

引用格式: ZHANG Xiao-han, YAO Li-bo, LÜ Ya-fei, *et al.* Center Based Model for Arbitrary-oriented Ship Detection in Remote Sensing Images[J]. *Acta Photonica Sinica*, 2020, 49(4):0410005

张筱晗,姚力波,吕亚飞,等.基于中心点的遥感图像多方向舰船目标检测[J].光子学报,2020,49(4):0410005

基于中心点的遥感图像多方向舰船目标检测

张筱晗¹,姚力波¹,吕亚飞¹,韩鹏²,李健伟³

(1 海军航空大学 信息融合研究所,山东 烟台 264000)

(2 海军 91039 部队,北京 102488)

(3 海军 92877 部队,浙江 舟山 316000)

摘 要:针对目前基于深度学习的舰船目标斜框检测方法存在计算量大、效率低的问题,提出一种基于目标中心点的单阶段检测模型.由于舰船中心点不受舰船分布方向影响,模型主要思想是以目标中心点检测为基础,回归中心点处目标斜框的尺度和方向.首先设计特征提取网络,将卷积神经网络细节信息丰富的底层特征与语义信息丰富的高层特征融合起来形成特征图;然后将特征图输入到三个检测分支,分别预测目标中心点、中心点偏移值以及斜框的尺度与方向;设计组合损失函数对网络进行训练,并改进非极大值抑制算法以适应目标斜框检测的需要.在公开的 SAR 图像舰船目标检测数据集与光学遥感图像上进行了实验,实验结果表明,测试集平均准确率达 0.906,检测精度与速度均优于其它检测模型,充分验证了所提算法的有效性.

关键词:目标检测;图像分析;深度学习;遥感图像;舰船目标;特征提取

中图分类号:TP751.1

文献标识码:A

doi:10.3788/gzxb20204904.0410005

Center Based Model for Arbitrary-oriented Ship Detection in Remote Sensing Images

ZHANG Xiao-han¹, YAO Li-bo¹, LÜ Ya-fei¹, HAN Peng², LI Jian-wei³

(1 *Information Fusion Institute, Naval Aviation University, Yantai, Shandong 264000, China*)

(2 *Troops of 91039, Beijing 102488, China*)

(3 *Troops of 92877, Zhoushan, Zhejiang 316000, China*)

Abstract: The recent proposed deep learning-based arbitrary-oriented objects detection algorithms increase extra computation burden and could not work efficiently. A one-stage model based on object centers detection is proposed for arbitrary-oriented ship detection. As the centers of objects are free from the influence their distribution directions, the key of the model is to regress the parameters of object's oriented bounding box on the basis of center detection. Firstly, a feature extracting network is designed to achieve feature map and a new feature fusion method is proposed which aggregates the low-level features rich in detailing information and high-level features rich in semantic information together. Then the feature map is entered to three detection branches, which predict of centers, offsets of centers, and size and direction of the oriented bounding boxes respectively. A combined loss function is proposed for the training of the network, and a modified non-maximum suppression algorithm is proposed for removing invalid oriented bounding boxes. The proposed model achieves state-of-art performance in public SAR ship detection dataset with mean average precision as 0.906, outstanding than other methods both in

基金项目:国家自然科学基金(Nos. 91538201, 61790554, 61971432, 61531020)

第一作者:张筱晗(1992-),女,博士研究生,主要研究方向为遥感图像目标检测与识别、深度学习. Email: Xhan_zhang@163.com

通讯作者:姚力波(1980-),男,副教授,博士,主要研究方向为卫星遥感图像处理、信息融合. Email: ylb_rs@126.com

收稿日期:2019-12-30;录用日期:2020-02-03

speed and precision.

Key words: Object detection; Image analysis; Deep learning; Remote sensing images; Ship; Feature extraction

OCIS Codes: 100.3008; 100.2960; 100.4996

0 引言

随着我国海洋战略逐步从近海走向远洋,提升海洋环境的感知和监视能力成为建设海洋强国的重要保障.其中,基于遥感图像的海上舰船目标提取是进行海洋监视的一项基础技术.如何实现遥感图像中舰船目标快速精确的检测成为研究热点.

目标分布呈多方向是遥感图像的一大特点.空天基遥感传感器获取的是地物顶部空间特征,因此在遥感图像中,舰船目标的分布是呈任意方向的^[1].从遥感图像中提取多方向舰船目标通常有三种方式:目标最小外接垂直矩形框^[2]、目标最小外接矩形框^[3]以及像素级目标分割^[4].通用目标检测任务中通常使用第一种方式,但这种检测框不能反映舰船的方向,对于密集排列的舰船,不同目标的检测框之间存在较大的重叠区域;像素级分割能够最精确地提取目标,但实例分割的标注代价远远大于目标检测,限制了其在实际中的应用.以目标外接倾斜矩形框作为目标检测框,可以视为通用目标检测任务到目标分割任务的过渡,不仅能更精准地框定出目标的边界,还能反映出目标的方向、尺度、长宽比等信息,用于后续目标识别等任务;此外,倾斜矩形框的标注代价与垂直矩形框的标注类似,远小于目标分割任务.因此,对遥感图像多方向舰船目标来说,采用倾斜矩形边界框(简称“斜框”)进行检测更加实用、合理.

2012年以来,深度学习技术在计算机视觉中大放异彩^[5-6],通用目标检测任务中出现了以 Faster Region Convolutional Neural Network (Faster R-CNN)^[7]、Single Shot Detector (SSD)^[8]、You Only Look Once (YOLO)^[9]等为代表的经典目标检测模型.作为目标检测任务的一大分支,多方向目标检测受到关注,其在场景文字识别、遥感目标检测等多领域得到应用^[10-12].基于深度学习的遥感图像舰船目标斜框检测算法多从通用目标检测模型发展而来,2017年 LIU Zi-kun 等^[3]基于 Faster RCNN 提出 Rotated Region based CNN (RR-CNN)模型,用于光学图像多方向舰船目标的检测,模型设计了旋转边框(Rotated Bounding Box, RBB)回归模块,用于回归目标斜框的五个要素;随后, XUE Yang 等^[13-14]在区域建议网络(Region Proposal Network, RPN)锚框的设置中增加了角度约束,用不同方向、不同尺度、不同长宽比的一组锚框在特征图上进行遍历,较 RR-CNN 提升了检测效果;WANG Ji-zhou 等^[15]将增加了角度约束的锚框设计应用在 SSD 模型中;李健伟^[2]则设计了循环金字塔特征融合网络与多比例可旋转感兴趣区域池化模块,实现合成孔径雷达(Synthetic Aperture Radar, SAR)图像多方向舰船检测.以上方法都基于目标检测中的锚框遍历机制,但对于多方向目标检测来说,锚框遍历机制并非最高效的选择.首先,锚框机制依赖于一定数量的预设好尺寸、比例等属性的锚框在特征图上进行遍历,生成潜在目标感兴趣区域,通用目标检测任务中设置的垂直锚框难以拟合水平、垂直方向以外的目标;其次,若增加锚框的角度属性,虽然能够解决部分多方向目标检测问题,但锚框总数成倍增加^[1],增加了运算负担.LIU 等^[16]在 YOLO 模型的基础上回归目标斜框参数,避免了锚框遍历机制的缺点,但 YOLO 模型将整幅图像均分为若干网格,每个网格仅负责预测固定数量的中心点落在网格中的目标,因此不能很好地检测密集分布目标,检测精度较 RR-CNN 提升有限.

遥感图像中目标分布虽然呈现多个方向,但是目标的中心点位置不受方向影响.受通用目标检测任务中的 Centernet 模型^[17]启发,目标中心点可以在卷积特征图上直接进行预测.因此,本文采用中心点检测这一思路实现多方向舰船目标的斜框检测.首先,设计网络预测目标的中心点;然后对中心点处目标斜框的尺度、方向等参数进行回归,从而实现目标的斜框预测;特征提取、参数预测均通过卷积操作完成,并设置组合损失函数对网络进行训练.本文模型摆脱了对锚框机制的依赖,不仅保持了单阶段目标检测模型的速度优势,还通过改进的特征表征实现了高精度检测;与 YOLO 模型相比,在预测中心点时不受网格分布以及各网格预测目标数量的限制,能够更灵活地对密集分布目标实施检测.

1 算法

1.1 整体框架

所提模型整体框架如图 1 所示.模型分为三部分,第一部分为特征提取与融合模块,利用卷积神经网络

提取图像特征并对分层特征进行融合,得到融合特征图;第二部分为斜框预测模块,融合特征图进入三个预测分支,分别预测目标中心点,目标的宽、高和方向以及中心点的坐标偏移,生成初步预测结果;第三部分对第二部分预测结果进行处理,主要采用了针对斜框的非极大值抑制(Non-Maximum Suppression, NMS)算法得到最终检测结果。

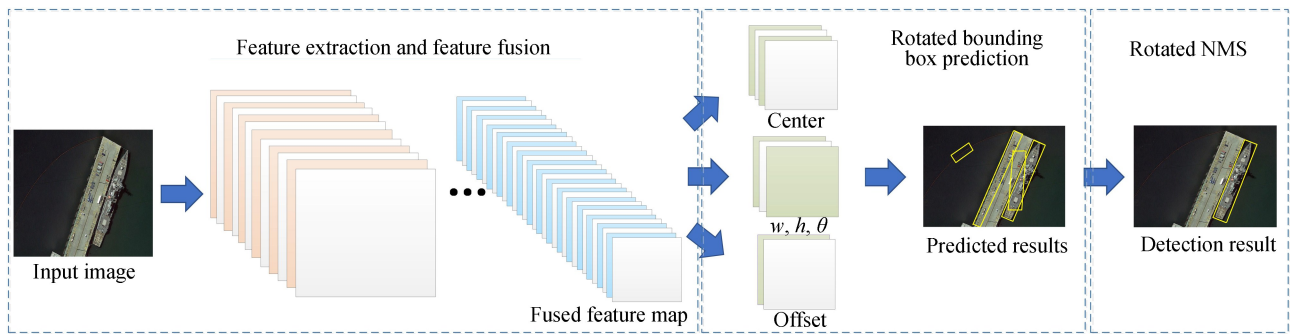


图 1 本文算法框架

Fig.1 The architecture of the proposed detection network

1.2 特征提取网络

整个模型的输入图像大小为 $W \times H \times N$, 经过特征提取网络得到特征图. 假设特征图大小为 $\frac{W}{R} \times \frac{H}{R} \times N_{out}$, R 为原图到特征图的降采样率. 卷积神经网络中的池化(pooling)操作^[5]实现了对特征的降采样, 减少了网络参数, 同时也增大了同尺度卷积核的局部感受野. 因此, 卷积神经网络底层特征空间分辨率高, 包含更多图像局部细节信息, 而高层获取的特征空间分辨率低, 通常代表图像的全局、抽象语义信息. 对于目标检测任务来说, 底层空间信息、高层语义信息都至关重要. 因此, 在基础特征提取网络中增加了特征融合模块, 聚合高低层特征的优势. 特征提取网络框架如图 2 所示(图中以 $R=4$ 为例).

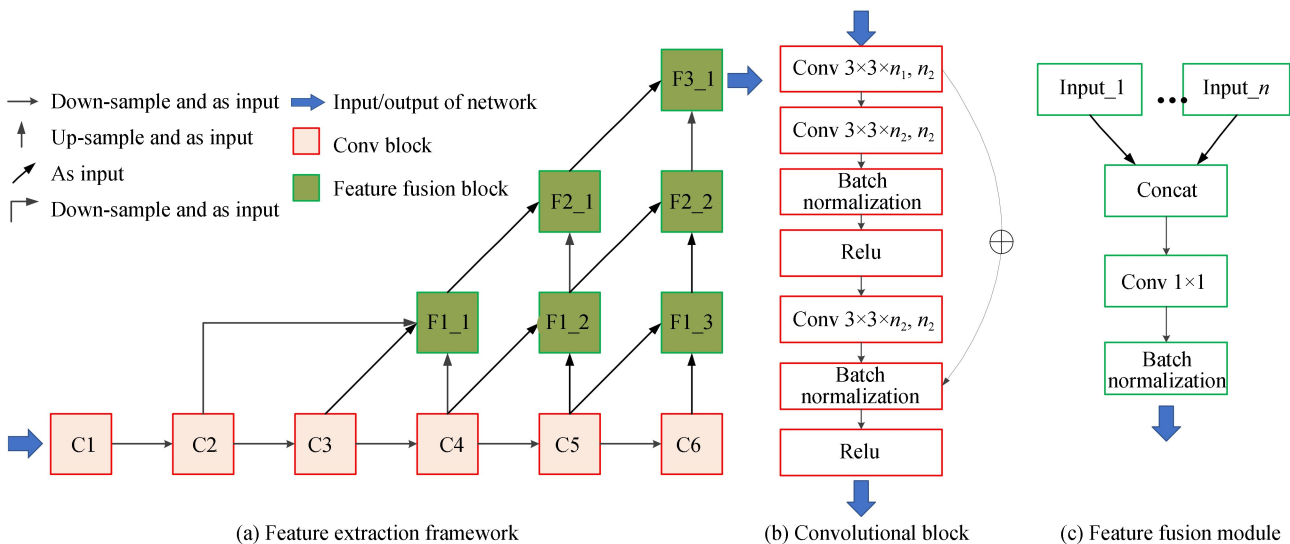


图 2 特征提取网络

Fig.2 The architecture of feature extracting network

图 2(a)底部框 $C1 \sim C6$ 与箭头代表基础特征提取网络, 其中每个框代表一个卷积模块, 结构借鉴 Resnet^[18] 跳层连接机制, 具体组成如图 2(b)所示; 横向箭头代表最大池化(max pooling), 将特征图空间分辨率降为输入的二分之一; F 框代表特征聚合块, 如图 2(c)所示, 首先将待融合的特征层通过降采样或上采样处理, 统一空间分辨率, 然后将它们在通道维串联起来, 通过 1×1 卷积层调整通道数量, 再经过批标准化处理后得到融合结果. 特征层距离越大, 其提取的特征差异性越大, 因此本文方法只对相邻卷积块输出的特征进行融合, 再对融合特征进行逐级融合, 具体过程如图 2(a)所示. 与经典的特征金字塔网络 (Feature Pyramid Network, FPN)^[19] 仅融合相邻特征层的方式相比, 本文方法融合了更多的卷积特征层, 从而使最终

用于检测分支的特征中包含更多底层局部细节信息与高层全局语义信息;此外,逐级融合的方式在保持一定特征空间分辨率的同时,避免了将差异较大的底层特征与高层特征直接进行聚合。

1.3 检测分支

在二维平面中确定一个倾斜矩形,最少需要 5 个要素,采用倾斜矩形框中心点坐标 (x, y) 、矩形框宽 w 、高 h 以及矩形框长边与水平轴夹角 θ ($\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$) 这五个要素来表示一个倾斜矩形框.特征提取网络得到的融合特征图分别进入三个分支,分别输出目标中心点、目标斜框尺寸方向以及中心点偏移的预测值.各分支均由 $k \times k$ 卷积、Relu 激活函数、 1×1 卷积层组成,其中 k 决定了预测分支卷积核的感受野, k 设置不同的值会对检测结果产生一定影响。

1.3.1 中心点预测分支

假设要检测的目标类别共有 C 类,则中心点预测分支的输出大小为 $\frac{W}{R} \times \frac{H}{R} \times C$, C 个通道中每个通道代表一个类别,空间维上每一点(对应原图同一位置处 $R \times R$ 个点的区域)的值代表该位置包含目标中心点的概率,大小在 $[0, 1]$ 区间.因此,该分支 1×1 卷积层后的输出需要经过 Sigmoid 激活函数进行归一化.Sigmoid 函数定义为

$$\text{sigmoid}(x) = \frac{1}{1 + \exp(-x)} \quad (1)$$

训练网络时借鉴文献 [20] CornerNet 训练的思想,每个类别目标各生成一个大小为 $\frac{W}{R} \times \frac{H}{R}$ 的真值热图.对于类别 C_i 下的目标 T ,其中心点在原图坐标为 (x_T, y_T) .将其映射到真值热图中,坐标为 $\left(\left\lfloor \frac{x_T}{R} \right\rfloor, \left\lfloor \frac{y_T}{R} \right\rfloor\right)$ (“ $\lfloor \cdot \rfloor$ ”表示向下取整操作),热图中对应坐标位置处的值为 1,在训练中为正样本.热图中不包含目标中心点的位置取值本应为 0,在训练中作为负样本.但是由于一幅图像中目标数量是有限的,正负样本数量通常极不平衡.为了平衡正负样本,以热图中目标的中心点为圆心,使用高斯核函数对热图中其余位置 (x, y) 的值 Y_{xyc} 进行处理,即

$$Y_{xyc} = \exp \left[-\frac{\left(x - \left\lfloor \frac{x_T}{R} \right\rfloor\right)^2 + \left(y - \left\lfloor \frac{y_T}{R} \right\rfloor\right)^2}{2\sigma_p^2} \right] \quad (2)$$

式中, σ_p 为自适应目标尺度因子,本文中取作目标短边的 0.1 倍.若同一类别多个目标的高斯圆有重叠,则重叠位置的值取其中的最大值。

中心点预测图与真值热图的像素一一对应,训练时采用带有惩罚项的逻辑回归损失^[21],该分支产生的损失为

$$L_c = -\frac{1}{N_T} \sum_{xyc} \begin{cases} (1 - \hat{Y}_{xyc})^\alpha \log(\hat{Y}_{xyc}) & \text{if } Y_{xyc} = 1 \\ (1 - \hat{Y}_{xyc})^\beta (\hat{Y}_{xyc})^\alpha \log(1 - \hat{Y}_{xyc}) & \text{else} \end{cases} \quad (3)$$

式中, α, β 为人为设定的惩罚系数, N_T 为图像中目标的总个数。

1.3.2 中心点偏移值预测分支

由于真实图像到中心点预测图空间分辨率下降了 R 倍,因此,将 1.3.1 节中心点位置映射到原图时,对应的是原图中相应位置处 $R \times R$ 大小区域,而不是中心点的精确位置.目标中心点从真值热图映射到原图像,水平、垂直方向的偏移值分别为 $p_{\text{off}_x} = x_T - \left\lfloor \frac{x_T}{R} \right\rfloor$, $p_{\text{off}_y} = y_T - \left\lfloor \frac{y_T}{R} \right\rfloor$.

中心点偏移值预测分支输出的大小为 $\frac{W}{R} \times \frac{H}{R} \times 2$,两个通道分别对应中心点预测图到原图中心点坐标在 x 轴、 y 轴的位移量预测 \hat{p}_{off_x} 、 \hat{p}_{off_y} .训练时,采用 L1 距离损失函数,即

$$L_{\text{off}} = \frac{1}{N_T} \sum (|\hat{p}_{\text{off}_x} - p_{\text{off}_x}| + |\hat{p}_{\text{off}_y} - p_{\text{off}_y}|) \quad (4)$$

其中,只有目标中心点位置处的偏移预测值才参与训练,其余位置的偏移预测值训练中不产生损失。

1.3.3 斜框尺度方向预测分支

该分支的输出大小为 $\frac{W}{R} \times \frac{H}{R} \times 3$,三个通道分别对应目标斜框预测值的宽 \hat{w} 、高 \hat{h} 以及方向角度 $\hat{\theta}$ 。训练中同样采用 L1 距离损失函数

$$L_{size} = \frac{1}{N_T} \sum_{k=1}^N (|\hat{w} - w_t| + |\hat{h} - h_t| + |\hat{\theta} - \theta_t|) \quad (5)$$

式中, w_t 、 h_t 和 θ_t 为目标真值斜框的宽、高、角度。与偏移预测值一样,只有目标中心点处的尺度方向预测值参与训练。

1.4 训练与推理

模型网络的整体训练损失为 1.3 节中三个损失函数的加权和,公式为

$$L = L_c + \lambda_{off} L_{off} + \lambda_{size} L_{size} \quad (6)$$

式中, λ_{off} 、 λ_{size} 为损失权重系数。利用梯度下降法反向传播更新网络参数,实现对整个网络的训练。

模型的测试推理中,首先选择中心点概率得分阈值最大的前 K 个点作为预测目标的中心点,对于某个预测中心点 (\hat{x}, \hat{y}) ,于偏移值预测分支、斜框尺寸方向预测分支的输出图对应位置获取各中心点对应的 \hat{p}_{off_x} 、 \hat{p}_{off_y} 、 \hat{w} 、 \hat{h} 及 $\hat{\theta}$,分别计算得到该中心点位置目标框四个角点的预测坐标值 (\hat{x}_1, \hat{y}_1) 、 (\hat{x}_2, \hat{y}_2) 、 (\hat{x}_3, \hat{y}_3) 、 (\hat{x}_4, \hat{y}_4) ,即

$$\begin{cases} \hat{x}_1 = \hat{x} + 4 \times \hat{p}_{off_x} - l \times \cos(\hat{\theta} - \rho) \\ \hat{y}_1 = \hat{y} + 4 \times \hat{p}_{off_y} + l \times \sin(\hat{\theta} - \rho) \end{cases} \quad (7)$$

$$\begin{cases} \hat{x}_2 = \hat{x} + 4 \times \hat{p}_{off_x} - l \times \cos(\hat{\theta} + \rho) \\ \hat{y}_2 = \hat{y} + 4 \times \hat{p}_{off_y} + l \times \sin(\hat{\theta} + \rho) \end{cases} \quad (8)$$

$$\begin{cases} \hat{x}_3 = \hat{x} + 4 \times \hat{p}_{off_x} + l \times \cos(\hat{\theta} - \rho) \\ \hat{y}_3 = \hat{y} + 4 \times \hat{p}_{off_y} - l \times \sin(\hat{\theta} - \rho) \end{cases} \quad (9)$$

$$\begin{cases} \hat{x}_4 = \hat{x} + 4 \times \hat{p}_{off_x} + l \times \cos(\hat{\theta} + \rho) \\ \hat{y}_4 = \hat{y} + 4 \times \hat{p}_{off_y} - l \times \sin(\hat{\theta} + \rho) \end{cases} \quad (10)$$

式中, $l = \frac{\sqrt{\hat{w}^2 + \hat{h}^2}}{2}$, $\rho = \arctan\left(\frac{\hat{w}}{\hat{h}}\right)$ 。

1.3 节介绍的检测方法可能存在对同一目标的重复检测。在通用目标检测模型中,通常采用基于矩形框交并比(Intersection over Union, IOU)的非极大值抑制(Non-Maximum Suppression, NMS)算法^[22]解决这类问题。借鉴 NMS 算法思想,提出倾斜矩形框的 NMS 算法对重叠检测框进行取舍。

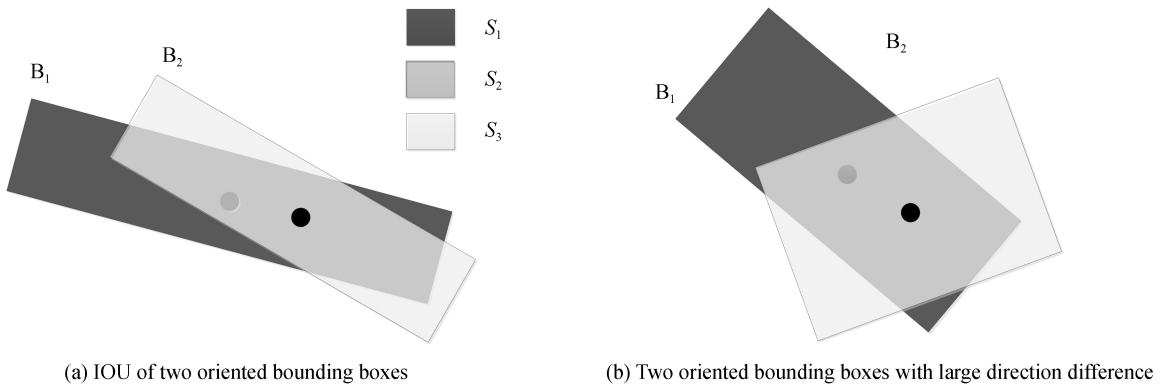


图3 两斜框间 IOU 图示

Fig.3 Illustration of the IOU of two oriented bounding boxes

首先定义两个斜框的交并比,如图 3(a)所示,斜框 IOU 定义为

$$\text{IOU}_{B_1, B_2} = \frac{S_{B_1 \cap B_2}}{S_{B_1 \cup B_2}} = \frac{S_2}{S_1 + S_2 + S_3} \quad (11)$$

其中, B_1 、 B_2 代表两个斜框, S 代表图 3(a) 对应区域的面积。

根据 NMS 算法原理,只要两个检测框的 IOU 大于一定阈值,即可认为二者对应的是同一个目标,需通过一定规则舍弃其中一个检测框.但如图 3(b)所示,即使两个斜框的 IOU 较大,当两个矩形框的方向差别比较大,不能视为检测的是同一目标.因此,设计斜框 NMS 规则如下:设置 IOU 阈值 φ_{IOU} 、角度差阈值 φ_θ ,对于两个存在重叠区域的检测框,首先计算其交并比,若交并比大于 φ_{IOU} ,则计算其方向角度差值,若差值小于 φ_θ ,则舍弃中心点概率预测得分较低的目标框。

2 实验验证与结果分析

2.1 数据集与实验环境

为验证本文算法的有效性,在 SAR 图像舰船目标检测数据集(SAR Ship Detection Dataset, SSDD)^[18] 进行实验.SSDD 数据集是国内第一个公开的用于深度学习算法验证的 SAR 舰船目标检测数据集,在很多文献^[15,23]中得到应用,是 SAR 遥感图像目标检测最有影响力的基准数据集之一.SSDD 中共有 1 160 幅多分辨率 SAR 图像,按照 7 : 2 : 1 的比例划分了训练集、测试集和验证集,包含 2 456 个舰船目标,标签中给出了每个目标最小外接矩形的四个顶点坐标.选择 SAR 图像来验证本文算法,是因为与光学图像相比,SAR 图像目标细节更少,实施检测时更有挑战性.此外,也在高分辨率光学图像、低分辨率多光谱图像上进行了验证。

实验平台为搭载了 64 位 Ubuntu16.04 操作系统的计算机,显卡型号为 NVIDIA GTX2080Ti,使用 CUDA10.0 和 cuDNN7.0 进行加速,编程语言为 Python,并使用 Pytorch 深度学习框架编写代码.在实验中,由于中心点预测分支中逻辑回归损失惩罚系数 α 、 β 以及训练损失中中心点偏移权重 λ_{off} 、斜框尺度权重 λ_{size} 设置不受使用的检测数据类型以及检测框类型影响,其设置参考文献通用检测模型 Centernet^[17] 的实验设置,取 $\alpha = 2$ 、 $\beta = 4$ 、 $\lambda_{\text{off}} = 0.1$ 、 $\lambda_{\text{size}} = 1$.参数 R 代表原图到预测图的降采样率,取值小有利于密集目标检测,但计算负担大,取值大能够减少计算量,但也降低了检测精度.为平衡检测效率与精度,实验中取 $R = 4$.此外,训练中迭代次数设为 300 次,初始学习率设为 0.001,每迭代 50 次学习率下降 50%。

2.2 评价指标

实验中,当检测结果与真值边界框的 IOU 超过 0.5 即可认为检测正确.采用准确率(precision)、召回率(recall)、平均准确率(mean Average Precision, mAP)作为算法评价指标.其中,准确率、召回率在中心点得分阈值为 0.5、目标与真值 IOU 阈值为 0.5、角度阈值为 $\pi/6$ 的设置下计算.mAP 的定义为

$$\text{mAP} = \int_0^1 P(R) dR \quad (12)$$

式中, P 为准确率, R 为召回率.mAP 指标能够更全面地衡量检测算法在多召回率下的精度, mAP 值越大,算法性能越好。

2.3 实验结果与分析

2.3.1 不同特征提取网络对结果的影响

为了验证 1.2 节所设计的特征提取网络的有效性,分别采用经典的 ResNet18^[19]、ResNet + FPN 网络^[20]、DLANet^[24] 以及本文方法进行特征提取,检测分支卷积核设为 $k = 3$.训练时,由于 SSDD 中图像长边长多为 400~600,因此训练中调整输入图像大小为 512×512 (降采样特征图空间维大小为 128×128),使用上述 4 类网络提取特征,其中,ResNet18 不进行特征融合,输出特征通过上采样的方式放大至指定尺度.将四个网络的特征图分别输入到三个检测分支,训练四个检测网络,然后在测试集上进行测试,测试结果如表 1 所示(最佳结果用粗体标出)。

由于检测网络在单一分辨率的特征图上实施检测,特征融合能够有效聚合不同分辨率特征信息从而提升检测精度.实验结果表明,FPN、DLANet 与本文方法的融合策略都是有效的,其中,本文设计的融合特征提取网络取得了最好的检测结果。

表 1 不同特征提取网络检测结果

Feature extraction network	Precision	Recall	mAP
ResNet18	0.525	0.725	0.653
ResNet+FPN	0.687	0.810	0.750
DLANet	0.847	0.893	0.867
Proposed method	0.852	0.901	0.887

2.3.2 检测分支不同卷积核对结果的影响

检测分支的卷积核大小决定了网络预测中心点位置、目标斜框参数时的感受野,分别设置 $k=3$ 、 $k=5$ 、 $k=7$ 三种参数进行实验,实验结果如表 2 所示。

表 2 检测分支不同卷积核设置下的检测结果

Settings of kernel size	Precision	Recall	mAP
$k=3$	0.852	0.901	0.887
$k=5$	0.856	0.928	0.905
$k=7$	0.861	0.899	0.882

从表中可以看出,卷积核的设置对结果有一定影响, k 取值太小,对大目标检测不利,而 k 取值太大又不利于小目标的检测,当 $k=5$ 时,检测模型取得了本文算法在 SSDD 数据集的最佳 mAP。

2.3.3 不同斜框检测方法的对比

为进一步验证本文方法的检测效果,将其与近年来典型斜框检测算法进行对比,包括 R-DFPN^[13]、M-R2CNN 模型^[14]、SAR-SDBRR 模型^[2] 以及 AMAR-SDD 模型^[15] 等方法,采用这几个模型在文献中的最佳设置,与本文方法的最佳设置进行对比,结果如表 3 所示。同时,为了比较各模型的运算速度,统计了各模型在单张图像上的平均训练、测试时间。

表 3 不同模型检测精度与速度比较

Models	Precision	Recall	mAP	Training time/ms	Testing time/ms
R-DFPN	0.828	0.863	0.835	1211	371
M-R2CNN	0.826	0.872	0.841	658	193
SAR-SDBRR	0.863	0.905	0.903	981	277
AMAR-SDD	0.835	0.861	0.842	251	71
Proposed method	0.856	0.928	0.905	103	65

从表中可以看出,本文方法与 SAR-SDBRR 模型在检测精度上有着明显优势,但是,SAR-SDBRR 与前两个模型都基于二阶段检测模型 Faster R-CNN,运行速度相对较低;AMAR-SDD 是基于单阶段检测模型 SDD 设计的,运行速度较前三个模型有了较大提升,但精度相对不高,他们都是基于锚框遍历的算法,本文方法在保持精度优势的同时,也保持了速度优势,综合来说,在 SSDD 数据集上进行目标斜框检测时取得了最佳性能。

2.4 检测结果展示

除了 SAR 图像,本文算法还利用 Google Earth 高分辨率光学图像训练模型并进行了测试,选取多分辨率下多场景典型图像的检测结果如图 4 所示,其中图 4(a)为多分辨率 SAR 图像,图 4(b)为高分辨率光学图像。从图中可以看出,本文检测算法对目标斜框有较高的回归精度,检测框能够准确地包围目标,分离目标与背景,斜框的角度能够代表舰船目标的方向,而斜框的长度、宽度也能反映舰船目标的实际尺度,检测结果反映模型能够适用于多分辨率图像、多场景下的多尺度舰船目标检测,并且对密集排列目标有着较好的检测效果,但是模型也存在一定的目标漏检情况。

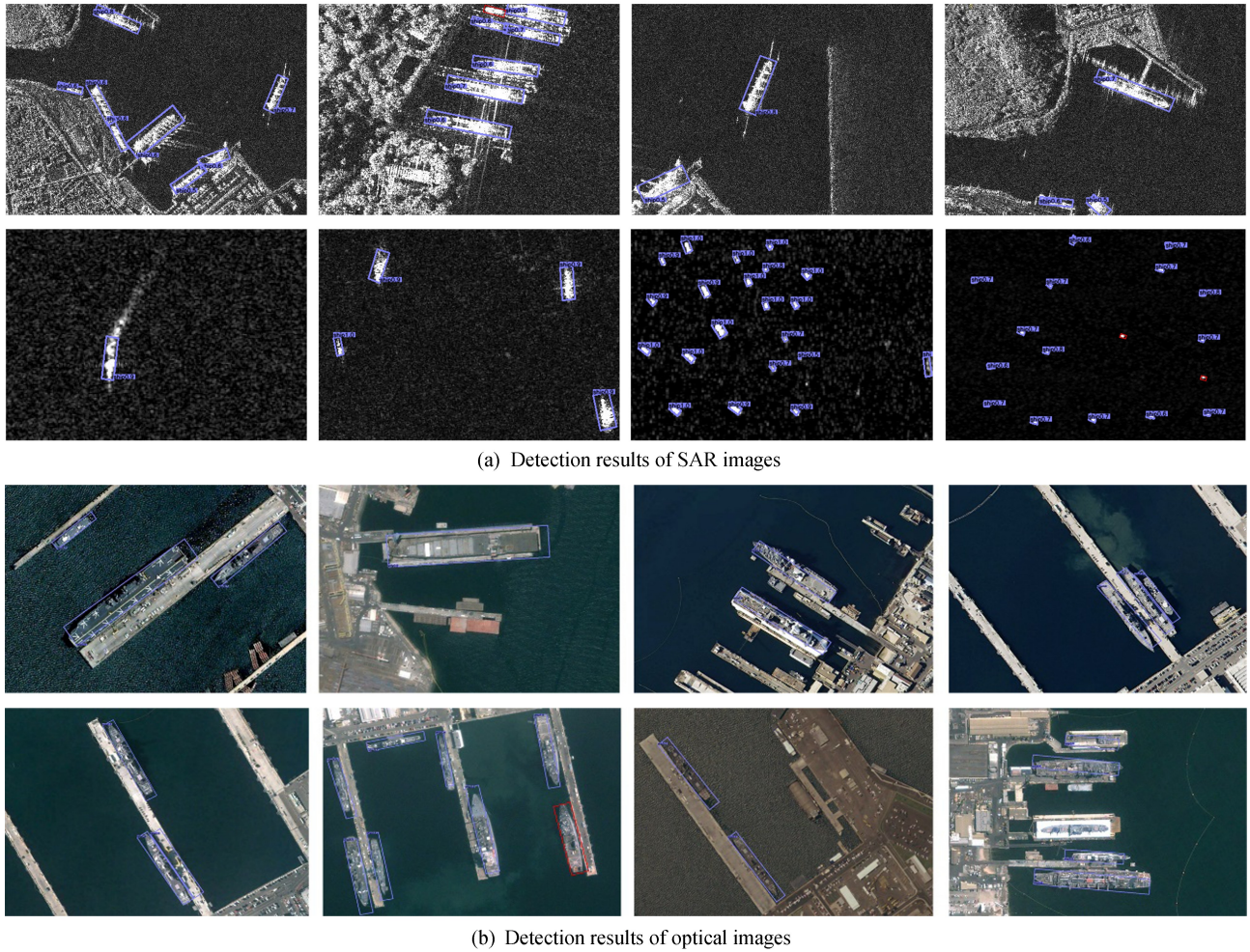


图4 本文方法在 SAR 与光学图像中目标检测结果(红色为漏检目标)

Fig.4 The detection results in SAR images and optical images using the proposed method (the red bounding boxes represent missing ships)

3 结论

针对遥感图像中多方向舰船目标检测的问题,本文提出了一种新的基于中心点的单阶段目标斜框检测模型.与现有的检测算法相比,该方法不依赖于深度学习方法目标检测模型中经典的锚框遍历机制,在检测目标中心点的基础上,对多方向舰船目标斜框的相关参数进行回归实现检测.针对斜框检测任务的特点设计了特征提取网络并对高低层特征进行融合,优化了深度卷积特征学习,通过实验验证了该网络提取的特征具有较强的表征能力.实验表明,与其他经典模型相比,该方法不仅有较高的检测精度,还保持了单阶段检测模型的速度优势.在 SAR 图像与光学遥感图像中的实验结果说明所提舰船目标斜框检测模型能够适用于多分辨率遥感图像多场景下多尺度舰船目标检测,具有实用价值.本文算法在 SSDD 数据集中测试中最佳召回率为 92.8%,高于其他模型,但仍存在一些漏检、误检情况.此外,受使用数据集标注限制,实验测试中没有统计各尺度下舰船目标的具体检测精度.如何进一步优化特征表征、提升模型在复杂场景遥感图像中的检测效果是后续工作的重点研究方向.

参考文献

- [1] DING Jian, XUE Nan, LONG Yang, *et al.* Learning RoI transformer for detecting oriented objects in aerial images[C]. 2019 Conference on Computer Vision and Pattern Recognition (CVPR 2019), USA, 2019:1-10.
- [2] LI Jian-wei. Research on ship detection and recognition in SAR images based on convolutional neural networks[D]. Yantai: Naval Aviation University, 2019.
李建伟. 基于卷积神经网络的 SAR 图像舰船目标检测与识别技术研究[D]. 烟台:海军航空大学, 2019.
- [3] LIU Zi-kun, HU Jian-gao, WENG Lu-bin, *et al.* Rotated region based CNN for ship detection[C]. IEEE International

- Conference on Image Processing, China, 2017: 1-10.
- [4] ZHANG Xiao-qing, XIONG Bo-li, DONG Gang-gang, *et al.* Ship segmentation in SAR images by improved nonlocal active contour model[J]. *Sensors*, 2018, **18**(12): 4220-4240.
- [5] SZEGEDY C, LIU Wei, JIA Yang-qing, *et al.* Going deeper with convolutions[C]. 2015 Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, USA, 2015: 1-9.
- [6] SENGUPTA A, YE Y, ROBERT W. Going deeper in spiking neural networks: VGG and residual architectures[J]. *Frontiers in Neuroscience*, 2018, **13**: 1-12.
- [7] REN Shao-qing, HE Kai-ming, 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.
- [8] LIU Wei, ANGUELOV D, ERHAN D, *et al.* SSD: single shot multibox detector[C]. 2016 European Conference on Computer Vision(ECCV2016), Amsterdam, The Netherlands, 2016: 21-37.
- [9] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection[C]. 2015 Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, USA, 2015: 779-788.
- [10] MA Jian-qi, SHAO Wei-yuan, YE Hao, *et al.* Arbitrary-oriented scene text detection via rotation proposals[J]. *IEEE Transactions on Multimedia*, 2018, **20**(11): 3111-3122.
- [11] CHENG Gong, ZHOU Pei-cheng, HAN Jun-wei. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2016, **54**(12):7405-7415.
- [12] XIA Gui-song, BAI Xiang, DING Jian, *et al.* Dota: A large-scale dataset for object detection in aerial images[C]. Conference on Computer Vision and Pattern Recognition (CVPR 2018), USA, 2018: 1-10.
- [13] XUE Yang, SUN Hao, FU Kun, *et al.* Automatic ship detection in remote sensing images from Google Earth of complex scenes based on multiscale rotation dense feature pyramid networks[J]. *Remote Sensing*, 2018, **10**(1):132-145.
- [14] XUE Yang, SUN Hao, SUN Xian, *et al.* Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network[J]. *IEEE Access*, 2018, **6**: 50839-50849.
- [15] WANG Ji-zhou, LU Chang-hua, JIANG Wei-wei. Simultaneous ship detection and orientation estimation in SAR images based on attention module and angle regression[J]. *Sensors*, 2018, **18**(9):2851-2868.
- [16] LIU Wen-chao, MA Long, CHEN He. Arbitrary-oriented ship detection framework in optical remote-sensing images [J].*IEEE Geoscience and Remote Sensing Letters*, 2018,15(6): 937- 941.
- [17] ZHOU Xing-yi, WANG De-quan, KRÄHENBÜHL P. Objects as points[J/OL]. 2019. Available: <https://arxiv.org/abs/1904.07850v1>.
- [18] HE Kai-ming, ZHANG Xiang-yu, REN Shao-qing, *et al.* Deep residual learning for image recognition[C]. 2016 Conference on Computer Vision and Pattern Recognition (CVPR 2016), USA, 2016:1-10.
- [19] LIN T, DOLLÁR P, GIRSHICK R, *et al.* Feature pyramid networks for object detection. 2016 Conference on Computer Vision and Pattern Recognition (CVPR 2016), USA, 2016:1-10.
- [20] LAW H, DENG Jia. Cornernet: detecting objects as paired keypoints[C]. 15th European Conference on Computer Vision (ECCV2018), Munich, Germany, 1-10.
- [21] LIN T, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, **99**: 2999-3007.
- [22] ALEXANDER N, LUC V. Efficient non-maximum suppression [C]. 18th International Conference on Pattern Recognition (ICPR2006), USA, 2006: 3850-855.
- [23] JIAO Jiao, ZHANG Yue, SUN Hao, *et al.* A densely connected end-to-end neural network for multiscale and multiscene SAR ship detection[J]. *IEEE Access*, 2018, **6**(99): 20881-20892.
- [24] YU F, WANG De-quan, SHELHAMER E, *et al.* Deep layer aggregation[C]. 2018 Conference on Computer Vision and Pattern Recognition (CVPR 2018), USA, 2016: 1-10.