.0	77.9	86.3	79.1	89.4	86.7	68.6	66.9	52.1	80.8	72.1	57.9	61.4	64.8	68.6	73.0
AFPN-DFA ^[32]	88.7	87.4	83.6	79.4	90.7	84.3	65.5	66.9	52.2	80.8	74.7	69.3	58.1	62.6	72.3	74.4
YOLOv10	90.3	89.4	74.9	80.4	91.5	50.7	55.3	81.3	52.4	84.4	57.0	33.4	64.8	40.1	80.0	68.6
S-RepPoints ^[33]	86.4	86.7	85.6	78.9	90.8	82.6	69.8	73.9	52.4	75.9	77.3	62.3	68.2	60.1	72.2	74.9
YOLO11	92.3	89.3	74.8	80.0	91.3	49.7	54.3	81.2	51.4	85.3	57.4	32.3	61.4	39.9	80.6	70.3
CFA ^[34]	89.2	84.4	85.5	81.7	90.7	83.4	67.1	73.0	51.8	78.2	79.9	64.9	67.7	54.8	70.2	75.0
S2A-Net ^[35]	89.1	86.9	85.5	81.5	90.8	83.5	72.8	66.5	48.7	76.7	78.2	56.2	61.6	62.7	68.9	74.0
GI-Vertex ^[36]	89.6	86.8	86.8	85.0	90.7	79.0	77.3	72.9	52.2	73.1	73.0	57.3	70.9	59.5	70.8	75.0
DMADet	93.1	89.8	86.9	83.2	91.2	50.1	54.3	82.1	52.9	86.2	58.3	40.6	72.2	43.8	85.7	75.9

为了进一步验证 DMADet 在不同目标物体上识别与分类的性能表现,将其与一些针对遥感图像的代表性的目标检测算法和改进的算法模型在 NWPU VHR-10 数据集上进行泛化性评估。如表 5 所示,DMADet 在 NWPU VHR-10 数据集上依旧保持着一定的优势,所有类别的 mAP@0.5 值达到了 96.8%,相对于位居第二的 GLSANet 和 DA2FNet 提高了 2.2%。DMADet 在多个类别上依旧保持着优越的检测性能,其中,针对 airplane(C1)的小目标检测上,mAP@0.5 达到了 100%。在多尺度目标的表现上,storage tank(C3),baseball diamond(C4),tennis court(C5),basketball court(C6)4 类目标相对于其他算法模型达到了最优。从 DOTAv1.0 数据集上的各类别实验以及 NWPU VHR-10 数据集的泛化性实验上分析,DMADet 在小目标以及背景复杂的多尺度目标的检测结果上都凸显出其算法模型的优越性能,证明了改进算法

模型的有效性。

4) 消融实验与分析

为了验证 DDCNet、MFDPN、LDC-Attention 和自适应旋转感知增强检测头的有效性,在 DOTAv1.0 数据集上进行了消融实验。为了消融实验的严谨性,将各个算法模型部署在 YOLOv8 模型的基础上进行消融实验验证。如图 11 所示,从实验曲线可以观察到,随着网络模块的逐步叠加,模型性能呈现稳定提升的趋势,并均达到收敛状态。

基准模型 YOLOv8 在 200 轮次 mAP@0.5 达到 68.4%,引入 DDCNet 模块后提升至 71.0%(+2.6%),得益于其空间-频域的双向交互既聚焦于空间域特征的局部细节刻画,又强化了频域特征的全局语义理解。随后,叠加 MFDPN 模块保持同等性能但收敛曲线更加平滑,说明多尺度特征增强了训练稳定性。LDC-Attention 模块

表 5 不同算法模型在 NWPU VHR-10 上的各类别检测对比
Table 5 Comparison of different algorithm models on NWPU VHR-10 (%)

Method	mAP@ 0. 5%										mAP@ 0. 50
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	(all)
YOLOv7	99. 5	80. 4	93. 5	99. 4	98. 2	92. 9	95. 4	89. 6	81. 1	87. 6	91. 8
GLSANet ^[37]	99. 9	95. 8	97. 1	99. 4	98. 8	86. 1	99. 5	97. 5	84. 8	86. 8	94. 6
YOLOv8	99. 1	87. 8	98. 2	97. 3	95. 8	94. 8	97. 3	92. 1	93. 4	94. 0	93. 2
DMNet ^[38]	99. 3	88. 5	87. 8	97. 9	96. 2	94. 5	98. 5	98. 7	88. 2	92. 6	94. 2
DA ² FNet ^[39]	99. 9	89. 5	87. 2	98. 7	95. 7	95. 2	99. 5	99. 0	89. 0	93. 6	94. 6
EfficientDet ^[40]	95. 1	68. 8	74. 7	81. 5	72. 1	83. 3	79. 9	85. 6	91. 1	83. 2	81. 5
HSP-Net ^[41]	99. 8	92. 5	97. 0	98. 6	90. 4	91. 5	99. 0	88. 9	87. 1	89. 1	93. 4
DMADet	100. 0	88. 1	99. 5	99. 7	98. 9	96. 7	98. 7	98. 3	97. 7	98. 2	96. 8

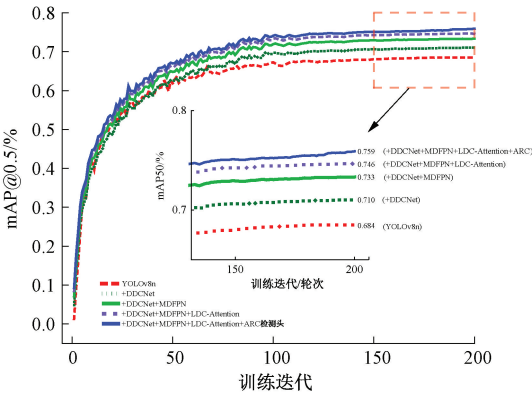


图 11 消融实验迭代
Fig. 11 Ablation experiment iteration

的加入进一步提升了 1.3% mAP@ 0. 5 至 74. 6%,最后附加自适应旋转感知增强检测头使模型达到最优性能 75.9% mAP@ 0. 5。各模块贡献具有累加特性,完整模型

的 mAP@ 0. 5 相比基准提升了 7.9%。总体来看,每个模块的引入都逐步提升了模型的性能,显示出各个模块的有效性。

此外,针对 LDC-Attention 注意力机制的多尺度表现,实验引入了 COCO 分层评价指标:小目标 (average precision for small objects, APS, 面积<32×32 pixels)、中目标 (average precision for large objects, APM, 32×32 ~ 96×96 pixels) 和大目标 (APL,>96×96 pixels) 的平均精度,以量化 LDC-Attention 对不同尺寸物体的关注程度,如表 6 所示,LDC-Attention 添加前后的效果,从表 6 可以看出,该模块显著提高了遥感图像中小目标检测的精度。

此外,增加了在网络特征融合阶段从上到下加入 LDC-Attention 后的特征响应的可视化图,如图 12 所示,颜色温度越高,表示网络对目标特征的关注度越高。

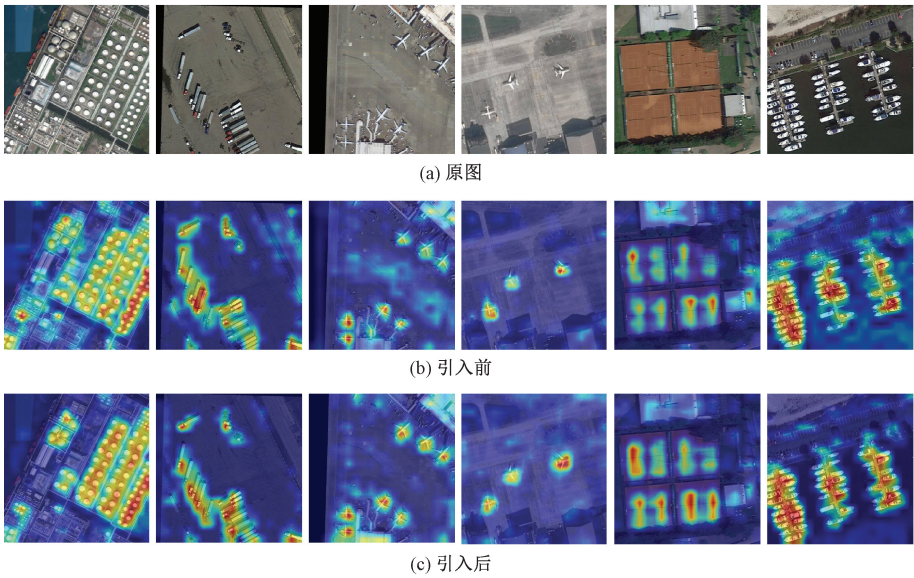


图 12 引入 LDC-Attention 前后热力对比

Fig. 12 A comparison of thermal maps before and after the introduction of LDC-Attention

表 6 注意力机制消融对比

Table 6 Comparison of attention mechanism ablation

模型	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	Parameters($\times 10^6$)	APS/%	APM/%	APL/%
未加入 LDC-Attention	73.3	48.3	9.1	3.48	45.3	64.2	92.3
加入 LDC-Attention	74.6	54.2	9.73	3.83	47.2	65.4	92.4

图 12(a) 为未加入 LDC-Attention 注意力机制,其中的热力图能量偏低且分布较为分散,表达效果较差,这使得在遥感图像复杂场景中难以准确捕捉目标区域。图 12(b) 为加入 LDC-Attention 后模型的可视化结果,针对密集小目标与遮挡物体的信息的表达能力更强,模型对目标特征的关注也更加精准。

5) 可视化分析与泛化性测试

为了直观地展示 DMADet 模型的优势,将其与最新的 YOLOv12 在 DOTA v1.0 数据集上测试了一些具有挑战性的遥感场景图像,并提供了相应的检测结果。如图 13 所示,图 13(a)、(c)、(e) 为 DMADet 的检测结果,图 13(b)、(d)、(f) 为 YOLOv12 的检测结果。

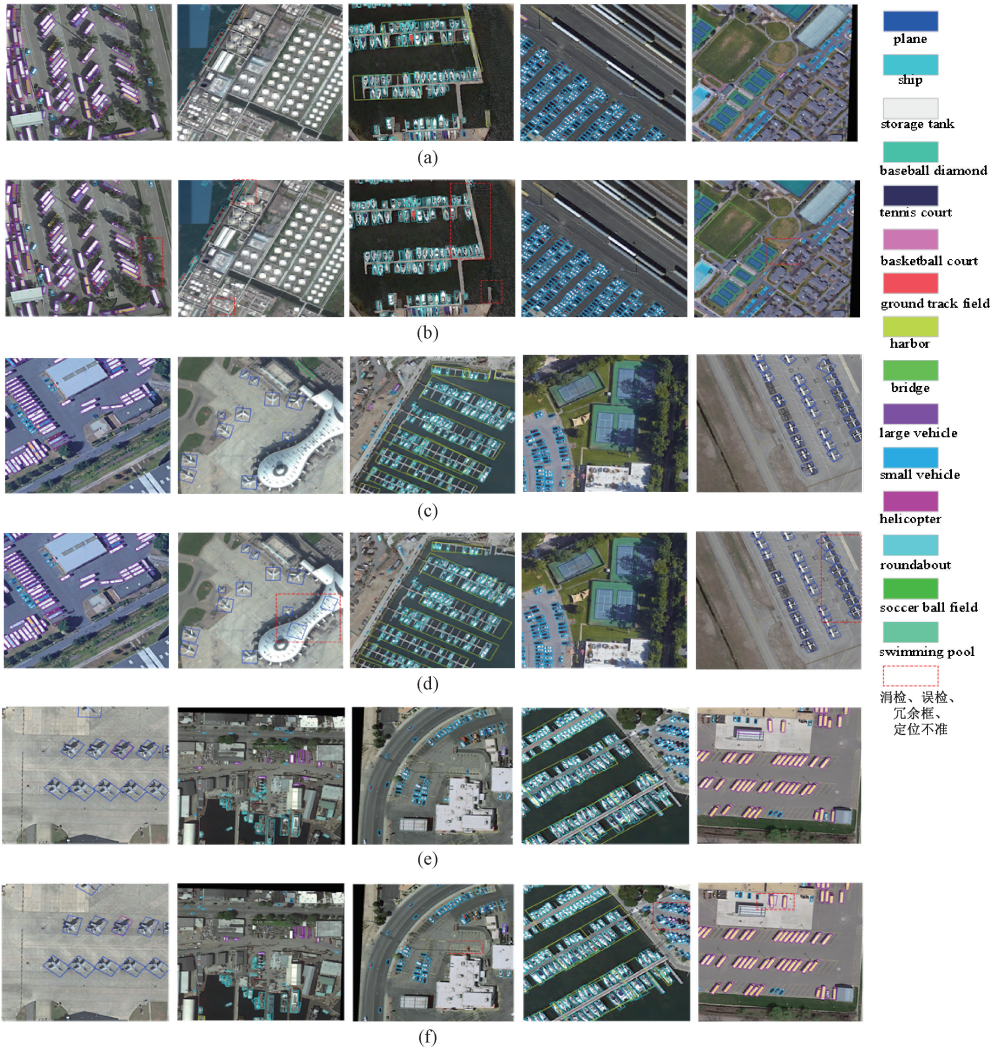


图 13 DOTA v1.0 检测结果对比

Fig. 13 DOTA v1. Comparison of detection results

从结果来看,DMADet 在结构化目标(如“bridge”、“helicopter”)的定位精度较高,其标注框更贴合物体几何轮廓,且对相似色系地物(如“swimming pool”与“basketball court”)具有更强的区分能力。YOLOv12 相较于 DMADet 更加轻量化,但对小尺度目标(如“small vehicle”)存在漏检现象,且在复杂纹理区域(如“ground

track field”)易产生重叠框误判。DMADet 在复杂场景下的检测能力更强,能够提供更详细和准确的检测结果,而 YOLOv12 因其轻量化的结构,丢失了部分有效的细节信息,在遮挡和小目标的检测定位上逊色于 DMADet。为了进一步验证 DMADet 在不同遥感图像数据集上的鲁棒性,本文还在 RSOD 数据集上进行了泛化实验分

析,并选取了一些典型的场景图像并提供了相应的检测结果如图 14 所示。

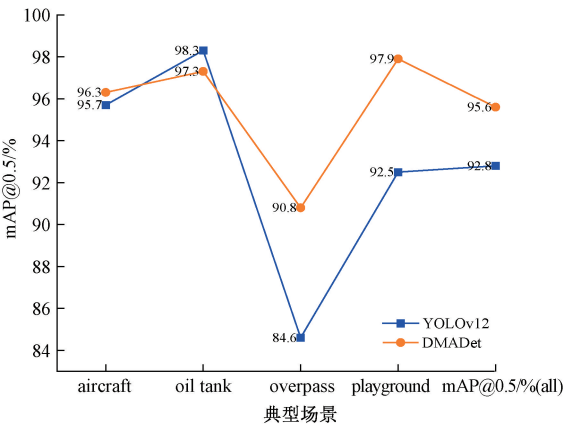


图 14 RSOD 各类别检测性能对比

Fig. 14 Comparison of various types of tests

其中 DMADet 的类别平均 mAP@0.5 比 YOLOv12 增加了 2.8%,在泛化性上依然保持着优越的性能,验证了 DMADet 的鲁棒性。具体来看,虽然 DMADet 在 oil tank 类别上比 YOLOv12 略低 1.0%,但在 aircraft、overpass 以及 playground 类别上保持领先,表明其对复杂空间结构的特征捕捉更为有效。

其次,如图 15 所示,可以更加直观地展示 DMADet 的鲁棒性能。图 15(a)、(c)、(e)为 DMADet 的检测结果,图 15(b)、(d)、(f)为 YOLOv12 的检测结果。从图 15(a)、(b)来看,本文提出的 DMADet 模型在密集小目标检测中展现出显著优势,虽然存在个别误检情况,但针对大部分的密集小目标依旧能够精准定位,而 YOLOv12 存在的漏检和误检情况较为严重。在背景复杂的田径场和操场等场景的检测中,图 15(e)和(f)验证了 DMADet 对多尺度场景语义关联特征的学习能力,能



图 15 RSOD 检测结果对比

Fig. 15 Comparison of the RSOD test results

够更准确地定位和识别这些目标,边界框更为贴合实际物体,且误检率和漏检率较低,进一步验证了DMADet的泛化性能。

3 结 论

DMADet算法通过引入DDCNet,有效整合了边缘细节与全局频域表征,显著提升了小目标特征提取能力。设计的MFDPN通过跨层特征交互与LDC-Attention,增强了多尺度特征的关联性和信息传递效率。ARC采用动态路由策略,显著提升了模型对旋转目标几何特征的捕捉能力,最终使模型在DOTAv1.0数据集上的mAP@0.5达到了75.9%,相较于最新的YOLOv12算法提升了4.5%。尽管DMADet在遥感图像目标检测任务中展现了卓越性能,但该算法仍存在一定的局限性:频域特征提取模块的计算复杂度较高,可能影响实时性需求;对于极端尺度目标(如超小密集目标)的检测精度仍有提升空间。未来的工作将致力于探索轻量化的频域特征压缩方法以及优化多尺度特征融合策略,以期推动遥感目标检测技术向更高效、更普适的方向发展。

参考文献

- [1] FANG J, JIANG H, ZHANG S, et al. BAF-Net: Bidirectional attention fusion network via CNN and transformers for the pepper leaf segmentation [J]. *Frontiers in Plant Science*, 2023, 14: 1123410.
- [2] GAO T, NIU Q, ZHANG J, et al. Global to local: A scaleaware network for remote sensing object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-14.
- [3] JI J Y, ZHAO Y F, LI A H, et al. Dense small object detection algorithm for unmanned aerial vehicle remote sensing images in complex backgrounds [J]. *Digital Signal Processing*, 2025, 158: 104938.
- [4] ZHAO Z H, LIU X Y, HE P. PSO-YOLO: A contextual feature enhancement method for small object detection in UAV aerial images [J]. *Earth Science Informatics*, 2025, 18(2): 258.
- [5] WANG L, SHI Y, MAO G J, et al. Consumer-centric insights into resilient small object detection: Sciou loss and recursive transformer network [J]. *IEEE Transactions on Consumer Electronics*, 2023, 70(1): 2178-2187.
- [6] GAO T, XIA S L, LIU M K, et al. Msnet: Multi-scale network for object detection in remote sensing images[J]. *Pattern Recognition*, 2025, 158: 110983.
- [7] JIANG L J, YUAN B X, DU J W, et al. Mffsodnet: Multi-scale feature fusion small object detection network for UAV aerial images [J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 1-14.
- [8] 王迎龙, 孙备, 丁冰, 等. BG-YOLO: 复杂大视场下低慢小无人机目标检测方法[J]. *仪器仪表学报*, 2025, 46(2): 255-266.
WANG Y L, SUN B, DING B, et al. BG-YOLO: A method for target detection in small, slow, and low-performance drones under complex wide field of view[J]. *Chinese Journal of Scientific Instrument*, 2025, 46(2): 255-266.
- [9] 冉宁, 施高朗, 张少康, 等. 基于YOLOv8的遥感小目标检测算法[J]. *电子测量与仪器学报*, 2025, 39(5): 197-207.
RAN N, SHI G L, ZHANG SH K, et al. YOLOv8-based remote sensing small object detection algorithm [J]. *Journal of Electronic Measurement and Instrumentation*, 2025, 39(5): 197-207.
- [10] 谢星星, 程堃, 姚艳清, 等. 动态特征融合的遥感图像目标检测[J]. *计算机学报*, 2022, 45(4): 735-747.
XIE X X, CHENG S, YAO Y Q, et al. Target detection in remote sensing images with dynamic feature fusion [J]. *Journal of Computer Science*, 2022, 45(4): 735-747.
- [11] LI D, LIANG Q H, LIU H Q, et al. A novel multidimensional domain deep learning network for SAR ship detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-13.
- [12] HAN K, WANG Y H, TIAN Q, et al. Ghostnet: More features from cheap operations [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 1580-1589.
- [13] PU Y F, WANG Y R, XIA Z F, et al. Adaptive rotated convolution for rotated object detection[C]. *IEEE/CVF International Conference on Computer Vision*, 2023: 6589-6600.
- [14] DING J, XUE N, XIA G S, et al. Object detection in aerial images: A large-scale benchmark and challenges[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(11): 7778-7796.
- [15] CHENG G, ZHOU P C, HAN J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(12): 7405-7415.
- [16] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images: A survey and a new benchmark [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 159: 296-307.

- [17] CHEN H T, WANG Y H, GUO J Y, et al. Vanillanet: The power of minimalism in deep learning[J]. Advances in Neural Information Processing Systems, 2023, 36: 7050-7064.
- [18] LIU X Y, PENG H W, ZHENG N X, et al. Efficientvit: Memory efficient vision transformer with cascaded group attention [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 14420-14430.
- [19] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [20] TANG Y H, HAN K, GUO J Y, et al. GhostNetv2: Enhance cheap operation with long-range attention[J]. Advances in Neural Information Processing Systems, 2022, 35: 9969-9982.
- [21] KOONCE B, KOONCE B. MobileNetV3 [J]. Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization, 2021: 125-144.
- [22] XU H, YAO L W, ZHANG W, et al. Auto-fpn: Automatic network architecture adaptation for object detection beyond classification [C]. IEEE/CVF International Conference on Computer Vision, 2019: 6649-6658.
- [23] CHEN J, MAI H S, LUO L B, et al. Effective feature fusion network in BIFPN for small object detection[C]. 2021 IEEE International Conference on Image Processing (ICIP). IEEE, 2021: 699-703.
- [24] XU X ZH, JIANG Y Q, CHEN W H, et al. Damo-yolo: A report on real-time object detection design[J]. arXiv preprint arXiv:2211.15444, 2022.
- [25] YANG Z Q, GUAN Q, ZHAO K, et al. Multi-branch auxiliary fusion YOLO with re-parameterization heterogeneous convolutional for accurate object detection [C]. Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Singapore: Springer Nature Singapore, 2024: 492-505.
- [26] CHAURASIA D, PATRO B D K. Detection of objects in satellite and aerial imagery using channel and spatially attentive YOLO-CSL for surveillance [J]. Image and Vision Computing, 2024, 147: 105070.
- [27] CHEN X, WANG C J, LI Z H, et al. Coupled Global-Local object detection for large VHR aerial images[J]. Knowledge-Based Systems, 2023, 260: 110097.
- [28] GAO F, CAI C X, TANG W T, et al. RA2DC-Net: A residual augment-convolutions and adaptive deformable convolution for points-based anchor-free orientation detection network in remote sensing images[J]. Expert Systems with Applications, 2024, 238: 122299.
- [29] 高卫峰, 易宇轩, 黄玲玲, 等. 一种高效的无人机航拍小目标检测算法[J]. 控制与决策, 2025, 40(8): 1-9.
GAO W F, YI Y X, HUANG L L, et al. An efficient UAV aerial photography small target detection algorithm [J]. Control and Decision, 2025, 40(8): 1-9.
- [30] XU Y H, SHEN J F, DAI M, et al. PARDet: Dynamic point set alignment for rotated object detection [J]. Pattern Recognition, 2024, 153: 110534.
- [31] XIAO J S, YAO Y T, ZHOU J, et al. FDLR-Net: A feature decoupling and localization refinement network for object detection in remote sensing images [J]. Expert Systems with Applications, 2023, 225: 120068-120069.
- [32] HOU L P, LU K, XUE J. Refined one-stage oriented object detection method for remote sensing images[J]. IEEE Transactions on Image Processing, 2022, 31: 1545-1558.
- [33] HOU L, LU K, XUE J, et al. Shape-adaptive selection and measurement for oriented object detection[C]. AAAI Conference on Artificial Intelligence, 2022, 36(1): 923-932.
- [34] GUO Z H, LIU C, ZHANG X S, et al. Beyond bounding-box: Convex-hull feature adaptation for oriented and densely packed object detection[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8792-8801.
- [35] HAN J M, DING J, LI J, et al. Align Deep Features for Oriented Object Detection [C]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-11.
- [36] XU Y C, FU M T, WANG Q M, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(4): 1452-1459.
- [37] HAN W, LI J, WANG S, et al. A context-scale-aware detector and a new benchmark for remote sensing small weak object detection in unmanned aerial vehicle images [J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 112: 102966.
- [38] LI C L, YANG T J N, ZHU S J, et al. Density map guided object detection in aerial images [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 190-191.
- [39] GUO Y Y, TONG X H, XU X, et al. An anchor-free network with density map and attention mechanism for multiscale object detection in aerial images [J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.

[40] TAN M X, PANG R M, LE Q V. Efficientdet: Scalable and efficient object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.

[41] XU C Y, LI C Z, CUI Z, et al. Hierarchical semantic propagation for object detection in remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(6): 4353-4364.

作者简介



周冬梅(通信作者),1999 年于兰州交通大学获得学士学位,2010 年于兰州交通大学获得硕士学位,现为兰州交通大学副教授,主要研究方向为通信与信息系统、人工智能等。

E-mail: zhoudongmeitd@mail.lzjtu.cn

Zhou Dongmei(Corresponding author) received her B. Sc. degree from Lanzhou Jiaotong University in 1999, and M. Sc. degree from Lanzhou Jiaotong University in 2010. Now she is an associate professor at Lanzhou Jiaotong University. Her main research interests include communication and information systems, and artificial intelligence.



王飞成,2023 年于兰州财经大学获得学士学位,现为兰州交通大学硕士研究生,主要研究方向为深度学习目标检测。

E-mail: 1272767280@qq.com

Wang Feicheng received his B. Sc. degree from Lanzhou University of Finance and Economics in 2023. Now he is a M. Sc. candidate at Lanzhou Jiaotong University. His main research interests include deep learning and object detection.