

融合注意力机制的遥感图像旋转目标检测算法

张宇, 马杰*, 崔静雯, 赵月华, 刘宏

河北工业大学电子信息工程学院, 天津 300401

摘要 针对光学遥感图像目标检测中存在的精度低及忽略目标方向性的问题,提出了一种基于改进 YOLOv5m 的遥感图像旋转目标检测算法。首先,融合注意力机制模块提升模型对重要特征的提取能力;其次,在特征融合模块部分考虑各节点特征融合的贡献度,并增加同一特征尺度的跳跃连接;最后,针对旋转检测中存在的角度边界问题,使用密集编码标签对角度进行离散化处理。实验结果表明,所提算法在 DOTA 数据集子集上的检测精度达到了 82.75%,在小幅降低模型计算量的情况下,较原有 YOLOv5m 提升了 11.73 个百分点,同时在 HRSC2016 舰船数据集上也取得了 88.89% 的检测精度。即该算法能有效提升光学遥感图像旋转检测的精度。

关键词 机器视觉; 遥感图像; 注意力机制; 旋转目标检测; 密集编码标签; YOLO

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP202259.2415005

Rotation Target Detection Algorithm for Remote Sensing Image Using Attention Mechanism

Zhang Yu, Ma Jie*, Cui Jinwen, Zhao Yuehua, Liu Hong

School of Electronics and Information Engineering, Hebei University of Technology, Tianjin, 300401, China

Abstract In this paper, we propose a rotating target detection algorithm using the improved YOLOv5m to solve the problems of low accuracy and poor target direction in the rotating target detection of optical remote sensing images. First, we integrate the attention mechanism module into the network to improve the ability of the model to extract important features. Second, we consider the contribution of the feature fusion at each node in the feature fusion module, with the addition of a skip connection with the same feature scale. Finally, the angles are discretized through densely coded labels due to the angles and boundaries in rotation detection. The experimental results show that this algorithm achieves a detection accuracy of 82.75% for a subset of DOTA data, indicating an improvement of 11.73 percentage points compared with the original YOLOv5m network when the model computation is reduced slightly. Furthermore, we achieved a detection accuracy of 88.89% in the HRSC2016 ship dataset. That is, the algorithm can effectively improve the accuracy of the rotation detection of optical remote sensing images.

Key words machine vision; remote sensing image; attention mechanism; rotating target detection; dense coding label; YOLO

1 引言

遥感图像是指利用航空技术采用高空鸟瞰的形式获取的图像。近年来,随着航空影像技术的不断发展,对遥感图像的处理与利用成为研究的热点^[1-2]。目标检测作为计算机视觉任务的重要分支之一,近年来在各类场景中都取得了不错的发展,但目前针对 RGB 光学

遥感图像场景下的目标检测仍然存在诸多挑战:一方面,遥感图像具有目标小、数量多、密集性强、背景复杂等特点,一般的模型很难精准检测出目标;另一方面,遥感图像由高空俯拍获取,目标的长宽比值较大,且本身具有方向性特征。而水平框目标检测器不考虑角度信息会忽略目标的方向性特征,同时水平框框选的目标包含大量背景信息,这也不利于目标的精确定位。

收稿日期: 2021-09-15; 修回日期: 2021-10-15; 录用日期: 2021-10-25

基金项目: 河北省自然科学基金(F2020202045)、河北省研究生创新项目(CXZZBS2020026)、天津市教委科研项目(2018KJ268)

通信作者: *jma@hebut.edu.cn

目前的深度学习目标检测算法主要分为单阶段目标检测和双阶段目标检测:单阶段目标检测算法如 SSD^[3]、YOLO 系列^[4-7],其特点是检测速度快,但检测的精度较低;双阶段目标检测算法如 RCNN 系列^[8-10],其特点是精度高,但双阶段处理导致时间复杂度提高,检测速度较慢。Yang 等^[11]针对遥感图像中任意方向密集排列的小目标,引入有监督的多维注意力网络,提升模型对复杂背景下目标的检测能力,同时对损失函数进行改进,提出了 IoU-Smooth L1 损失,有效提升了角度回归的效果。Guo 等^[12]将改进后的 RCNN 用于遥感舰船目标的识别,引入平衡特征金字塔,整合高级特征的语义信息和低级特征的详细信息,使得模型能够同时学习一幅图像中的大目标和小目标,但模型检测的速度和精度还有待提高。

YOLO 系列的目标检测算法检测速度快,同时随着版本的不断更新,检测精度也不断提高。最新的 YOLOv5 检测模型在多个场景检测任务都取得了不错的效果^[13],其中更封装了 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 等 4 种不同大小的模型框架,以适应不同场景检测任务的需求。针对光学遥感图像检测本身具有的特点及通用目标检测模型在光学遥感图像检测存在的问题,综合考量模型的计算量和检测精度,本文以 YOLOv5m 模型框架为基础,提出了一种遥感图像旋转目标检测算法。首先,通过融合金字塔分割注意力机制 (PSA)^[14]和高效通道注意力机制

(ECA)^[15],提升网络对于重要特征的提取能力;其次,对原网络中的路径聚合网络 (PAN)^[16]进行改进,考虑各节点特征信息的贡献度,增加同一尺度特征信息的跳跃连接;最后,针对旋转检测中存在的角度边界问题,使用密集编码标签 (DCL)^[17]对角度进行离散化处理,将角度回归问题转化为分类问题,提升角度预测效果。

2 算法描述

2.1 YOLOv5 目标检测算法

YOLOv5 整个网络结构由输入端、主干网络、Neck 和预测端等 4 个部分组成,如图 1 所示。输入端通过自适应缩放将输入图像缩放到统一大小。主干网络由 Focus、Convolution-Batch Normalization-Leak ReLU (CBL)、BottleneckCSP 及空间金字塔池化模块 (SPP)^[18]组成,用以提取图像的特征。Focus 模块接近于下采样的过程,在保证图像特征信息不丢失的情况下,将宽和高的信息集中到通道中去。BottleneckCSP 结构融合了残差结构的思想,通过两个不同的分支进行连接,实现不同层次的特征融合,大大提升了网络的特征提取能力。SPP 模块由不同卷积核大小的池化层组成,通过多重感受野融合提升了模型对于复杂场景的检测能力。Neck 部分由 PAN 构成,通过上采样和下采样的过程实现了特征图在相同尺度大小的融合,丰富了特征信息。预测端则得到了 3 个不同尺度的特征向量用于预测最终的结果。

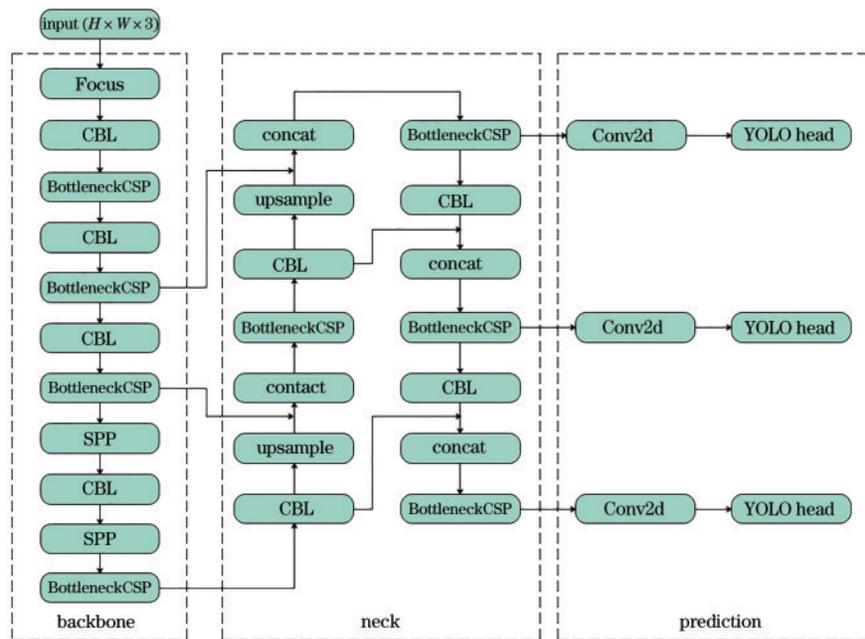


图 1 YOLOv5 网络结构

Fig. 1 YOLOv5 network structure

YOLOv5 虽然在自然场景下取得了不错的检测效果,但将其应用于光学遥感图像检测时仍存在一些问題:首先相比于自然场景,遥感图像目标多为高长宽比、具有方向性意义的目标,采用水平框会忽略目标的

方向性信息,同时也不利于目标的精确定位;其次遥感图像背景复杂且夹杂有噪声,YOLOv5 中缺乏有效的注意力机制模块,导致其在复杂背景下的检测精度不高;最后,特征融合模块中的 PAN 结构虽然提升了对

于低层信息的利用率,但其只是对各节点无差别地进行融合,忽略了特征融合中各个节点的贡献度。

2.2 增加角度回归

为了描述出目标的方向性信息,与水平框目标检测算法不同,旋转框目标检测的标注框多了一个角度信息

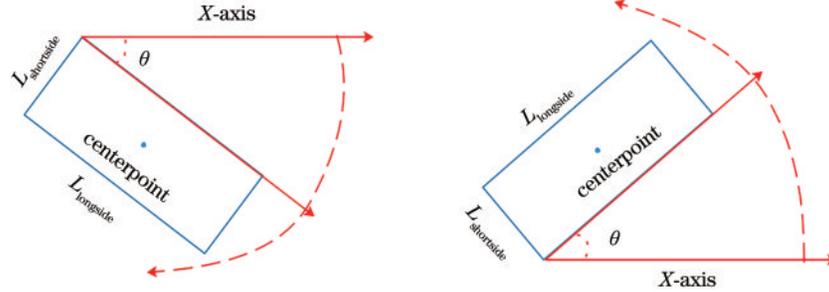


图2 长边定义法

Fig. 2 Long edge definition method

2.3 融合注意力机制

注意力机制模块指的是通过向模型中引入可学习权重,将模型的关注点更多地聚焦在重要区域,能够有效降低噪声在遥感图像特征提取中的影响。目前主流的注意力机制模块可分为两类:一类是以 squeeze-and-excitation(SE)Layer^[19]为代表的通道注意力机制;另一类是则以 convolutional block attention module (CBAM)^[20]为代表的结合通道与空间注意力机制。SELayer 单一地考虑通道,忽略了空间信息的重要性。CBAM 模块虽然兼具通道与空间注意力机制模块,但受其结构限制,往往只能捕捉局部信息,不能有效利用不同尺度特征图的空间信息,缺乏建立远程通道依赖关系的能力。

2.3.1 PSANeck 模块

针对遥感图像目标自身特点及各注意力机制模块存在的问题,借鉴 Bottleneck 中残差边的设计思想,将原有 Bottleneck 模块中的 3×3 卷积层替换成了金字塔分割注意力机制(PSA)模块形成 PSANeck 模块,如图 3 所示。PSANeck 模块由主边和残差边构成,主边需要依次经过卷积核为 1×1 的二维卷积、BN 层和激活函数 hardswish,最后经过 PSA 模块。当主边与残差输出的通道数一致时,将两边的特征信息相加输出。

对于输入特征信息 $\mathbf{X} \in \mathbf{R}^{C \times H \times W}$,PSA 模块首先通过分组卷积将特征信息提取到不同的分支,并采用不同大小的卷积核进行处理,如图 4 所示。

分组大小与卷积核大小的对应关系为

$$G = 2^{\frac{K-1}{2}}, \quad (1)$$

式中: K 为多维卷积核的大小; G 为对应卷积核尺寸的分组大小。进而,输入特征信息 $\mathbf{X} \in \mathbf{R}^{C \times H \times W}$ 经过不同分支多尺度卷积核处理过后的数学表达式为

$$\mathbf{F}_i = \text{Conv}(K_i \times K_i, G_i)(\mathbf{X}), \quad i = 0, 1, 2, \dots, S-1, \quad (2)$$

θ 用以表示方向性。采用长边定义法来表示一个中心坐标为 (x, y) 、长边为 L_{longside} 、短边为 $L_{\text{shortside}}$ 的旋转四边形,其标注格式为 $(x, y, L_{\text{longside}}, L_{\text{shortside}}, \theta)$,其中 θ 为 x 轴旋转到长边所在边的角度,角度范围为 $[-90^\circ, 90^\circ)$,顺时针旋转为正,逆时针旋转为负,如图 2 所示。

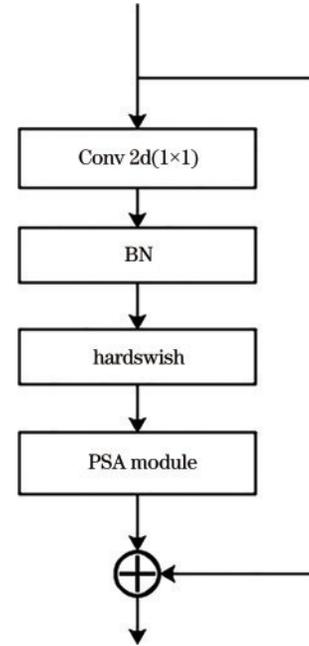


图3 PSANeck 结构

Fig. 3 PSANeck structure

式中: $\mathbf{F}_i \in \mathbf{R}^{C \times H \times W}$ 表示不同尺度的特征图; K_i 表示当前尺度的卷积核大小; G_i 表示当前尺度的分组大小; K_i 与 G_i 的对应关系如式(1)所示。接着将不同尺度的特征图对应于通道维度连接起来,得到的输出为

$$\mathbf{F} = \text{Concat}[\mathbf{F}_0, \mathbf{F}_1, \dots, \mathbf{F}_{S-1}], \quad (3)$$

式中: $\mathbf{F} \in \mathbf{R}^{C \times H \times W}$ 为得到的多尺度特征图。接着将先前得到的不同尺度特征图 $\mathbf{F}_i \in \mathbf{R}^{C \times H \times W}$ 分别通过 SE 模块进行处理,可得到对应不同尺度的注意力权重向量:

$$\mathbf{Z}_i = \text{SEWeight}(\mathbf{F}_i), \quad (4)$$

式中: $\mathbf{Z}_i \in \mathbf{R}^{C \times 1 \times 1}$ 为当前尺度下的注意力权重。将各尺度下的注意力权重通过 Softmax 进行归一化处理,提升其自适应选择不同尺度空间的能力,可得到软性

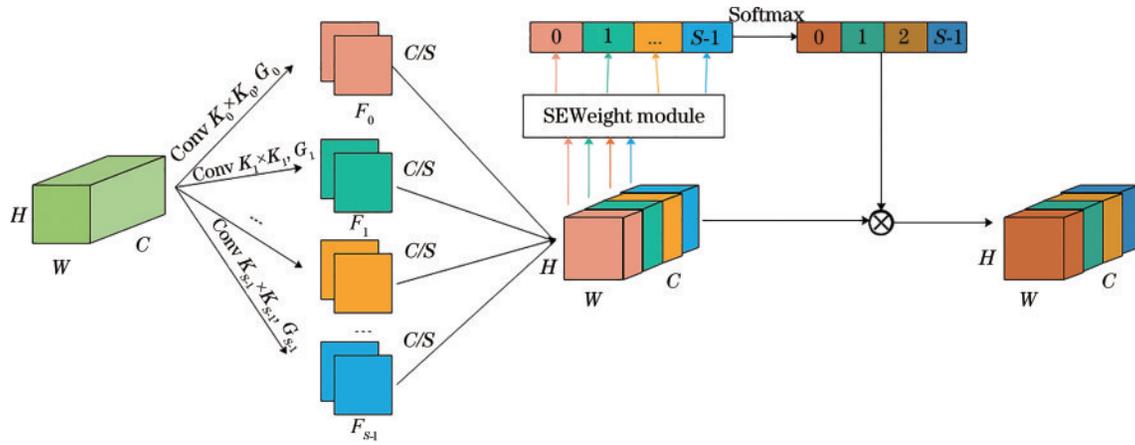


图 4 PSA module 结构

Fig. 4 PSA module structure

注意力权重:

$$W_{sw,i} = \text{Softmax}(Z_i) = \frac{\exp(Z_i)}{\sum_{i=0}^{S-1} \exp(Z_i)} \quad (5)$$

将归一化处理得到的不同尺度注意力权重与其相对应的特征图相乘, 并进行 Concat 处理得到最终的输出:

$$X' = \text{Concat} [F_0 \odot W_{sw,0}, F_1 \odot W_{sw,1}, \dots, F_{S-1} \odot W_{sw,S-1}] \quad (6)$$

式中: \odot 表示对应通道维度的乘法; $X' \in \mathbf{R}^{C \times H \times W}$ 为 PSA 模块最终的输出。

2.3.2 ECALayer 模块

考虑到模型检测性能的要求, 向网络中插入高效通道注意力机制 ECALayer 能有效提升模型性能, 同时由于其轻量化设计, 不会造成模型参数的大量增加。ECALayer 模块中, 将输入特征信息 $F \in \mathbf{R}^{H \times W \times C}$ 先进行全局平均池化, 得到对应于通道的权重信息 $F_{avg} \in \mathbf{R}^{1 \times 1 \times C}$ 。随后再通过一维卷积进行提取, 并经过 Sigmoid 激活函数进行处理, 最后将得到的通道注意力权重与原输入特征信息相乘得到输出特征信息 $F' \in \mathbf{R}^{H \times W \times C}$, 如图 5 所示。一维卷积的卷积核大小 K 由具体的通道参数自适应决定, 其计算公式为

$$K = \varphi(C) = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (7)$$

式中: C 表示通道数; $\lfloor \cdot \rfloor_{\text{odd}}$ 为取最邻近奇数; γ 和 b 为超参数, 这里取 $\gamma=2, b=1$ 。

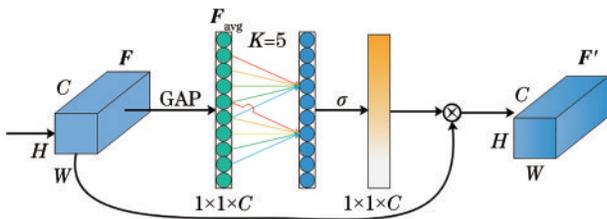


图 5 ECALayer 结构

Fig. 5 ECALayer structure

2.4 改进特征融合

高效的特征融合模块能够实现高低层次信息的充分融合, 提升模型性能。YOLOv5 的特征融合模块采用了 PAN 结构, 相比于传统的 FPN 结构, 增加了一条自下而上的特征融合层, 提高了低层特征信息的利用效率, 如图 6 所示。但 PAN 等传统的特征融合结构只是对各节点无差别地进行融合, 忽略了特征融合中各个节点的贡献度。借鉴 BiFPN^[21] 特征融合的思想, 对 YOLOv5 中的特征融合模块 PAN 结构进行改进, 增加了同一特征尺度的跳跃连接, 同时去除了贡献度较低的节点, 提升特征融合的效率, 如图 7 所示。

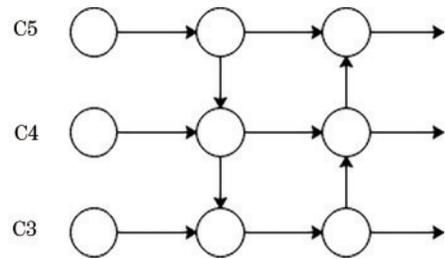


图 6 YOLOv5 PAN 结构

Fig. 6 YOLOv5 PAN structure

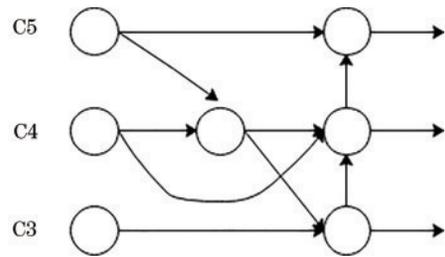


图 7 改进的特征融合结构

Fig. 7 Improved feature fusion structure

2.5 改进角度回归

与通用水平框检测算法不同, 旋转目标检测算法中由于旋转角度具有周期性, 当建议框角度处在边界位置时, 回归会存在边界问题^[11], 致使一些处于竖直方

向的检测框回归效果不好。对此,本实验组引入DCL将角度回归问题转化分类问题,有效地解决了角度回归存在的周期性问题。同时密集编码,相对于循环平滑标签(CSL)^[22],降低了编码位数。

通过密集编码标签的方式,将旋转角度 $\theta \in [-90^\circ, 90^\circ]$ 分成 N 类,相当于对角度进行了离散化处理,如图8所示。这里将角度范围边缘值 -90° 和 89° 连在了一起,能够有效避免预测框角度处在边界时的角度值突变。根据角度分类的结果需要进行编码,由于具体编码方式的不同,可分为稀疏编码和密集编码。稀疏

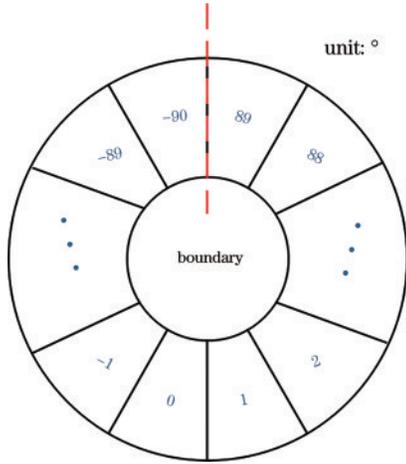


图8 密集编码标签

Fig. 8 Densely coded label

编码采用one-hot方式进行编码,每个角度分类编为一位码,当角度分类数较大时,会导致编码位数过多。本实验组采用密集编码中的二进制编码方式,大大减少了编码位数,有利于模型的轻量化。具体编解码流程如下:

1) 根据旋转角度范围 R_{AR} 、角度离散化粒度 ω ,可以得到二进制编码位数:

$$n = \text{cell}[\log_2(R_{AR}/\omega)], \quad (8)$$

式中: n 为编码位数; cell 表示向上取整操作。

2) 根据角度范围 R_{AR} 和编码位数 n ,对范围内的角度进行编码,编码公式为

$$\text{Bin}\{-\text{Round}[(\theta_{gt} - 90)/\omega]\}, \quad (9)$$

式中: Bin 表示二进制编码操作; Round 为四舍五入运算; θ_{gt} 为角度真实值。

3) 对实际网络输出的预测值进行解码,解码公式为

$$90 - \omega \text{Int}\{\text{Round}[\text{Sigmoid}(O_{\text{logits}})]\}, \quad (10)$$

式中: O_{logits} 为网络预测输出; Int 表示取整操作。

2.6 改进后的网络结构

改进后的网络结构如图9所示。将YOLOv5m中部分BottleneckCSP结构替换成了融合金字塔切割注意力机制的PSAneckCSP结构,同时在每一个CSP结构后插入高效通道注意力机制ECALayer提升其对重要特征的提取能力。针对原本PAN结构特征融合存

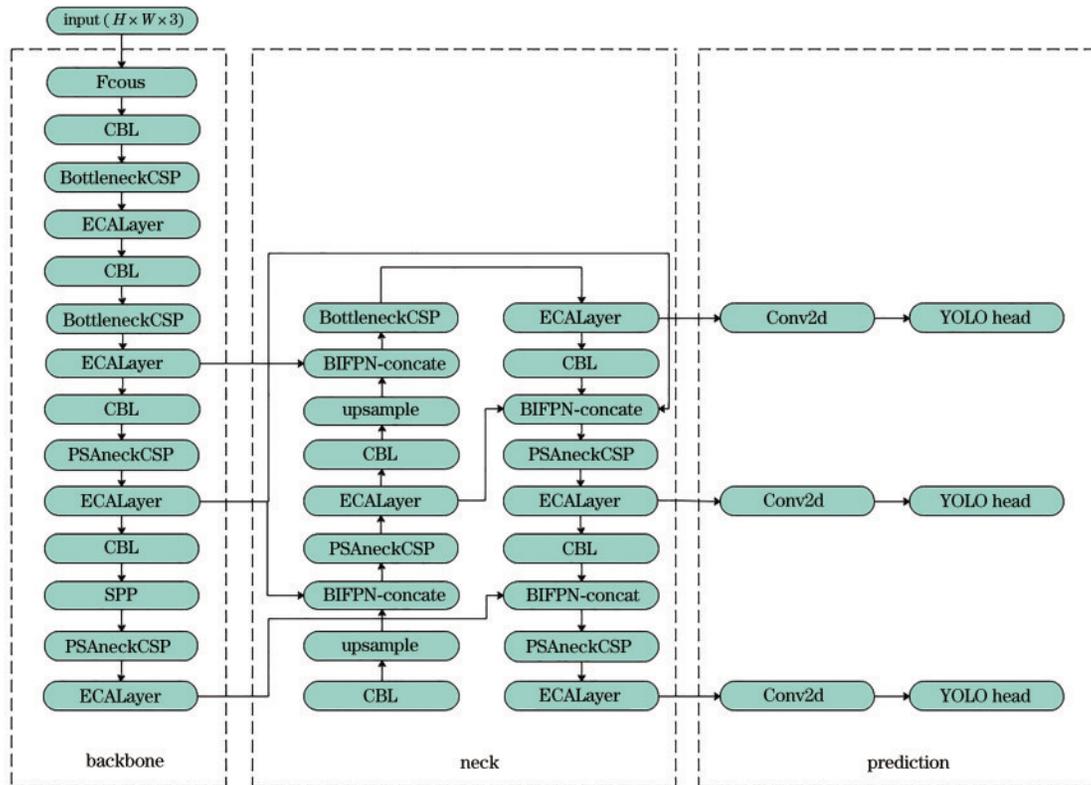


图9 改进的网络结构图

Fig. 9 Improved network structure diagram

在的不足,改进了特征融合模块,考虑了各节点特征融合的贡献度,增加了同一特征尺度的跳跃连接。

3 实验与结果分析

为评估所提改进 R-YOLOv5m 模型的性能,在 DOTA 多目标数据集和 HRSC2016 舰船数据集上进行训练和测试。实验使用 NVIDIA RTX3070 显卡, Windows 10 操作系统,采用的深度学习框架为 PyTorch 1.8.1。

3.1 实验数据集处理

DOTA 数据集^[23]是大型高分辨遥感图像数据集,包括采集自 Google Earth 等多个平台的 2806 张 RGB 图像,像素尺寸在 800×800~4000×4000 之间,分为水平与旋转两种标注方式,目标总数为 188282 个,一共包含 15 类目标,分别为飞机(PL)、轮船(SH)、小型车辆(SV)、大型车辆(LV)、储油罐(ST)、网球场(TC)、操场跑道(GTF)、桥梁(BR)、环路(RA)、游泳池(SP)、棒球场(BD)、篮球场(BC)、港口(HA)、直升机(HC)和足球场(SBF)。考虑到各类目标的特点及所提改进算法主要针对任意方向排列的小目标,从中抽取了 4 类目标进行训练和测试,分别为飞机、轮船、小型车辆和大型车辆。这 4 类目标具有高长宽比和方向性意义,同时多以密集排列的形式出现,更适合作为所提算法的改进目标。DOTA 数据集由于其输入图像尺寸较大,难以直接输入网络进行训练,所以需要 DOTA 数据集进行切割处理。为了避免边缘信息的丢失,还会在图像切割处保留一定的重叠区域。采用 DOTA_devkit 将 DOTA 数据集切割成 1024×1024, GAP 设为 50%。

HRSC2016 舰船数据集的主要数据来自于 Google Earth,采集地点为 6 个重要港口,场景复杂。根据舰船所在场景的不同,分为海面和近岸两种,共有 1061 张 RGB 图像,包含 2976 个任意角度的舰船目标,目标特征丰富,采用旋转框标注法。图像的像素尺寸在 300×300~1500×900 之间。

3.2 评价标准

实验采用平均准确率均值(mAP)来评估模型的性能:

$$P_{\text{precise}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (11)$$

表 2 各改进模块消融实验对比

Table 2 Comparison of ablation experiments of each improved module

Group	PSA	ECA	BiFPN	DCL	PL	SH	SV	LV	GFlops	mAP
G1	×	×	×	×	78.79	79.46	64.61	61.23	53.7	71.02
G2	✓	×	×	×	86.53	79.47	64.60	69.34	49.8	74.99
G3	✓	✓	×	×	87.45	79.60	71.59	69.64	49.8	77.07
G4	✓	✓	✓	×	86.54	86.89	69.94	76.82	52.9	80.05
G5	✓	✓	✓	✓	87.46	87.34	76.00	80.20	52.9	82.75

$$R_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (12)$$

式中: N_{TP} 表示真正例的数量; N_{FP} 表示假正例的数量; N_{FN} 表示假反例的数量。当单个对象的真实框与锚框之间的交并比(IOU)大于 0.5 时,则将锚框定义为真正例,否则将其定义为假正例。根据公式计算得到 P_{precise} 和 R_{recall} ,并绘出 PR 曲线,设置步长为 0.1,取 $R_{\text{recall}}=[0,0.1,0.2,\dots,1]$ 对应的 P_{precise} 值,这些 P_{precise} 的平均值为平均准确率(AP),将各类的 AP 求和取平均即可得到 mAP。

3.3 消融实验

3.3.1 DCL 中角度离散化粒度消融实验

密集编码标签中角度离散化粒度 ω 的不同会直接影响检测器的最终性能: ω 设置过大会导致分类精度上限下降,提升了理论误差; ω 设置过小,预测精度上限提高,但会提升角度分类的难度。为了评估不同离散化粒度 ω 对模型性能的影响,设置了 5 组不同离散化粒度 ω 的消融实验,分别为 180/8、180/32、180/64、180/128、180/180。实验模型以改进过后的 YOLOv5m 模型 R-YOLOv5m 作为基准,在 DOTA 数据集子集中进行训练并测试。具体结果如表 1 所示,其中 PL、SH、SV、LV 分别表示对应类别的 AP 值。从表 1 可以看出,当 ω 取 180/64 时可以使模型达到较好的检测效果。

表 1 不同角度离散化粒度 ω 下模型性能对比

Table 1 Comparison of model performance under different angle discretization granularity ω unit: %

ω	PL	SH	SV	LV	mAP
180/8	86.55	86.90	73.58	76.82	80.96
180/32	87.45	87.33	76.00	77.40	82.04
180/64	87.46	87.34	76.00	80.20	82.75
180/128	87.44	87.32	76.01	80.18	82.73
180/180	87.39	87.33	76.04	79.90	82.67

3.3.2 各改进模块消融实验

为了评估各改进模块对模型整体性能的影响,共设置了 5 组对照实验,在同样的条件下采用 DOTA 数据集子集进行训练和测试。5 组消融实验结果如表 2 所示,其中 ✓ 表示模型中含有该结构,× 表示模型中不含有该结构,Gflops 表示浮点运算数。

从表 2 可以看出:第 1 组实验是在以 YOLOv5m 模型基础上增加角度回归参数进行的,并以此作为基准模型, mAP 达到了 71.02%;第 2 组实验是将模型中部分 Bottleneck 结构替换成为改进的 PSANeck 结构,在大幅减少计算量的情况下,将 mAP 提升了 3.97 个百分点;第 3 组实验是在第 2 组的基础上加入了通道注意力机制 ECA Layer,在几乎不增加模型计算量的情况下,将 mAP 提升了 2.08 个百分点;第 4 组在第 3 组的基础上对特征融合中 PAN 结构进行改进,在大幅增加模型计算量的同时,将 mAP 提升了 2.98 个百分点;第 5 组实验在第 4 组基础上使用密集编码标签解决角度回归问题,虽然实际训练时间会增加,但检测精度得到

了有效提升。

热力图(CAM)是对网络检测效果进行可视化的重要工具。从 CAM 中可以观察出为了实现正确的分类,网络应更加侧重的区域。首先得到 YOLOv5m 与改进算法网络预测层 3 个不同尺度的输出,再将其关于大型车辆(LV)的类别分值分别映射到原图,进而得到最终的热力图显示结果,如图 10 和图 11 所示,其中红色越深表示得分越高,蓝色越深表示得分越低。从图中可以看出,相比于原 YOLOv5m 算法,所提改进算法能够有效抑制无用的背景及噪声信息,使网络更聚焦于重要的特征,提升网络在复杂背景下的检测能力。

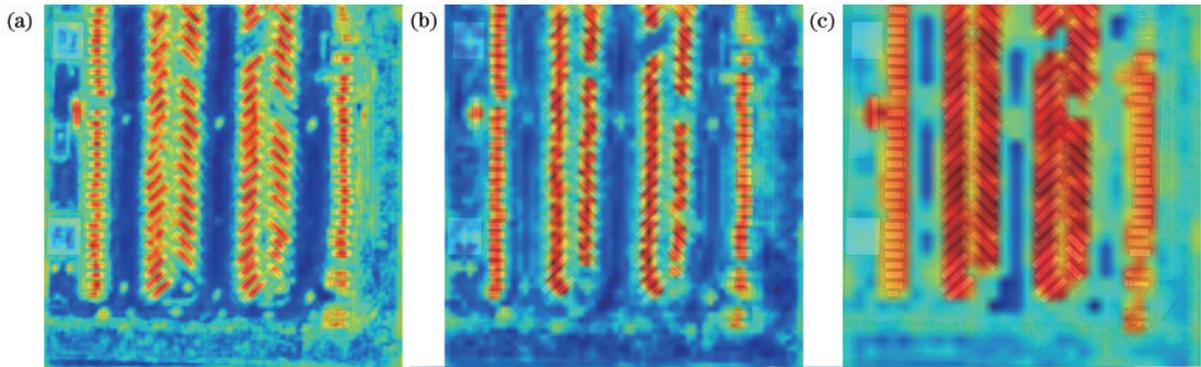


图 10 YOLOv5m 多尺度特征热力图。(a)高分辨率分支;(b)中分辨率分支;(c)低分辨率分支

Fig. 10 YOLOv5m multi-scale characteristic thermodynamic diagram. (a) High resolution branch; (b) medium resolution branch; (c) low resolution branch

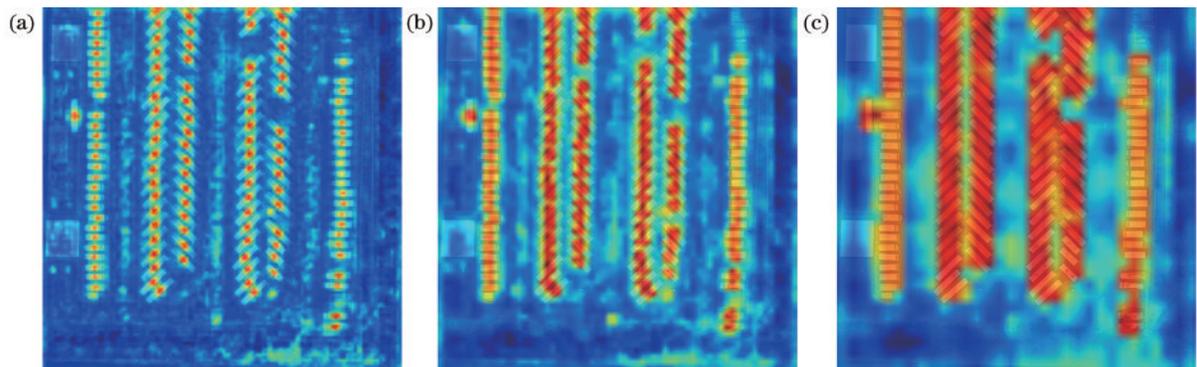


图 11 改进算法多尺度特征热力图。(a)高分辨率分支;(b)中分辨率分支;(c)低分辨率分支

Fig. 11 Improved algorithm multiscale characteristic thermodynamic diagram. (a) High resolution branch; (b) medium resolution branch; (c) low resolution branch

比较了原 YOLOv5m 算法与改进算法在密集排列、目标小、数量多及复杂背景下的检测效果,结果如图 12 所示。从图中可以看出,原 YOLOv5m 算法存在目标角度回归不理想和错检等问题,而所提改进算法在这些方面表现更好。

3.4 与其他算法对比

为了进一步验证所提改进算法 R-YOLOv5m 的有

效性,将其与目前主流旋转目标检测算法在 DOTA 数据集子集及 HRSC2016 舰船数据集上的进行了对比,具体结果如表 3 和表 4 所示。

从表 3 和表 4 可以看出,与目前主流的旋转目标检测算法相比,所提改进算法在 DOTA 数据集子集及 HRSC2016 舰船数据集上均能取得不错的效果,能够有效提升遥感图像旋转目标检测的精度。

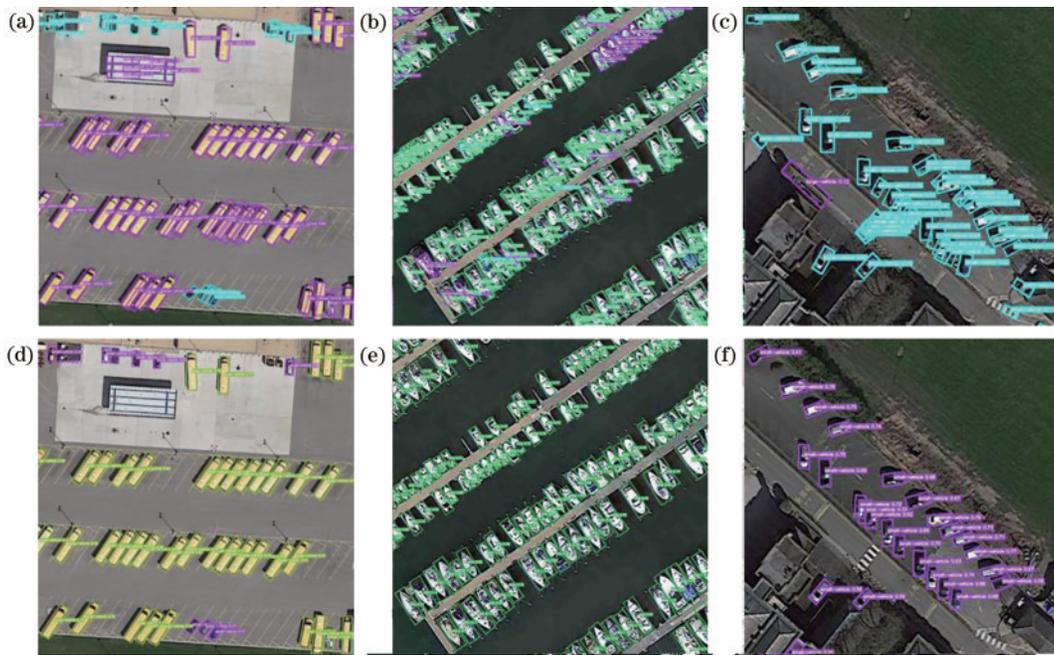


图 12 YOLOv5m 与改进算法检测效果对比。(a)(b)(c) YOLOv5m 算法检测结果; (d)(e)(f) 改进算法检测结果

Fig. 12 Comparison of detection effect between YOLOv5m and improved algorithm. (a)(b)(c) detection results of YOLOv5m algorithm; (d)(e)(f) detection results of improved algorithm

表 3 不同算法在 DOTA 数据集子集结果对比

Table 3 Comparison of results of different algorithms on DOTA dataset subset unit: %

Method	PL	SH	SV	LV	mAP
FR-O ^[23]	79.42	37.16	35.30	38.02	47.48
IE-Net ^[24]	80.20	52.58	49.71	65.01	61.88
SCRDet	89.98	72.41	68.36	60.32	72.77
RSDet ^[25]	90.10	73.60	70.20	78.70	78.15
R-YOLOv5m	87.46	87.34	76.00	80.20	82.75

表 4 不同算法在 HRSC2016 舰船数据集结果对比

Table 4 Comparison of different algorithms on HRSC2016 ship dataset

Method	R ² CNN ^[26]	R ² PN ^[27]	Gliding Vertex ^[28]	R-YOLOv5m
mAP / %	73.07	79.60	88.20	88.89

4 结 论

针对遥感图像目标存在的问题,分别从网络结构和优化角度回归这两个方面进行改进。在网络结构改进方面,融合了注意力机制模块 PSANeck 和 ECALayer,考虑了各节点特征对于网络的贡献度,增加同一特征尺度的跳跃连接。在优化角度回归方面,对角度进行离散化处理,通过密集编码标签的方式将角度回归转化为分类问题。实验结果表明,所提改进算法在大幅降低模型计算量的情况下,有效提升了检测精度,更能适应遥感图像目标检测任务。本研究同时也提供了将一种通用目标检测应用于遥感图像旋转检测的思路,具有一定的通用性。但考虑实际部署的

情况,当前模型的计算量和参数量还是较大,未来的工作可以探索在保证模型检测精度的情况下,对模型进行轻量化改进。

参 考 文 献

- [1] 薛俊达,朱家佳,张静,等. 基于 FFC-SSD 模型的光学遥感图像目标检测[J]. 光学学报, 2022, 42(12): 1210002.
Xue J D, Zhu J J, Zhang J, et al. Object detection in optical remote sensing images based on FFC-SSD model [J]. Acta Optica Sinica, 2022, 42(12): 1210002.
- [2] 农元君,王俊杰. 基于嵌入式的遥感目标实时检测方法[J]. 光学学报, 2021, 41(10): 1028001.
Nong Y J, Wang J J. Real-time object detection in remote sensing images based on embedded system[J]. Acta Optica Sinica, 2021, 41(10): 1028001.
- [3] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [4] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [5] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [6] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-04-05]. <https://arxiv.org/abs/1804.02767>.

- [7] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-04-05]. <https://arxiv.org/abs/2004.10934>.
- [8] 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, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [9] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [10] Ren S Q, He K M, 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.
- [11] Yang X, Yang J R, Yan J C, et al. SCRDet: towards more robust detection for small, cluttered and rotated objects[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8231-8240.
- [12] Guo H Y, Yang X, Wang N N, et al. A rotational libra R-CNN method for ship detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(8): 5772-5781.
- [13] 薛芳芳, 王月明, 李琦. 基于特征部位空间关系的牛日常行为识别[J]. 激光与光电子学进展, 2021, 58(22): 2215007.
Xue F F, Wang Y M, Li Q. Recognition of cattle daily behavior based on spatial relationship of feature parts[J]. Laser & Optoelectronics Progress, 2021, 58(22): 2215007.
- [14] Zhang H, Zu K K, Lu J, et al. EPSANet: an efficient pyramid split attention block on convolutional neural network[EB/OL]. (2021-05-30) [2021-04-05]. <https://arxiv.org/abs/2105.14447>.
- [15] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [16] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.
- [17] Yang X, Hou L P, Zhou Y, et al. Dense label encoding for boundary discontinuity free rotation detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 15814-15824.
- [18] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [19] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [20] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [21] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10778-10787.
- [22] Yang X, Yan J C. Arbitrary-oriented object detection with circular smooth label[M]//Vedaldi A, Bischof H, Brox T, et al. Computer Vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12353: 677-694.
- [23] Xia G S, Bai X, Ding J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3974-3983.
- [24] Lin Y T, Feng P M, Guan J, et al. IENet: interacting embranchment one stage anchor free detector for orientation aerial object detection[EB/OL]. (2019-12-02) [2021-04-05]. <https://arxiv.org/abs/1912.00969v2>.
- [25] Qian W, Yang X, Peng S L, et al. Learning modulated loss for rotated object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(3): 2458-2466.
- [26] Jiang Y Y, Zhu X Y, Wang X B, et al. R2CNN: rotational region CNN for arbitrarily-oriented scene text detection[C]//2018 24th International Conference on Pattern Recognition (ICPR), August 20-24, 2018, Beijing, China. New York: IEEE Press, 2018: 3610-3615.
- [27] Ma J Q, Shao W Y, Ye H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.
- [28] 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, 2021, 43(4): 1452-1459.