

基于多尺度特征提取的遥感旋转目标检测

吴洛冰¹, 谷玉海^{1,2*}, 吴文昊¹, 范帅鑫¹

¹北京信息科技大学现代测控技术教育部重点实验室, 北京 100089;

²北京信息科技大学机电工程学院, 北京 100089

摘要 针对高分辨率遥感图像具有物体尺度差异大、小目标排列密集且方向性强的问题,提出一种基于多尺度特征提取的旋转遥感目标检测算法。选用CenterNet作为基准模型,对其进行重新设计。首先,为增强上下文信息提取能力,结合多尺度空洞卷积提出并使用感受野扩展模块;其次,结合自适应特征融合,增强算法对多尺度目标的提取能力;最后,重新设计CenterNet检测头并更新损失函数,增强模型对旋转物体的检测性能。设计的模型命名为CenterNet for remote sensing images(CenterNet-RS)。在DOTA数据集上进行实验,CenterNet-RS的平均精度均值(mAP)达73.01%,相较于基准模型,提高了9.45个百分点。实验结果表明,设计的方法可有效提高算法检测遥感图像目标的精度。

关键词 遥感图像; 旋转目标检测; 多尺度; 卷积神经网络; 特征融合

中图分类号 TP753

文献标志码 A

DOI: 10.3788/LOP221716

Remote Sensing Rotating Object Detection Based on Multi-Scale Feature Extraction

Wu Luobing¹, Gu Yuhai^{1,2*}, Wu Wenhao¹, Fan Shuaixin¹

¹Key Laboratory of Modern Measurement and Control Technology, Ministry of Education, Beijing Information Science & Technology University, Beijing 100089, China;

²Mechanical Electrical Engineering School, Beijing Information Science & Technology University, Beijing 100089, China

Abstract A rotation remote sensing target detection algorithm based on multi-scale feature extraction is proposed, because high-resolution remote sensing images have large object scale differences, dense small-object arrangements, and strong orientation. In this study, CenterNet was chosen as the benchmark model and redesigned. First, to improve the context information extraction ability, we proposed and applied the receptive field expansion module combined with multi-scale cavity convolution. Second, the extraction ability of the algorithm for multi-scale targets was improved in combination with adaptive feature fusion. Finally, we redesigned the CenterNet detection head and updated the loss function to improve the detection performance of the model for rotating objects. The designed model is named CenterNet for remote sensing images (CenterNet-RS). Experiments were performed on the DOTA dataset, and the mean average precision (mAP) of CenterNet-RS reaches 73.01%, which is 9.45 percentage points higher than the baseline model. Thus, the experimental findings demonstrate that the proposed method can significantly increase the target detection accuracy for remote sensing images.

Key words remote sensing image; rotating object detection; multi-scale; convolutional neural network; feature fusion

1 引言

基于遥感图像的目标识别算法是近年来深度学习领域的研究重点。该算法可广泛应用于交通、环境、军事等领域,在数量统计、目标定位等任务中有着广泛的

应用前景。因此,设计一款可靠性高、实时监测的遥感图像目标检测算法具有重要意义。

遥感图像具有背景环境复杂、目标尺度变化大、小目标密集且偏转明显的特点。传统方法通过对特定特征进行人工提取,可快速完成对特定目标的检测,但只

收稿日期: 2022-05-27; 修回日期: 2022-06-26; 录用日期: 2022-07-14; 网络首发日期: 2022-07-24

基金项目: 国家自然科学基金(51975058)、学科建设专项资助项目(5112011015)

通信作者: *guyuhai@bistu.edu.cn

能提取低层次的特征,缺乏自适应性和泛用性,难以胜任遥感图像的目标检测工作^[1]。而基于卷积神经网络的目标检测算法在遥感领域表现出越来越强的自适应性,已成为近年来的研究热点。其中,在遥感目标检测领域,基于先验框的检测算法较为流行^[2-4],如近年的 R³Det^[5]、RSDet^[6]、Gliding Vertex^[7]等。该类算法通过 Selective Search^[8]等方法,提取出少量的候选区域,再对区域进行特征提取、分类和线性回归微调;或者将图像预先划分为 $n \times n$ 个网格,每个网格中包含数个之前根据所检测目标预设尺寸的先验框(anchor box),并对每个网格进行回归运算^[9-10]。以上两种方法在检测精度和速度上有所差别,却均使用了锚框机制,在检测小目标时,需要更小的划分网格和先验框,这使得计算成本大幅上升。另外,小目标和大目标的正样本数量差距较大,使网络倾向于检测大目标而丢失小目标。为避免先验框检测所带来的问题,有学者采用基于改进 Transformer 的 DETR 模型^[11-12],或直接采用基于关键点的检测算法。DETR 类算法将自然语言处理中常用的自注意力机制迁移至目标识别领域,相较于低级信息,更善于提取高级语义信息。该类算法在技术上仍有较大的提升潜力。而基于关键点的目标检测算法根据高分辨率热力图推断物体的位置和种类,避免了先验框机制难以适应小目标尺寸而对准确率造成影响的问题,很大程度上提高了模型的检测能力^[13-14]。该类算法的研究重点是基于关键点旋转密集目标的回归问题,如 CHPDet^[15]和 BBAVectors^[16]等算法,都是在关键点检测的基础上对角度感知方法进行创新的。

基于关键点的目标检测算法在检测小目标上具有很大潜力,但也存在一些缺陷:在检测前的上采样过程中,容易丢失低级特征;在检测密集目标时,水平检测框容易出现重叠现象,造成目标漏检。针对以上缺陷,本文以 CenterNet 为基础,在物体信息提取和检测方式上进行改进:设计结合多尺度空洞卷积的感受野扩展模块,以适应不同种类目标的语义尺度;采用自适应的方式对不同尺度特征进行融合,保留小目标的低级特征并过滤环境噪声;在检测阶段加入角度信息图并增加角度分类损失项,以提升对旋转目标和密集目标的检测性能。

2 CenterNet 算法

CenterNet 是一种基于关键点推断目标的端对端目标检测算法。该算法将目标检测问题转换为关键点预测问题,避免了锚框(anchor box)机制^[17]在小目标检测中的困难。CenterNet 通过主干网络提取特征,得到高维特征图,并对特征图进行三次上采样操作增加其分辨率。通过对高分辨率特征图进行卷积,得到物体的中心点位置、目标中心点偏置和预测框宽高信息,对这些信息进行解码,得到物体的预测框。图 1 为 CenterNet 算法的流程。

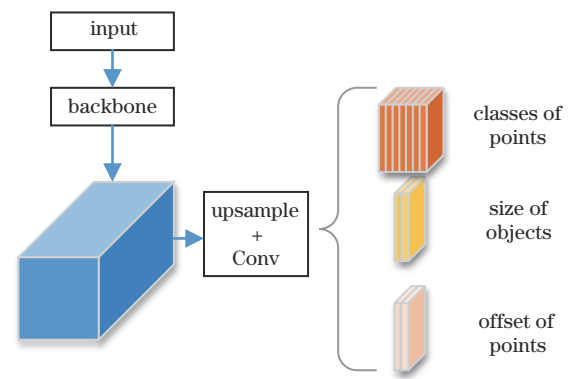


图 1 CenterNet 模型的结构图

Fig. 1 Structure diagram of CenterNet model

CenterNet 在单一特征图上即可完成多尺度目标检测,但主干网络的多次卷积会造成低级特征的丢失。因此,该模型在特征提取和融合方面仍有改进的空间。同时,CenterNet 的检测头结构具有功能扩展的可能。

3 算法改进

本文对 CenterNet 的整体结构、特征提取方法和模型检测端进行重新设计,提出 CenterNet-RS。以 ResNet101^[18]作为主干网络提取特征,对主干网络中通道数分别为 128 和 2048 的特征图进行上采样,使其与通道数为 64 的特征图在维度上保持相同。选择通道数为 128 和 64 的特征图,是因为其高分辨率带来了更多低级特征,选择 2048 的特征图是因为参考了 CenterNet 结构中主干网络的输出特征图维度。接下来,统一维度,为自适应特征融合做准备。

接下来整合不同尺度的特征,并进行自适应融合。在特征融合结构中结合多尺度空洞卷积扩展感受野,以提取并融合多尺度目标特征;利用 3×3 和 1×1 的卷积,使检测头解码得到目标的种类、位置及预测框大小;另外,在检测头中加入角度预测分支,采用分类的方式对目标偏转角度进行预测。为方便区分,将改进后的算法命名为 CenterNet for remote sensing images (CenterNet-RS)。图 2 为 CenterNet-RS 算法的流程。

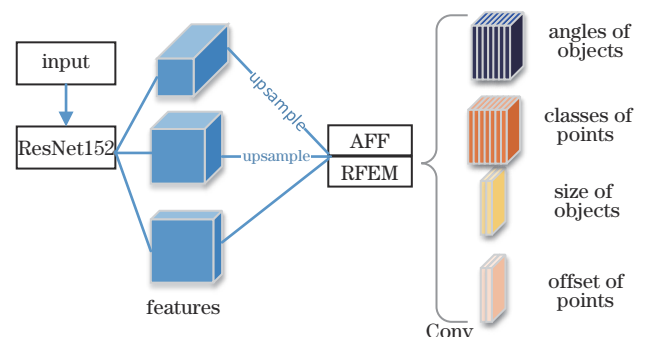


图 2 CenterNet-RS 模型的结构图

Fig. 2 Structure diagram of AFF RFEM of CenterNet-RS model

其中, AFF 代表自适应特征融合, RFEM 代表感受野扩展模块。

3.1 感受野扩展模块

遥感图像中目标尺度差异较大, 不同尺度的感受野能提取不同的语义信息。借助多尺度的上下文信息, 提升算法对遥感图像中大尺度差异性目标的检测能力^[19]。根据卷积神经网络中感受野的定义^[20], 相较于单一尺度的卷积层, 使用多尺度的卷积层能有效地提取不同尺度感受野的信息。为此, 设计感受野扩展模块 (RFEM), 通过多尺度的卷积层结合不同膨胀率的空洞卷积, 实现对多尺度上下文信息的采集。

首先, 使用 3×3 和 1×1 的卷积层调整通道数; 其次, 为了添加多尺度上下文信息用于小目标检测, 采用

三个平行的 1×1 卷积层对特征图进行通道调整, 并进行多尺度卷积和对应空洞卷积的处理; 最后, 对空洞卷积得到的特征图进行拼接和通道调整, 再与输入特征图进行相加, 并采用 ReLU 激活函数对最终结果进行非线性处理。该过程中, 设置的 3 个平行卷积层的参数为 $(k=1, p=0, s=1)$ 、 $(k=3, p=1, s=1)$ 和 $(k=5, p=2, s=1)$; 3 个空洞卷积的参数分别为 $(k=3, p=1, d=0)$ 、 $(k=3, p=2, d=2)$ 和 $(k=3, p=4, d=4)$ 。其中, k 代表卷积核大小; p 代表卷积过程中特征图边缘填充的像素数量; s 代表卷积核的步长; d 代表空洞卷积的膨胀率。通过这样的参数设置, 可以在避免网格效应的前提下, 对多尺度的语义信息进行提取。图 3 展示了 RFEM 的原理和结构。

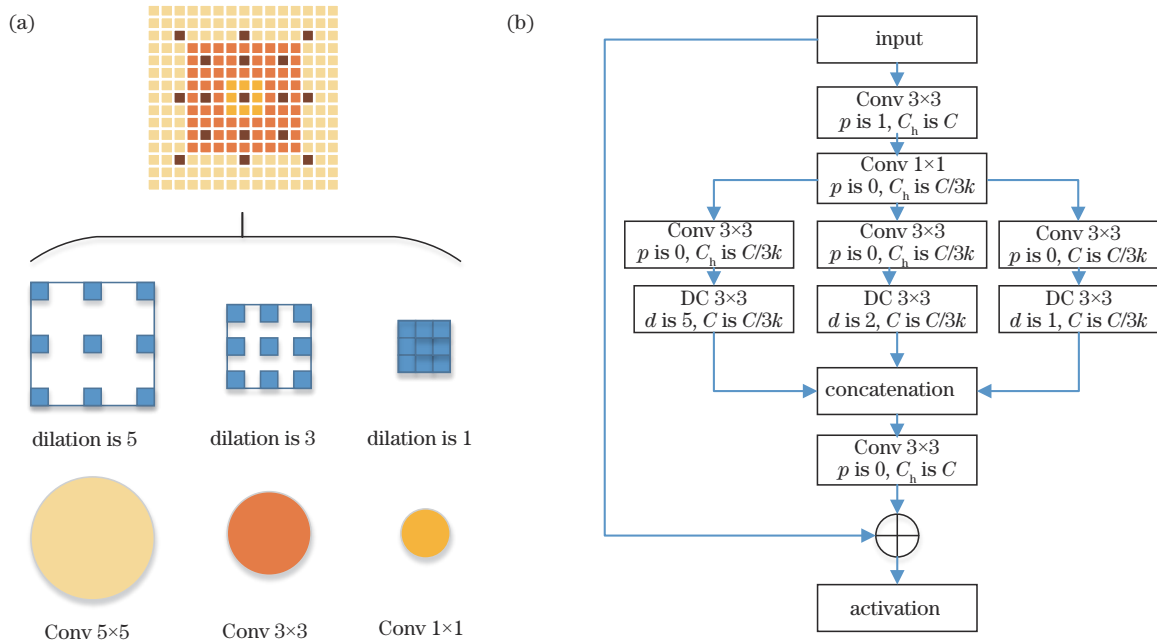


图 3 感受野扩展模块的原理图与结构图。(a)原理图;(b)结构图

Fig. 3 Schematic of principle and structure of receptive field expansion module. (a) Schematic of principle; (b) schematic of structure

将输入模块的特征图通道数为 C , 通过 C/k 调整通道数, 为了方便拼接操作, 设置 $k=3$ 。将通道数相除的结果向上取整, 并在特征图拼接后利用 1×1 的卷积调整输出通道数与输入通道数一致。得益于三个不同尺度的平行 3×3 空洞卷积层, RFEM 得以获得更多尺度的感受野, 并借此提高对小目标的检测性能。然后, 通过拼接操作合并特征信息, 并结合 1×1 的卷积调整特征通道并减少拼接操作引起的混叠现象, 以保持特征的稳定性和信息的相对丰富度。最后, 为了防止梯度发散影响模型效果, 使用 shortcut 操作将结果与初始特征结合。在对遥感图像目标进行检测的过程中, 特征图的浅层信息比大尺度的语义信息更重要。为了保留特征图浅层信息, 对特征提取网络中维度为 64 和 128 的高分辨率特征图进行感受野扩展操作。

3.2 自适应特征融合

遥感图像存在复杂的背景干扰且目标尺度变化

大, 如果仅采用高维通道特征图的信息, 则会丢失低级特征。为了整合低级特征和高级特征, 需要对不同尺度的特征图进行信息融合。而该过程可能导致不同尺度间信息的干扰, 因此需要对不同尺度的特征进行自适应的加权融合。本文受文献 [21] 的启发, 为 CenterNet 加入特征融合结构, 其基本思路在于对不同级别的特征进行空间滤波, 从而对可利用特征进行组合。

自适应特征融合需要通过缩放统一特征图的维度。该模型中, 取维度为 $(152, 152, 64)$ 的特征图作为基准, 对 $(76, 76, 128)$ 和 $(19, 19, 2048)$ 尺寸特征图分别进行 2 次和 3 次步长等于 2 的上采样。维度统一后, 需要对特征图和二维权重矩阵进行相乘, 并通过反向传播训练权重矩阵。设某一特征图的任一元素为 $x_{ij}^n (n \in \{1, 2, 3\})$, 该元素对应权重为 $\alpha_{ij}, \beta_{ij}, \gamma_{ij}$, 则在输出特征图中的对应映射可表示为

$$y_{ij} = \alpha_{ij} \cdot x_{ij}^1 + \beta_{ij} \cdot x_{ij}^2 + \gamma_{ij} \cdot x_{ij}^3, \quad (1)$$

式中: α_{ij} 在单一通道中共享, 目的是过滤空间信息, 故规定 $\alpha_{ij} + \beta_{ij} + \gamma_{ij} = 1$ ($\alpha_{ij}, \beta_{ij}, \gamma_{ij} \in [0, 1]$), 使权值能更准确地表示信息的重要程度。 $\alpha_{ij}, \beta_{ij}, \gamma_{ij}$ 由对应的控制参数 $\lambda_{\alpha_{ij}}, \lambda_{\beta_{ij}}, \lambda_{\gamma_{ij}}$ 确定。故权重 α_{ij} 的表达式为

$$\alpha_{ij} = \frac{e^{\lambda_{\alpha_{ij}}}}{e^{\lambda_{\alpha_{ij}}} + e^{\lambda_{\beta_{ij}}} + e^{\lambda_{\gamma_{ij}}}}. \quad (2)$$

通过对特征图 X^1, X^2, X^3 进行 1×1 的卷积运算, 获取对应的权重映射矩阵 $\mathbf{A}_\alpha, \mathbf{A}_\beta$ 和 \mathbf{A}_γ 。由此, 可根据链式传播法则确定该部分的梯度:

$$\frac{\partial l}{\partial x_{ij}^1} = \alpha_{ij} \frac{\partial y_{ij}^1}{\partial x_{ij}^1} \cdot \frac{\partial l}{\partial y_{ij}}, \quad (3)$$

式中: $\frac{\partial y_{ij}^1}{\partial x_{ij}^1}$ 表示激活函数求导, 激活函数固定的情况下, 该项可用常数 1 代替^[11]。因此, 梯度计算公式可简化为

$$\frac{\partial l}{\partial x_{ij}^1} = \alpha_{ij}^1 \frac{\partial l}{\partial y_{ij}}. \quad (4)$$

图 4 为 CenterNet-RS 的自适应特征融合的结构。

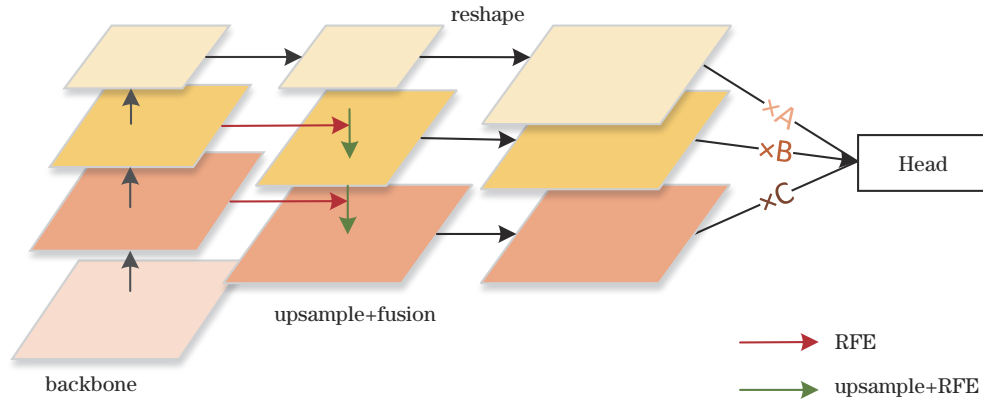


图 4 自适应特征融合结构

Fig. 4 Structure of adaptive feature fusion

3.3 基于角度标签的可旋转检测头

遥感图像包含大量排列密集的偏转目标。若采用传统的水平检测框进行检测, 则可能由于预测框排列密集导致 NMS 算法^[22]剔除一部分正确的预测框, 从而导致目标漏检。此外, 水平的检测框难以反映物体的真实大小, 有较大的定位误差。因此, 需要在 CenterNet 的检测部分加入角度预测分支, 并对损失函数进行改进, 使其具备检测物体偏转角度的功能。

原本的 CenterNet 的检测头分为三个部分, 分别对应热力图、中心点图和宽高信息图。在此基础上, 采用对角度进行分类的方法获取角度信息。将旋转框定义的角度范围划分为不同类别, 每变化 1° 归为一类。因此, 在检测头新增一个维度为 $(128, 128, T)$ 的角度信息图, T 代表所取的角度数量。同时, 在角度分类过程中, 采用 circular smooth label (CSL)^[23] 进行角度标签平滑处理, 得到每个物体的近似取整的角度标签。由于在训练和检测任务中加入了角度分类任务, 故在损失函数中加入角度回归和角度分类项。角度回归项的公式为

$$L_{\text{reg}} = \frac{1}{N} \sum_{n=1}^N |\hat{\theta}_{p_n} - \theta_n|, \quad (5)$$

式中: $\hat{\theta}_{p_n}$ 为预测角度值; θ_n 为真实角度值。而角度分类损失则采用 Sigmoid 交叉熵损失进行计算, 公式为

$$L_{\text{CSL}} = -\frac{1}{N} \sum_{n=1}^N \log [p_n \log \hat{p}_n + (1 - p_n) \log (1 - \hat{p}_n)], \quad (6)$$

式中: p_n 和 \hat{p}_n 分别是物体角度真实值和预测值的概率分布。取 $\lambda_{\text{size}} = 0.1, \lambda_{\text{off}} = 1, \lambda_{\text{reg}} = 1, \lambda_{\text{CSL}} = 0.5$, 用以调整损失。考虑以上两种损失后, 模型的损失函数为

$$L = L_k + \lambda_{\text{size}} L_{\text{size}} + \lambda_{\text{off}} L_{\text{off}} + \lambda_{\text{reg}} L_{\text{reg}} + \lambda_{\text{CSL}} L_{\text{CSL}}, \quad (7)$$

式中: L_k 是中心点分类预测损失; L_{size} 是尺度预测损失; L_{off} 是中心点偏置损失。以上三个损失分支均参照 CenterNet, 定义分别为

$$L_k = \frac{-1}{N} \sum_{\text{xyrc}} \begin{cases} (1 - \hat{Y}_{\text{xyrc}})^\alpha \log \hat{Y}_{\text{xyrc}}, & \hat{Y}_{\text{xyrc}} = 1 \\ (1 - Y_{\text{xyrc}})^\beta \hat{Y}_{\text{xyrc}} \log (1 - \hat{Y}_{\text{xyrc}}), & \hat{Y}_{\text{xyrc}} \neq 1 \end{cases}, \quad (8)$$

$$L_{\text{size}} = \frac{1}{N} \sum_{k=1}^N |\hat{S}_{p_k} - s_k|, \quad (9)$$

$$L_{\text{off}} = \frac{1}{N} \sum_p \left| \hat{O}_{\hat{p}} - \left(\frac{p}{R} - \tilde{p} \right) \right|, \quad (10)$$

式中: Y_{xyrc} 表示真实值; \hat{Y}_{xyrc} 为预测值; \hat{S}_{p_k} 为目标的预测尺寸; s_k 为目标在图像上的真实尺寸; $\hat{O}_{\hat{p}}$ 为网络预测的偏移量数值; p 为图像中心点坐标; \tilde{p} 表示缩放后中心点的近似整数坐标; R 为热力图的缩放因子; α 与 β 是两个超参数, 用来均衡难易样本。

4 实验结果与分析

4.1 数据集和实验准备

使用 dataset for object detection in aerial (DOTA) 航空影像数据集^[24]对算法性能进行测试,该数据集包含 2806 张图片,包含飞机(plane, PL)、轮船(ship, SH)、储油罐(storage tank, ST)、棒球场(baseball diamond, BD)、网球场(tennis court, TC)、游泳池(swimming pool, SP)、田径场(ground track field, GTF)、港口(harbor, HA)、桥梁(bridge, BR)、大型车辆(large vehicle, LV)、小型车辆(small vehicle, SV)、直升机(helicopter, HC)、环形交叉路口(roundabout, RA)、足球场(soccer-ball field, SBF)及篮球场

(basketball court, BC)共 15 个物体类别。图片尺寸为 800 pixel×800 pixel~4000 pixel×4000 pixel。图 5 即为 DOTA 数据集的部分图像显示,可见数据集的目标差异大且多为小目标,因此处理该数据集有较高的难度。准备 2000 张图片作为训练集,806 张图片作为测试集。由于 DOTA 数据集中图像像素较多,无法直接输入模型,故将图像裁剪成尺寸为 608 pixel×608 pixel 的子图像,并输入网络。利用 K-means 聚类算法^[25]对样本中大目标进行尺寸统计,参考最大的聚类尺寸,图像裁剪过程中保留宽度为 150 pixel 的重叠区域。检测完成后,对子图像进行拼接还原,重叠区域的检测结果通过 NMS 算法筛选,避免重复检测影响模型性能评估。



图 5 DOTA 数据集展示

Fig. 5 Presentation of DOTA dataset

为了更客观地描述对比算法和所提算法性能的差异性,使用 UCAS-AOD 数据集^[26]进行对比测试。UCAS-AOD 数据集专用于对遥感图像中飞机和车辆的检测,包含飞机实例 7482 个和车辆实例 7114 个,共计 2420 幅图像,取 1694 张图片作为训练集,726 张图片作为测试集。UCAS-AOD 数据集的图像尺寸均为 1280 pixel×659 pixel,以 1280 pixel×1280 pixel 作为输入网络的尺寸。

每批次数量设置为 64, epoch 设置为 100。将样本分批次进行迭代训练,并对迭代后的模型进行参数更新。迭代过程中,采用余弦退火算法更新学习率^[27],初始学习率设置为 1×10^{-3} 。表 1 是实验和模型搭建过程中采用的软件和硬件环境。

4.2 评价指标

采用检测平均精度(AP)和平均精度均值(mAP)作为评估模型的指标。AP 值根据召回率(R)和准确率(P)得到。召回率和准确率的计算公式分别为

$$R = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (11)$$

$$P = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (12)$$

式中:TP 代表真正样本;FP 代表假性正样本;FN 为假

表 1 实验和模型搭建所采用的软硬件环境

Table 1 Hardware and software environments for experimental and model construction

Experiment	Content
CPU	Intel Core i5
GPU	Nvidia RTX 2080 Ti
RAM	16 GB
OS	Ubuntu 18.04
Platform	CUDA 10.0
Library	PyTorch

性负样本。AP 值由单一类别物体的 TP 和 FP 构成的曲线积分确定,表达式为

$$P_{AP} = \int_0^1 p(r) dr, \quad (13)$$

式中: $p(\cdot)$ 表示以召回率 r 为参数的函数。mAP 为所有类别检测的平均 AP 值,即 $P_{mAP} = \frac{1}{n} \sum_{i=1}^n P_{AP_i}$,其中, n 代表所检测物体的类别数。

4.3 算法性能实验

4.3.1 测试效果

图 6 为所提算法在 DOTA 数据集上的可视化结

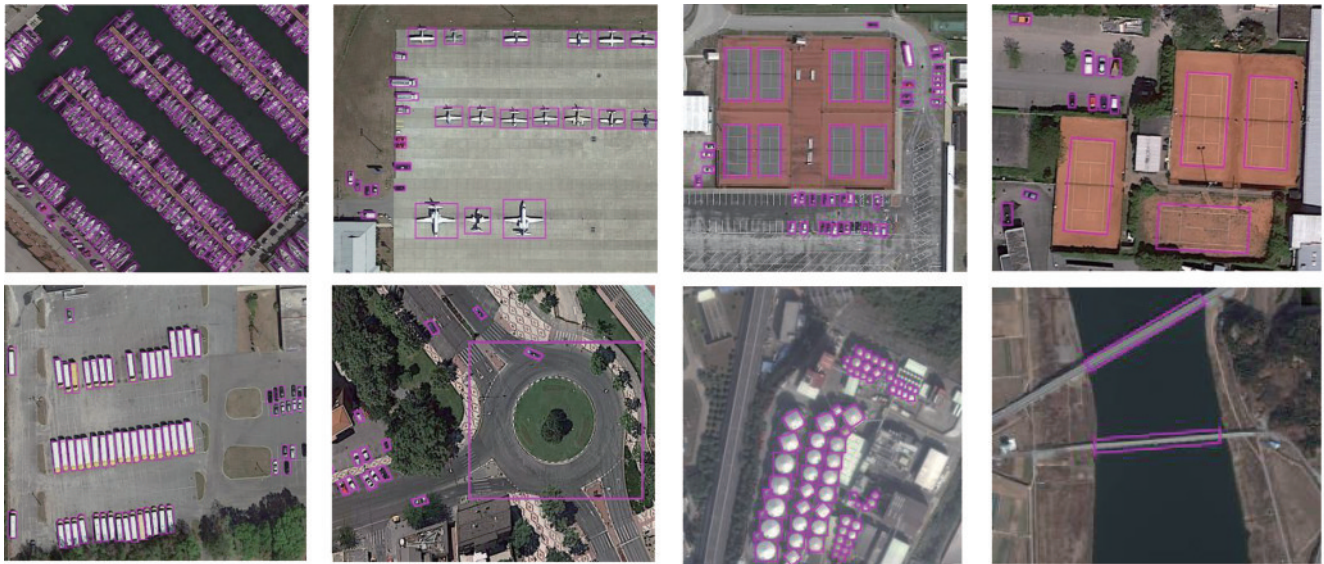


图 6 CenterNet-RS 在 DOTA 数据集上的效果展示

Fig. 6 Effect presentations of the CenterNet-RS on DOTA dataset

果,可见所提算法对不同尺度的目标都具备较好的识别效果。同时,所提算法可以有效地预测目标的位置和旋转角度,避免了检测框重叠现象,这对识别排列密集的旋转物体很有帮助。

4.3.2 实验数据与结果分析

为检测模型性能,对 CenterNet-RS 和 R2CNN、ICN^[28]、CAD-Net^[29]、RoI-Trans 和 BBAVectors^[30] 遥感目标检测模型在 DOTA 数据集进行测试,观察它们在 DOTA 数据集下的准确性表现。此外,还对 CenterNet 和 RetinaNet^[31] 两种单阶段目标检测模型进行同等条件测试,以观察所提改进对模型实时性的影响。表 2 记录了这些模型的性能表现。

表 2 CenterNet-RS 和其他模型在 DOTA 数据集上的性能对比
Table 2 Comparison of performance of CenterNet-RS and other models on DOTA dataset

Model	Backbone	Rotatable-Head	mAP / %	Running time / s
R2CNN	VGG-16	✓	60.67	0.263
ICN	ResNeXt101	✓	68.16	0.221
CAD-Net	ResNet101	✓	69.90	0.172
RoI-Trans	ResNet101	✓	69.59	0.175
BBAVectors	ResNet101	✓	75.36	0.086
CenterNet	ResNet101	×	63.56	0.072
RetinaNet	DLA-34	×	62.76	0.081
CenterNet-RS	ResNet101	✓	73.01	0.078

在使用 ResNet101 作为主干网络的前提下,CenterNet-RS 的 mAP 为 73.01%,相较于使用相同主干网络的原版算法,提高了 9.45 个百分点。同时,CenterNet-RS 对单张图像的检测用时为 0.078 s,在检

测速度上领先于除 CenterNet 之外的对比模型,所提改进在提高检测精度的前提下保证了算法的速度。与现有方法^[32-33]相比,CenterNet-RS 在检测精度上有一定差距,但作为一款单阶段算法,在准确性和实时性上有更好的平衡。

表 3 为不同算法检测不同类别物体的 AP 值。从表 3 可得,所提算法对飞机(PL)、桥梁(BR)、网球场(TC)、储存罐(ST)、足球场(SBF)、环岛(RA)、码头(HA)有最高的识别精度,AP 值分别为 92.37%、50.42%、93.96%、87.30%、62.72%、69.16%、

表 3 DOTA 数据集上不同算法检测各类物体的 AP 值
Table 3 AP of different algorithms detecting various objects on DOTA dataset unit: %

Category	R2CNN	ICN	CAD-Net	RoI-Trans	BBA-Vectors	CenterNet-RS
PL	80.89	81.36	87.80	88.64	88.63	92.37
BD	65.75	74.30	82.40	78.52	84.06	74.86
BR	35.34	47.70	49.40	43.44	52.13	50.42
GTF	67.44	70.32	73.50	75.92	69.56	66.43
SV	59.93	64.89	71.10	68.81	78.26	64.49
LV	50.91	67.82	63.50	73.68	80.40	79.63
SH	55.81	69.98	76.70	83.59	88.06	76.74
TC	90.67	90.76	90.90	90.74	90.87	93.96
BC	66.92	79.06	79.20	77.27	87.23	75.27
ST	72.39	78.20	73.30	81.46	86.39	87.30
SBF	55.06	53.64	48.40	58.39	56.11	62.72
RA	52.23	62.90	60.90	53.54	65.62	69.16
HA	55.14	67.02	62.00	62.83	67.10	66.76
SP	53.35	64.17	67.00	58.93	72.08	59.60
HC	48.22	50.23	62.20	47.67	63.96	56.05
mAP	60.67	68.16	69.90	69.56	75.36	73.01

66.76%。所提算法对飞机、网球场等小型目标及棒球场、环岛等大型目标均有较高的检测精度,而对小型车辆(SV)等缺乏特征的密集目标的检测精度低于BBAVectors和RoI-Trans。

考虑到上述实验中较多地引用了其他模型的实验数据,为了增加模型性能评估的客观性,在UCAS-AOD数据集上进行模型性能对比。表4为不同算法在UCAS-AOD数据集上的对比测试结果,其中带*的为引用数据。可以发现,CenterNet-RS的mAP可达91.10%,对飞机和车辆类的密集小型目标有着较好的检测效果。

为了探究使用方法的有效性和对模型性能的影响,进行消融实验。分别在原版CenterNet模型中加入FPN^[31]、FPN+RFEM、自适应特征融合结构(AFF)及FPN+可旋转检测头(Rotatable-Head),并与原版CenterNet进行对比。通过算法检测对各类物体的AP值,观察这些改进的效果。在其中三组消融实验的对照模型中加入FPN结构,是为了部署感受野扩展模块的同时控制模型结构上的变量,并检测AFF是否比传

表4 UCAS-AOD数据集上不同算法检测各类物体的AP值
Table 4 AP of different algorithms detecting various objects on UCAS-AOD dataset unit: %

Method	Backbone	AP		mAP
		Car	Airplane	
R2CNN*	VGG-16	78.89	89.76	84.32
ICN	ResNeXt101	85.02	90.32	87.67
CAD-Net	ResNet101	88.35	89.93	89.14
RoI-Trans*	ResNet101	87.99	89.90	88.95
BBAVectors	ResNet101	90.27	91.41	90.84
RetinaNet *	DLA-34	83.64	89.51	86.57
CenterNet	ResNet101	79.90	90.61	85.25
CenterNet-RS	ResNet101	89.37	92.82	91.10

统特征融合方法更具有效性。此外,CenterNet并非针对旋转目标检测而设计,作为对照的原版模型在训练和测试过程中剔除了数据集中目标的角度信息。为了探究角度识别对模型性能的直接影响,对仅安装Rotatable-Head的CenterNet进行测试。图7展示了测试模型的结构。

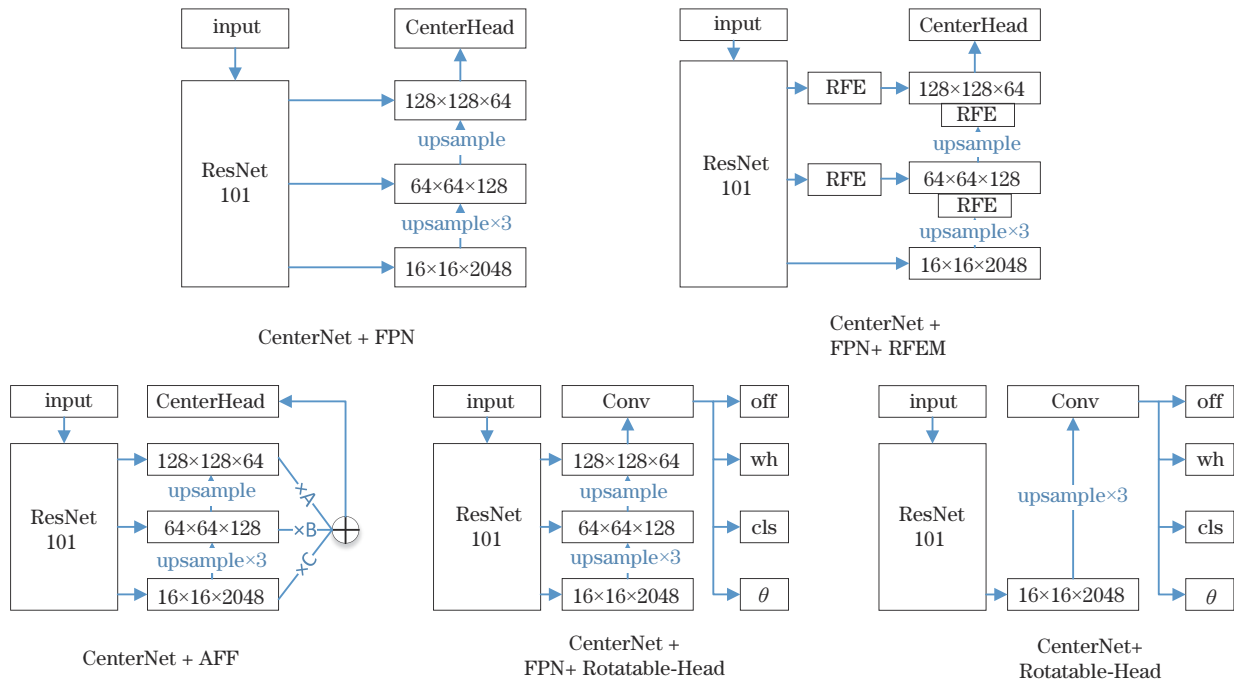


图7 消融实验中的模型结构

Fig. 7 Structure of models in the ablation experiment

表5记录了使用不同方法后在DOTA数据集下模型检测不同种类物体的AP值。可以看出:在检测过程中加入角度预测分支后,相对于CenterNet, Rotatable-Head对小型车辆(SV)、大型车辆(LV)和船舶(SH)等密集小目标的检测精度提升明显,分别提升了9.37个百分点、4.30个百分点、7.57个百分点,mAP提升了3.60个百分点;采用FPN进行特征融合后,模型的mAP提高了3.01个百分点,采用AFF进行

自适应特征融合后,mAP提高了4.21个百分点,说明这两种方法具备有效性;而感受野扩展模块更多地提升了对大型目标的检测效果,将RFEM部署到特征融合结构中,模型对游泳池(SP)和直升机(HC)的检测能力有所提升,相较于FPN,AP分别提升了6.37个百分点和6.51个百分点。总体而言,多尺度特征提取提升了模型对不同大小目标的检测性能,配合角度检测分支很好地弥补了模型在密集小目标上的不足。

表 5 消融实验中不同方法的 AP 值对比

Table 5 Comparison of AP values of different methods in ablation experiment

unit: %

Category	CenterNet	FPN	FPN+ RFEM	AFF	Rotatable-Head	FPN+Rotatable-Head	CenterNet-RS
PL	90.20	91.24	91.94	90.81	90.99	91.13	92.37
BD	68.13	71.12	73.44	72.75	69.77	71.86	74.86
BR	44.72	47.28	46.01	45.18	48.23	49.89	50.42
GTF	57.49	59.98	64.98	62.98	58.39	59.98	66.43
SV	54.44	54.69	56.47	56.44	63.81	63.16	64.49
LV	71.37	75.72	75.29	72.35	75.67	79.12	79.63
SH	66.39	68.54	69.93	68.80	73.96	75.21	76.74
TC	89.82	90.58	91.79	93.76	92.17	92.03	93.96
BC	63.22	66.90	71.55	70.86	69.96	72.74	75.27
ST	76.59	84.25	85.98	86.77	76.45	83.21	87.3
SBF	60.05	61.16	64.56	59.87	60.61	60.82	62.72
RA	62.82	65.48	67.63	65.43	64.95	66.71	69.16
HA	57.28	60.23	62.83	63.66	63.76	65.81	66.76
SP	47.97	51.26	57.63	55.00	50.92	53.31	59.60
HC	40.93	48.17	54.68	49.99	45.72	52.06	56.05
mAP	63.43	66.44	68.98	67.64	67.03	69.14	73.01

5 结 论

针对高分辨率遥感图像具有物体尺度差异大、小目标排列密集且方向性强的问题,提出了一种 CenterNet-RS 目标检测算法。首先,结合空洞卷积设计感受野扩展模块,并将其嵌入自适应特征融合结构中;利用多尺度感受野和特征融合,提取不同层次的语义信息,以提升模型对复杂环境中多尺度目标的检测能力;另外,在模型的检测部分加入角度分类分支,以加强对密集目标的识别能力。实验结果表明:与基线算法对比,CenterNet-RS 有效提升了对遥感图像目标检测的精度;且相较于其他经典的遥感目标检测算法,CenterNet-RS 在检测精度和实时性上均有一定的优势。单阶段检测算法带来了速度上的优势,但其在检测精度上具有一定局限性。相对于当今推出的遥感目标检测算法,CenterNet-RS 在 mAP 上仍有上升空间。接下来的研究中,将着重探索背景环境降噪和强化浅层语义提取的内容,以进一步提升对遥感图像中密集小目标的检测精度。

参 考 文 献

[1] 李旭冬, 叶茂, 李涛. 基于卷积神经网络的目标检测研究综述[J]. 计算机应用研究, 2017, 34(10): 2881-2886, 2891.
Li X D, Ye M, Li T. Review of object detection based on convolutional neural networks[J]. Application Research of Computers, 2017, 34(10): 2881-2886, 2891.

[2] Jiang Y Y, Zhu X Y, Wang X B, et al. R2CNN: rotational region CNN for orientation robust scene text detection[EB/OL]. (2017-06-29) [2022-02-05]. <https://arxiv.org/abs/1706.09579>.

[3] 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.

[4] 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.

[5] Yang X, Yan J, Feng Z, et al. R3det: refined single-stage detector with feature refinement for rotating object [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(4): 3163-3171.

[6] Qian W, Yang X, Peng S L, et al. Learning modulated loss for rotated object detection[EB/OL]. (2019-11-19) [2022-02-01]. <https://arxiv.org/abs/1911.08299>.

[7] 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.

[8] Uijlings J R R, van de Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

[9] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[EB/OL]. (2015-12-08) [2022-02-04]. <https://arxiv.org/abs/1512.02325>.

[10] 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.

[11] Ding J, Xue N, Long Y, et al. Learning RoI transformer for oriented object detection in aerial images[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern

- Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 2844-2853.
- [12] Ma T L, Mao M Y, Zheng H H, et al. Oriented object detection with transformer[EB/OL]. (2021-06-06)[2022-02-04]. <https://arxiv.org/abs/2106.03146>.
- [13] Law H, Deng J. CornerNet: detecting objects as paired keypoints[J]. International Journal of Computer Vision, 2020, 128(3): 642-656.
- [14] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points [EB/OL]. (2019-04-16)[2022-02-01]. <https://arxiv.org/abs/1904.07850>.
- [15] Zhang F, Wang X Y, Zhou S L, et al. Arbitrary-oriented ship detection through center-head point extraction[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5612414.
- [16] Yi J R, Wu P X, Liu B, et al. Oriented object detection in aerial images with box boundary-aware vectors[EB/OL]. (2020-08-17)[2022-02-04]. <https://arxiv.org/abs/2008.07043>.
- [17] 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.
- [18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [19] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015-11-23)[2022-02-04]. <https://arxiv.org/abs/1511.07122>.
- [20] Luo W J, Li Y J, Urtasun R, et al. Understanding the effective receptive field in deep convolutional neural networks[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain. New York: ACM Press, 2016: 4905-4913.
- [21] Liu S T, Huang D, Wang Y H. Learning spatial fusion for single-shot object detection[EB/OL]. (2019-11-21)[2022-02-04]. <https://arxiv.org/abs/1911.09516>.
- [22] Neubeck A, Van Gool L. Efficient non-maximum suppression[C]//18th International Conference on Pattern Recognition, August 20-24, 2006, Hong Kong, China. New York: IEEE Press, 2006: 850-855.
- [23] Yang X, Yan J C. Arbitrary-oriented object detection with circular smooth label[EB/OL]. (2020-03-12)[2021-02-04]. <https://arxiv.org/abs/2003.05597>.
- [24] 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.
- [25] MacQueen J. Some methods for classification and analysis of multivariate observations[EB/OL]. <http://www.cs.cmu.edu/~bhiksha/courses/mlsp.fall2010/class14/macqueen.pdf>.
- [26] Zhu H G, Chen X G, Dai W Q, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]//2015 IEEE International Conference on Image Processing, September 27-30, 2015, Quebec City, QC, Canada. New York: IEEE Press, 2015: 3735-3739.
- [27] Loshchilov I, Hutter F. SGDR: stochastic gradient descent with warm restarts[EB/OL]. (2016-08-13)[2022-02-04]. <https://arxiv.org/abs/1608.03983>.
- [28] Azimi S M, Vig E, Bahmanyar R, et al. Towards multi-class object detection in unconstrained remote sensing imagery[EB/OL]. (2018-07-07) [2022-02-03]. <https://arxiv.org/abs/1807.02700>.
- [29] Zhang G J, Lu S J, Zhang W. CAD-net: a context-aware detection network for objects in remote sensing imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(12): 10015-10024.
- [30] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.
- [31] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2117-2125.
- [32] 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.
- [33] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.