doi: 10.6046/zrzyyg.2020372
引用格式: 刘万军,高健康,曲海成,等. 多尺度特征增强的遥感图像舰船目标检测[J]. 自然资源遥感,2021,33(3):97 - 106. (Liu W J, Gao J K, Qu H C, et al. Ship detection based on multi - scale feature enhancement of remote sensing images[J]. Remote Sensing for Natural Resources, 2021, 33(3): 97 - 106.)

多尺度特征增强的遥感图像舰船目标检测

刘万军, 高健康, 曲海成, 姜文涛
(辽宁工程技术大学软件学院,葫芦岛 125105)

摘要: 针对背景复杂的遥感图像中,舰船方向任意、密集排列造成的漏检问题,基于旋转区域检测网络,提出多尺度特征增强的遥感图像舰船目标检测算法。在特征提取阶段,利用密集连接感受野模块改进特征金字塔网络,选用不同空洞率的卷积获取多尺度感受野特征,增强高层语义信息的表达; 为了抑制噪声并突出目标特征,在特征提取后设计基于注意力机制的特征融合结构,根据各层在空间上的权重值融合所有层,得到兼顾语义信息和位置信息的特征层,再对该层特征进行注意力增强,将增强后的特征融入原金字塔特征层; 在分类和回归损失基础上,增加注意力损失,优化注意力网络,给予目标位置更多关注。在 DOTA 遥感数据集上的实验结果表明,该算法平均检测精度可以达到 71.61%,优于最新的遥感图像舰船目标检测算法,有效地解决了目标漏检问题。

关键词: 卷积神经网络; 多尺度特征融合; 注意力机制; 遥感图像; 舰船目标检测
中图法分类号: TP 751.1 **文献标志码**: A **文章编号**: 2097 - 034X(2021)03 - 0097 - 10

0 引言

遥感图像舰船目标检测一直是遥感图像处理的研究热点,核心任务是定位和识别图像中的舰船目标,在渔业管理、海上运输、船只救援、保卫领土等领域有着重要的现实意义[1]。在遥感图像中,舰船存在被复杂的背景包围,目标小且密集排列的现象,这导致舰船漏检现象严重,是遥感图像解译面临的挑战性问题。

传统的舰船目标检测算法通过先验信息和纹理特征对图像进行海陆分离[2],选择水域作为感兴趣区域,使用模板匹配、形态学比对算法、监督分类[3]在感兴趣区域检测舰船目标。由于舰船检测受到雾气、云层、光照的干扰,传统的算法检测精度较低、鲁棒性差,很难满足实用性需求。卷积神经网络[4]在目标检测中的应用,使得更多高效的目标检测算法被提出,检测算法可分为单阶段和双阶段两类,主流的单阶段检测模型有 YOLO 系列[5]、SDD 算法[6],该类方法基于回归的思路,直接预测类别置信度,并且在图像上定位出目标位置,但是单阶段检测对于多尺度、小目标的检测效果较差。双阶段模型提出了区域建议网络结构,生成一系列包含潜在目标的候选框,再进一步确定目标类别和校正边界框。以 Faster R - CNN[7]为代表发展出了特征金字塔网络(feature pyramid networks, FPN)[8]、Mask R - CNN[9]等基于多尺度特征融合的算法。单阶段的模型检测速度更优,达到了实时检测的效果,双阶段的检测准确率更占优势。

虽然基于深度学习的检测算法不断地应用在遥感图像舰船检测中,但都是基于水平区域的检测,遥感图像中存在方向角任意的舰船,目标角度一旦倾斜,水平检测框的冗余区域与船只的重叠部分会变大,不利于后期非极大值抑制操作。为了提高方向任意目标的检测效果,Ma[10]提出采用旋转锚,引入角度变量控制检测框方向,有效地提高了候选框的质量; Yang 等[11]基于旋转框目标检测,提出密集连接特征金字塔结构(dense feature pyramid networks, DFPN),高层语义信息不仅和相邻层进行融合,还要和其余特征层进行融合,增强了语义信息的传播。基于旋转区域检测算法适应舰船目标旋转特性,有效解决了检测区域冗余问题,不过对于背景复杂的小目标,检测性能有待提高。

为了突出复杂背景下的舰船目标,本文提出多

收稿日期: 2020 - 11 - 23; **修订日期**: 2021 - 05 - 16
基金项目: 国家自然科学基金青年基金项目"面向宽幅高光谱遥感影像的高效压缩方法研究"(编号: 41701479)和辽宁工程技术大学学科创新团队资助项目"智慧农业遥感监测创新团队"(编号: LNTU20TD - 23)共同资助。
第一作者: 刘万军(1959 -),男,教授,主要研究方向为数字图像处理、运动目标检测与跟踪。Email: liuwanjun@lntu.edu.cn。
通信作者: 高健康(1996 -),男,硕士研究生,主要研究方向为深度学习、遥感图像目标检测。Email: 1554797460@qq.com。

尺度特征增强的遥感图像舰船目标检测算法，命名为 MFEDet。考虑到舰船目标尺度多变，提出密集连接感受野模块（densely connected receptive field，DCRF），不同空洞率的卷积，涵盖更密集的不同感受野的特征，可以丰富高层语义特征的多尺度表达；为抑制遥感图像的背景干扰，设计基于注意力机制的特征融合结构（attention - guided feature fusion，AFF），旨在一次使用特征金字塔所有层，通过尺度调整、加权融合、注意力增强的方式，突出目标位置，减少目标漏检现象。

1 旋转区域检测网络原理

为了实现方向任意的舰船目标的检测，本文选择旋转区域检测网络（rotation region detection network，RRDN）作为算法原始框架，该框架是在 Faster R - CNN 基础上改进的，同样包含 3 个阶段：特征提取网络、区域候选网络（region proposal networks，RPN）和 Fast R - CNN 阶段。

RRDN 在特征提取阶段采用 FPN 获取多尺度特征，FPN 结构如图 1 所示。将图像输入主干网络进行特征提取，得到特征层{C2，C3，C4，C5}，高层特征语义信息丰富，适用于目标种类的判别，低层特征具有较高分辨率和位置信息，适用于目标位置的回归，以自顶向下的方式将高层特征信息融入低层特征中，得到{P2，P3，P4，P5} 4 个融合特征层。FPN 以简单的方式，使得高层特征为低层特征的目标检测提供指导信息。

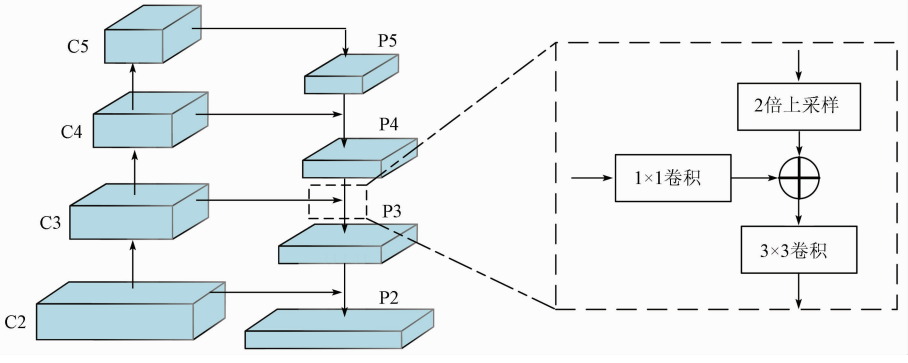


图 1 FPN 结构

Fig. 1 Feature pyramid network structure

与传统的检测不同，RPN 阶段需要利用方向包围框（oriented bounding box，OBB）重新定义锚框，从而适应旋转目标。OBB 采用五元组（x，y，w，h，θ）表示旋转锚框，其中（x，y）表示旋转锚框中心点坐标，旋转角 θ 表示水平轴逆时针旋转遇到旋转框第一条边所成的夹角，同时标记该边为 w，另一条边为 h，旋转角 θ 的范围为（0°，90°]。OBB 的表示如图 2 所示。

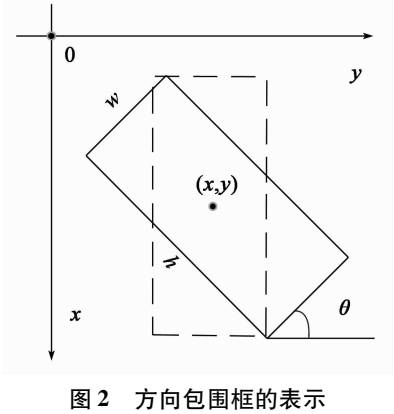


图 2 方向包围框的表示

Fig. 2 The representation of oriented bounding box

OBB 提供了（$-\pi/6$，0，$\pi/6$，$\pi/3$，$\pi/2$，$2\pi/3$）6 个旋转角，能更准确地描述旋转锚框之间的角度差。此外，RPN 会初步判断锚框属于目标还是背景，为下一阶段提供更精确的候选框。

在 Fast R - CNN 阶段，ROI Align 利用候选框的位置坐标，在特征图上提取固定大小的感兴趣区域（regions of interest，ROI），为了匹配 ROI 的对齐操作，该阶段提取的是候选区域水平外接矩形，不是候选框本身。接着传送至后续全连接层，实现候选框更精确的分类和回归，输出最好的检测结果。

基于 RRDN 的检测模型，利用 FPN 结构，将高层语义信息自顶向下传播至低层特征，实现了特征的重用，对尺度不一的目标检测很有帮助；RPN 中旋转角的设置，更适应旋转目标方向任意的特性，输出的检测框冗余区域小。因此，选择该模型用于舰船类的目标，检测效果更佳。

2 多尺度特征增强检测方法（MFEDet）

遥感图像包含的背景信息会干扰目标特征的表达，造成目标位置的模糊，使用 RRDN 模型依然存

在漏检现象。为解决上述问题,本文对提取到的特征信息增强,丰富多尺度特征的表达,使目标特征获得更多关注。基于 RRDN 模型,对提取的特征信息增强,首先利用 DCRF 模块的不同空洞率卷积,感知多尺度感受野语义特征;其次设计基于注意力机制的特征融合结构,融合高层语义信息和低层位置信息,使用注意力网络减弱背景信息的干扰,突出目标位置。本文方法的总体结构如图 3 所示。将处理好的数据送入特征提取网络,基础网络选用 Resnet_101 提取特征,提取的特征送入 FPN 进行特征融合;其次,考虑到最高层 C5 感受野单一,对高层语义信息的感知不充分,增加 DCRF 模块,不同空洞率的卷积会获取多尺度感受野特征,经过密集连接的方式,丰富 P5 层的多尺度特征;接着,将融合后的特征层{P2,P3,P4,P5}送入 AFF 中,根据层级权重进行自适应特征融合,对融合后的特征做注意力增强,给与目标位置更多关注,融合后的特征在与之前各层叠加,组成新的特征层{A2,A3,A4,A5},每个新层都融合了高层语义信息和低层位置信息;最后,在 RPN 中根据设定的旋转锚框(anchor)选定到高质量的候选框(proposals),Fast R-CNN 阶段经过两个全连接层(fc),实现目标的分类回归,输出最终的检测结果。图 3 中 cls 表示分类分支,其作用是判别检测框所属类别是否为目标;Score2× 代表模型输出的目标和非目标的两种概率,当目标概率更大时,系统判定该检测框的类别为目标;回归分支 reg 预测的是目标的参数化坐标;(t_x,t_y) 为预测框的中心点坐标;t_w 和 t_h 为目标框的长和宽;t_θ 为目标框相对于水平轴的旋转角度。

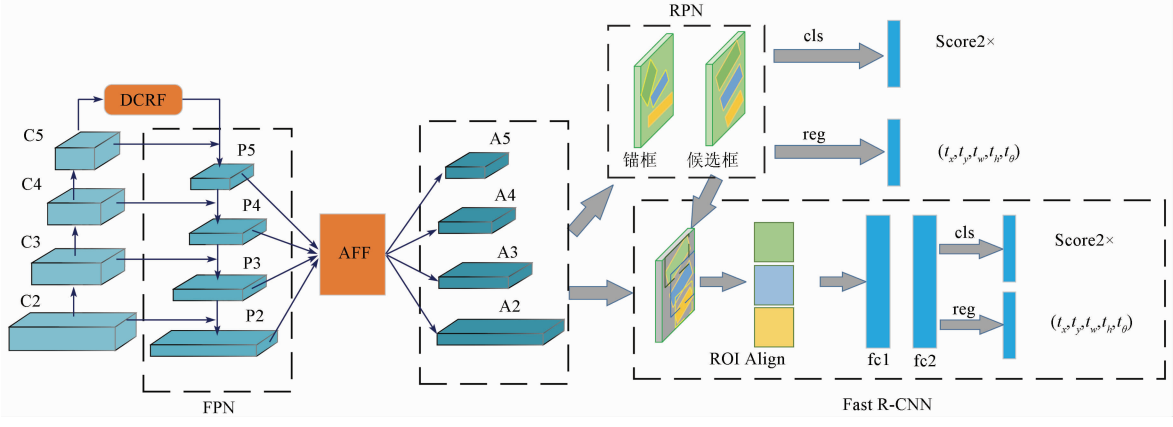


图 3　总体结构

Fig. 3　Overall framework

2.1　密集连接感受野模块(DCRF)

卷积神经网络被用来提取图像特征信息,越深的卷积网络,提取到的语义信息越丰富。FPN 将最高层 P5 语义信息,采用自顶向下的传播方式,融入进较低层级{P2,P3,P4},利用高层特征的语义信息对低层特征进行增强,这对小目标检测效果的提升非常重要。低层特征可以融合高层特征的信息,但最高层 P5 直接由 C5 降维得到,没有融合任何上下文信息,且通道数由 2 048 骤减至 256,信息损失严重,因此,需要对高层特征层进一步强化,利用不同感受野的卷积丰富 P5 层语义信息。

RFBNet[12] 在 Inception[13] 网络基础上,提出了 RFB_S 结构,选用不同空洞率的卷积,可覆盖多尺度的感受野,对提取上下文的信息非常有用,但是该结构中的每个分支都是独立存在的,提取的特征相互之间缺少依赖。受 DenseNet[14] 密集思想的启发,本文改进 RFB_S 结构,提出了 DCRF 模块。DCRF 结构如图 4 所示,图中ⓒ表示串联(concat)操作。该模块采用的两个策略:级联模式和并行模式。级联模式采用密集连接方式,较大空洞率的卷积层接收较小空洞率的卷积层的输出,可以充分利用上下文信息,产生更大的感受野。并行模式使得多个卷积层接收相同的输入,经过不同卷积核的卷积层以及不同空洞率的空洞层后,输出多尺度的感知特征。另外,为了保持原始输入的全局信息,将全局平均池化层(global average pool,GAP)和上采样层(up sample)连接,与串联后的信息相加,实现全局信息与局部信息的融合。该模块不仅继承了 RFB_S 结构多空洞卷积的优点,而且更好地利用了卷积层之间的内部联系。

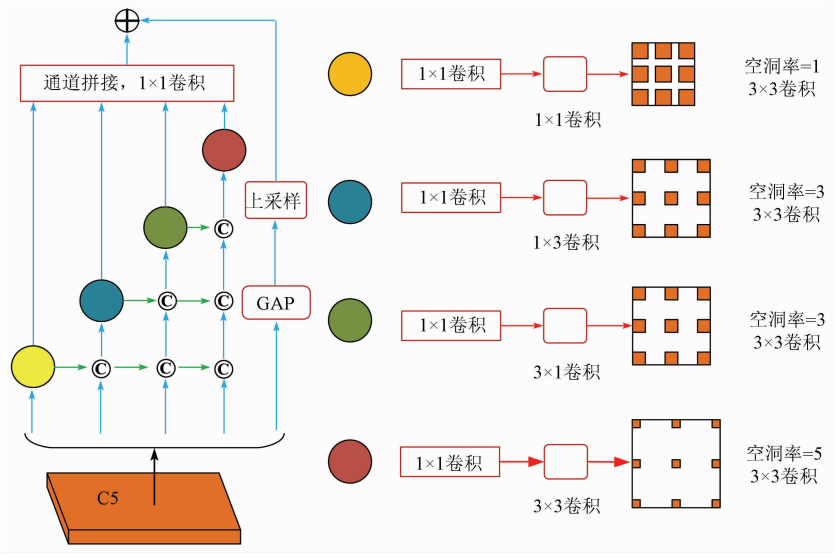


图 4 密集连接感受野模块
Fig. 4 The module of DCRF

2.2 基于注意力机制的特征融合结构(AFF)

遥感图像存在背景信息复杂的问题,导致后续 RPN 生成的候选框会引入噪声信息,众多噪声信息会淹没目标,使目标区域变得模糊,出现误检漏检现象。因此,在特征送入 RPN 层之前,有必要对特征层进行注意力增强,更多地关注目标特征,弱化非目标特征。若是对所有特征层{P2,P3,P4,P5}单独进行注意力增强,会导致计算量激增,并且每一层只做自身注意力增强,高层特征缺乏低层位置信息,低层特征缺乏高层语义信息,层级之间信息缺少有效的沟通,表现出不平衡状态。对此,设计 AFF 结构,加权融合所有高低特征层信息,从整体上增强目标特征的表达,特征融合结构如图 5 所示。

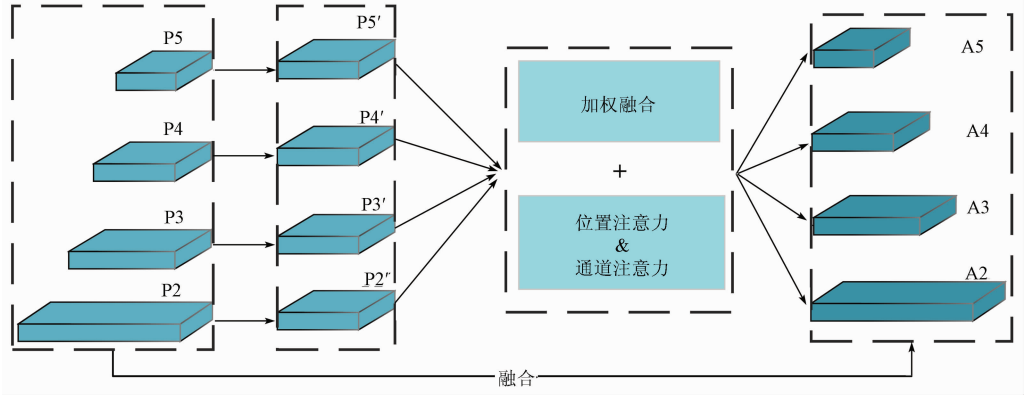


图 5 特征融合结构
Fig. 5 Feature fusion structure

在 FPN 后,进一步对提取到的特征层{P2,P3,P4,P5}进行特征增强。AFF 结构与 Libra R-CNN 中 BFP 思想类似,一次利用所有 FPN 层,通过分辨率调整、加权融合的方式,得到一层融合后的特征,对该层做双重注意力增强,将增强后的特征再和原始层相加,实现特征强化和高低层信息充分融合,得到了增强后的多尺度特征层{A2,A3,A4,A5}。

2.2.1 加权融合层级特征

FPN 的 P5 层获取到更多是的是语义信息,P2 层为高分辨率层,学习到更多的是细节特征,适合小目标检测,但是缺乏语义信息指导,小目标容易产生漏检现象,高低层特征融合能够很好地解决这个问题。P4 层的分辨率更适合语义信息和细节信息的融合[15],所以将 4 层特征尺寸调整至 P4 大小进行特征融合。通常的融合方式是各层相加取平均,将各层空间信息差异较大的特征直接相加,会削弱多尺度特征表达能力。本文通过获取不同特征层在空间位置(i,j)上的权重,对 4 层特征进行自适应融合,融合方式定义为:

$$w_{ij}^{l} = \frac{\exp(\omega_{ij}^{l})}{\sum_{k=2}^{5}\exp(\omega_{ij}^{k})}, \tag{1}$$

$$b_{ij} = \sum_{l=2}^{5} w_{ij}^{l} \cdot \mu_{ij}^{l}, \tag{2}$$

式中:l 为当前特征层;k 为遍历的 P2—P5 层;ω_{ij}^{l} 为特征经过卷积网络学习到当前位置在层级间的权重值;w_{ij}^{l} 为通过指数函数 exp( )获得归一化后的权重值;μ_{ij}^{l} 为当前层级(i,j)位置的像素值;b_{ij} 为所有层加权融合后的特征值。

2.2.2　双重注意力网络

注意力机制[16-17]的提出,有效地解决了目标遮挡、模糊问题。遥感图像中的舰船目标容易被复杂的背景信息淹没,目标位置的模糊容易导致漏检现象,所以,使用注意力机制对特征增强是十分必要的。本文设计的位置和通道双重注意力网络如图 6 所示。上半部分为位置注意力,融合后的特征图 Px 经过一系列不同卷积核的卷积运算,得到了双通道显著图,双通道分别映射了前景和背景的概率,Softmax 函数会将显著图的值映射到[0,1]之间,选择显著图的一个通道与 Px 相乘,生成新的特征图,可以抑制噪声信息,强化目标信息。下半部分是通道注意力机制,使用 SEnet[18] 的通道注意力辅助增强特征层,顺着通道维度对 Px 进行全局平均池化压缩,获取全局感受野,经过全连接层和 Sigmid 非线性处理,将输出结果作为每个通道的权重值。为了使通道注意力更轻便,用比例 r 减少全连接层尺寸,选择合适的比例 r 能兼顾模型的计算效率和检测性能($r=4$)。通道注意力获得的权重值也与 Px 相乘,生成的特征图与位置注意力特征图做融合,得到新的注意力特征图 Ax。

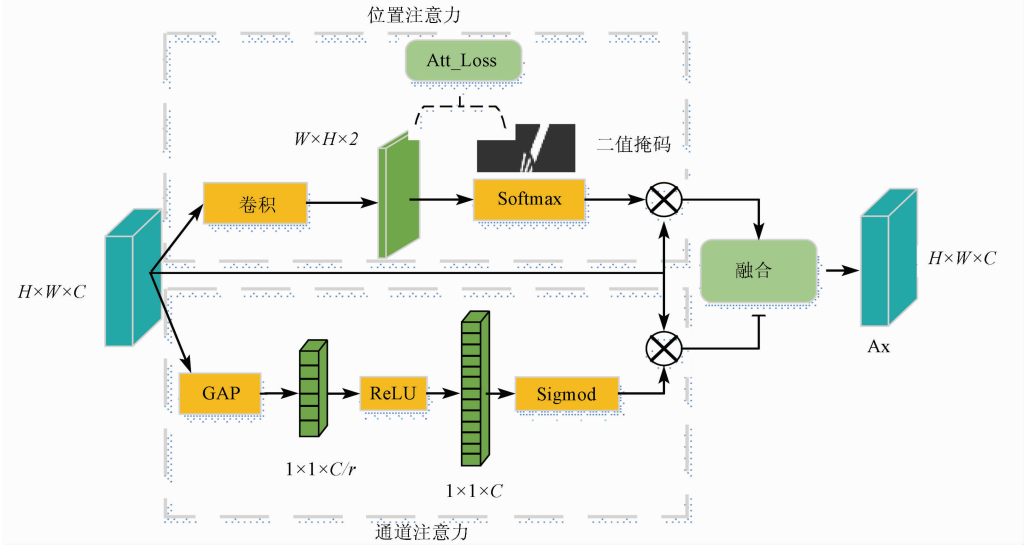
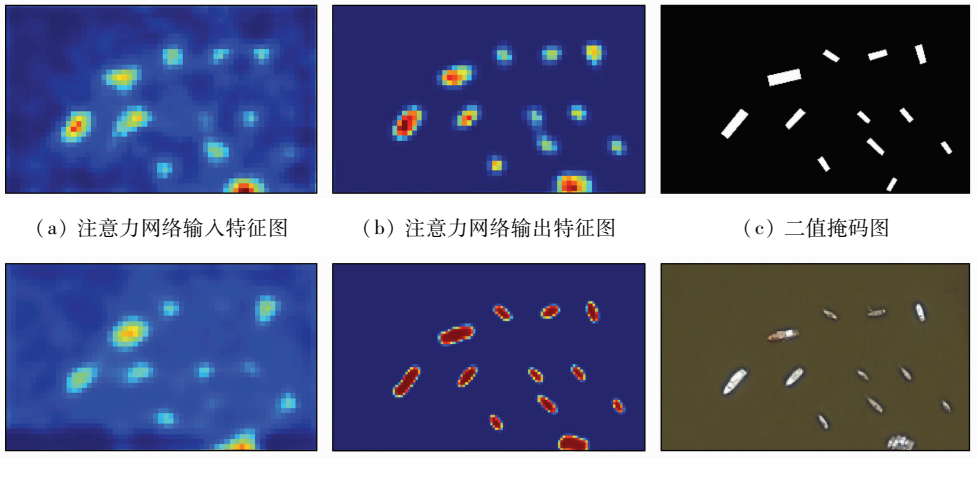


图 6　双重注意力网络

Fig. 6　Dual attention network

注意力网络可视化如图 7 所示。图 7(a)为加权融合后的特征图,出现目标位置模糊现象,说明融合前 FPN 提取到的高低层特征目标特征已经被背景信息淹没,目标位置不够显著。图 7(b)为注意力



(a) 注意力网络输入特征图　(b) 注意力网络输出特征图　(c) 二值掩码图

(d) 无监督的显著图　(e) 有监督的显著图　(f) 原图像

图 7　注意力网络可视化

Fig. 7　Visualization of the attention network

增强后的特征图,注意力机制的引入,突出目标位置,抑制噪声信息对特征图的干扰,很好地解决了待检测目标模糊问题。目前大多数的注意力网络都是非监督的,不能更好地关注目标位置,本文设置可监督机制,即在训练阶段,根据真值图生成的二值掩码如图 7(c)所示,将掩码和位置注意力中双通道显著图的交叉熵损失,作为注意力网络损失,优化注意力网络。没有添加注意力损失的无监督显著图,目标位置标记不精确,目标区域模糊,如图 7(d)所示;有监督的显著图目标位置更突出,能够很好地引导网络关注目标信息,显著图如图 7(e)所示。

2.3 损失函数

为了训练 RPN,提取高质量的候选框,需要在所有旋转框中挑选正负样本,每个框会分配一个二值类别标签和 5 个参数化坐标。正样本的旋转框需要满足以下两个条件之一即可:①旋转框与真实目标框之间交并比(intersection over union,IOU)重叠大于 0.5,且角度差小于 15°;②旋转框与目标框的 IOU 重叠最高。同样的负样本也是两个条件:①IOU 重叠小于 0.2;②IOU 的重叠大于 0.5,但角度差大于 15°。小批量总数是 512,正负样本比例是 1∶1,其余不满足条件的候选框会被摒弃。

同 RPN 阶段类似,Fast R-CNN 阶段也以同样方式选择正负样本,不同的是小批量总数变为 256。该阶段也会对每个候选框分类,分配 5 个参数化坐标,回归出最终预测框,加入了角度信息后,旋转框可以更精准的定位目标,参数化坐标的回归定义为:

$$\begin{cases} t_x = (x - x_a)/w_a \\ t_y = (y - y_a)/h_a \\ t_w = \ln(w/w_a) \\ t_h = \ln(h/h_a) \\ t_\theta = \theta - \theta_a + k\pi/2 \end{cases}, \tag{3}$$

$$\begin{cases} t'_x = (x' - x_a)/w_a \\ t'_y = (y' - y_a)/h_a \\ t'_w = \ln(w'/w_a) t'_h = \ln(h'/h_a) \\ t'_\theta = \theta' - \theta_a + k\pi/2 \end{cases}, \tag{4}$$

式中:变量 x,x_a,x' 分别为预测框、旋转框和真值框的中心点 x 坐标(y,w,h,θ 同样);$k(k \in Z)$ 为保持旋转角在(0°,90°]、令旋转框保持在相同位置的参数,当 k 为奇数时,边框的 w 与 h 需要互换。损失函数采用多任务损失,新增注意力损失后,其定义为:

$$L(p_n, l_n, t'_n, t_n, \mu'_{ij}, \mu_{ij}) = \frac{\lambda_1}{N}\sum_{n=1}^{N} L_{cls}(p_n, l_n) + \frac{\lambda_2}{N}\sum_{n=1}^{N} l_n L_{reg}(t'_n, t_n) + \frac{\lambda_3}{h \times w}\sum_{i}^{h}\sum_{j}^{w} L_{att}(\mu'_{ij}, \mu_{ij}) \tag{5}$$

式中:N 为候选框数目;l_n 为真值的标签,$l_n = 1$ 代表目标,$l_n = 0$ 代表背景,背景不参与回归;p_n 为经过 Softmax 函数后的目标概率分布;t_n 为预测的 5 个参数化坐标向量;t'_n 为旋转框与真值框的偏移量;u_{ij},u'_{ij} 分别为掩模像素的类别和预测概况;超参数 λ_1,λ_2,λ_3 用来权衡多任务损失,分别为 $\lambda_1 = 1$,$\lambda_2 = 2$,$\lambda_3 = 1$;分类损失 L_{cls} 和注意力损失 L_{att} 均为交叉熵损失,其描述的是预测的概率分布与真实的概率分布之间的差距,交叉熵值越小,预测值越接近真实值。回归损失 L_{reg} 采用 $smooth_{L1}$ 函数,定义为:

$$L_{reg}(t, t') = smooth_{L1}(t - t'), \tag{6}$$

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & x < 1 \\ x - 0.5 & 其他 \end{cases}。 \tag{7}$$

$smooth_{L1}$ 函数避开了 L_1 和 L_2 损失函数的缺陷,解决了梯度爆炸问题,强鲁棒性使得该函数更适合目标框的回归。

3 实验结果与分析

3.1 参数设置

在 RPN 阶段采用旋转框作为锚框,使用 5 个变量$\{x, y, w, h, \theta\}$来唯一确定旋转框,(x, y)表示目标框中心点坐标,旋转角 θ 是由 x 轴逆时针旋转与框所成的夹角,并记框的这条边为 w,另一条边为 h,旋转角范围是$[-90°, 0°)$,这与 OpenCV 中的定义保持一致。本文从多尺度、多角度、多长宽比 3 个参数生成各式各样的旋转锚框。为特征层{A2,A3,A4,A5,A6}(A6 是由 A5 下采样得到)分配单一尺度,尺度大小分别为{32,64,128,256,512}像素,设计 6 个角度$\{-15°, -30°, -45°, -60°, -75°, -90°\}$预测舰船的方向,可以多角度覆盖目标,根据舰船形状,设置锚框有{1∶2,2∶1,1∶3,3∶1,1∶5,5∶1,1∶7,7∶1}的长宽比。每层上每个特征点产生 48 个旋转框(8×6),输出 240 个回归参数(5×48)和 96 个分类分(2×48)。

本文实验是在 Ubuntu16.04 系统、NVIDIA GeForce GTX 1080Ti 的计算机上,深度学习实验环境为 TensorFlow。为加快模型收敛,使用预训练模型 ResNet101 对网络进行初始化。实验经历 100k 次迭代,前 40k 次迭代的学习率为 0.001,再 40k 次学习率降为 0.000 1,最后 20k 次迭代为 0.000 01,权重

衰减为 0.000 1,动量为 0.9,优化器选择 Momentum。

3.2　实验数据和评估指标

DOTA 是用于遥感图像目标检测的大型数据集,每个实例都由一个任意的四边形标记,包含 2 806 张来自不同平台的图像[19],每张图像分辨率大小从 800 ×800 到 4 000 ×4 000 不等,囊括了多尺度、任意方向和形状各异的目标。从该数据集中提取出包含舰船的图像,并以 256 像素点的步幅,裁剪出 1 000 ×600 的子图像,再经过 180°旋转、水平翻转对数据进行增强。

使用平均精度(average precision,AP)评定不同方法在舰船类目标检测的性能[19],它是反映全局性能的指标,由精确率 P 和召回率 R 积分得出,定义为:

$$AP = \int_0^1 P(R)\,\mathrm{d}R \quad , \tag{8}$$

$$P = \frac{TP}{TP + FP} \quad , \tag{9}$$

$$R = \frac{TP}{TP + FN} \quad , \tag{10}$$

式中:TP 为舰船样本被正确标记为舰船个数;FP 为非舰船样本被标记为舰船个数;TN 为非舰船样本被正确标记为非舰船目标个数;FN 为舰船样本被标记为非舰船目标个数。

3.3　自身模块对比实验

本文选用旋转区域检测网络作为基础网络(Baseline),包括特征提取网络(ResNet_101),FPN,RPN,ROI Align 以及旋转非极大值抑制等。将以上基础网络和所有的实验参数保持一致,使用平均精度衡量性能,自身模块实验结果见表 1,表中加粗部分为同类指标中的最佳值。

表 1　不同模块的消融实验结果

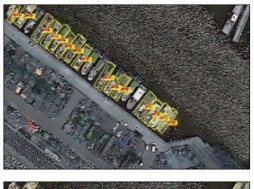
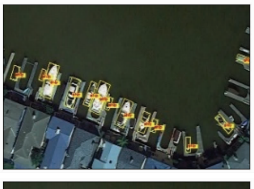
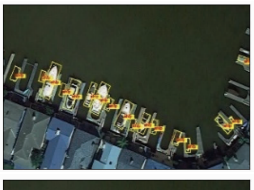
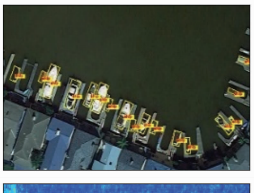
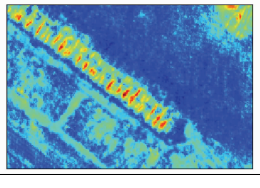
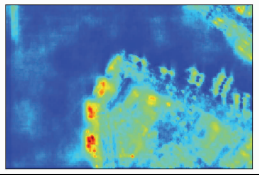
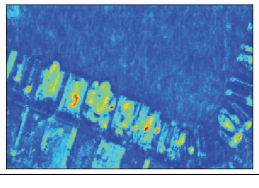
Tab.1　Results of ablative experiments of different module

实验方法	召回率/%	精确率/%	*AP*/%	检测时间/s
基础网络	76.56	**87.86**	67.37	**0.21**
+ DCRF	80.82	85.92	69.52	**0.21**
+ AFF	80.99	85.82	69.66	0.22
本文方法	**81.74**	87.04	**71.61**	0.22

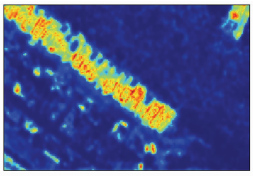
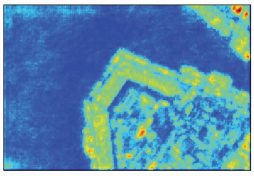
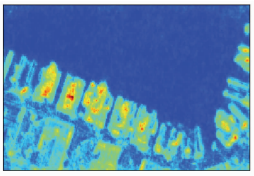
基础网络的精确率虽高,但其召回率只有 76.56%,目标漏检现象严重;DCRF 可以感知到不同感受野信息,减少高层语义信息的损失,将框架的 AP 提高到 69.52%,召回率明显增加了 4.26 百分点;AFF 模块融合增强高低层特征,高层信息能指导低层特征检测小目标,低层信息丰富高层特征的空间信息,单独结合 AFF,同样会提高召回率,相比于基础网络,AP 增长至 69.66%;本文方法 MFEDet 在基础网络之上,结合 DCRF 和 AFF 模块,进一步改善框架性能,AP 达到 71.61%。不同模块的对比结果见表 2。本文选取 3 类图片验证各模块的有效性:示例 1 为背景复杂的舰船图像,示例 2 中

表 2　不同模块的结果展示

Tab.2　Show the results of different modules

类型	示例 1	示例 2	示例 3
基础网络结果图			
DCRF 结果图			
MFEDet 结果图			
基础网络特征图			

（续表）

类型	示例 1	示例 2	示例 3
MFEDet 特征图			

图像舰船目标密集排列，示例 3 的图像中存在小目标。

1）DCRF 模块有效性分析。表 2 第二行为基础网络的结果，在背景杂乱、目标密集的图像中，该方法的漏检现象严重。第三行为增加 DCRF 模块的结果，DCRF 利用不同空洞率卷积，获得多尺度感受野特征，增强高层语义信息的提取和传播，使目标漏检现象明显减少，大目标和小目标均能被准确的标记。

2）AFF 模块有效性分析。第四行为 MFEDet 检测结果，可以看出，对于密集排列的目标，模型给出了更准确的目标框进行标记，舰船和背景相似的困难目标也被精确检测到，证明 AFF 模块的引入，抑制了背景信息的干扰，目标位置受到更多关注，改善了困难样本的漏检问题。

3）表 2 最后两行分别为基础检测模型的 3 幅场景的特征图和本文提出的 MFEDet 模型下图像的特征图。前者提取的特征目标位置不够显著，边缘出现模糊现象，并且部分目标被背景淹没，目标特征丢失，相比之下，MFEDet 模型对提取到的特征进行多尺度增强，抑制了背景信息的表达，目标位置更清晰准确，有效地解决了舰船漏检问题，更适用于复杂场景下的遥感图像目标检测。

3.4 对比实验

为了进一步验证 MFEDet 的有效性，本文方法还和 FR－O[20]，RRPN，R－DFPN 以及 RADet[21] 作比较，不同方法的对比结果见表 3。

表 3 不同方法的对比结果

Tab. 3 Different methods comparison results （%）

对比方法	召回率	精确率	*AP*
FR－O	58.53	65.60	39.24
R－DFPN	67.85	87.67	59.78
RRPN	69.35	89.68	63.42
RADet	—	—	68.86
本文方法	81.74	87.04	71.61

表中 FR－O 代表 Faster R－CNN OBB 检测器，是 DOTA 官方给出的旋转检测方法，可以看出该方法的 AP 相对较差。R－DFPN 和 RRPN 等旋转区域的检测法，虽然舰船检测准确率相对较高，但是舰船召回率低，漏检现象严重，准确率和召回率不能很好地平衡，导致其 AP 较低。此外，本文还与最新的 RADet 检测器做了对比，由二者的 AP 可知，本文方法相对于最新的检测算法，检测性能依然有优势。

本文方法在 RRDN 算法基础上新增两个模块，提高检测性能的同时，测试速度依然处于较快水平。不同方法训练时间和测试时间的比较见表 4。

表 4 不同方法的训练时间和测试时间

Tab. 4 Training time and test time for each method （s）

方法	训练时间	测试时间
FR－O	0.34	0.10
RRPN	0.85	0.35
R－DFPN	1.15	0.38
Baseline	0.58	0.21
本文方法	0.64	0.22

4 结论

本文提出的多尺度特征增强的舰船目标检测方法，针对方向任意、场景复杂、小目标聚集的遥感舰船图像。

1）设计了两个新的结构，在最高层添加密集连接感受野模块（DCRF），改进 FPN 网络，有效地增强了高层语义信息的表达。

2）设计基于注意力机制的特征融合结构（AFF），加权融合了高低层信息，同时对融合后的特征进行双重注意力增强，抑制噪声信息并突出目标位置，对于复杂场景中的小目标检测十分重要。

3）针对舰船不同长宽比的特点，设置不同长宽比例和不同旋转角的锚框，改善检测区域的冗余问题。在传统的多任务损失中新增注意力损失，不断优化注意力网络，使整个检测模型达到最佳。在 DOTA 公开遥感数据集上，本文方法取得了较好的检测效果。

参考文献（References）：

[1] 王彦情，马 雷，田 原. 光学遥感图像舰船目标检测与识别综述[J]. 自动化学报，2011，37(9)：1029－1039.
Wang Y Q，Ma L，Tian Y. Overview of ship target detection and recognition based on optical remote sensing image[J]. Acta Automatica Sinica，2011，37(9)：1029－1039.

[2] 谢奇芳，姚国清，张 猛. 基于 Faster R－CNN 的高分辨率图像目标检测技术[J]. 国土资源遥感，2019，31(2)：38－43. doi：10.6046/gtzyyg.2019.02.06.

Xie Q F, Yao G Q, Zhang M. Research on high resolution image object detection technology based on Faster R - CNN[J]. Remote Sensing for Land and Resources, 2019, 31(2): 38 - 43. doi: 10.6046/gtzyyg. 2019. 02. 06.

[3] 史文旭,江金洪,鲍胜利. 基于特征融合的遥感图像舰船目标检测方法[J]. 光子学报,2020,49(7):57 - 67.
Shi W X, Jiang J H, Bao S L. Ship target detection in remote sensing image based on feature fusion[J]. Acta Photonica Sinica, 2020, 49(7): 57 - 67.

[4] Szegedy C, et al. Going deeper with convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Boston, MA, 2015: 1 - 9.

[5] 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(CVPR), Las Vegas, NV, 2016: 779 - 788.

[6] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European Conference on Computer Vision, Springer, Cham, 2016: 21 - 37.

[7] Ren S, He K, Girshick R, et al. Faster R - CNN: Towards real - time object detection with region proposal networks[C]//Advances in neural information processing systems, 2015: 91 - 99.

[8] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117 - 2125.

[9] He K, Gkioxari G, Dollár P, et al. Mask R - CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961 - 2969.

[10] Ma J. Arbitrary - oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111 - 3122.

[11] Yang X, Sun H, Fu K, et al. Automatic ship detection in remote sensing images from google earth of complex scenes based on multi-scale rotation dense feature pyramid networks[J]. Remote Sensing, 2018, 10(1): 132.

[12] Zhu Y, Mu J, Pu H, et al. FRFB: Integrate receptive field block into feature fusion net for single shot multibox detector[C]//2018 14th International Conference on Semantics, Knowledge and Grids (SKG), Guangzhou, China, 2018: 173 - 180.

[13] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Las Vegas, NV, 2016: 2818 - 2826.

[14] Huang G, Liu Z, Der Maaten L V, et al. Densely connected convolutional networks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017: 2261 - 2269.

[15] Pang J, Chen K, Shi J, et al. Libra R - CNN: Towards balanced learning for object detection[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019: 821 - 830.

[16] Wang X, Girshick R, Gupta A, et al. Non - local neural networks[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018: 7794 - 7803.

[17] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[J]. Lecture Notes in Computer Science, 2018: 3 - 19.

[18] Hu J, Shen J and Sun G, Squeeze - and - excitation networks[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018: 7132 - 7141.

[19] Han J, Zhou P, Zhang D, et al. Efficient, simultaneous detection of multi - class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding[J]. ISPRS Journal of Photogrammetry & Remote Sensing, 2014, 89: 37 - 48.

[20] Xia G, et al. 2018. DOTA: A large - scale dataset for object detection in aerial images[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018: 3974 - 3983.

[21] Li Y, Huang Q, Pei X, et al. RADet: Refine feature pyramid network and multi - layer attention network for arbitrary - oriented object detection of remote sensing images[J]. Remote Sensing, 2020, 12(3): 389.

Ship detection based on multi - scale feature enhancement of remote sensing images

LIU Wanjun, GAO Jiankang, QU Haicheng, JIANG Wentao

(*College of Software, Liaoning Technical University, Huludao* 125105, *China*)

Abstract: Aiming at the omission in the ship target detection from remote sensing images with complex background caused by the arbitrary and dense arrangement of ships, this study, based on the rotation region generation network, proposes a ship target detection algorithm using the multi - scale feature enhancement of remote sensing images. The detailed steps are as follows. Firstly, improve the feature pyramid network using the receptive field module with dense connection at the feature extraction stage. Then obtain the characteristics of multi - scale receptive fields using the convolution of different dilate rates. In this way, the expression of high - level semantic information can be enhanced. Then design a feature fusion structure based on attention mechanisms to restrain noise

and highlight the target characteristics. Afterward, fuse all layers according to the spatial weight value of each layer to obtain a feature layer that takes both semantic and position information into account. Then conduct attention enhancement to the features of this layer, and integrate the enhanced features into the original feature layer in the pyramid network. Consequently, pay more attention to target locations by increasing attention loss and optimizing the attention network according to the classification and regression loss. As indicated by the experiment results of DOTA remote sensing dataset, the average precision of this algorithm is as high as 71.61%, which is higher than the latest ship target detection algorithm based on remote sensing images. In this manner, the omission in ship target detection can be effectively solved.

Keywords: convolution neural network; multi – scale feature fusion; attention mechanism; remote sensing image; ship target detection

(责任编辑: 张 仙)