



基于提案增强的解耦特征挖掘旋转检测器

赵振博, 付天怡, 董红斌, 张小平

引用本文:

赵振博, 付天怡, 董红斌, 等. 基于提案增强的解耦特征挖掘旋转检测器[J]. *智能系统学报*, 2025, 20(5): 1123-1135.

ZHAO Zhenbo, FU Tianyi, DONG Hongbin, et al. Decoupled feature mining rotational detector based on proposal enhancement[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(5): 1123-1135.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202410017>

您可能感兴趣的其他文章

改进Faster R-CNN的汽车仪表指针实时检测

Improved Faster R-CNN vehicle instrument pointer real-time detection algorithm
智能系统学报. 2021, 16(6): 1056-1063 <https://dx.doi.org/10.11992/tis.202011003>

舰载机位姿实时视觉测量算法研究

Research on real-time vision measurement algorithm of shipborne aircraft pose
智能系统学报. 2021, 16(6): 1045-1055 <https://dx.doi.org/10.11992/tis.202103014>

基于改进FCOS的拥挤行人检测算法

Crowded pedestrian detection algorithm based on improved FCOS
智能系统学报. 2021, 16(4): 811-818 <https://dx.doi.org/10.11992/tis.202010012>

采用相关滤波的水下海参目标跟踪

Underwater sea cucumber target tracking algorithm based on correlation filtering
智能系统学报. 2019, 14(3): 525-532 <https://dx.doi.org/10.11992/tis.201711037>

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.202410017

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20250624.1222.002>

基于提案增强的解耦特征挖掘旋转检测器

赵振博¹, 付天怡¹, 董红斌¹, 张小平²

(1. 哈尔滨工程大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001; 2. 中国中医科学院 中医药数据中心, 北京 100700)

摘要: 小而杂乱的物体交织在一起, 在遥感图像中尤为常见, 给目标检测带来了巨大挑战。在旋转目标检测任务中这个困难更加突出。鉴于此, 本文提出了基于提案增强的解耦特征挖掘旋转检测器 (decoupled feature mining rotational detector based on proposal enhancement, PDMDet)。首先, 采用单阶段检测器取代两阶段检测器的提案生成网络, 通过生成高质量提案以减少背景冗余。其次, 在相同维度使用自注意力, 不同维度使用交叉注意力, 通过对相同维度特征增强, 不同维度特征交错融合提升检测器对不同尺寸目标的识别能力。最后, 鉴于分类和定向边界框回归任务对特征的敏感性不同, 本文提出解耦特征细化处理两个不同任务。通过实验, PDMDet 在 DOTA-v1.0、DOTA-v1.5 和 HRSC2016 这 3 个数据集上分别取得单尺度 78.37%、72.35%、98.60% 的平均精度均值, 检测准确率高于其他算法, 在复杂的旋转目标检测场景具有一定的竞争力。

关键词: 遥感图像; 目标检测; 旋转目标检测; 小目标; 高密度; 倾斜边界框; 跨尺度融合; 两阶段检测器

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2025)05-1123-13

中文引用格式: 赵振博, 付天怡, 董红斌, 等. 基于提案增强的解耦特征挖掘旋转检测器 [J]. 智能系统学报, 2025, 20(5): 1123-1135.

英文引用格式: ZHAO Zhenbo, FU Tianyi, DONG Hongbin, et al. Decoupled feature mining rotational detector based on proposal enhancement[J]. CAAI transactions on intelligent systems, 2025, 20(5): 1123-1135.

Decoupled feature mining rotational detector based on proposal enhancement

ZHAO Zhenbo¹, FU Tianyi¹, DONG Hongbin¹, ZHANG Xiaoping²

(1. School of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China; 2. Traditional Chinese Medicine Data Center, China Academy of Chinese Medical Sciences, Beijing 100700, China)

Abstract: In remote sensing images, small and cluttered objects often appear intertwined, presenting considerable challenges for object detection. These challenges are even further amplified in rotational object detection tasks. Aiming to address these challenges, this paper proposes a decoupled feature mining rotational detector based on proposal enhancement (PDMDet). First, a single-stage detector is employed to replace the region proposal network of the two-stage detector, generating high-quality proposals to reduce background redundancy. Second, self-attention is applied within the same feature dimensions and cross-attention across different dimensions, aiming to enhance intradimensional features and fuse interdimensional features, thereby improving the capability of the detector to identify objects of varying sizes. Finally, recognizing that classification and oriented bounding box regression tasks have different feature sensitivities, this paper proposes a decoupled feature refinement strategy that processes the two tasks separately. Experiment results demonstrate that PDMDet achieves single-scale mAP scores of 78.37%, 72.35%, and 98.60% on DOTA-v1.0, DOTA-v1.5, and HRSC2016 datasets, respectively, outperforming existing algorithms in terms of detection accuracy and demonstrating strong competitiveness in complex rotational object detection scenarios.

Keywords: remote sensing image; object detection; rotated object detection; small object; high density; oriented bounding box; cross-scale fusion; two-stage detector

收稿日期: 2024-10-24. 网络出版日期: 2025-06-24.

基金项目: 国家自然科学基金项目 (82374621).

通信作者: 董红斌. E-mail: donghongbin@hrbeu.edu.cn.

©《智能系统学报》编辑部版权所有

目标检测是计算机视觉中一项基本任务, 近些年随着深度学习的发展, 各种通用检测器层出不穷。其中遥感检测因其在智慧农业、智慧交

通、军事国防等领域广泛应用而备受关注。人类已经进入智能化、现代化时代。交通中车辆、船舶数量急剧增加,且遥感图像是从高空视角拍摄,这导致遥感目标检测有更加突出的难点:

1) 目标较小且排列密集。航拍图像中包含大量的被复杂周围场景淹没的小物体。检测小目标存在两个主要障碍,分别是物体信息不足和锚样本不足。原因是网络对特征进行多层池化,小目标物体在深层网络中信息丢失严重;同时高层特征图的较大采样步频往往会直接跳过小目标导致特征对小目标的表达能力下降。

2) 物体尺寸差异较大。航拍图像另一特征是物体尺寸变化大,长宽比较大的物体对于角度更加敏感,对于两个细长形检测框,旋转较小的角度就会导致交并比的巨大差异进而使损失变大。而类似圆或方形物体对于角度并不敏感。

3) 方向任意。航拍图像中的物体呈现不同方向,且小目标和尺寸变化进一步增加了检测的难度。此外,现有大多数检测器将分类和回归任务耦合进行。水平框检测任务的效果尚可,但旋转目标检测的回归分支需要提取对角度敏感的旋转特征,而耦合的方式严重限制了旋转检测器的精度。

遥感目标检测质量直接影响到经济和军事发展^[1],人们对遥感检测器的性能要求越来越高,因此寻找精度高可靠性强的遥感目标检测器尤为重要。

现有的检测器包括 Fast RCNN(fast region based convolutional neural network)^[2]、YOLO(you only look once)^[3]和 SSD(single shot multibox detector)^[4]等,这些通用的检测器都是基于水平检测框构建的。在进行遥感图像检测时,小而杂乱的目标交织在一起,水平框检测器会产生重合度很高的候选框。较高的边框重叠率会导致检测框的错误消除^[5],这种冲突在非极大抑制(non maximum suppression, NMS)中尤为突出。此外,遥感目标检测中物体的朝向包含着重要的语义信息^[6],使用旋转框能细粒度提取遥感图像中的信息。

然而,尽管取得了如此多的进展,仅有较少的工作考虑到了遥感图像中提案质量的重要性以及回归分支中旋转等变性的提取。鉴于此,本文提出基于提案增强的解耦特征挖掘旋转检测器(decoupled feature mining rotational detector based on proposal enhancement, PDMDet),包含以下 3 个部分:

1) 提案增强网络(proposal enhancement network, PEN)。使用一阶段无锚检测器 FCOS(fully

convolutional one-stage object detection)^[7]替换两阶段检测器的 RPN(region proposal network)生成高质量提案。通过高质量提案,感兴趣区域(region of interest, ROI)将特征与目标更好对齐,有效解决了物体尺寸变化较大的问题。

2) 多维特征挖掘网络(multi-level feature mining network, MFMN)。通过相同维度使用自注意力、不同维度使用交叉注意力,增强旋转信息,去除模糊噪声,增强网络对小目标的表达能力。

3) 解耦特征细化网络(decoupled feature refinement network, DFRN)。引入特征细化解耦方法,通过单独的网络分支来解决分类和回归任务的不一致性,且解耦方法具有一定的通用性能能够与不同的检测器集成。

在 DOTA-v1.0、DOTA-v1.5 和 HRSC2016 这 3 个公开数据集上进行的实验验证了所提模型的有效性。

1 相关工作

1.1 旋转目标检测框架

许多现有的旋转目标检测方法都是基于水平框的改进。Lin 等^[8]提出的 RetinaNet 使用一个主干特征金字塔网络连接两个子网络,一个用于分类,一个用于回归旋转框。Tian 等^[7]提出 FCOS 一种无锚点全卷积单阶段检测器,避免了锚点框之间的复杂计算。Yang 等^[9]使用一种完全基于点表示的检测器 RepPoint(point set representation for object detection),通过自适应点表示进而捕获任意方向的集合信息。此外,高效率的 YOLO^[3]系列把图片分为不同的子区域,进而在每个区域内部进行目标检测。以上经典的目标检测框架,通过添加角度回归的分支实现旋转框目标检测。

现阶段高准确率目标检测多依赖于 RCNN(region based convolutional neural network)^[2]框架,该框架由 RPN、ROI 和不同种类检测头组成。Ding 等^[10]提出一种两阶段检测器 RoI Transformer,在感兴趣区域进行空间变换,并且在水平框的监督下学习变换参数。Yang 等^[11]的 SCRDet (multicategory rotation detector for small, cluttered and rotated objects)使用注意力机制减弱图像的背景噪声,使得模型适用于密集小目标的检测。Han 等^[12]提出的 S²Anet(single shot alignment network)作为一种单样本对齐网络通过生成高质量锚框和滤波器对位置信息进行编码,产生旋转不变性。R3Det(refined rotation retinanet)^[13]采用渐进回归的方法,从

粗到细粒度回归旋转边界框。O-RCNN(oriented RCNN)^[14] 是一个更为通用的两阶段旋转检测器, 第一阶段面向 ROI 以极少的代价生成高质量提案, 第二阶段是 RCNN 旋转头部用于细化 ROI 进行目标的识别与检测框的回归。除此之外, AO2-DETR(arbitrary oriented object detection Transformer)^[15] 将 Transformer 应用到遥感图像, 增加了模型多样性。R3DetGauss(refined rotated RetinaNet detector)^[16] 将高斯 Wasserstein 距离作为回归损失来进一步提高估计边界框的质量。基于椭圆方程的遥感有向目标检测模型(elliptical equation based remote sensing oriented object detection network, EllipticNet)^[17] 将椭圆方程应用到旋转目标检测, 克服五参数旋转框表示法的边界不连续问题。

1.2 旋转目标检测中的旋转不变性和旋转等变性

旋转不变性和旋转等变性一直是旋转目标检测中具有挑战性问题^[18]。旋转不变性是指图像旋转过程中语义信息保持不变, 旋转等变性是指图像旋转过程中物体空间位置和方向发生变化。目前解决上述问题大多是基于设计旋转敏感性卷积网络^[12], 例如 ReDet(rotation equivariant detector)^[19] 将旋转等变网络融入到检测器中以提取等变特征, 准确预测方向变化。近期还有工作使用弱监督来解决旋转不变性和等变性问题。例如 RINet(rotation invariant aerial object detection network)^[20] 使用灵活的多分支检测器进行特征细化, 进而使模型对定向对象具有更强的旋转感知能力。方婷婷等^[21] 提出视图一致性网络下的弱监督遥感影像旋转目标检测模型, 通过利用不同旋转视图下的旋转检测框的空间位置一致性约束来调整角度预测, 达到更优的检测效果。

此外, 将主动旋转滤波器(active rotating filters, ARF) 嵌入到卷积网络也能较好捕捉到物体旋转特征, 如 ORConv(oriented response networks)^[22]。ARF 是一种在卷积过程中主动旋转以生成具有多个方向通道的特征图的滤波器。当 ARF 充当滤波器组, 只有一个滤波器被具体化和学习, 借助 ARF, ORConv 需要的网络参数明显减少, 计算开销可以忽略不计, 并且能够实现显式分层方向信息编码。

1.3 Transformer 和金字塔池化 Transformer

Transformer 模型在计算机视觉领域的成功表明, 捕获远程依赖关系不仅有助于自然语言理解, 还能简化图像特征抽取。Transformer 模型工

作原理可分为 3 部分: 位置编码、自注意力、非线性变换。其中自注意力部分多使用多头自注意力(multi head self-attention, MHSA)。

自 Vision Transformer^[23] 被提出以来, 基于 MHSA 的视觉检测器层出不穷。其中 Swin Transformer^[24] 通过窗口内计算来减少 MHSA 的计算量, 通过窗口平移逐渐实现全局依赖建模, 整个过程与卷积神经网络的卷积池化级联操作类似, 通过堆叠更多的 Transformer 层来扩大感受野。然而对于图像处理, 无论是各类型 ViT 模型还是 Swin Transformer, 计算成本都是二次方量级, 较大的序列长度导致了较高的计算成本一直是 Transformer 应用到视觉模型中难以解决的问题。

受到金字塔池化在计算机视觉任务中成功应用的启发, Wu 等^[25] 提出了金字塔池化 Transformer(pyramid pooling Transformer) 旨在解决图像序列计算成本较高的问题。该方法将金字塔池化和多头自注意力相结合, 借助 Transformer 架构, 减少序列长度并捕获丰富的上下文特征。

2 PDMDet 旋转目标检测算法

针对遥感影像中微小而杂乱密集的物体, 本文在两阶段旋转目标检测算法 Oriented R-CNN 基础上提出了改进算法 PDMDet。

本文提出的基于提案增强的解耦特征挖掘旋转检测器(PDMDet) 结构如图 1 所示。PDMDet 中, 被处理图像作为输入, 首先由骨干网络进行特征提取, 将提取特征送入特征金字塔网络(feature pyramid networks, FPN), 使用卷积进行初步特征融合得到特征序列 $\{P_2, P_3, P_4, P_5, P_6\}$ 。接着把得到的特征序列送入具有 Transformer 编码器和自上而下、自下而上解码器架构的多维特征挖掘网络, 增强旋转不变性, 捕捉旋转等变性, 增强对小目标的表达能力。然后, 使用一阶段无锚检测器 FCOS 替换两阶段 RPN 生成高质量提案, 高质量提案 ROI 特征能够更精准更少冗余地实现背景对齐。最后, 设计了一个解耦特征细化网络, 使得分类分支和回归分支能够分开处理, 回归分支基于旋转卷积能够捕捉到旋转特征实现旋转框回归, 分类分支通过基于池化注意力获得强大的上下文抽象能力实现精细分类。

2.1 旋转提案增强网络

本文提出提案增强网络(PEN), 旨在为第二阶段提供高质量提案, 以尽量减少提案中的负样本。PEN 将一阶段检测器替换成二阶段的 RPN,

FCOS^[7]中说明了这一替换的可行性, CenterNet^[26]给出替换的一般步骤。本文所使用的 FCOS 是一种无锚单阶段检测器。采用无锚检测器不仅可以避免与锚框相关的复杂计算, 还能消除通常难以

优化的与锚框相关的超参数。随后, PEN 使用 PAA(probabilistic anchor assignment)^[27]通过概率分布动态分配标签, 而无需设置固定阈值, 从而能够以更少的超参数实现自适应样本选择。

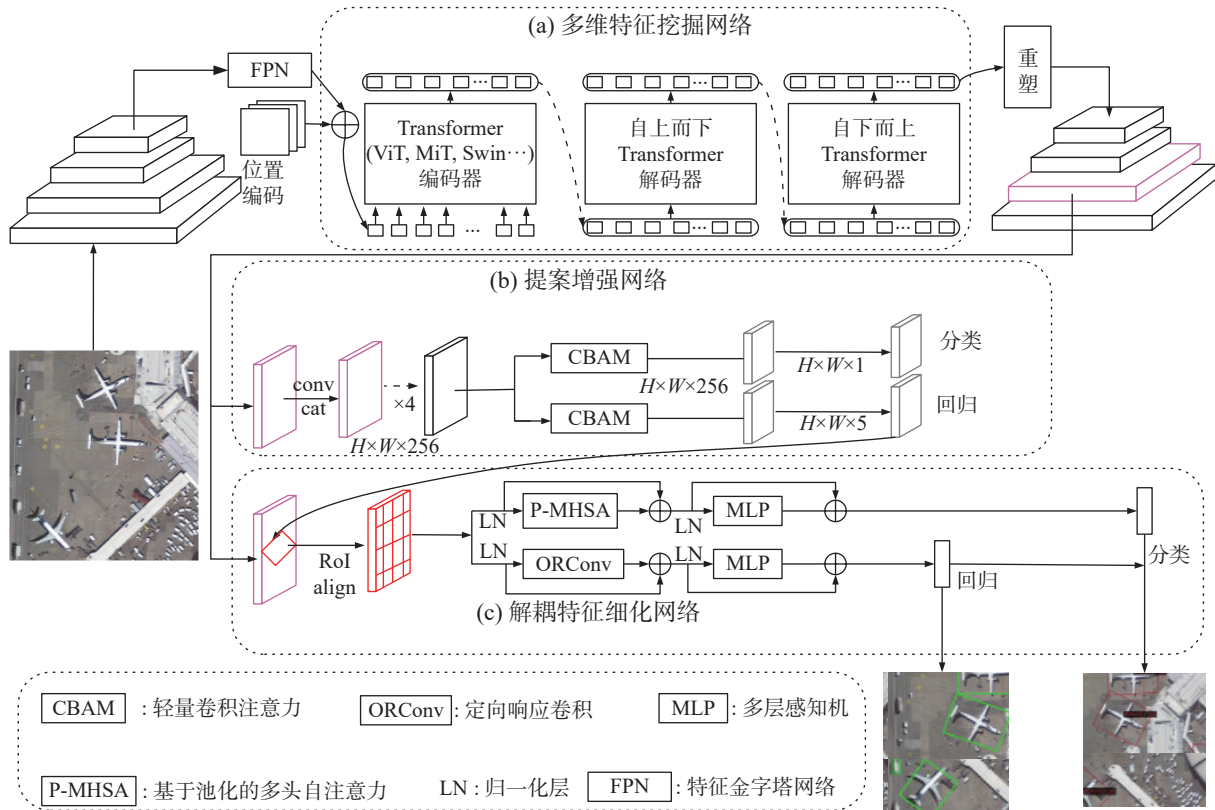


图 1 PDMDet 框架
Fig. 1 PDMDet framework

检测头是检测器的重要组成部分。现阶段分类和回归任务之间共享特征, 也就是说分类和回归任务是耦合的, 耦合结构如图 2 所示。耦合检测头不能够充分体现不同任务对特征的敏感性。

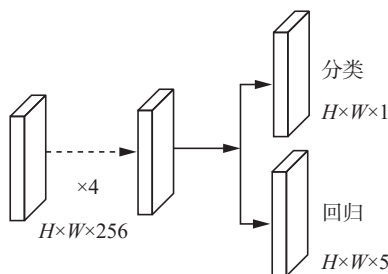


图 2 耦合检测头
Fig. 2 Coupled detection head

为了进一步增强一阶段检测器性能, 本文受 YOLO^[3]与深度重塑增强网络 (depth reshaping enhanced network, DR-ENet)^[28]启发, 设计了一种轻量级解耦模块, 以较少的额外参数分解回归和分类任务。具体做法是在分类和回归分支中分别添

加卷积注意力模块 (convolutional block attention module, CBAM), 结构如图 3 所示。

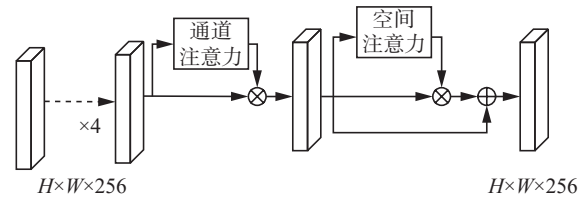


图 3 CBAM 结构
Fig. 3 Structure of CBAM

利用 CBAM 提取的解耦特征, 分类分支输出维度 $H \times W \times 1$ 与类别无关的前景预测, 而回归分支预测 4 个偏移量和角度 (x, y, h, w, θ) 。在训练过程中, 为了最大限度地利用样本, PEN 放弃随机采样, 并使用 Focal Loss^[8]来减少分类良好的样本的权重。

2.2 多维度旋转特征挖掘网络

在图像高度旋转场景下, 如何提取不受物体旋转影响的图像信息, 以及精确提取旋转位置信息, 是两个最重要的问题。普通的卷积神经网络

缺乏旋转不变性, 而大多数 FPN 也采用卷积融合特征, 这存在两个缺陷: 首先, 同一个对象在不同深度特征图上都有表示, FPN 在进行上采样和卷积融合时难以统一同一对象的表示; 其次, 由于卷积感受野的限制, FPN 整体感知能力较差。

本文设计的 Transformer 解码器通过位置编码能够感知物体方向, 从而提取与旋转无关的图像语义信息, 并借助 MHSA 获得全局感知能力。MFMN 主要包含 3 个部分: Transformer 编码器、自上而下解码器和自下而上解码器。

2.2.1 Transformer 编码器

考虑到旋转特征对于位置敏感的特性, MFMN 位置编码使用正弦编码^[29]。为了获得更大感受野, 使用自注意力对每一层特征进行增强:

$$\begin{aligned} F^i &= \text{self_attention}(\text{LN}(F^{i-1})) + F^{i-1} \\ F^i &= \text{FFN}(\text{LN}(F^i)) + F^i \\ F^i &= \text{self_attention}(\text{LN}(F^i)) + F^i \\ F^{i+1} &= \text{FFN}(\text{LN}(F^i)) + F^i \end{aligned}$$

式中: F^{i-1} 和 F^{i+1} 分别表示输入和输出特征; self_attention(\cdot) 是自注意力, 使用 MHSA 实现; FFN(\cdot) 是前馈神经网络; LN(\cdot) 是归一化层。

2.2.2 自上而下解码器

深层特征包含更为丰富的语义信息, 而浅层特征则包含丰富的图片细节和纹理。本文思路与 FPN 相似, 旨在将深层语义信息融合到浅层特征中。但与 FPN 不同的是, 不采用上采样加卷积的方式, 而是在不同层特征之间应用交叉注意力 (cross attention), 解码器结构如图 4 所示, 具体计算过程为

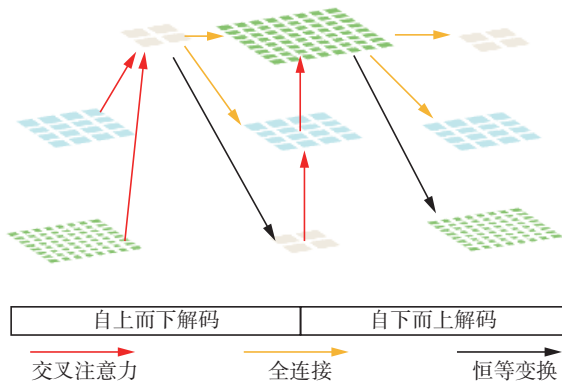


图 4 MFMN 解码器结构
Fig. 4 Decoder structure of MFMN

$$\begin{aligned} Q_i &= f_i W_Q + E_i \\ K_j &= f_j W_K + E_j \\ V_j &= f_j W_V \end{aligned}$$

式中: f_i 、 f_j 是输入特征矩阵, 且 $i \neq j$, 即特征来自 FPN 不同层; E_i 、 E_j 是位置编码; W_Q 、 W_K 、 W_V 是

Q 、 K 、 V 对应投影矩阵。最终得到第 j 层经过交叉注意力计算后输出 f_{out} :

$$f_{\text{out}} = \text{softmax} \left(\frac{(Q_i \times K_j)^T}{\sqrt{d}} \right) V_j$$

由于在输入 MFMN 之前 FPN 与相邻特征已经进行了一定程度的融合, 因此 MFMN 在自上而下解码层中进行跨层融合。具体做法是对于输入特征, P_4 和 P_6 计算交叉注意力, 同时对 P_5 和 P_6 计算交叉注意力。跨层融合不仅能够一定程度上减少计算量, 而且 P_6 还能提供丰富的语义特征, 融合后的特征具有更强的表示能力。

2.2.3 自下而上解码器

自下而上解码器则采用层层递进的融合方式。自下而上解码器和自上而下解码器实现方式类似, 计算注意力的 Q 、 K 和 V 来自不同的特征层。MFMN 整体架构如图 5 所示。

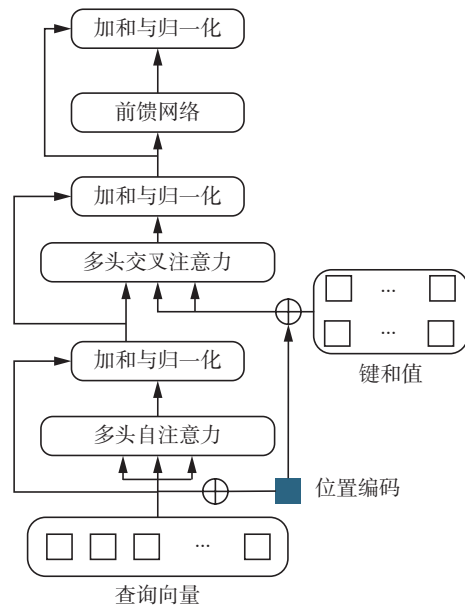


图 5 MFMN 整体架构
Fig. 5 Overall architecture of MFMN

2.3 解耦特征细化网络

由于分类任务和回归定位任务对特征性质有不同的要求。回归分支偏好旋转敏感性特征, 而分类分支偏好旋转不变性特征。因此, 在解耦特征细化网络 (DFRN) 中, 将特征分为分类分支和回归分支, 分功能进行细化处理。

2.3.1 分类分支

分类任务的关键在于提取类别之间的细微差异。金字塔池化在分类任务中已被广泛验证为有效方法, 因此本文使用基于池化的多头注意力 (pooling based MHSA, P-MHSA) Transformer 结构, 以增强上下文特征提取能力, 从而提升分类效果。

分类分支仅使用 P-MHSA 替换掉基础 Transformer 中的缩放点积注意力, 其余部分保持不变。P-MHSA 整体架构如图 6 所示, 具体做法是先将特征经过池化操作, 然后进行拼接操作。

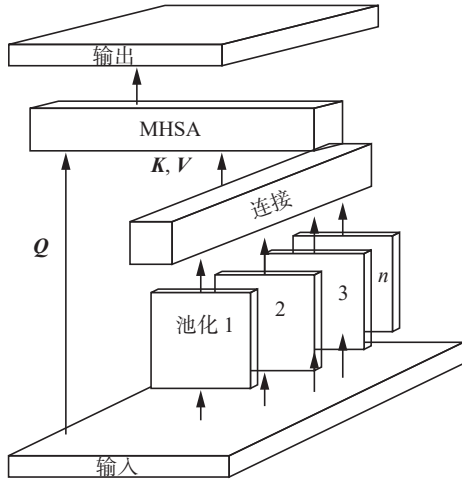


图 6 P-MHSA 整体架构
Fig. 6 Overall architecture of P-MHSA

使用 P-MHSA 的优势在于将金字塔池化的思想应用到 Transformer, 通过不同尺度池化操作提升分类模型全局建模能力。在精确度与 MHSA 相当的情况下, P-MHSA 能减少模型参数和计算量。

2.3.2 旋转定位分支

旋转框的回归任务在于提取旋转敏感性特征。本文启发性结合 Transformer 和 ORConv 来提取特征的旋转等变性。

与分类分支类似, 使用 ORConv 替换基础 Transformer 中缩放点积注意力。由于 ORConv 能够感知位置, 因此无需最初的位置编码。ORConv 通过 $N-1$ 次主动旋转, 旋转后的特征图拥有 N 个方向通道。超参数 N 采用 ORConv 原始设置 $N=8$, 其余操作和基础 Transformer 保持一致。旋转卷积计算记为 $\tilde{M} = \text{ORConv}(\mathbf{F}, \mathbf{M})$, \mathbf{F} 为滤波器, \mathbf{M} 为通道特征。其中第 k 个通道计算方式可以表示为

$$\tilde{M}^{(k)} = \sum_{n=0}^{N-1} \mathcal{F}_{\theta_k}^{(n)} \times \mathcal{M}^{(n)}, \theta_k = k \frac{2\pi}{N}, k = 0, 1, \dots, N-1$$

式中: \mathcal{F}_{θ_k} 是 \mathbf{F} 顺时针旋转 θ_k 角度的版本, 而 $\mathcal{F}_{\theta_k}^{(n)}$ 和 $\mathcal{M}^{(n)}$ 分别是 \mathcal{F}_{θ_k} 和 \mathbf{M} 的第 n 个方向通道。在每次进行旋转卷积操作时 ARF 会主动旋转特定角度, 如图 7 所示, 从而产生具有明确位置和方向编码的特征图。在反向传播过程中, ARF 将使用来自所有旋转角度的特征图误差进行集体更新。

通过分类分支和旋转定位分支可将分类和定位任务分成两个独立的子任务, 如图 3 所示。

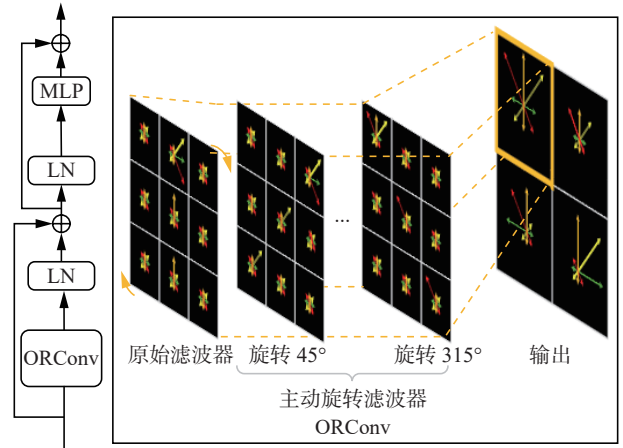


图 7 旋转分支示意
Fig. 7 Schematic of the rotation branch

3 实验与结果分析

3.1 实验平台

实验硬件平台设置 GPU 为一块 NVIDIA GeForce GTX 4090 24 GB。软件平台选择 Ubuntu 18.04、PyTorch 1.13.1, 在商汤开源的工具箱 MM-Rotate 0.3.2^[30] 进行模型的训练和推理。

3.2 数据集

模型在 3 个主流遥感目标检测数据集上进行实验验证。

DOTA-v1.0^[31], 是一个用于遥感定向目标检测的大型数据集, 包含 2 806 张图像 (其中 1 411 张用于训练, 937 张用于验证, 458 张用于测试), 188 282 个实例, 以及 15 个方向和尺度变化较大的类别: 飞机 (plane, PL)、棒球场 (baseball diamond, BD)、桥梁 (bridge, BR)、地面田径场 (ground track field, GTF)、小型车辆 (small vehicle, SV)、大型车辆 (large vehicle, LV)、船 (ship, SH)、网球场 (tennis court, TC)、篮球场 (basketball court, BC)、储罐 (storage tank, ST)、足球场 (soccer-ball field, SBF)、环形交叉路口 (roundabout, RA)、港口 (harbor, HA)、游泳池 (swimming pool, SP) 和直升机 (helicopter, HC)。

DOTA-v1.5^[31], 是一个更具有挑战性的遥感定向目标检测的大型数据集, 和 DOTA-v1.0 使用相同的图像, 但是标注了更小的实例 (小于 10 像素)。此外, 该数据集还添加了一个名为集装箱起重机 (container crane, CC) 的新类别, 数据集共包含 403 318 个实例。

HRSC2016^[32], 是一个船舶遥感检测数据集, 包含 1 061 张航拍图像, 其中 436 张用于训练, 181 张用于验证, 444 张用于测试。图片大小范围在 300 像素×300 像素~1 500 像素×900 像素。

3.3 模型设置

模型选用学习率设置为 0.01、动量项设置为 0.9、权重衰减系数为 0.000 1 的 SGD(stochastic gradient descent) 优化器。在训练开始时最先进入模型预热 (Warmup) 阶段, 预热起始学习率为初始学习率的 1/3, 预热阶段有助于模型在初始训练阶段逐渐适应学习率, 避免初始学习率过大导致的训练不稳定。模型训练共迭代 12 个轮次。

对于 DOTA-v1.0 和 DOTA-v1.5 数据集, 将原始图像裁剪为 1024 像素×1024 像素, 重叠度为 200 像素, 在推理过程中, 将所有图像块的检测结果进行合并以评估性能。DOTA 数据集的测试集并没有提供标签, 则需要将训练结果上传 DOTA 服务器获取各类别精度。对于 HRSC2016 数据集, 将图像裁剪为 800 像素×800 像素。为了防止过拟合以及增加数据多样性, 在训练过程中采用水平、垂直和对角线 3 种翻转方式, 概率都为 0.25。旋转框角度定义方法为长边定义法, 取值 $\theta_{lc} \in [-\pi/2, \pi/2)$ 。

3.4 评价指标

实验选取平均精度 (average precision, AP) 作为主要的评价指标, 每个类别的 AP 综合考虑了预测框和真实框之间的准确率 (precision, P), 召回率 (recall, R):

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

式中: P 代表准确率, R 代表召回率, N_{TP} 、 N_{FP} 和 N_{FN} 分别代表真正例、假正例和假反例。正例和反例取决于定义的 IoU 阈值, 本文实验设置阈值为 0.5。通过使用不同阈值改变召回率可以得到不同的 P-R 曲线, 计算下方面积即为 AP:

$$I_{AP}^i = \int_0^1 P_i(R_i) dR$$

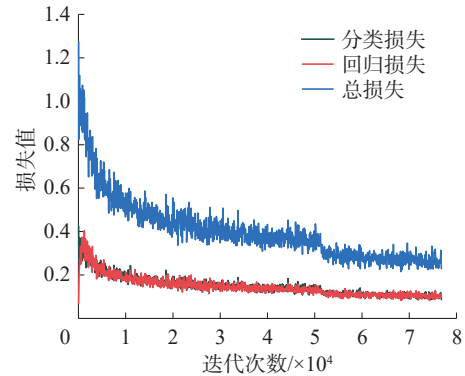
$$I_{mAP} = \frac{1}{N_{cls}} \sum_{i=0}^{N_{cls}} I_{AP}^i$$

式中: I_{AP}^i 代表平均精度, I_{mAP} 代表均值平均精度 (mean average precision, mAP), N_{cls} 为目标类别数目。

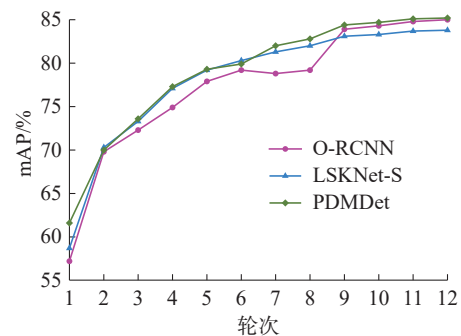
3.5 结果分析

模型训练损失如图 8(a) 所示, 本文将损失细

化为模型总损失、分类损失和边框的回归损失, 3 种损失的趋势相同。在训练的初期, 损失下降较快, 到 30 000 次迭代时, 模型逐渐收敛, 50 000 次迭代时调小学习率, 使整体损失进一步收敛。



(a) PDMDet 训练损失



(b) 训练集验证集各轮次 mAP

图 8 PDMDet 训练损失和训练集验证集 mAP

Fig. 8 PDMDet training loss and mAP of the training and validation sets

图 8(b) 给出了 PDMDet 与基线模型在 DOTA 训练集和验证集上的性能比较。在 12 轮训练过程中, PDMDet 表现出更平滑的增长和最高的 mAP。

考虑到 DOTA 数据集的类别多样性, 为了确保实验的鲁棒性和稳定性, 实验采用广泛有效的 O-RCNN^[14] 作为基线模型。为了验证模型的泛化性, 实验分别使用经典骨干网络 ResNet-50 和较新的骨干网络 LSKNet (large selective kernel network)^[33] 进行对比, 在 DOTA-v1.0 数据集上实验结果如表 1 所示。

表 1 DOTA-v1.0 数据集上逐类性能比较
Table 1 Comparison of per-class performance on the DOTA-v1.0 dataset

检测器类别	检测器	骨干网络	AP														mAP	
			PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP		HC
	R3Det-GWD ^[34]	ResNet-50	88.82	82.94	55.63	72.75	78.52	83.10	87.46	90.21	86.36	85.44	64.70	61.41	73.46	76.94	57.38	76.34
	R3Det-KLD ^[35]	ResNet-50	88.90	84.17	55.80	69.35	78.72	84.08	87.00	89.75	84.32	85.73	64.74	61.80	76.62	78.49	70.89	77.36
单阶段	Rotated FCOS ^[7]	ResNet-50	88.52	77.54	47.06	63.78	80.42	80.50	87.34	90.39	77.83	84.13	55.45	65.84	66.02	72.77	49.17	72.45
	R3Det ^[13]	ResNet-50	89.00	75.60	46.64	67.09	76.18	73.40	79.02	90.88	78.62	84.88	59.00	61.16	63.65	62.39	37.94	69.70
	S ² ANet ^[12]	ResNet-50	89.11	82.84	48.37	71.11	78.11	78.39	87.25	90.83	84.90	85.64	60.36	62.60	65.26	69.13	57.94	74.12

续表 1

检测器类别	检测器	骨干网络	AP														mAP	
			PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP		HC
两阶段	ReDet ^[19]	ResNet-50	88.79	82.64	53.97	74.00	78.13	84.06	88.04	90.89	87.78	85.75	61.76	60.39	75.96	68.07	63.59	76.25
	Roi Trans. ^[10]	ResNet-50	89.01	77.48	51.64	72.07	74.43	77.55	87.76	90.81	79.71	85.27	58.36	64.11	76.50	71.99	54.06	74.05
	FR-O ^[1]	ResNet-50	89.40	81.81	47.28	67.44	73.96	73.12	85.03	90.90	85.15	84.90	56.60	64.77	64.70	70.28	62.22	73.17
	O-RCNN ^[14]	ResNet-50	89.46	82.12	54.78	70.86	78.93	83.00	88.20	90.90	87.50	84.68	63.97	67.69	74.94	68.84	52.28	75.87
	O-RCNN ^[14]	ARC ^[36]	89.40	82.48	55.33	73.88	79.37	84.05	88.06	90.90	86.44	84.83	63.63	70.32	74.29	71.91	65.43	77.35
	O-RCNN ^[14]	LSKNet-S ^[33]	89.66	85.52	57.72	75.70	74.95	78.69	88.24	90.88	86.79	86.38	66.92	63.77	77.77	74.47	64.82	77.49
	O-RCNN ^[14]	GRA ^[37]	89.27	81.71	53.44	74.17	80.01	85.07	87.97	90.90	86.08	85.51	66.92	68.37	74.19	72.58	68.48	77.64
本文方法	PDMDet	ResNet-50	89.20	82.36	52.32	78.60	78.74	83.64	88.43	90.84	86.97	85.17	67.39	69.38	75.54	73.23	65.74	77.84
	PDMDet	LSKNet-S ^[33]	89.83	82.32	52.04	78.29	77.33	84.47	88.44	90.89	87.26	86.67	65.69	64.60	76.60	77.57	73.62	78.37

注: 加粗表示本列最优结果。

对于单阶段的检测器, 实验使用 ResNet-50 在模型 R3Det^[13] 和 S²ANet^[12] 上分别取得了 69.70% 和 74.12% 的 mAP, 在使用相同骨干网络的情况下 PDMDet 相较于 R3Det 和 S²ANet 分别提升 8.14 和 3.72 百分点。

在两阶段检测器中, PDMDet 在使用骨干网络 ResNet-50 时达到 77.84% 的 mAP, 相较于使用

相同骨干网络的 Faster-O-RCNN^[1] 和 O-RCNN 分别提升 4.67 和 1.95 百分点。

DOTA-v1.5 数据集上实验结果如表 2 所示。PDMDet 在该数据集上的 mAP 也达到了 72.35%。与相同基线的 LSKNet 和 PKINet(poly kernel inception network)^[38] 相比分别提升了 2.09 和 0.88 百分点。

表 2 DOTA-v1.5 数据集上逐类性能比较
Table 2 Comparison of per-class performance on the DOTA-v1.5 dataset

检测器	AP																mAP
	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	CC	
Mask R-CNN ^[39]	76.84	73.51	49.90	57.80	51.31	71.34	79.75	90.46	74.21	66.07	46.21	70.61	63.07	64.46	57.81	9.42	62.67
HTC ^[40]	77.80	73.67	51.40	63.99	51.54	73.31	80.31	90.48	75.12	67.34	48.51	70.63	64.84	64.48	55.87	5.15	63.40
ReDet ^[19]	79.20	82.81	51.92	71.41	52.38	75.73	80.92	90.83	75.81	68.64	49.29	72.03	73.36	70.55	63.33	11.53	66.86
LSKNet-S ^[33]	72.05	84.94	55.41	74.93	52.42	77.45	81.17	90.85	79.44	69.00	62.10	73.72	77.49	75.29	55.81	42.19	70.26
PKINet-S ^[38]	80.31	85.00	55.61	74.38	52.41	76.85	88.38	90.87	79.04	68.78	67.47	72.45	76.24	74.53	64.07	37.13	71.47
PDMDet	80.70	83.05	52.86	77.70	59.43	80.72	88.56	90.86	83.33	76.89	58.67	74.28	74.88	73.24	72.04	30.35	72.35

注: 加粗表示本列最优结果。

DOTA-v1.5 包含许多非常小的实例, 证明了 PDMDet 在小实例下效果较好。实验结果还表明, PDMDet 的精度优于目前大多数旋转目标检测器, 其性能的提升并不仅仅依赖于特定的骨干网络。PDMDet 能够适应不同的骨干网络, 并在各类骨干网络上均表现出相对于基线模型的提升, 从而以更高的精度和稳定性用于遥感旋转目标检测。

在 HRSC2016 数据集上实验结果如表 3 所示。在 VOC2007 指标下, 取得了 90.68% 次优的结果; 在 VOC2012 指标下, 取得了对比实验中最佳结果。PDMDet 相较于 O-RCNN、LSKNet 和 PKINet 分别有 0.98、0.12 和 0.04 百分点提升。PDMDet 在 HRSC2016 数据集上相较于对比模型优势较小, 甚至在 VOC2007 指标下不及当前主流模型, 主要原因是该数据集为船舶单一类别, 数

据量不足, 这导致模型在处理目标时无法充分发挥类别间对象尺寸变化的优势。

表 3 与其他模型在 HRSC2016 上比较
Table 3 Comparison with other models on the HRSC2016 dataset

检测器	骨干网络	mAP(07)	mAP(12)
ReDet ^[19]	ResNet-101	90.46	97.63
O-RepPoints ^[41]	ResNet-101	90.38	97.26
Roi Trans. ^[10]	ResNet-101	86.20	—
GWD ^[34]	ResNet-101	89.85	97.37
O-RCNN ^[14]	ResNet-50	90.40	96.50
O-RCNN ^[14]	ResNet-101	90.50	97.60
LSKNet ^[33]	LSKNet-S	90.65	98.46
PKINet ^[38]	PKINet-S	90.70	98.54
PDMDet	LSKNet-S ^[33]	90.68	98.58

注: mAP(07/12)表示VOC2007^[42]/VOC2012^[43]指标, “—”表示数据缺失, 加粗表示本列最优结果。

3.6 消融实验

消融实验在 DOTA-v1.0 数据集上进行, 其余实验设置均与模型训练保持一致。消融实验结果如表 4 所示。由表中数据分析可知, PEN 能够有效提高模型的 mAP, 与 MFMN 组合后 mAP 也有明显提高。最后, 在加入 DFRN 模块后, 模型的 mAP 达到实验最优值 78.37%。

表 4 模块消融实验
Table 4 Module ablation experiment %

检测器	mAP
baseline	75.87
+ PEN	77.14
+ PEN +MFMN	77.53
+ PEN +MFMN + DFRN	78.37

注: 加粗表示本列最优结果。

3.7 有效性分析

3.7.1 提案增强有效性分析

为了直观展示提案增强的效果, 对原始提案和增强后的提案进行可视化对比。图 9(a) 是未经提案增强效果, 图 9(b) 是经过提案增强效果。从图 9 可以看到, 通过增强后的提案不仅可以去除冗余边框, 还能有效去除分类错误的边框。通过增强后的提案不仅数量更少而且质量更高, 更少的提案能够提高模型的效率, 更高质量提案能够提升模型精准度。

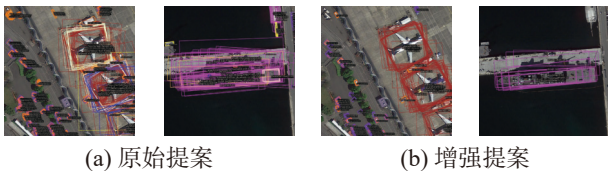


图 9 增强提案对比

Fig. 9 Enhanced proposal comparison

为了验证 CBAM 在一阶段检测器中的作用, 本文对 CBAM 有效性实验如表 5 所示。分析表中数据可知, CBAM 可以在几乎不增加模型参数数量和运算复杂度的情况下提高 mAP。这得益于 CBAM 将一阶段检测器中分类和回归分离且 CBAM 是轻量级注意力。

表 5 CBAM 有效性实验
Table 5 Effectiveness study of CBAM

检测器	参数量/ 10 ⁶	帧率/ (帧/s)	浮点运算 次数/10 ⁹	mAP/%
PDMDet	59.97	10.1	248.12	78.37
PDMDet-CBAM	59.96	10.1	248.12	78.30

注: 加粗表示本列最优结果。

3.7.2 多维特征挖掘有效性分析

为了清晰展示 MFMN 特征融合的效果, 本文对未经特征融合特征、FPN 特征融合特征和 MFMN 特征融合特征进行了可视化, 如图 10 所示。通过特征可视化图分析, 未经融合的特征图缺乏明显的层次感, 难以有效区分背景和检测目标。尽管 FPN 融合特征在一定程度上突出了检测特征, 但仍然包含大量背景噪声, 这会显著降低检测器的性能。而经过 MFMN 融合的特征能够有效去除背景信息, 更精准地呈现被检测物体的特征。

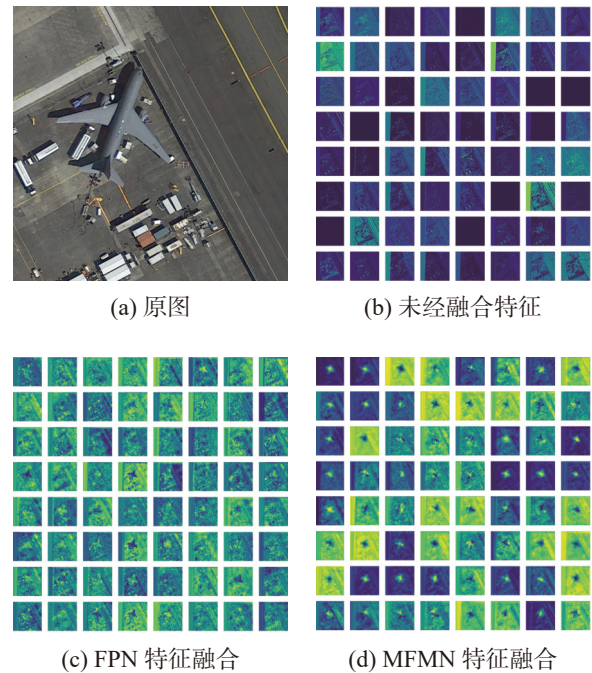


图 10 特征融合对比

Fig. 10 Feature fusion comparison

MFMN 在自上而下解码器使用跨层融合, 自下而上解码器使用渐进式融合。为了验证这种融合策略的有效性, 本文对不同融合策略进行了实验, 结果如表 6 所示。

表 6 MFMN 融合策略
Table 6 MFMN fusion strategy %

自上而下解码器	自下而上解码器	mAP
渐进融合	渐进融合	77.26
渐进融合	跨层融合	76.88
跨层融合	渐进融合	78.37
跨层融合	跨层融合	78.01

注: 加粗表示本列最优结果。

分析表中数据可知, 自上而下的解码器更适合进行跨层融合, 而自下而上的解码器则更适合渐进融合。这是因为在 MFMN 之前, 模型已经通过 FPN 进行了特征融合, FPN 通过上采样和卷积

实现自下而上的渐进融合。因此,结合 FPN 的自下而上的渐进融合、自上而下的解码器进行跨层融合以及自下而上的解码器进行渐进式融合,更有效挖掘物体的细节特征。

3.7.3 解耦特征细化有效性分析

为了验证 P-MHSA 的有效性,实验对比了基于 MHSA 和 P-MHSA 实现的 PDMDet,实验结果如表 7 所示。分析数据表明,在检测精度相当的情况下,P-MHSA 显著减少了模型的参数量和计算量。更重要的是,MHSA 在检测效率方面表现较差,使用 MHSA 构建的模型平均每秒只能处理 6.9 帧三通道图像。而通过池化操作,可以明显降低 MHSA 的计算负载,并捕获丰富的上下文信息,从而提高帧率。基于 P-MHSA 的模型平均每秒能处理 10.1 帧图像,比 MHSA 提高了 3.2 帧/s。

表 7 P-MHSA 模块有效性实验
Table 7 P-MHSA module effectiveness

检测器	参数量/ 10 ⁶	帧率/ (帧/s)	浮点运算 次数/10 ⁹	mAP/%
baseline	41.14	16.2	211.43	75.87
PDMDet(MHSA)	70.80	6.9	278.32	78.35
PDMDet(P-MHSA)	59.97	10.1	248.12	78.37

注:加粗表示本列最优结果。

最后,为了验证旋转卷积能够有效提取图像中的旋转信息,本文对比了普通卷积和旋转卷积的两种编码策略:ORPool 和 ORAlign。热力图可视化结果如图 11 所示。

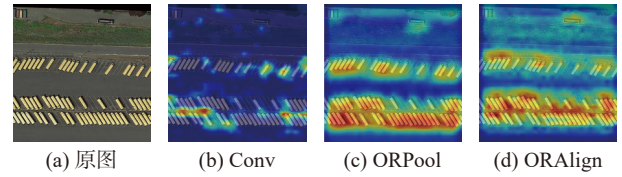


图 11 不同卷积热力图

Fig. 11 Different convolution heatmaps

通过热力图对比,两种策略的旋转卷积在旋转物体上显示出较高的热度,表明旋转卷积能够更有效地学习到旋转物体的特征。然而,ORAlign 方式实现的旋转卷积在检测效果上明显优于 ORPool,具体体现在 ORAlign 能够更好地集中注意力于车头等能够刻画物体方向的位置。这是因为 ORAlign 保留了方向信息,通过通道对齐实现旋转不变,而 ORPool 通过池化简化特征,损失了方向结构。

为了量化 PDMDet 模型复杂度,本文在理论层面考虑了 3 个指标,分别是:每层的计算复杂度、并行化计算量和网络中长程依赖之间的路径长度。结果如表 8 所示,部分数据引自文献 [29]。其中 n 代表序列长度、 d 代表维度、 k 是卷积核大小、 r 是窗口大小、 m 是所有池化特征的串联序列长度。对于本文默认的池化比例 {12, 16, 20, 24}, 可以计算得到 $m \approx \frac{n}{66.3} \approx \frac{n}{8^2}$ 。注意力和卷积仅仅使用常数时间内顺序执行所有连接位置,而循环神经网络则需要 $O(n)$ 次顺序操作。此外,影响模型学习能力的一个关键因素是梯度张量在模型中前向传播和反向传播的最大长度。模型中任意输入输出序列组合之间的路径越短,学习到长程依赖就越容易^[44]。

表 8 模块复杂度分析
Table 8 Complexity analysis of modules

层类型	每层复杂度	并行化计算量	长程依赖最大路径长度
自注意力	$O(n^2 \cdot d)$	$O(1)$	$O(1)$
循环神经网络	$O(n \cdot d^2)$	$O(n)$	$O(n)$
卷积网络	$O(k \cdot n \cdot d^2)$	$O(1)$	$O(\log_k(n))$
基于窗口自注意力	$O(r \cdot n \cdot d)$	$O(1)$	$O(n/r)$
基于池化的自注意力	$O((n + 2m) \cdot d^2 + 2n \cdot m \cdot d)$	$O(1)$	$O(1)$

3.7.4 模型复杂度分析

为了验证改进后的 PDMDet 模型的运行速度,实验采用单张 4090 显卡对模型的复杂度进行评估,主要指标包括推理速度和参数量,实验结果如表 9~10 所示。分析表中数据可知在使用相同的骨干网络的情况下 PDMDet 没有损失过多的推理速度,也没有显著增加参数量,基本符合两

阶段检测器对于精度和效率的平衡。

表 9 改进前后速度对比
Table 9 Comparison of speed before and after improvement

骨干网络	O-RCNN ^[14]	PDMDet
ResNet-50	16.2	14.8
LSKNet-S ^[33]	13.6	10.1

表 10 改进前后模型参数量对比
Table 10 Comparison of model parameters before and after improvement 10^6

骨干网络	O-RCNN ^[14]	PDMDet
ResNet-50	41.14	45.89
LSKNet-S ^[33]	55.31	59.97

3.8 可视化实验及混淆矩阵

本文对 PDMDet 和 O-RCNN 在 DOTA-v1.0 数据集上的检测结果进行了可视化, 结果如图 12 所示。从图中可以看出, PDMDet 在物体高密度场景(如密集停放的车辆)以及长宽比较大的场景(如桥梁和港口)中, 旋转框更贴合目标物体。此外, 相较于 O-RCNN, PDMDet 在小目标检测方面能够有效减少漏检。由于分类分支的改进, PDMDet 的分类效果也优于 O-RCNN。

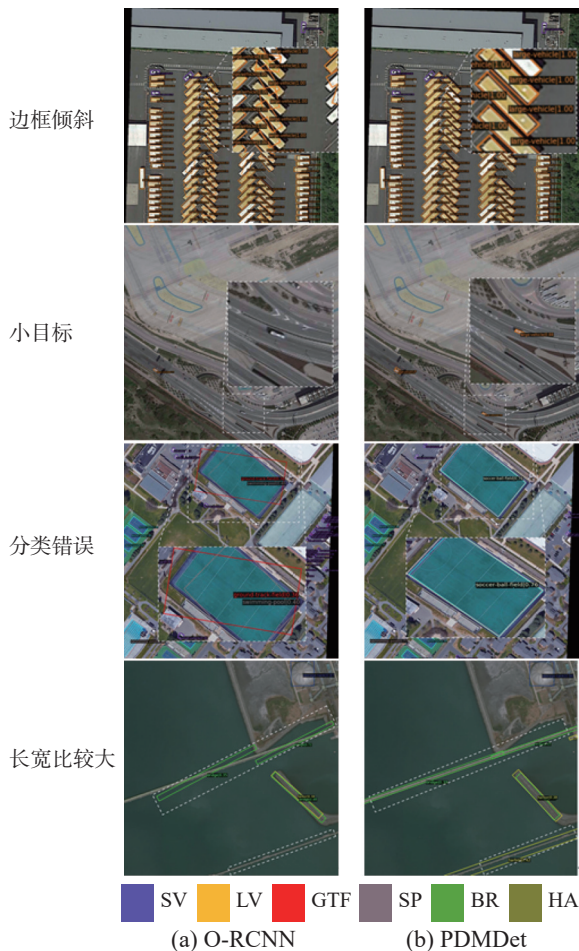
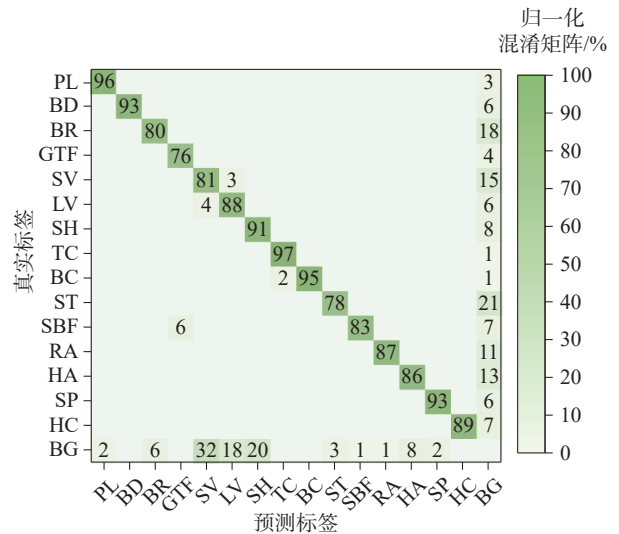


图 12 结果可视化对比

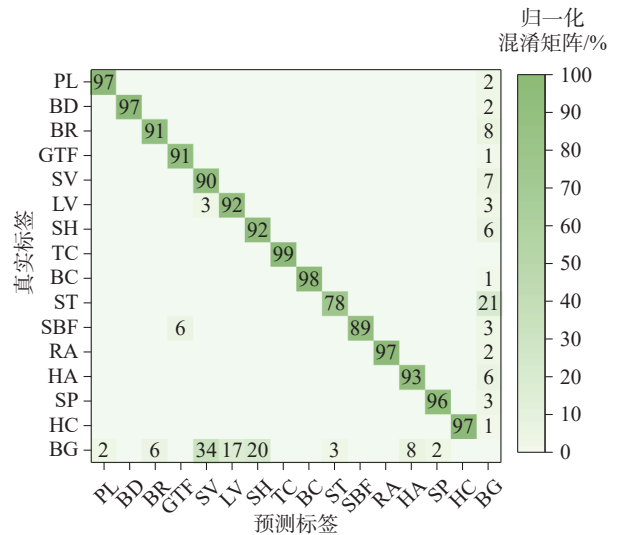
Fig. 12 Results visualization comparison

为了量化 PDMDet 的分类表现, 本文分别计算了 O-RCNN 和 PDMDet 在 DOTA-v1.0 上的混淆矩阵, 如图 13 所示, 其中 BG 代表背景。对比混淆矩阵可以看出, 通过添加增强提案、特征融合以及解耦分支, 分类精准度得到了有效提升, 特别是桥梁和田径场等长宽比较大的物体, 其分类

精度有明显改善, 同时, 其他类别的分类精度也在一定程度上有所提升。



(a) O-RCNN 混淆矩阵



(b) PDMDet 混淆矩阵

图 13 DOTA-v1.0 训练集和验证集混淆矩阵

Fig. 13 Confusion matrix of the DOTA-v1.0 training and validation sets

4 结束语

本文针对目标检测中小且杂乱、长宽比例较大以及方向任意排列的物体检测问题, 提出了 PDMDet 模型。该模型通过使用一阶段检测器替换 RPN 来提升提案质量, 并引入多维特征挖掘网络以有效融合多尺度特征, 同时, 模型对分类和回归任务进行了解耦优化。与 O-RCNN 相比, PDMDet 在参数量略微增加的前提下, 显著提升了旋转目标检测的精度。

参考文献:

[1] 蓝鑫, 吴淞, 伏博毅, 等. 深度学习的遥感图像旋转目标

- 检测综述[J]. 计算机科学与探索, 2024, 18(4): 861–877.
- LAN Xin, WU Song, FU Boyi, et al. Survey on deep learning in oriented object detection in remote sensing images[J]. *Journal of frontiers of computer science and technology*, 2024, 18(4): 861–877.
- [2] 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.
- [3] 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.
- [4] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 21–37.
- [5] 徐红明, 王兴华, 方诚, 等. 基于旋转不变性的高分辨率遥感影像船舶检测[J]. *中国航海*, 2024, 47(2): 120–127.
- XU Hongming, WANG Xinghua, FANG Cheng, et al. Rotation invariant object detection assisted ship identification from high-resolution remote sensing imagery[J]. *Navigation of China*, 2024, 47(2): 120–127.
- [6] 牛为华, 郭迅. 基于改进 YOLOv8 的舰船遥感图像旋转目标检测算法[J]. *图学学报*, 2024, 45(4): 726–735.
- NIU Weihua, GUO Xun. Rotating target detection algorithm in ship remote sensing images based on YOLOv8 [J]. *Journal of graphics*, 2024, 45(4): 726–735.
- [7] TIAN Zhi, SHEN Chunhua, CHEN Hao, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9626–9635.
- [8] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2020, 42(2): 318–327.
- [9] YANG Ze, LIU Shaohui, HU Han, et al. RepPoints: point set representation for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9656–9665.
- [10] DING Jian, XUE Nan, LONG Yang, et al. Learning RoI Transformer for oriented object detection in aerial images[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 2844–2853.
- [11] YANG Xue, YANG Jirui, YAN Junchi, et al. SCRDet: towards more robust detection for small, cluttered and rotated objects[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 8231–8240.
- [12] HAN Jiaming, DING Jian, LI Jie, et al. Align deep features for oriented object detection[J]. *IEEE transactions on geoscience and remote sensing*, 2021, 60: 5602511.
- [13] YANG Xue, YAN Junchi, FENG Ziming, et al. R3det: refined single-stage detector with feature refinement for rotating object[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S. l.]: AAAI Press, 2021: 3163–3171.
- [14] XIE Xingxing, CHENG Gong, WANG Jiabao, et al. Oriented R-CNN for object detection[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 3500–3509.
- [15] DAI Linhui, LIU Hong, TANG Hao, et al. AO₂-DETR: arbitrary-oriented object detection Transformer[J]. *IEEE transactions on circuits and systems for video technology*, 2023, 33(5): 2342–2356.
- [16] 杨学, 严骏驰. 基于特征对齐和高斯表征的视觉有向目标检测[J]. *中国科学: 信息科学*, 2023, 53(11): 2250–2265.
- YANG Xue, YAN Junchi. Visual oriented object detection via feature alignment and Gaussian parameterization[J]. *Scientia sinica (informationis)*, 2023, 53(11): 2250–2265.
- [17] 涂可龙, 卿雅娟, 李真强, 等. EllipticNet: 基于椭圆方程的遥感有向目标检测[J]. *遥感学报*, 2025, 29(3): 713–727.
- TU Kelong, QING Yaxian, LI Zhenqiang, et al. EllipticNet: equation-based remote sensing oriented object detection network[J]. *National remote sensing bulletin*, 2025, 29(3): 713–727.
- [18] 毛泽勇, 陈欣易, 丁俊峰, 等. 基于不确定性感知旋转目标检测的二次接线质检[J]. *计算机技术与发展*, 2024, 34(10): 178–185.
- MAO Zeyong, CHEN Xinyi, DING Junfeng, et al. Uncertainty-aware oriented object detection for trustworthy quality inspection of secondary wiring[J]. *Computer technology and development*, 2024, 34(10): 178–185.
- [19] HAN Jiaming, DING Jian, XUE Nan, et al. ReDet: a rotation-equivariant detector for aerial object detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2021: 2786–2795.
- [20] FENG Xiaoxu, YAO Xiwen, CHENG Gong, et al. Weakly supervised rotation-invariant aerial object detection network[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 14126–14135.
- [21] 方婷婷, 刘斌, 陈春晖, 等. 视图一致性网络下的弱监督遥感影像旋转目标检测[J]. *遥感学报*, 2024, 28(12): 3213–3230.
- FANG Tingting, LIU Bin, CHEN Chunhui, et al. View-consistency network for weakly supervised oriented object detection in remote sensing images[J]. *National remote sensing bulletin*, 2024, 28(12): 3213–3230.
- [22] ZHOU Yanzhao, YE Qixiang, QIU Qiang, et al. Oriented response networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 4961–4970.
- [23] ALEXEY D, LUCAS B, ALEXANDER K, et al. An image is worth 16x16 words Transformers for image recognition at scale[C]//International Conference on Learning Representations. [S. l.]: OpenReview.net, 2021: 1–21.
- [24] LIU Ze, LIN Yutong, CAO Yue, et al. Swin Transformer: hierarchical vision Transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 9992–10002.
- [25] WU Yuhuan, LIU Yun, ZHAN Xin, et al. P2T: pyramid pooling Transformer for scene understanding[J]. *IEEE*

- transactions on pattern analysis and machine intelligence, 2023, 45(11): 12760–12771.
- [26] ZHOU Xingyi, KOLTUN V, KRÄHENBÜHL P. Probabilistic two-stage detection[EB/OL]. (2021–3–12)[2024–11–21]. <https://arxiv.org/abs/2103.07461v1>.
- [27] KIM K, LEE H S. Probabilistic anchor assignment with IoU prediction for object detection[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 355–371.
- [28] 付天怡, 杨本翼, 董红斌, 等. 基于深度重塑的航拍目标检测增强网络[J]. 模式识别与人工智能, 2024, 37(7): 652–662.
FU Tianyi, YANG Benyi, DONG Hongbin, et al. Depth-reshaping based aerial object detection enhanced network[J]. Pattern recognition and artificial intelligence, 2024, 37(7): 652–662.
- [29] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Long Beach: Springer, 2017: 1–11.
- [30] ZHOU Yue, YANG Xue, ZHANG Gefan, et al. MMRotate: a rotated object detection benchmark using PyTorch[C]//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa: ACM, 2022: 7331–7334.
- [31] XIA Guisong, BAI Xiang, DING Jian, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3974–3983.
- [32] LIU Zikun, YUAN Liu, WENG Lubin, et al. A high resolution optical satellite image dataset for ship recognition and some new baselines[C]//Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods. Porto: SCITEPRESS-Science and Technology Publications, 2017: 324–331.
- [33] LI Yuxuan, HOU Qibin, ZHENG Zhaohui, et al. Large selective kernel network for remote sensing object detection[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 16748–16759.
- [34] YANG Xue, YAN Junchi, MING Qi, et al. Rethinking rotated object detection with Gaussian Wasserstein distance loss[C]//International Conference on Machine Learning. [S. l.]: PMLR, 2021: 11830–11841.
- [35] YANG Xue, YANG Xiaojiang, YANG Jirui, et al. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence[C]//Advances in Neural Information Processing Systems. [S. l.]: Curran Associates Inc., 2021: 18381–18394.
- [36] PU Yifan, WANG Yiru, XIA Zhuofan, et al. Adaptive rotated convolution for rotated object detection[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 6566–6577.
- [37] WANG Jiangshan, PU Yifan, HAN Yizeng, et al. GRA: detecting oriented objects through group-wise rotating and attention[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 298–315.
- [38] CAI Xinhao, LAI Qiuxia, WANG Yuwei, et al. Poly kernel inception network for remote sensing detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 27706–27716.
- [39] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980–2988.
- [40] CHEN Kai, PANG Jiangmiao, WANG Jiaqi, et al. Hybrid task cascade for instance segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 4969–4978.
- [41] LI Wentong, CHEN Yijie, HU Kaixuan, et al. Oriented RepPoints for aerial object detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 1819–1828.
- [42] EVERINGHAM M, VAN G, WILLIAMS C K I, et al. The PASCAL visual object classes challenge 2007 (VOC-2007) results[EB/OL]. (2007–06–07)[2024–12–20]. <http://www.pascalnetwork.org/challenges/VOC/voc2007/workshop/index.html>.
- [43] EVERINGHAM M, VAN G, WILLIAMS C K I, et al. The PASCAL visual object classes challenge 2012 (VOC-2012) results[EB/OL]. (2012–05–18)[2024–12–20]. <http://www.pascalnetwork.org/challenges/VOC/voc2012/workshop/index.html>.
- [44] KOLEN J F, KREMER S C. Gradient flow in recurrent nets: the difficulty of learning longterm dependencies[J]. A field guide to dynamical recurrent networks, 2001, 237–243.

作者简介:



赵振博, 硕士研究生, 主要研究方向为深度学习、旋转目标检测。E-mail: zhao_zhenbo@hrbeu.edu.cn。



付天怡, 博士研究生, 主要研究方向为深度学习、计算机视觉。E-mail: futianyi@hrbeu.edu.cn。



董红斌, 教授, 博士生导师, 中国计算机学会高级会员, 主要研究方向为多智能体系统、机器学习。主持和完成国家自然科学基金项目、工信部基础研究项目、黑龙江省自然科学基金项目, 荣获黑龙江省高校科学技术奖和黑龙江省优秀高等教育科学成果奖。发表学术论文 90 余篇, 主编教材 2 部。E-mail: donghongbin@hrbeu.edu.cn。