面向遥感图像的多阶段特征融合目标检测方法

陈 立[1],张 帆[2],郭 威[2],黄 赟[1]
(1. 信息工程大学,河南郑州 450001;2. 国家数字交换系统工程技术研究中心,河南郑州 450002)

摘 要: 遥感图像目标具有多尺度、大横纵比、多角度等特性,给传统的目标检测方法带来了新的挑战.针对现有方法应用于目标尺度小、横纵比例不均衡的遥感图像时存在的精度下降问题,提出一种基于多阶段特征融合的目标检测方法MF2M(Multi-stage Feature Fusion Method).该方法在一阶段对特征图通道进行组合拆分,再采用卷积拼接的融合方式聚合通道维度的特征,从而强化输出的目标空间轮廓信息;二阶段设计多比例的非对称卷积块,增强大横纵比目标的高维全局特征,改善目标与检测框匹配粗糙的问题,同时利用串并行相结合的处理方式减少冗余卷积参数,加速网络收敛.在DOTA(Dataset for Object deTection in Aerial images)数据集上的实验结果表明,基准方法引入MF2M后,在保证检测速度的前提下精度指标mAP提高至76.44%,结果验证了所提算法的有效性与可靠性.

关键词: 遥感图像;目标检测;多阶段特征融合;通道拼接;非对称卷积
基金项目: 国家自然科学基金(No.61521003)
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2023)12-3520-09
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20211421

Multi-Stage Feature Fusion Object Detection Method for Remote Sensing Image

CHEN Li[1], ZHANG Fan[2], GUO Wei[2], HUANG Yun[1]
(1. *Information Engineering University*, *Zhengzhou*, *Henan* 450001, *China*;
2. *National Digital Switching System Engineering Technology Research Center*, *Zhengzhou*, *Henan* 450002, *China*)

Abstract: The remote sensing image objects has the characteristics of multi-scale, large aspect ratio, multi-angle and so on, which brings new challenges to traditional object detection methods. To solve the problem of loss of accuracy when existing methods are applied to remote sensing images with small object scales and unbalanced aspect ratios, an object detection method based on dual-stage feature fusion—MF2M (Multi-stage Feature Fusion Method) is proposed. This method combines and splits the feature map channels in first stage, and then adopts the fusion method of convolution splicing to aggregate the characteristics of the channel dimensions, thereby enhancing the output object spatial contour information; in the second stage, we design a multi-scale asymmetric convolution blocks, enhancing the high-dimensional global features of large aspect ratio targets, improving the problem of rough matching between the target and the detection frame, and using a combination of serial and parallel processing to reduce redundant convolution parameters. Finally, we achieve the effect of accelerating network convergence. The experimental results on the DOTA (Dataset for Object deTection in Aerial images) dataset show that after the benchmark method is introduced into MF2M, the accuracy index mAP is increased to 76.44% under the premise of ensuring the detection speed. The results verify the effectiveness and reliability of the algorithm.

Key words: remote sensing image; object detection; double-stage feature fusion; channel splicing; asymmetric convolution
Foundation Item(s): National Natural Science Foundation of China (No.61521003)

1 引言

目标检测(object detection)是计算机视觉领域的基础性任务之一,通过在计算机及相关设备上运行检测算法,实现对图像中目标对象的分类与定位.近年来,随着卫星遥感技术的广泛应用,针对遥感图像的目标检测技术也逐渐成为了研究热点.由于具有宏观性、多

收稿日期:2021-10-18;修回日期:2021-12-18;责任编辑:梅志强

样性、周期性、经济性等特点,遥感图像目标检测技术可以用于灾害控制、交通规划、海上救援等方面[1],为军事任务演习、地表覆盖勘测、地物特征绘图以及自然资源管理等提供极大便利,在民用方面和军事方面都有着十分重要的意义.

传统的目标检测算法,例如边缘检测、阈值分割或是浅层的机器学习算法已经难以达到当前遥感图像检测任务的精度指标. 随着计算机硬件的不断发展以及大规模标注图像训练数据集的涌现,训练大规模神经网络用以检测遥感图像中的目标成为如今目标检测任务的热门趋势. 基于深度学习的卷积神经网络[2](Convolutional Neural Networks,CNN)通过结构化参数以及自动化图像预处理,在MS COCO(Microsoft Common Objects in Context)[3]、PASCAL VOC(PASCAL Visual Object Classes)[4]等主流标注数据集上取得良好的性能,从而在计算机视觉任务中得到广泛应用. 目前基于CNN的预设检测算法按照检测步骤的不同,大致分为双阶段检测器和单阶段检测器[5]. 以SPPNet(Spatial Pyramid Pooling Net)[6]、Faster R-CNN(Region Convolutional Neural Network)[7]等为代表的双阶段检测器使用感兴趣区域(Region of Interest,ROI)来连接两个阶段. 第一个阶段,利用区域候选网络(Region Proposal Network,RPN)生成感兴趣的候选区域框;第二个阶段则是通过ROI Pooling[7]或者ROI Align[8]从每个候选框中提取特征,并进行下一步的对象分类和候选框回归操作. 双阶段检测器通常具有较高的检测精度,但是由于在二阶段需要单独对每一个候选框做分类与回归操作,导致检测速度缓慢;一阶段检测器抛弃提取候选框的过程,只使用一个阶段便完成目标的分类和回归工作,代表模型有基于anchor-based的YOLO(You Only Look Once)[9]系列、SSD(Single Shot MultiBox Detector)系列[10]和基于anchor-free的CenterNet[11]、ExtremeNet等[12]等. 一阶段检测器的检测速度较快,但是准确率却难以与双阶段检测器相媲美.

然而,尽管基于CNN的目标检测方法在自然图像方面取得显著成果,将其直接转移至光学遥感图像仍然是困难的. 图1显示了自然图像与遥感图像的呈现区别:(1)前者按照垂直地平面角度呈现物体. 后者则源于高空航拍成像,使得图像整体分辨率更大,涵盖范围广,背景复杂;(2)遥感图像内的检测目标尺度更小,且按照任意的倾斜角度密集分布;(3)同一遥感图像内不同目标实体之间的尺度与横纵比差异明显. 遥感图像的上述特性大幅增加目标特征的复杂度,不论是单阶段或是双阶段检测器,都会在检测采样过程中丢失目标的部分特征信息,造成检测误差. 随着卷积层数的加深以及反向传播机制的影响,检测误差也会被逐渐放大. 因此,找到对遥感目标特征有效融合的方法是当前遥感图像目标检测任务值得探索的研究方向.

(*a*) 自然图像　　(*b*) 航拍图像

图1　不同拍摄场景下的飞机

因此,为解决目标尺度小、横纵比例不均衡所导致特征提取困难的问题,本文提出一种适用于检测遥感图像目标的多阶段特征融合方法MF2M(Multi-stage Feature Fusion Method). 该方法的主要贡献如下:(1)在一阶段,提出邻间通道特征融合思想,按照特定组数对通道拆分并进行邻间重组,有效构建纵向特征的同时丰富小尺度目标的语义信息;(2)在二阶段,提出对特征图采用多种比例的非对称卷积方案,获得更多的语义信息,并细化大横纵比检测目标的边缘特征. 由于该方法采用端到端训练方式,并且可以即插即用,本文在基准方法[13]的网络基础上引入该模块,并将融合后的网络称为MF2Net(Multi-stage Feature Fusion Net).

2　相关工作

2.1　遥感目标检测算法

遥感检测的算法来源于文本检测领域,包括Ma等[14]提出的RRPN(Rotated Region Proposal Network)、Zhou等[15]提出的EAST(Efficient and Accuracy Scene Text)以及Jiang等[16]提出的R^{2}CNN(Rotational Region Convolutional Neural Network),上述算法通过设计不同的角度变换方式,对旋转文本具有良好的检测效果,也为遥感目标提供了图2(*a*)类型的旋转框检测方式,避免了图2(*b*)类型的水平检测框导致的框重叠问题. 但是将其迁移至高分辨率遥感图像时,会出现不同程度的漏检与误检,检测效果不佳. 对此,现阶段一些学者致力于研究遥感图像目标检测的专用算法. Liu等[17]在提出的RR-CNN网络中设计了具有角度信息的感兴趣区域池化层(RRoI Pooling Layer),用于增强旋转目标的底层特征;Ding等[18]针对图像中朝向不同的密集小目标,设计了旋转感兴趣区域学习器(RRoI learner)和基于旋转敏感位置的区域对齐方法(rotated position sensitive RoI align),用来提取旋转无关的特征;Yang等[13]提出一种从粗粒度到细粒度的渐进回归方法R^{3}Det(Refined Rotated Retinanet Detector),通过精修再重构目标中心点达到检测精度的快速提升;为解决稀疏型编码

层数过于厚重的问题，Yang等[19]再对角度信息设计密集型编码标签(densely coded labels)，使得预测层更加轻量化，并且更易于从复杂背景中提取类正方形目标.

(a) 水平包围框 (b) 定向包围框

图2 遥感图像检测框的两种形式

2.2 多维度特征融合

特征融合属于多维度的空间范畴，包括多尺度特征、上下文关系特征、多通道特征、多粒度特征[20]以及时空特征等等. 在目标检测领域，特征融合的目的是为了更好的区分目标样本与背景样本. 深度卷积神经网络为研究者获取目标特征提供了极大的帮助，但动辄数十层的卷积层也导致目标原有特征的细粒度受到损失. 为解决上述问题，2017年，Lin等[21]提出特征金字塔网络(Feature Pyramid Network，FPN)，利用自底向上的分支生成多尺度特征，再利用自顶向下的分支传递顶层的语义信息. 由于FPN有效融合高层的语义特征与底层的细粒度特征，同时具有端到端的便捷优势，因此它成为多数目标检测算法的基本配置. 之后，研究者对FPN作了陆续的许多改进，例如：PANet(Pyramid Attention Network)[22]在FPN主干上增加新的自底向上分支来增强底层特征；ASFF(Adaptively Spatial Feature Fusion)[23]利用学习权重参数的方法自适应融合多层特征，较为优秀的改进方式还有AugFPN[24]以及BiFPN[25]等等.

然而，在遥感图像检测领域，由于目标存在类别众多、排列疏密不定以及前背景样本差距明显等特点，FPN固然可以进行高低层特征融合，但由于遥感图像包含的语义信息和位置信息极为丰富，导致对图像进行特征金字塔构建的过程中，仍然会丢失相当一部分的对象信息，影响检测效果. 虽然已有研究证明通道特征存在冗余性[26]，削减通道也成为模型轻量化的常用方式，但对于遥感领域的目标检测方法而言，通道之间存在的语义关联使得特征图需要生成大量通道来对复杂特征进行分解刻画. 因此，对通道维度的特征采取合适方式进行融合是很有必要的.

3 MF²Net网络

3.1 算法基础结构

图3给出了MF2Net算法的网络整体结构. 网络分为如下5个部分：残差网络、双阶段特征融合网络、特征金字塔网络、RPN精修网络以及分类与回归子网络.

残差网络采用ResNet-152[27]负责初始特征的提取工作，相比深度学习常用的ResNet-50[27]，深度神经网络ResNet-152能够提取出更为丰富的语义信息. MF2M网络负责将残差网络生成的特征提取层C_3、C_4和C_5，进行拆分重组与非对称卷积操作，进一步增强特征的表达能力. 得到增强后的特征后，特征金字塔网络进行纵向特征的整合生成. RPN优化网络负责生成可能包含检测对象的候选区域，并对检测框进行角度的精调与优化. 最后，分类与回归子网络实现检测对象的类别预测和边框的生成，并对结果进行可视化.

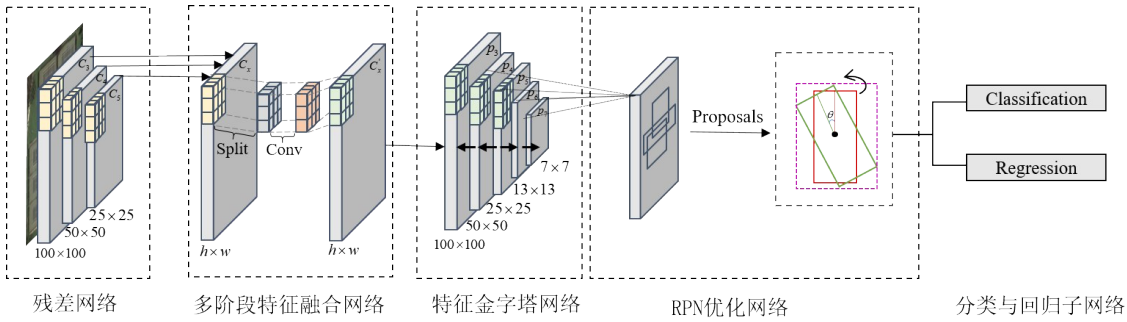


图3 MF2Net网络结构图

3.2 多阶段特征融合

3.2.1 特征拆分重组

在CNN中，卷积过程可以认为是对单个通道下的图像进行特征提取，提取后的通道数等于该卷积层的卷积核数. 由于遥感图像较为复杂，在神经网络中需要使用大量卷积核来尽可能保留细粒度特征，MF2Net在残差网络生成的C_3、C_4和C_5层，它们的卷积核数(通道数)分别达到了512、1 024和2 048. 如此多的通道，虽然会存在一定的冗余度，但是通道是表现特征的直接形式，通道之间也一定存在相关的联系. 为了证明多通道特征之间的联系，本文对遥感图像(检测目标是篮球场)通过C_3层的第一层卷积所生成的特征图进行可视化，并截取前12个通道特征，如图4所示. 可以观察到，除去特征度不高的冗余通道，其余通道均表征着检测

目标特性,包括边缘轮廓、内部纹理、色彩区分等等. 因此,不同于现有的多尺度特征变换,为进一步丰富小目标特征,双阶段特征融合方法MF2M在一阶段构建拆分重组模块,对相邻通道组特征进行融合.

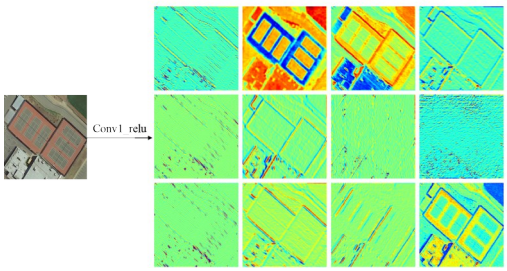


图4　相邻通道特征可视化

图5展示了该模块的基本流程:给定输入数据$\boldsymbol{X}\in\mathbb{R}^{c\times h\times w}$,$c$是输入的通道数,$h$和$w$是输入数据$\boldsymbol{X}$的宽和高,将$\boldsymbol{X}$经过$1\times1$卷积得到$L$组具有相同通道数$c/L$的特征图$\boldsymbol{x}_i, i\in[0,L-1]$,$\boldsymbol{x}_i$再进行$3\times3$卷积得到具有更大感受野信息的输出特征$\boldsymbol{x}'_i$,此时,再将$\boldsymbol{x}'_i$按照通道数分成两组输出特征$\boldsymbol{x}'_{i,1}$与$\boldsymbol{x}'_{i,2}$,每组通道数为$c/(2L)$. 在通道聚合过程中,本文以两组特征为操作单元,将相邻组的通道特征进行交换拼接,即:

$$\boldsymbol{x}'_{is}=\begin{cases}\boldsymbol{x}'_{i,1}\oplus\boldsymbol{x}'_{i+1,1}\ ,\ i=2k,k=0,1,\cdots,\dfrac{c}{2L}\\ \boldsymbol{x}'_{i,2}\oplus\boldsymbol{x}'_{i-1,2}\ ,\ i=2k+1,k=0,1,\cdots,\dfrac{c}{2L}\end{cases}\tag{1}$$

其中,$\oplus$表示在通道维度上对相邻组特征进行拼接,$\boldsymbol{x}'_{is}$表示完成两组特征拼接的特征图,通道数为c/L.

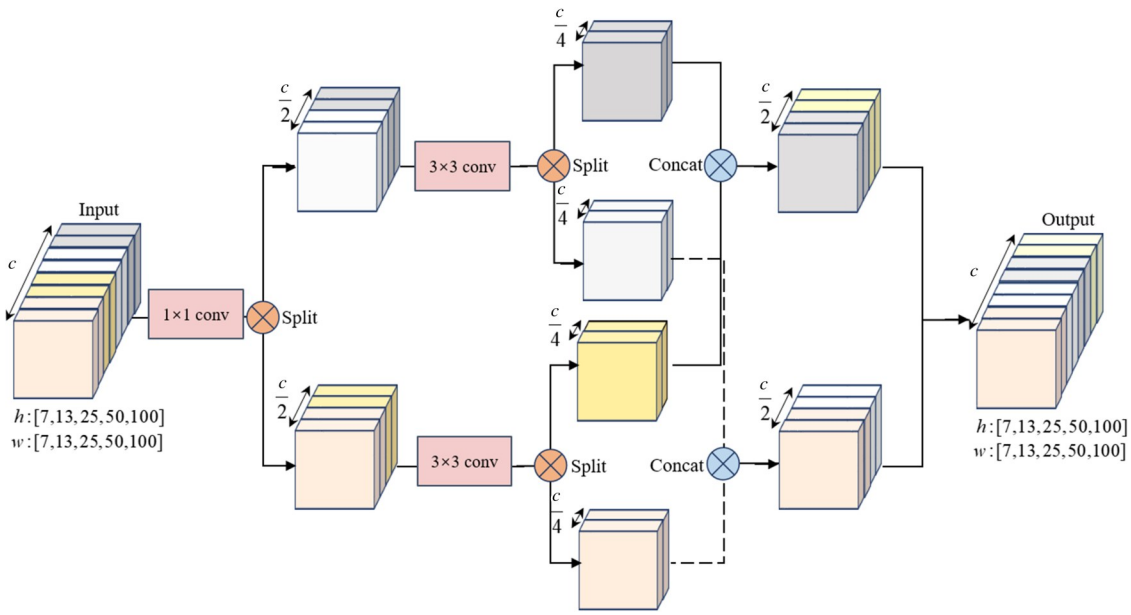


图5　特征拆分重组模块流程

对相邻组间的通道特征进行拆分重组,可以在更好的融合纵向特征的同时丰富语义信息,实现通道之间的信息聚合,并且由于交换操作简单且易于实现,减少了卷积过程中的参数量,更有利于检测性能的提升.

3.2.2　非对称卷积块设计

受仿生学影响,计算机视觉领域的研究者往往会采用1×1,3×3,5×5等卷积核大小的对称卷积方式进行对象特征提取,从而贴合人眼所接受的球状视觉区域范围. 然而遥感图像中多数检测对象呈现的是统一的矩形形状,例如船舶、桥梁、球场、车辆等等,对称卷积往往只能得到目标的局部特征,而非对称卷积则能够根据自身形状有效提取相应比例特征. 此外,非对称卷积会显著减少卷积参数量,在模型的压缩加速工作上也被研究者所青睐,例如Inception-v3[28]使用1×3以及3×1大小的卷积核,在精度略受损失的情况下减少了近33%的计算量. 因此,为了适应遥感图像目标特点,本文对ACNet[29]中的卷积融合方法进行改进. 依据二维卷积的可加性[29],在输入及步幅相同的前提下,若需得到相同输出,可以将大小不同但是兼容的卷积核进行相加,从而得到一个相同输出的等效卷积核. 即:

$$\boldsymbol{U}*\boldsymbol{T}^{(1)}+\boldsymbol{U}*\boldsymbol{T}^{(2)}=\boldsymbol{U}*(\boldsymbol{T}^{(1)}\oplus\boldsymbol{T}^{(2)})\tag{2}$$

其中,$\boldsymbol{U}$是特征图,通常以矩阵形式表示,$\boldsymbol{T}^{(1)}$和$\boldsymbol{T}^{(2)}$是两个二维卷积核,$\oplus$表示卷积核在相应位置的求和操作.

根据对遥感图像目标横纵比系数的分析,本文设计的非对称卷积块由两部分组成,如图6所示. 第一个部分使用2个串联的3×3卷积核$\boldsymbol{T}_1^{(j)}$和$\boldsymbol{T}_2^{(j)}$(由5×5卷积核分解而来)、一个1×5大小卷积核$\tilde{\boldsymbol{T}}_1^{(j)}$以及一个$5\times1$大小的卷积核$\hat{\boldsymbol{T}}_1^{(j)}$来替代方形卷积核;第二个部

分使用3个串联的3×3卷积核 $\boldsymbol{T}_3^{(j)}$、$\boldsymbol{T}_4^{(j)}$ 和 $\boldsymbol{T}_5^{(j)}$（由7×7卷积核分解而来）、一个1×7大小卷积核 $\tilde{\boldsymbol{T}}_2^{(j)}$ 以及一个7×1大小的卷积核 $\hat{\boldsymbol{T}}_2^{(j)}$. 三条通道融合后的结果按卷积块顺序依次表示为

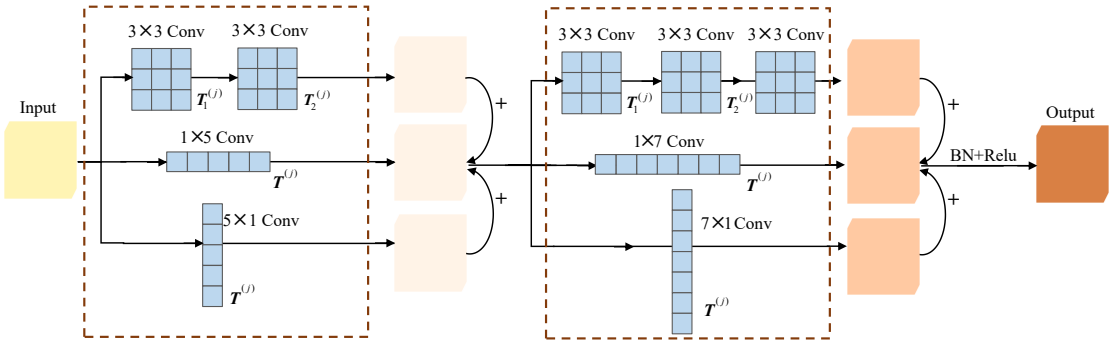


图6 非对称卷积的3条卷积通道

$$\boldsymbol{T}_1'^{(j)}=\frac{\eta_j}{\mu_j}(\boldsymbol{T}_1^{(j)}\oplus\boldsymbol{T}_2^{(j)})\oplus\frac{\tilde{\eta}_j}{\tilde{\mu}_j}\tilde{\boldsymbol{T}}_1^{(j)}\oplus\frac{\hat{\eta}_j}{\hat{\mu}_j}\hat{\boldsymbol{T}}_1^{(j)} \tag{3}$$

$$\boldsymbol{T}_2'^{(j)}=\frac{\eta_j}{\mu_j}(\boldsymbol{T}_3^{(j)}\oplus\boldsymbol{T}_4^{(j)}\oplus\boldsymbol{T}_5^{(j)})\oplus\frac{\tilde{\eta}_j}{\tilde{\mu}_j}\tilde{\boldsymbol{T}}_1^{(j)}\oplus\frac{\hat{\eta}_j}{\hat{\mu}_j}\hat{\boldsymbol{T}}_2^{(j)} \tag{4}$$

其中，η_j、$\tilde{\eta}_j$、$\hat{\eta}_j$ 是梯度学习的缩放因子，μ_j、$\tilde{\mu}_j$、$\hat{\mu}_j$ 是批量归一化层的标准差.

4 实验评估

实验硬件环境为配有Intel(R) Xeon(R) Gold 5218 CPU @ 2.30 GHz×64、内存为251.3 GiB以及装载有NVIDIA Tesla V100-SXM2-16GB×4的服务器，软件环境为Cent OS 7.6操作系统、Cuda10.2、Cudnn7.6.5、Tensorflow1.13以及Python3.7. 为验证特征融合方法的可靠性，本文使用公开数据集DOTA(Dataset for Object deTection in Aerial images)[30]进行实验分析，并采用平均精度均值(mean Average Percision，mAP)与每秒处理图像张数作为精度与速度评估标准. mAP是多类别检测精度(Average Percision，AP)的均值，综合了检测准确率(persion)和召回率(recall)，是目标检测领域最重要的评估指标.

4.1 数据集介绍

DOTA是大型公开的遥感图像基准数据集，采集来自谷歌地球、GF-2卫星、JL-1卫星等不同平台及传感器的数千张图像，并根据检测对象数目的不同分为v1.0、v1.5、v2.0版本，本文使用的DOTA-v1.0版本涵盖了大小范围从800×800像素到4 000×4 000像素的2 806张图像，数据集按照3∶1∶2的数量比例划分为训练集、验证集以及测试集. 图像共标记188 282个检测对象，囊括15种遥感图像常见类别，包括直升机(Helicopter，HC)、游泳池(Swimming Pool，SP)、港口(Harbor，HA)、环岛(Roundabout，RA)、足球场(Soccer Ball Field，SBF)、储罐(Storage Tank，ST)、篮球场(Basketball Court，BC)、网球场(Tennis Court，TC)、船舶(Ship，SH)、大型车辆(Large Vehicle，LV)、小型车辆(Small Vehicle，SV)、田径场(Ground Track Field，GTF)、桥梁(Bridge，BR)、棒球场(Baseball Diamond，BD)和飞机(Plane，PL).

DOTA的对象标注格式采用顶点边界框进行标注，顶点按照顺时针顺序进行排列，依次表示为$(x_1,y_1,x_2,y_2,x_3,y_3,x_4,y_4)$，除此之外，每个对象注释有类别和检测难度两个属性，类别即为上述15种常见类别，检测难度分为困难与不困难，分别用1和0进行表示. DOTA是类别极不均衡的长尾数据集，同一类别也存在多种尺度，如图7所示.

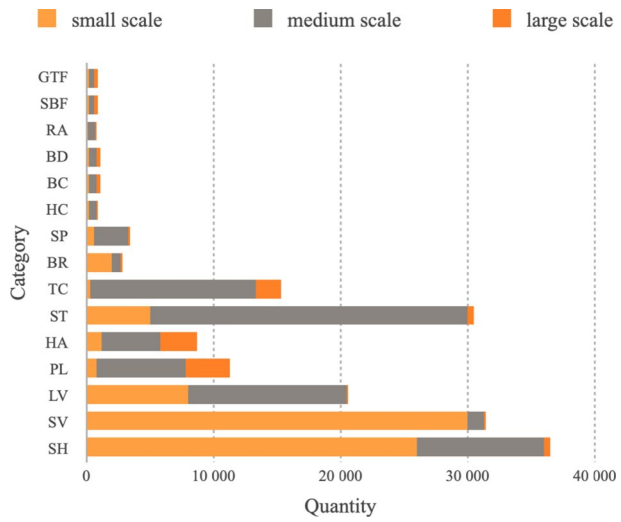


图7 DOTA数据集不同检测类别的数量

4.2 参数设置

训练参数方面，实验同时使用4块GPU进行模型训练与评估，批处理大小(Batchsize)设置为4，由于受GPU显存限制，将数据集图像按从左至右、从上至下顺序统一裁剪为800×800像素的子图像作为模型输入，设置子图像重叠步长为150像素. 模型进行36轮(共183 600次)训练迭代，初始学习率设置为5×10^{-4}，并在

20轮(第108 000次)和26轮(第140 400次)迭代之间指数衰减至5×10^{-6}.

模型超参数方面,在对不同参数进行重复训练与结果对比之后,为适应遥感图像中检测对象的尺度多样性,特征金字塔网络中的5层特征图(p_3、p_4、p_5、p_6、p_7)均在每一个像素点预设15个锚框,基准锚框尺度依次设置为32、64、128、256和512,一层特征图的锚框横纵比为{1/5, 1/3, 1/2, 1, 2, 3, 5},缩放比为$\{1, 2^{1/3}, 2^{2/3}\}$.

4.3　对比实验

为准确评估本文算法性能,实验训练图像采用DOTA训练集与验证集,推理图像采用测试集,产生的mAP值递交至DOTA官方服务器获得. 表1列举了5种不同算法在DOTA上的实验结果,包括属于双阶段检测器的R^{2}CNN[16]、RRPN[14]、SCRDet[31]和属于单阶段检测器的Rotation-RetinaNet[32]与R^{3}Det[13]. 其中,对比算法均采用Resnet-50+FPN作为骨干网络,符合控制变量的要求. 从表1中的数据得知,相较于基准算法与其余对比算法,MF2Net的检测精度达到最优效果,mAP值达到76.44%. 检测速度虽相较于R^{3}Det略受损失,达到19.5 fps,但相比其余算法仍具备速度优势.

因此,本文算法在保证检测速度的同时显著提升了检测精度,具备良好的可靠性. 图8示了MF2Net在DOTA上的旋转检测框效果.

4.4　消融实验

消融实验[7](Ablation Experiment)是量化目标检测算法模块有效程度的常用做法. 本文消融实验进行的步骤如下:建立基准模型、分离算法模块、进行模型训练和生成测试结果. 其中,基准模型仍采用R^{3}Det[13],基础网络采用ResNet-152[27]与FPN网络[21],损失函数采用分类子网络的交叉熵损失与回归子网络的SmoothL$_1$损失[7]. 算法模块分离为特征拆分重组与非对称卷积模块. 为保证结果数据的可靠准确,模型训练与测试过程均在DOTA数据集[30]上进行,实验初始参数均互相保持严格一致. 表2展示了不同模块对模型检测精度的影响,其中,"×"和"√"分别表示删除和添加模块. 两个模块都促进了检测精度的提高,非对称卷积模块的促进作用最为明显,mAP值相较于未采用状态提高了2.08%,这证明了本文所改进的算法模块可以准确提升

表1　不同算法在DOTA数据集上的检测结果

类别	双阶段检测算法			单阶段检测算法		
	R^{2}CNN[16]	RRPN[14]	SCRDet[31]	Retina Net-R[32]	R^{3}Det[13]	MF2Net
PL	80.94%	88.52%	**89.98%**	88.92%	89.42%	89.03%
BD	65.67%	71.20%	80.65%	67.67%	81.03%	**84.23%**
BR	35.34%	31.66%	52.09%	33.55%	50.41%	**58.29%**
GTF	67.44%	59.30%	**68.36%**	56.83%	65.93%	68.17%
SV	59.92%	51.85%	68.36%	66.11%	**70.90%**	69.74%
LV	50.91%	56.19%	60.32%	73.28%	78.63%	**83.45%**
SH	55.81%	57.25%	72.41%	75.24%	78.03%	**87.99%**
TC	90.67%	90.81%	90.85%	90.87%	90.67%	**94.84%**
BC	66.92%	72.84%	87.94%	73.95%	85.24%	**89.23%**
ST	72.39%	67.38%	**86.86%**	75.07%	84.10%	84.16%
SBF	55.06%	59.69%	**65.02%**	43.77%	61.64%	62.35%
RA	52.23%	52.84%	**66.68%**	56.72%	63.52%	66.28%
HA	55.14%	53.08%	66.25%	51.05%	68.15%	**68.83%**
SP	53.35%	51.94%	68.24%	55.86%	69.80%	**70.67%**
HC	48.22%	53.58%	65.21%	21.46%	67.09%	**69.38%**
mAP	60.67%	61.01%	72.61%	62.02%	73.63%	**76.44%**
Speed	2.1 fps	3.5 fps	18.4 fps	12.7 fps	**20.1** fps	19.5 fps

图8　MF2Net在DOTA上的可视化结果

网络性能.

表 2 MF2Net 模块消融实验

算法模块	MF^{2}Net		
基准模型	√	√	√
拆分重组	×	√	√
非对称卷积	×	×	√
mAP/%	73.65	74.37	**76.44**

最后,本文在消融实验的基础上,对特征融合模块的性能进行了更深入的探究,得出以下2点结论.

(1)拆分重组模块:观察表3结果可以发现,该模块让检测精度相比基准模型得到了提升,这说明通道特征具有关联性,适当的对通道特征进行融合有助于增强相关语义信息. 然而,本文在对特征拆分的不同组数进行实验中发现,随着组数的增加,检测精度在上升到某个阈值后会出现波动甚至下降,如表3所示. 针对出现精度损失的原因,本文认为,特征图的每个通道代表不同的特征类别,若对不同通道进行特征的深度耦合,反而会破坏原有特征的显著性,从而导致特征弱化并降低神经网络的性能. 因此,选择适合的通道数进行特征的拆分融合是十分必要的.

表 3 分组数量对检测精度的影响

分组数量	检测精度 mAP/%
2	73.11
4	74.37
8	**75.69**
16	73.09
32	71.57

(2)非对称卷积模块:图9展示了该模块消融实验前后,DOTA数据集中的15个类别检测平均精度的变化. 可以直观的发现,横纵比例较大的类别,包括船舶、桥梁、球场以及大型车辆等,它们的平均精度相较消融后有不同程度的明显提升,船舶与桥梁的检测精度分别提升了9.96%和7.88%,图10对比了检测示例消融实验的可视化结果(对比效果见红色虚框),左侧与右侧分别为基准实验与增加非对称卷积模块实验的检测效果. 同样观察到,加入非对称卷积模块实验后的检测框与目标贴合程度更完善. 这清晰的表明非对称卷积可以更好的检测长条矩形状的遥感图像目标.

5 结束语

针对遥感场景下目标具有小尺度、分布密集、横纵比不均衡而难以检测的问题,本文结合通道特征的内在联系与多比例非对称卷积的作用,提出一种基于双阶段特征融合的遥感目标检测方法MF2M. 结合MF2M的MF2Net网络在公开数据集DOTA上的实验结果表明,相较于基准方法,改进后的算法能在保证检测速度的同时显著提升检测精度,取得了较好的整体检测性能. 本文方法是值得继续研究探索的一个方向,事实上,MF2M属于即插即用的通识模块,因此在后续研究中,可以将本文方法移植到不同网络,在其他视觉任务中发挥作用;还可以重新设计对通道特征的拼接方式,例如结合注意力机制下的特征权重对通道进行加权,更好的聚合目标高维特征;并且可以继续探索其他形状的卷积核对大横纵比遥感目标的作用,并设计不同的卷积次序提升检测性能. 此外,遥感图像目标检测算法普遍存在硬件资源消耗过多,检测时延较高等缺陷,因此更好的对模型进行压缩,实现工业化需求同样是后续工作的研究方向.

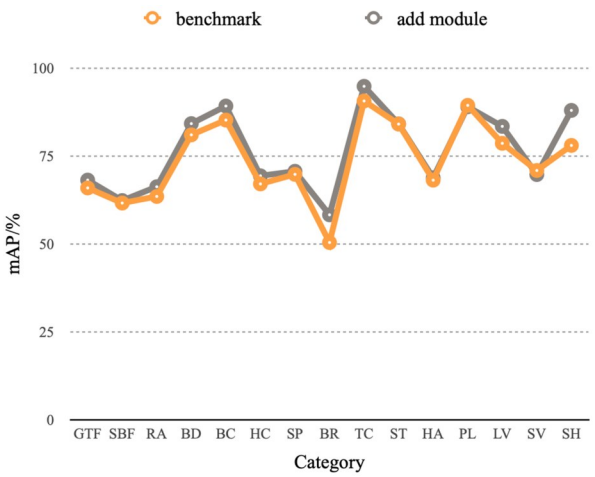


图9 非对称卷积消融实验前后的检测精度对比

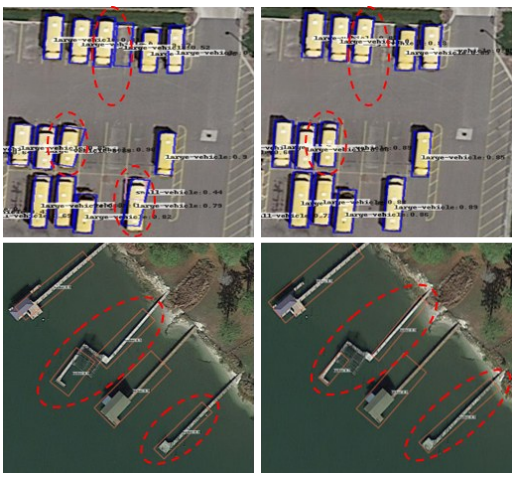


图10 消融实验的检测框对比(车辆与港口)

参考文献

[1] 罗会兰, 陈鸿坤. 基于深度学习的目标检测研究综述[J]. 电子学报, 2020, 48(6): 1230-1239.
LUO H L, CHEN H K. Survey of object detection based on deep learning[J]. Acta Electronica Sinica, 2020, 48(6): 1230-1239. (in Chinese)

[2] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[3] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]//European Conference on Computer Vision. Cham: Springer, 2014: 740-755.

[4] EVERINGHAM M, VAN GOOL L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International journal of computer vision, 2010, 88(2): 303-338.

[5] OKSUZ K, CAM B C, KALKAN S, et al. Imbalance problems in object detection: A review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43 (10): 3388-3415

[6] 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.

[7] 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.

[8] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2980-2988.

[9] 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). Piscataway: IEEE, 2016: 779-788.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[C]//Computer Vision—ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[11] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as points[EB/OL]. (2019-04-16) [2021-10-01]. https://arxiv.org/abs/1904.07850.

[12] ZHOU X Y, ZHUO J C, KRÄHENBÜHL P. Bottom-up object detection by grouping extreme and center points [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 850-859.

[13] YANG X, YAN J C, FENG Z M, et al. R3Det: Refined single-stage detector with feature refinement for rotating object [EB/OL]. (2019-08-15)[2021-10-01]. https://arxiv.org/abs/1908.05612.

[14] 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.

[15] ZHOU X Y, YAO C, WEN H, et al. EAST: An efficient and accurate scene text detector[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 2642-2651.

[16] 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) [2021-10-01]. https://arxiv.org/abs/1706.09579.

[17] LIU Z K, HU J G, WENG L B, et al. Rotated region based CNN for ship detection[C]//2017 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2018: 900-904.

[18] DING J, XUE N, LONG Y, et al. Learning RoI transformer for detecting oriented objects in aerial images[EB/OL]. (2018-12-01)[2021-10-01]. https://arxiv.org/abs/1812. 00155.

[19] YANG X, HOU L P, ZHOU Y, et al. Dense label encoding for boundary discontinuity free rotation detection[EB/OL]. (2020-11-19)[2021-10-01]. https://arxiv.org/abs/2011.09670.

[20] 王子晔, 苗夺谦, 赵才荣, 等. 基于多粒度特征的行人跟踪检测结合算法[J]. 计算机研究与发展, 2020, 57(5): 996-1002.
WANG Z Y, MIAO D Q, ZHAO C R, et al. A pedestrian tracking algorithm based on multi-granularity feature[J]. Journal of Computer Research and Development, 2020, 57(5): 996-1002. (in Chinese)

[21] 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 (CVPR). Piscataway: IEEE, 2017: 936-944.

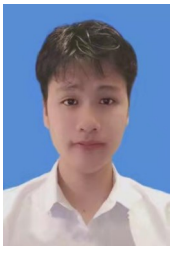
[22] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8759-8768.

[23] LIU S T, HUANG D, WANG Y H. Learning spatial fusion for single-shot object detection[EB/OL]. (2019-11-27)[2021-10-01]. https://arxiv.org/abs/1911.09516.

[24] GUO C X, FAN B, ZHANG Q, et al. AugFPN: Improving multi-scale feature learning for object detection[C]// 2020 IEEE/CVF Conference on Computer Vision and Pat-

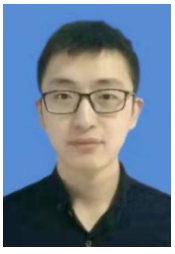
tern Recognition (CVPR). Piscataway: IEEE, 2020: 12592-12601.
[25] TAN M X, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[EB/OL]. (2019-05-28)[2021-10-01]. https://arxiv.org/abs/1905.11946v5.
[26] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 1577-1586.
[27] 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 (CVPR). Piscataway: IEEE, 2016: 770-778.
[28] 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). Piscataway: IEEE, 2016: 2818-2826.
[29] DING X H, GUO Y C, DING G G, et al. ACNet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 1911-1920.
[30] 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. Piscataway: IEEE, 2018: 3974-3983.
[31] 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). Piscataway: IEEE, 2020: 8231-8240.
[32] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2999-3007.

作者简介



陈 立 男，1997 年 2 月出生于浙江省义乌市．信息工程大学硕士生．主要研究方向为计算机视觉．
E-mail: 2464863136@qq.com

张 帆(通讯作者) 男，1981 年 9 月出生．博士．现为国家数字交换系统工程技术研究中心副研究员、硕士生导师．主要研究方向为主动防御、人工智能、高性能计算．
E-mail: 17034203@qq.com



郭 威 男，1990 年 8 月出生．博士．现为国家数字交换系统工程技术研究中心助理研究员．主要研究方向为主动防御、人工智能、高性能计算．
E-mail: guowjss@126.com

黄 赟 男，1993 年 9 月出生于江西省新余市．信息工程大学硕士生．主要研究方向为神经网络模型量化压缩．
E-mail: yyhuangz@163.com