

文章编号 1004-924X(2023)16-2465-18

多元特征提取与表征优化的遥感多尺度目标检测

孟月波, 王 菲, 刘光辉*, 徐胜军

(西安建筑科技大学 信息与控制工程学院, 陕西 西安 710055)

摘要: 遥感目标具有较大的尺度差异性, 针对其在复杂背景干扰下易导致细粒度级别多尺度特征提取困难、预测部分有效表征较弱的问题, 本文基于无锚框思想, 提出一种多元特征提取与表征优化的遥感多尺度目标检测方法 (Multivariate Feature extraction and Characterization optimization, MFC)。在特征提取部分, 设计多元特征提取模块 (Multivariate Feature Extraction, MFE) 挖掘细粒度级别的多尺度特征, 通过分组操作及跨组连接的方式扩大感受野、增强多个特征尺度的组合效应, 并联合上下文信息进一步加强对小目标的关注; 采用深层聚合结构对深浅层特征进行充分融合, 以获得更全面的特征表达。在预测部分, 提出一种表征优化策略 (Characterization Optimization Strategy, COS), 利用椭圆型映射进行标签优化以适应具有较大纵横比的遥感目标, 设计坐标像素注意力组合关注多尺度目标通道、位置及像素信息, 减少复杂背景干扰, 使有效信息得以突出表征。在 DIOR, HRRSD, RSOD 数据集上进行消融及对比实验, 实验结果表明: MFC 模型的 mAP 分别达到了 70.9%, 90.2% 和 96.9%, 优于大多现有方法, 有效改善了误检、漏检问题, 适应性和鲁棒性较强。

关键词: 遥感多尺度目标; 无锚框; 多元特征; 表征优化; 注意力

中图分类号: TP751 **文献标识码:** A **doi:** 10.37188/OPE.20233116.2465

Remote sensing multi-scale object detection based on multivariate feature extraction and characterization optimization

MENG Yuebo, WANG Fei, LIU Guanghui*, XU Shengjun

(College of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China)

* Corresponding author, E-mail: guanghuil@163.com

Abstract: Remote sensing objects have large scale differences. In order to solve the problems that they are prone to lead to difficulties in fine granularity multi-scale feature extraction and weak prediction part of effective representation under complex background interference, a multi-scale remote sensing object detection method (MFC) for multivariate feature extraction and characterization optimization based on the idea of anchor-free is proposed. In the feature extraction part, a multivariate feature extraction module (MFE) is designed to mine multi-scale features at the fine granularity level, expand the receptive field through grouping operation and cross group connection, enhance the combination effect of multiple feature scales, and further strengthen the focus on small objects by combining context information; The deep and shallow features are fully integrated by the deep layer aggregation structure to obtain a more comprehensive feature

收稿日期: 2022-11-10; 修订日期: 2022-11-28.

基金项目: 陕西省重点研发计划资助项目 (No. 2021SF-429); 国家自然科学基金面上项目资助 (No. 52278125)

expression. In the prediction part, a characterization optimization strategy (COS) is proposed, which uses elliptical mapping to optimize tags to adapt to remote sensing targets with large aspect ratio. And a Coordinate-Pixel attention is designed to focus on multi-scale object channels, positions and pixel information, reduce complex background interference, and make effective information prominent. Ablation and contrast experiments were conducted on DIOR, HRRSD and RSOD datasets. The experimental results showed that the mAP of MFC model reached 70.9%, 90.2% and 96.9% respectively, which was superior to most existing methods. It effectively improved the problems of false detection and missing detection, and had strong adaptability and robustness.

Key words: remote sensing of multi-scale targets; anchor-free; multivariate feature; characterization optimization; attention

1 引言

随着卫星、遥感及传感器技术的日益发展,遥感图像数据不断增长,涵盖的信息愈加丰富、地域跨度极大,包含的地物目标也更加复杂多样,目标类内及类间的尺度差异性也愈加显著,遥感目标检测迎来新阶段蓬勃发展的同时也面临着更加艰巨的挑战。遥感目标检测技术广泛应用于发展国防军事侦查、环境监测及城市规划等领域中,在对地观测系统中有着至关重要的研究意义与应用价值。

在遥感领域,基于传统目标检测的方法在提取目标特征时,往往过分依赖大量先验知识来设计手工特征,过程繁琐、精度不佳且易受外界环境干扰导致鲁棒性较差,难以适应复杂多样的遥感任务场景。近年来,基于深度学习的目标检测算法在精度和速度上均远优于传统方法,逐渐占据了主流位置,根据预测框生成方式可分为基于锚框的目标检测方法(Anchor-based)和无锚框(Anchor-free)的目标检测方法。

基于锚框的方法依据分类回归的方式通常包含以精度占优的两阶段检测方法,如Faster R-CNN^[1]系列,Mask R-CNN^[2]等,以及以速度著称的单阶段检测方法,如YOLO^[3]系列,RetinaNet^[4]等。而遥感图像不同于自然图像,基于其目标尺度变化剧烈、尤其小目标较难检测且背景复杂多样等难点,不少研究学者据此提出了针对性的改进算法。两阶段网络中,文献[5]和文献[6]分别通过引入并行设计和多级融合结构来对应不同尺度目标进行检测,可提高一定精度但鲁棒性不强。文献[7]针对遥感目标尺寸变化大及类

间相似性高的特点,通过通道注意力模块和跨尺度特征融合模块,增加更具判别性的多尺度特征表达改善检测效果,但适应性不强。文献[8]提出多尺度密集特征融合模型,通过含有不同尺度卷积核的多尺度特征融合残差模块提取不同尺度图像特征并融合,以丰富多尺度特征。两阶段方法检测精度较高,但普遍设计繁杂且伴随计算量的显著增加,仍以牺牲效率为代价提高检测效果。一阶段网络中,FFC-SSD^[9],FE-YOLO^[10]可保证一定的计算速度,分别利用目标框分组聚类、反池化高效多尺度特征融合模块和特征增强网络,聚合上下文信息获取丰富特征表达;关注多尺度目标中的小目标及复杂背景等难点进行优化的算法^[11-13]有一定效果提升,但在改善误检、漏检问题上与两阶段算法仍存在一定的差距。上述基于锚框的检测方法虽各具优势,但往往较难实现检测精度与速度的平衡,同时涉及大量锚框及超参数等复杂计算,面对遥感复杂任务时往往受锚框数量、大小及宽高比等众多因素制约,参数设定也对最终检测性能十分敏感,导致网络效率和适应性普遍较低。

无锚框思想由CornerNet^[14]首次提出后凭借灵活简洁的框架开启了目标检测新方向,自此涌现的一系列方法^[15-17]也为遥感目标检测领域带来了新思路。在实际应用时,网络难以全面提取到遥感场景下丰富多样的多尺度目标特征信息,缺乏对通道、位置等有效信息的突出表征,较难适应遥感复杂场景下的多尺度目标检测。无锚框的方法中,借助角点检测的CornerNet及其优化网络CornerNet-Lite^[15]摒弃了锚框设计,分别通过角点池化和引入注意力及紧凑骨干两种变体

组合等方式提高检测结果,但会存在角点归属问题且易受外界复杂环境影响,从而产生误检等问题。而通过极值点、角点去联合中心点的 ExtremeNet^[16]和 CenterNet^[17]针对角点检测的不足有着一定改善,不仅加强了对物体内部信息关注也考虑一定全局信息的补充,但由于关键点分组检测及后处理等操作导致实时性不高,上下文信息间交互不足。基于全卷积的 FCOS^[18]主要利用类似语义分割的思想,通过结合特征金字塔的方式获得多尺度特征图,但对预测部分需要的有效信息关注不足,在遥感任务中检测效果较弱。而文献[19]将目标中心点视为关键点,以回归的方式获取目标宽高等信息进行预测,更加灵活快速,但较为单一的感受野难以实现对不同尺度、纵横比等目标特征的全面捕获。文献[20]基于关键点检测可灵活应对复杂多样的遥感目标,但精度提升甚微,对尺度变化剧烈等目标检测效果并不突出。文献[21]和文献[22]在关键点检测的基础上,分别通过在编解码结构中加入水平连接以解决多尺度问题,以及通过设计特征选择模块和动态增强检测头操作,实现对形状、尺度差异大的遥感目标的自适应调整从而优化检测效果,但在复杂背景干扰下,对位置等有效信息会造成不同程度的损失,影响检测结果准确输出,依然会产生不少漏检、误检等问题。

上述无锚框的检测算法虽然摆脱了锚框诸多限制而更简洁灵活,但只关注部分局部信息且感受野不足,导致多尺度特征提取不全面、不充分,无法有效应对遥感目标巨大的尺度差异;预测阶段易受背景干扰导致有效表征不突出,影响检测结果。本文认为充足的感受野以及对全局信息的关注十分重要,加强多尺度特征提取能力、抑制背景对重要表征的干扰,可有效改善遥感多尺度目标检测效果。

鉴于此,本文基于无锚框思想,提出一种多元特征提取与表征优化的遥感多尺度目标检测方法(Multivariate Feature extraction and Characterization optimization, MFC)。在特征提取部分,通过设计多元特征提取模块(Multivariate Feature Extraction, MFE)深挖细粒度级别多尺度特征,增大感受野以适应多尺度目标,并联合上下文信息重点加强对小目标的关注,提取全面且丰

富的多元特征;采用深层聚合^[23](Deep Layer Aggregation, DLA)结构实现深浅层空间、语义信息充分聚合,获得更全面的多尺度特征表达。在预测部分,提出一种表征优化策略(Characterization Optimization Strategy, COS),利用椭圆型映射优化标签和坐标像素注意力(Coordinate-Pixel Attention, CPA)组合优化并加强有效信息的突出表征,有效降低复杂背景干扰的影响。实验结果表明,本文方法显著提高了在复杂遥感场景下对多尺度目标的检测效果。

2 多元特征提取与表征优化的遥感多尺度目标检测模型 MFE

2.1 网络结构

多元特征提取与表征优化的遥感多尺度目标检测 MFC 网络结构如图 1 所示,主要由特征提取网络以及预测网络两部分构成。

在特征提取网络中,通过多元特征提取模块 MFE(图 1 中简化为 M)扩大感受野提高对多尺度目标特征的全面捕获,联系上下文着重关注小尺度目标,而使用 16 个 MFE 模块可随着网络的不断加深挖掘不同层次中的多尺度特征,获得含有丰富多元信息的高质量特征图。采用的深层聚合结构包括分层深度聚合(Hierarchical Deep Aggregation, HDA)以及迭代深度聚合(Iterative Deep Aggregation, IDA)两部分,分别从层级和分辨率不同侧重出发对获取到的多元特征进行全面且充分地融合。具体来说,由 {H1, H2, H3, H4} 四个分层深度聚合 HDA(图 1 中简化为 H)在各个层级内对 MFE 模块提取到的多尺度特征信息进行初步聚合并保留聚合结果;由 {I1, I2, I3} 三个迭代深度聚合 IDA(图 1 中简化为 I)从浅层低分辨率特征开始逐步与深层高分辨率特征进行跨越连接操作,实现跨尺度、跨层级的充分融合,得到全面且饱含多尺度特征信息的高质量特征图。

预测网络中,提出的表征优化策略 COS 先对输入特征图进行标签优化,将圆形映射的标签优化为椭圆形映射,提高网络对纵横比较大遥感目标的适应能力;在关键点分类预测、偏置预测和目标宽高预测的三条分支中均加入设计的坐

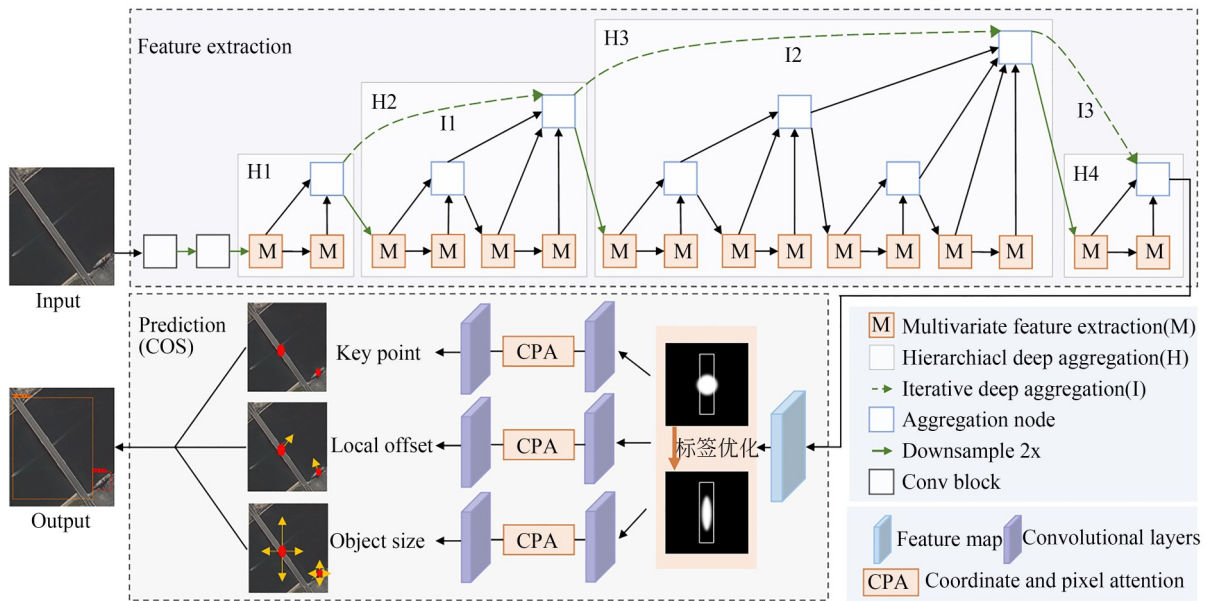


图1 多元特征提取与表征优化的遥感多尺度目标检测MFC网络结构

Fig. 1 Network structure of multivariate feature extraction and characterization optimization

标像素注意力CPA,目的是增强有效信息表征能力,减少复杂背景干扰,组合优化后得到更准确、精细的遥感图像多尺度目标预测结果。

2.2 特征提取网络

对遥感多尺度目标而言,扩大感受野以适应不同尺度目标是获取其全面特征信息的关键,而通过挖掘不同层级和深度的多尺度特征,是获得含有丰富细节信息的多元特征的关键,对其进行充分融合可为预测网络送入高质量特征图,提高预测结果。

2.2.1 骨干网络DLA-60

通常,浅层特征包含更丰富的空间细节而缺少语义信息,深层特征包含更丰富的语义信息而缺少空间信息,对深浅层特征进行充分融合是提升检测效果的重要环节之一。一般将其直接融合的方式过于简单,忽视了部分重要特征在后续融合操作中的重要性,抑制了互为补充的浅层空间信息和深层语义信息的有效发挥,融合效果较差。

深度聚合结构中的HAD和IDA可从层级和分辨率等不同方向出发,实现对多尺度特征的充分融合,常用的有DLA-34,DLA-60,DLA-102等。通常,网络结构的适当加深更有利于像遥感多尺度目标检测等复杂任务的进行,但过深的网络亦会对训练及整体效率等方面造成额外且不

必要的损失,在综合遥感检测任务需求以及考量各类成本及检测性能等因素后,本文采用DLA-60作为骨干网络提取初步特征同时保证多尺度特征信息的充分融合。

2.2.2 多元特征提取模块MFE

残差结构及其改进^[24]在自然场景中表现较佳,可有效针对多尺度特征获取提升检测性能而被广泛使用,但面对遥感场景更加剧烈的尺度变化情况,对可显著提高检测效果的多样化特征的挖掘能力稍显不足。具体来说,缺少对遥感大尺度目标全局信息的获取,易导致误检、漏检等问题,缺少对遥感小尺度目标的有效关注,会产生较大规模的漏检现象。

考虑到网络单一感受野和不充分的特征提取带来的局限性,为使网络更好地检测到遥感大、小尺度目标,提取到更多有效的多尺度特征,本文在残差结构基础上设计了多元特征提取模块MFE,如图2所示。MFE模块通过对输入进行分组操作和跨组连接的方式扩大感受野以适应不同尺度的遥感目标,其中,跨组连接可加强多尺度特征的组合效应和多层次信息间紧密联系,组合增强网络对多尺度细节特征的提取能力;内嵌的上下文模块^[25](Context Module,CM)可捕获小目标更丰富的上下文信息,为占比较小导致自身可利用特征不足这类小目标,可以最

大程度地补充周围背景信息以辅助检测;最后将不同分组输出进行拼接融合后,结合通道注意力

模块 SE^[24] 强化重要特征的通道信息,为预测部分提供较高质量特征图。

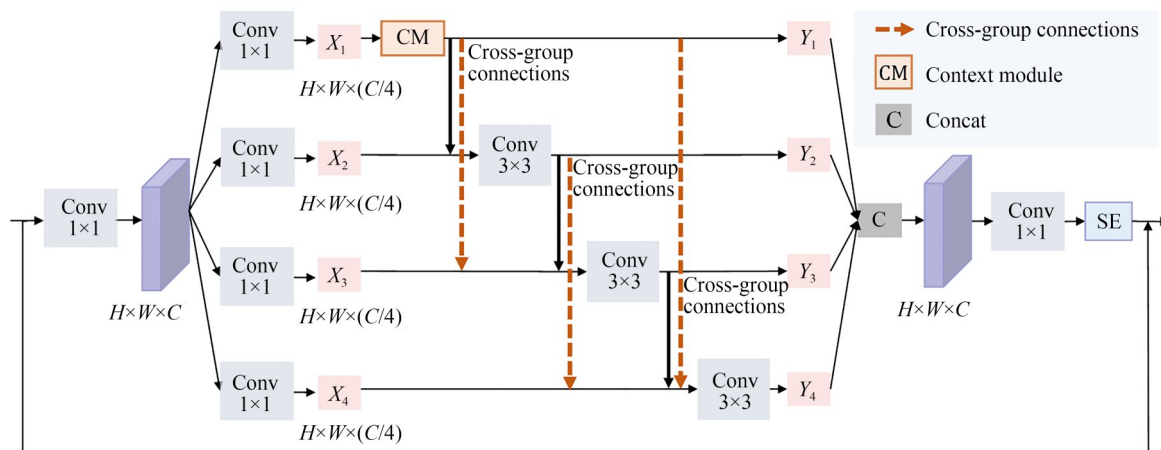


图2 MFE模块结构

Fig.2 MFE module structure

2.2.2.1 分组操作及跨组连接

在 MFE 模块中,对输入 $X \in R^{W \times H \times C}$, 其中 W 和 H 为特征图宽和高, C 为特征通道数, 将其经过一个 1×1 卷积得到空间大小不变的特征图后, 对此特征图进行分组操作, 分别通过四组 1×1 卷积得到四个特征子集, 即 $X_i (i = 1, 2, 3, 4)$, 每个特征子集的空间分辨率一致但对有效特征信息的关注能力不同, 其通道数均降为输入通道数的四分之一。四个特征子集并行操作增大网络感受野挖掘丰富的多元特征信息, 特征子集中的相邻组间可通过向下传输补充其他子集的特征多样性, 不相邻组则通过跨组连接加强不同组间的紧密联系进行多元特征提取, 最终得到各自输出 $Y_i (i = 1, 2, 3, 4)$ 。其中, 为加强对小目标的重点关注, 仅对特征子集 X_1 单独通过上下文模块 CM 以获取其周围丰富的上下文信息; 对特征子集 X_2 仅存在相邻组的输入, 即加入 X_1 的输出后再进行 3×3 卷积; 而特征子集 $X_i (i = 3, 4)$ 则除了相邻组的输入, 还需加入与之不相邻组的输出后, 再经过 3×3 卷积进行特征提取, 各子集输出 $Y_i (i = 1, 2, 3, 4)$ 在叠加中不断增大, 加强了不同尺度输出的组合效应。

在具体实现中, 令特征子集 X_1 通过 CM 模块得到补充小目标上下文信息的输出 Y_1 ; 对子集 X_2 需加上子集 X_1 的输出 Y_1 , 得到对应 3×3 感受

野的输出 Y_2 ; 对于子集 X_3 需加上子集 X_1 的输出 Y_1 和子集 X_2 的输出 Y_2 , 得到对应 5×5 感受野的输出 Y_3 ; 对于子集 X_4 需加上子集 X_1 的输出 Y_1 、子集 X_2 的输出 Y_2 以及子集 X_3 的输出 Y_3 , 得到对应 7×7 感受野的输出 Y_4 , 此过程可用公式表示如下:

$$Y_i = \begin{cases} K_i(X_i), i = 1; \\ K_i(X_i + Y_{i-1}), i = 2; \\ K_i(X_i + Y_{i-1} + Y_{i-2}), i = 3; \\ K_i(X_i + Y_{i-1} + Y_{i-2} + Y_{i-3}), i = 4. \end{cases} \quad (1)$$

其中, K_i 对特征子集 X_i 表示经过上下文模块 CM 的卷积操作, 对特征子集 $X_i (i = 2, 3, 4)$, K_i 则表示第 i 个 3×3 卷积操作。

分组操作和跨组连接的方式弥补了感受野过小导致全局信息不足等问题, 丰富了最终输出特征所包含的多尺度信息, 提高了网络恢复细节信息的能力, 挖掘多元特征的能力显著加强。

2.2.2.2 上下文模块 CM

为获取小目标丰富的上下文信息, 弥补其自身可利用特征较少的不足, 在浅层的特征子集 X_1 中嵌入了 CM 模块, 具体结构如图 3 所示, 着重关注在遥感图像中数量占比多、密集排列的众多小尺度目标, 改善大量漏检问题。

CM 模块主要由两个分支组合构成, 第一个分支主要对子集 X_1 中小目标特征实现初步提取,

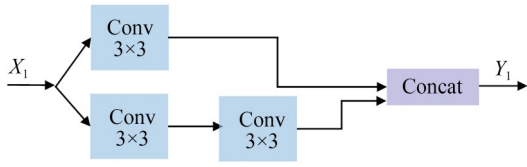


图3 CM模块结构

Fig. 3 CM module structure

仅包含一个 3×3 卷积；另一分支提取子集 X_1 更深层特征同时扩大感受野，由两个连续 3×3 卷积构成。具体实现中，经过两个分支后得到的特征图通道数均降为输入子集 X_1 通道数的一半，在分别获取不同深度的特征信息后再进行通道维度上的拼接，使输出 Y_1 在空间分辨率大小和通道数上均与输入保持一致。此时， Y_1 则既包含了小目标自身特征信息也获取了其周围丰富的上下文信息，可有效改善小目标大量漏检现象，进而提高整体检测精度。

2.2.2.3 组合输出

对经过分组操作及跨组连接后的各组输出 $Y_i \in R^{W \times H \times c} (i = 1, 2, 3, 4)$ ，将其在通道维度上进行拼接，使用 1×1 卷积对不同尺度特征信息进行融合，使各特征子集在继承了原图特征的基础上，还增强了各自输出感知范围，得到具有不同感受野组合的特征信息，获取到细粒度级别的多元特征信息。之后，为重新赋予通道特征以不同的权值，使得网络具有更多的非线性需引入注意力进行信息聚合，而CBAM注意力计算复杂性较高，选用通道注意力模块SE可更好地拟合特征通道间的相关性，剔除少部分无效特征，使特征表述更为细致清晰，从而得到更加全面且充分的多元特征信息。

2.3 表征优化策略 COS

本文网络基于无锚框思想，采用关键点检测可灵活适应多尺度目标，但遥感目标自身的尺度差异以及恶劣的天气、光照、成像质量参差不齐等复杂背景在最终关键点预测环节造成的干扰，势必会对前期获取到的丰富特征信息的充分表达产生影响，导致检测效果大打折扣，产生误检、漏检等问题。COS策略通过针对性地优化网络对纵横比较大目标的适应性，同时强化对通道、位置、像素等所需信息的有效表征，弱化背景信息干扰，减少特征损失，使其在预测部分显著

表达，组合优化预测网络的输出效果。

2.3.1 标签优化

本文通过对遥感常用数据集和初期实验结果的观察与分析，发现遥感场景中的部分常见目标，如图4所示的桥梁、水坝、大型卡车及船舰等，其自身长宽固有尺寸存在纵横比较为显著的特点，对预测部分的检测效果会产生一定影响，故本文针对此问题在预测部分进行了相关优化。

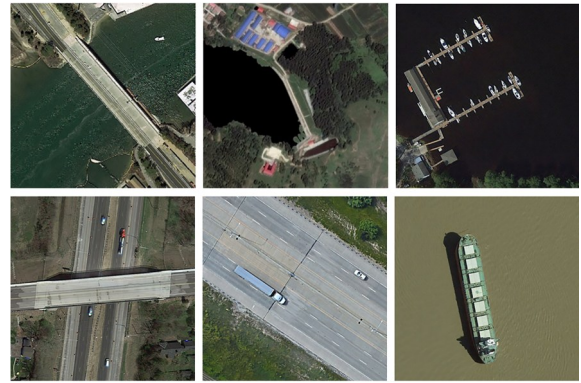


图4 遥感大纵横比目标部分示例图

Fig. 4 Example of remote sensing large aspect ratio target

一般而言，在关键点预测网络中主要将输入通过高斯核映射函数得到预测关键点热力图 $\hat{Y} \in [0, 1]^{\frac{W}{R} \times \frac{H}{R} \times c}$ ， R 为下采样倍数取为4； c 为目标类别数，也是关键点热力图通道数； W 和 H 分别为输入图像宽度和高度； \hat{Y} 为1表示检测到目标关键点， \hat{Y} 为0表示背景。

上述过程中，在将标签(ground truth)即关键点真实值，以圆形高斯核映射得到关键点热力图时，其高斯核半径为目标宽高的标准差，为一固定常数，如公式(2)所示，较难适应具有狭长外形的遥感目标。而本文认为，将标签关键点以椭圆形高斯核映射进行优化，可提高对此类纵横比较大目标的适应性，输出更准确的目标框，使高斯核半径处于目标宽高尺寸范围内，优化后标签映射如式(3)所示：

$$Y = \exp\left(-\frac{(x - \tilde{p}_x)^2 + (y - \tilde{p}_y)^2}{2\sigma_p^2}\right), \quad (2)$$

$$Y = \exp\left(-\frac{(x - \tilde{p}_x)^2}{2\sigma_x^2} - \frac{(y - \tilde{p}_y)^2}{2\sigma_y^2}\right), \quad (3)$$

其中: σ_p 为圆型映射时高斯核的宽高标准差, σ_x , σ_y 分别为椭圆形映射时高斯核宽、高各自的标准差, (x, y) 为标签对应的真实关键点坐标, $(\tilde{p}_x, \tilde{p}_y)$ 为预测所得的关键点坐标。

高斯核映射形状呈圆形时,在目标边界框外还会有部分权重非零点的存在,而网络会将这些点倾向于正样本进行学习,但通常定义边界框外的点为严格负样本,容易造成正负样本模糊,网络难以收敛等问题。而标签优化后除了更能适应纵横比较大的目标外,还可将正样本限制在目标边界框范围内,能够极大地减少模糊样本,使预测网络的适应性和准确性更好,标签优化对比如图 5 所示。

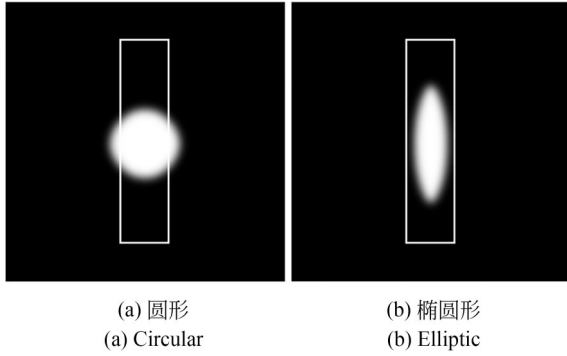


图 5 标签优化对比

Fig. 5 Label optimization comparison

2.3.2 坐标像素注意力 CPA

注意力机制是遥感领域较常见的方法,而位置信息则是对目标准确定位的关键,通过注意力机制可对有效信息重点关注同时抑制无关背景干扰,应用在预测网络中能起到有效优化特征表达的作用。而通常使用的 SE 注意力机制等虽然有一定的性能提升,但通过二维全局池化操作将特征张量转换为单个特征向量仅能获得通道信息不利于位置信息的获取,而坐标注意力 CA^[26]则在通道信息的基础上考虑到了位置信息在遥感目标定位中的重要性。此外考虑到对于即将输入至检测模块的特征,依然受到不少如光照、天气、阴影等因素干扰,影响检测输出。文献[27]所提特征注意力可弱化薄雾等背景,包含通道注意力 CA 以及像素注意力 PA,本文仅保留 PA 以提高网络对有效特征的突出表达,在预测网络中,将其与坐标注意力 CA 组合,提出坐标像

素注意力模块 CPA,如图 6 所示,可使重要信息组合输出并有效表达,一定程度上抑制遥感复杂背景干扰,获得显著增强的表征优化结果。

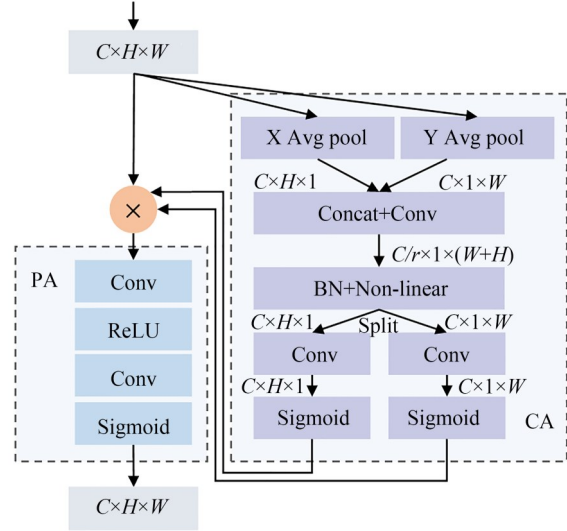


图 6 CPA 模块结构

Fig. 6 CPA module structure

CPA 在不增加额外计算开销的情况下将空间位置信息嵌入到通道注意力中,对输入通道为 C , 高度为 H , 宽度为 W 的特征图,将特征编码分解为两个并行的一维特征编码过程,在其高度和宽度两个方向上进行全局平均池化,获得高度和宽度两个方向上的特征图,公式分别如下:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_{c(h,i)}, \quad (4)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_{c(j,w)}, \quad (5)$$

其中: z_c^h , z_c^w 分别表示第 c 通道在具体高度和宽度方向上的输出; x_c 表示第 c 通道的输入。可沿一个方向捕获远程依赖关系,同时沿另一方向得到精确的位置信息,对生成的特征图编码为一对方向感知和位置敏感的注意力特征图进行互补。

将式(2)获得的特征图进行拼接和卷积变换,维度变为原来的 C/r , r 为超参数,经过批量归一化和非线性激活操作得到中间特征图,对其沿高度和宽度方向分解,使用 2 个 1×1 卷积使其通道数与输入特征图通道数保持一致,再使用 Sigmoid 函数对 W 和 H 两个维度的特征重新分配权重,将其与输入进行相乘加权操作,以增强对通道及位置信息的表征。同时,为重点优化由光

照、阴影、云雾等外界因素引起图片质量参差不齐,易使部分非高清区域对检测效果造成一定影响,在获取到通道及位置信息后,通过关注像素信息使网络更加关注模糊背景下的多尺度目标,进一步弱化光线等影响,提升整体检测效果。CPA一定程度上抑制了背景干扰,有效增强了预测部分目标关键点、偏置及目标尺寸三个分支有效信息充分表征。

2.3.3 损失函数

由图 1 可知,网络的预测部分主要包含了中心关键点(Key Point)、偏置(Local Offset)及目标尺寸(Object Size)三部分。故此,总的损失函数 L 分别由上述三部分对应的关键点分类损失 L_k 、关键点偏置损失 L_{off} 和目标框宽高损失 L_{size} 加权总和构成,相应偏置损失函数和宽高损失函数的权重系数 λ_{off} 和 λ_{size} 分别为 1 和 0.1,具体公式依次如下:

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

$$L_k = -\frac{1}{N} \begin{cases} \sum_{x_{yc}} \left(1 - \hat{Y}_{x_{yc}}\right)^\alpha \log\left(\hat{Y}_{x_{yc}}\right), Y_{x_{yc}}=1 \\ \sum_{x_{yc}} \left(1 - \hat{Y}_{x_{yc}}\right)^\beta \left(\hat{Y}_{x_{yc}}\right)^\alpha \log\left(1 - \hat{Y}_{x_{yc}}\right), Y_{x_{yc}} \neq 1 \end{cases}, \quad (7)$$

$$L_{off} = \frac{1}{N} \sum_p \left| \hat{O}_p - \left(\frac{P}{R} - \tilde{P}\right) \right|, \quad (8)$$

$$L_{size} = \frac{1}{N} \sum_{k=1}^N \left| \hat{S}_{pk} - S_k \right|, \quad (9)$$

其中: α 和 β 为关键点分类损失函数中超参数; $Y_{x_{yc}}$ 即为式(3)标签优化后以椭圆形映射的高斯核函数, (x, y) 为真实标签关键点坐标, c 为目标类别数; N 为输入图像内的目标总数; \hat{O}_p 为关键点偏置量, P 为真实标签关键点, \tilde{P} 为预测所得关键点, R 为下采样倍数; \hat{S}_{pk} 为预测所得目标的宽和高, S_k 为真实目标的宽和高。

3 实验及分析

3.1 数据集及实验设置

为评估本文所提方法的有效性,在 DIOR^[28], HRRSD^[29] 以及 RSOD^[30] 遥感图像公开检测数据

集上分别进行实验验证。本文及相关对比实验均在 Ubuntu 系统上进行,采用的 GPU 型号为 GeForce RTX 2080 Ti,环境配置为 CUDA10.0+python3.6+pytorch1.7。

3.2 评价指标

遥感目标检测领域普遍采用平均精度(Average Precision, AP)、平均精度均值(mean Average Precision, mAP)作为检测结果的主要评价指标,采用算法的浮点运算次数(Floating Point Operations, FLOPs)及模型每秒处理的图片数(Frame Per Second, FPS)作为衡量算法计算复杂性和检测速度的主要指标,因此本文也采用上述指标衡量所提网络 MFC 对遥感多尺度目标的检测能力。单个目标的 AP 和整体目标的 mAP 值越高,表示检测器的性能越好,其公式如下:

$$AP = \int_0^1 P(R) dR, \quad (10)$$

$$mAP = \frac{1}{c} \sum_i^c AP_i, \quad (11)$$

其中: c 为目标类别数; P 为精确率, R 为召回率,公式分别如下:

$$P = \frac{T_p}{T_p + F_p}, \quad (12)$$

$$R = \frac{T_p}{T_p + F_N}, \quad (13)$$

其中: T_p 表示检测正确的正样本数, F_p 表示检测错误的正样本数, F_N 表示检测错误的负样本数

算法的浮点运算次数 FLOPs 可理解为计算量,用于衡量算法的计算复杂性,该值越小表示算法复杂性越低。模型每秒处理的图片数 FPS 又称每秒帧率,可以理解为模型检测速度,该值越大则表示模型的检测速度越快,越能实现实时检测,其公式如下:

$$FLOPs = 2HW(C_{in}K^2 + 1)C_{out}, \quad (14)$$

$$FPS = \frac{N}{S}, \quad (15)$$

其中: H 和 W 表示输出特征图尺寸, C_{in} 和 C_{out} 分别为输入和输出的通道数, K^2 表示核尺寸; N 为被测样本数, S 为测试全部样本所需的时间。

3.3 DIOR 数据集实验与分析

DIOR 数据集(部分示例如图 7 所示)是 2019 年提出的一个比较新颖、内容复杂多样且极具挑战性的大型公开遥感数据集,包含 23 463 幅遥感图像和 192 472 个实例样本,图像空间分辨率在

0.5 m 到 30 m, 尺寸为 800×800 。

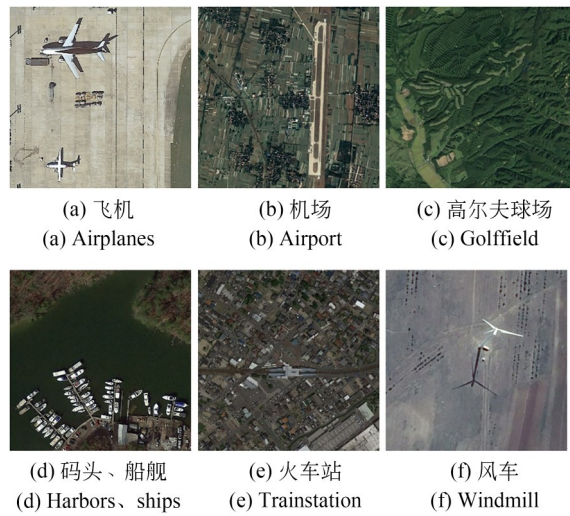


图 7 DIOR 数据集部分示例图

Fig. 7 Some examples images from DIOR dataset

该数据集覆盖的 20 类地物目标依次记为 D1-D20, 分别为飞机、机场、棒球场、篮球场、桥梁、烟囱、水坝、高速公路服务区、高速公路收费站、高尔夫球场、田径场、码头、立交桥、船舶、体育馆、储油罐、网球场、火车站、车辆、风车。其中, 桥梁、水坝、码头、立交桥、船舶及部分大型车辆、火车站等目标的外形均存在巨大纵横比特性, 共计约 24 226 个训练样本, 约占总体训练样本数的 35.6%, 对网络有着较高的考验。

从图 7 中可以看出, 该数据集在目标类内、类间, 不论是拍摄高度引起的相对尺度差异还是物体的固有尺度差异都极大, 同时目标所处特定环境的地域跨度性、复杂性也较为突出, 存在植被、

裸地、阴影、光线等大量复杂背景及成像质量等外界因素的干扰, 具有一定代表性, 目前遥感领域在该数据集上的研究工作仍未有较大突破。

在进行对比实验时, 选取 CenterNet(DLA-60) 作为 Baseline, 另外与部分基于锚框的两阶段网络 Faster RCNN^[1], Mask R-CNN^[2], PANet^[31], 一阶段网络 YOLOx^[3], RetinaNet^[4], 和无锚框网络 CornerNet^[14], 遥感网络 CSFF^[7], CBD-E^[32] 在该数据集上进行比较及定量、定性分析。

定量分析: 表 1 中可看到本文方法 mAP 相较于对比算法表现均更优, 对绝大多数目标的 AP 如飞机、码头、立交桥、船舶、车辆、风车等 14 类目标的检测结果也更占优势, 其中有 12 类目标 AP 达到了 70% 以上, 充分表现了本文方法对多尺度目标的有效性。本文在机场、水坝、体育馆、田径场、高速公路服务区和高速公路收费站目标的检测上略有不足, 稍逊于文献[7]和[14], 一方面是此类目标训练样本数不如其他目标充足, 另一方面则是部分目标存在一定的类间相似性干扰, 如水坝和桥梁、高速公路服务区和高速公路收费站等在外形和背景上均较为相似, 而文献[7]和[14]重点关注特征信息自身的判别性表达精度略优, 本文方法则重点关注目标间的尺度差异而存有微小差距, 但相较 Baseline 均有不同程度的精度提升表现较为突出。在计算复杂性和检测速度上, 可看到本文检测速度远优于大部分对比算法, 虽因引入了 MFE 模块以及 COS 策略不可避免地增加了一定的计算复杂性, 导致在速度上比 Baseline 降低了 2.4 FPS, 与追求速度的文献[3]相差 5.1 FPS, 但差距并不明显, 综合而言, 本文在检测精度上保持了绝对优势, 有良好的速度、精度折衷。

表 1 DIOR 数据集上不同算法性能对比

Tab. 1 Performance comparison of different algorithms on DIOR dataset

	Faster RCNN ^[1]	Mask R-CNN ^[2]	YOLOx ^[3]	RetinaNet ^[4]	PANet ^[31]	CSFF ^[7]	CBD-E ^[32]	CornerNet ^[14]	Baseline	本文方法 MFC
D1	54.0	53.9	72.5	53.3	60.2	57.2	54.2	58.8	78.1	87.4*
D2	74.5	76.6	76.3	77.0	72.0	79.6	77.0	84.2*	71.2	83.6
D3	63.3	63.2	74.1	69.3	70.6	70.1	71.5	72.0	72.6	74.3*
D4	80.7	80.9	77.4	85.0	80.5	87.4	87.1	80.8	87.1	87.8*
D5	44.8	40.2	40.8	44.1	43.6	46.1	44.6	46.4	38.8	47.2*
D6	72.5	72.5	70.5	73.2	72.3	76.6	75.4	75.3	75.3	77.3*

续表 1 DIOR 数据集上不同算法性能对比

Tab. 1 Performance comparison of different algorithms on DIOR dataset

	Faster RCNN ^[1]	Mask R-CNN ^[2]	YOLOx ^[3]	Retina- Net ^[4]	PANet ^[31]	CSFF ^[7]	CBD- E ^[32]	Corner- Net ^[14]	Baseline	本文方法 MFC
D7	60.0	60.4	53.1	62.4	61.4	62.7	63.5	64.3*	51.4	54.5
D8	75.6	76.3	55.6	78.6	72.1	82.6*	76.2	81.6	62.3	78.1
D9	62.3	62.5	57.2	62.8	66.7	73.2	65.3	76.3*	60.0	67.8
D10	76.0	76.0	77.5	78.6	72.0	78.2	79.3	79.5	75.3	80.5*
D11	76.8	75.9	73.3	76.6	73.4	81.6*	79.5	79.5	74.0	76.6
D12	46.4	46.5	48.2	49.9	45.3	50.7	47.5	26.1	51.3	54.3*
D13	57.2	57.4	53.7	59.6	56.9	59.5	59.3	60.6	55.8	61.5*
D14	71.8	71.8	83.9	71.1	71.7	73.3	69.1	37.6	84.7	87.4*
D15	68.3	68.3	62.2	68.4	70.4	63.4	69.7	70.7*	56.2	65.3
D16	53.8	53.7	61.4	45.8	62.0	58.5	64.3	45.2	68.9	72.6*
D17	81.1	81.0	86.3	81.3	80.9	85.9	84.5	84.0	86.3	88.2*
D18	59.5	62.3	61.4	55.2	57.0	61.9	59.4	57.1	53.4	62.4*
D19	43.1	43.0	47.3	44.4	47.2	42.9	44.7	43.0	45.1	47.9*
D20	81.2	81.0	84.5	85.5	84.5	86.9	83.1	75.9	75.7	87.2*
mAP/% ↑	65.1	65.2	65.9	66.1	66.1	68.0	67.8	64.9	66.2	70.9*
FPS ↑	11.2	12.9	33.2	16.4	14.5	—	—	26.7	30.5	28.1
FLOPs ↓	38.7 G	38.4 G	11.4 G	25.2 G	26.9 G	—	—	12.8 G	12.5 G	12.7 G

注:表中粗体加*表示最佳检测精度。

定性分析:本文与 Baseline 在 DIOR 数据集的检测结果对比如图 8 所示。从整体上来说,本文在不同检测环境及外界条件下,对多尺度目标的检测精度均有着大幅提高,对大尺度目标如棒

球场、网球场及田径场等运动场地的检测效果明显提升,对小尺度目标如储油罐、飞机及车辆等检测效果也更为突出。具体地,本文在前两组图存在植被遮挡及薄雾等外界干扰下,对其中密集

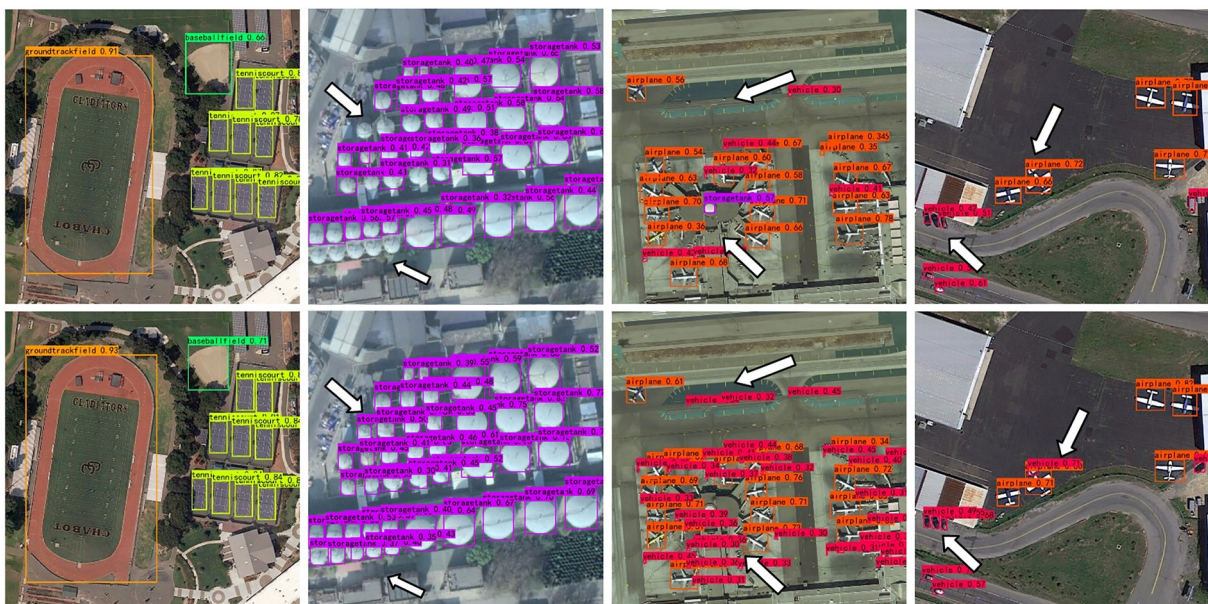


图 8 本文方法与 Baseline 在 DIOR 数据集的检测结果(图中上排为 Baseline,下排为本文方法)

Fig. 8 Experimental results of our proposed method (shown on the down) and Baseline (shown on the up) on DIOR dataset

排列、尺度不一的多类目标,在保证一定精度提升的同时也明显改善了大量漏检现象;在后两组图中,对呈现任意、交错排列的飞机、车辆目标,修正了误检现象并显著提高了检测框的数量及质量。特别在箭头所指处,可直观看到在遥感复杂的检测场景下,本文方法的检测效果十分显著,检测到的目标数明显增加,极大地减少了漏检、误检问题。实验表明,本文方法可有效应对尺度多变、密集排列、背景干扰等复杂因素对遥感多尺度目标造成的影响,得到更为精准的检测输出。

3.4 HRRSD 数据集实验与分析

HRRSD 数据集(部分示例如图 9 所示)是 2019 年中国科学院大学发布的大尺度航空遥感数据集,共有 21 758 幅遥感图像,包含约 55 740 个实例样本,图像空间分辨率在 0.15 m 到 1.2 m 之间,尺寸在 800×800 至 $4\,000 \times 4\,000$ 之间。

该数据集覆盖的 13 类地物目标依次记为 H1-H13,分别为飞机、船舰、储油罐、棒球场、网球场、篮球场、田径场、港口、桥梁、汽车、十字路口、T 型路口、停车场。其中,船舰、港口、桥梁等目标往往呈现较大的纵横比差异,共计约 7 957 个训练样本,约占总体训练样本数的 29.4%,易干扰网络对目标全貌的获取而导致误检等问题。从图 9 示例图中可看到该数据集目标尺度变化较为显著,目标所处地形及地貌间区别性较强,同

时因天气、图像来源及成像方式、传感器设备等不同,导致数据集成像质量参差不齐,检测难度大。

在 HRRSD 数据集上,Baseline 保持不变,再对比了基于锚框的两阶段算法 Faster RCNN, PANet、此数据集原作者所提方法 GACL Faster RCNN^[29],和一阶段算法 YOLOx, RetinaNet 及无锚框检测算法 CornerNet。

定量分析:从表 2 可以看出本文方法 mAP 在该数据集上明显优于其他对比算法。不论是在大尺度目标如田径场、桥梁等,还是小尺度目标如飞机、船舰等,对绝大多数遥感目标都有着十分优越的检测结果,其中 10 类目标 AP 已达到了 91% 以上,充分说明了本文算法的有效性;其中,对停车场目标的检测结果略低于文献[4],主要是该目标易与其他背景混淆,如任意停放车辆的空地、道路或建筑物场馆周边等场地,本文在针对目标尺度差异进行检测时,相较文献[4]重点对正负样本失衡问题进行研究,没有过多关注相似背景的影响而导致该目标检测结果较低,但精度相较 Baseline 提升了 4.4%,也明显优于其他对比算法结果。在计算复杂性和检测速度上,本文方法虽与 Baseline 和以速度著称的文献[3]均存在一定差距,但检测速度和其他对比方法相比表现更佳,针对多尺度问题引入了 MFE 模块和 COS 策略大大提升了检测效果,不可避免的在效率上产生相应的速度损耗,但整体来说仍能保持良好的速度和精度折衷。实验结果表明,本文方法能有效缓解薄雾、阴影及成像质量参差不齐等外界环境干扰所带来的影响,能够极大地改善对大目标的误检及对小目标的漏检等问题,对遥感图像多尺度目标检测任务具有着一定的适应性,在遥感复杂场景下依然可以得到较为优异的检测效果。

定性分析:本文与 Baseline 在 HRRSD 数据集的部分检测结果对比如图 10 所示。从整体上看,本文方法对田径场、受大量停泊船舰影响的港口、尺度不一且密集排列的储油罐等难度各异的多尺度目标检测都有着明显的改善,能够很好地缓解不同背景环境及成像差异等干扰因素影响,输出更精确的检测包围框,大幅提升检测精度。具体来说,从图 10(a),10(e)和 10(f)中不难

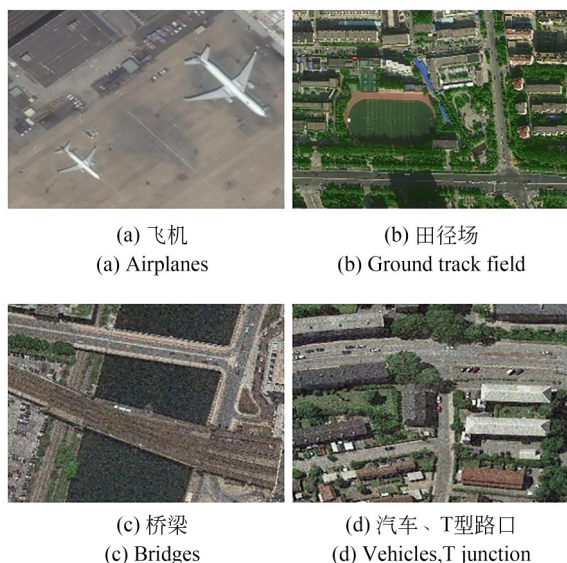


图 9 HRRSD 数据集示例图

Fig. 9 Example images from HRRSD dataset

表 2 HRRSD 数据集上不同算法性能对比

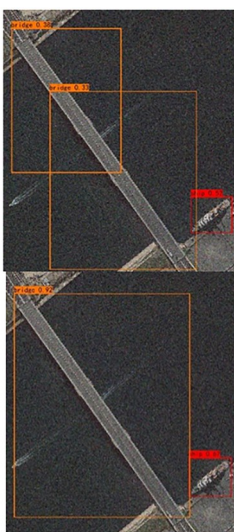
Tab. 2 Performance comparison of different algorithms on HRRSD dataset

	Faster RCNN ^[1]	GACL FasterRCNN ^[29]	YOLOx ^[3]	RetinaNet ^[4]	PANet ^[31]	CornerNet ^[14]	Baseline	本文方法 MFC
H1	90.5	90.8	91.3	92.6	93.7	88.5	96.6	98.4*
H2	86.1	88.5	85.6	86.7	87.2	89.2	90.7	92.2*
H3	87.3	89.2	86.7	89.1	90.5	93.3	94.1	95.0*
H4	86.7	87.2	86.3	87.5	89.3	89.5	88.4	91.1*
H5	80.7	80.8	85.3	84.9	87.7	90.6	91.8	93.8*
H6	48.2	49.7	71.7	51.3	58.2	72.6	71.4	75.7*
H7	89.9	90.7	87.4	89.6	92.3	94.0	94.2	96.9*
H8	90.1	89.7	90.2	89.9	89.7	89.9	91.5	94.8*
H9	86.2	85.6	85.1	86.2	88.4	88.5	88.7	92.3*
H10	86.5	86.9	85.4	87.7	88.9	89.2	90.3	93.8*
H11	89.4	88.2	87.5	87.4	89.4	90.8	91.9	94.9*
H12	74.7	75.0	74.1	75.3	79.3	79.4	80.0	85.8*
H13	66.2	65.3	65.3	68.3*	66.5	66.8	63.2	67.6
mAP/% ↑	81.7	82.1	83.2	82.8	84.7	86.3	87.1	90.2*
FPS ↑	14.5	—	48.9	23.2	20.9	42.2	45.5	43.7
FLOPs ↓	38.7 G	—	11.4 G	25.2 G	26.9 G	12.8 G	12.5 G	12.7 G

注:表中粗体加*表示最佳检测效果。

看出对尺度差异大的桥梁、船舰以及港口目标的检测结果更加准确;而图 10(b)可直观看到标签优化在纵横比较大目标上的作用,在输出检测框方面表现了对此类目标的改善情况,对提升检测效果也有一定助益;从图 10(c),10(d),10(f),10(g),10(h)中箭头所指处可以看到,除了对排列

较为密集且存在停泊船舰影响的大尺度的港口目标有效提高了检测效果,也增加了对一些排列更为密集的小尺度目标如储油罐以及占比相对较小的飞机、网球场的检测,对存在较多漏检、误检及定位不准确而导致检测框重叠等问题有着较为突出的优化效果。实验结果说明本文方法



(a) 尺度差异大
(a) Large different-scales



(b) 纵横比较大
(b) Large aspect ratio



(c) 尺度不一、密集
(c) Different scales, dense



(d) 多尺度目标
(d) Multi-scales objects

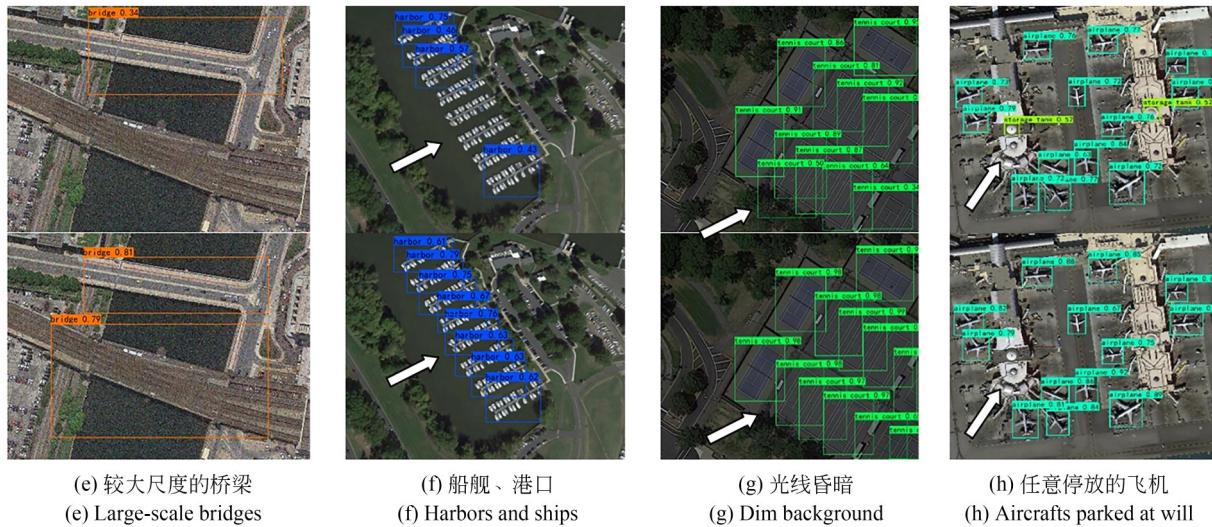


图 10 本文方法与 Baseline 在 HRRSD 数据集的检测结果(每组中上图为 Baseline,下图为本文方法)
 Fig. 10 Experiential results of our proposed method (shown on the down) and Baseline (shown on the up) on HRRSD dataset

对各类多尺度目标、大纵横比目标以及存在复杂外界干扰等情况的针对性更强,检测效果也更好。

3.5 RSOD 数据集实验与分析

RSOD 数据集是 2015 年由武汉大学发布用于遥感检测的标准数据集,共有 976 张遥感图像,包含 6 950 个实例样本,其中油罐 165 张,共 1 586 个实例,立交桥 176 张,共 180 个实例,操场 189 张,共 191 个实例,飞机 446 张,共 4 993 个实例,图像空间分辨率在 0.3 m 到 3 m 之间。

该数据集共包含四类目标,飞机、操场、立交桥和油桶依次表示为 R1-R4,示例图如图 11 所示。其中,立交桥、操场及部分大型飞机等目标的纵横比差异较为突出,共计约 865 个训练样本,约占总体训练样本数的 14.7%。特别的,立交桥目标由于延伸范围较广检测难度较大,对输出更精确的检测包围框有一定的考验。该数据集是从 Google Earth 和 Tianditu 下载整理得到,涉及拍摄高度、数据来源和传感器图像处理等不同因素,其目标尺度跨度性、地域复杂性和多样性也为遥感目标检测领域带来了较全面的性能挑战。

在 RSOD 数据集上,仍选取 CenterNet (DLA-60)作为 Baseline,同时还与基于锚框的二阶段算法 Faster RCNN,Mask R-CNN,及一阶段算法 YOLOx,RetinaNet,无锚框算法 CornerNet 和遥感算法 CBD-E 在该数据集上进行了算法对比。

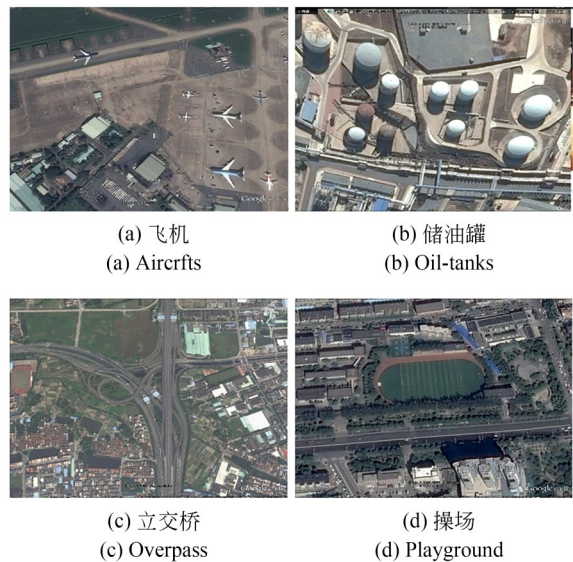


图 11 RSOD 数据集示例图
 Fig. 11 Example images from RSOD dataset

定量分析:从表 3 可以看出本文所提方法的 mAP 在该数据集上的精度表现达到了最优,在操场、立交桥和油桶目标的检测上均有着突出表现,对四类遥感目标的 AP 均达到了 94% 以上,说明了算法对多尺度目标的有效性;对飞机目标的检测结果略逊于文献[32],是由于同一类飞机目标由于自身用途不一导致在外形、外观上体现出的差异较大,本文相较文献[32]对所获取目标特征的增强不足,导致共性特征的适应性较弱,从而影响检测效果,但相对于其他对比方法依然有着领先优势。在计算复杂性和

检测速度上,本文方法依然优于大多数算法,整体性能较佳,且在该数据集上与速度占优的文献[3]间差距明显拉进,较 Baseline 降低了 2.3 FPS,是由于增加了模型计算复杂性而导致速度上的小幅度下降。实验结果充分验证了本文在

保证一定检测效率的同时,所提检测方法对尺度不一、外形不同的各类遥感多尺度目标均能有效处理,同时在地形复杂,植被遮挡、薄雾或昏暗等较恶劣检测条件下也能进行良好的目标定位及分类。

表 3 RSOD 数据集上不同算法性能对比

Tab. 3 Performance comparison of different algorithms on RSOD dataset

	Faster RCNN ^[1]	Mask R-CNN ^[2]	YOLOx ^[3]	RetinaNet ^[4]	CBD-E ^[32]	CornerNet ^[14]	Baseline	本文方 法 MFC
R1	79.6	81.2	87.5	82.2	95.8*	92.1	91.4	94.4
R2	94.7	93.4	95.3	99.7	99.7	96.8	96.3	99.7*
R3	85.0	85.9	83.6	93.8	88.8	90.0	90.1	94.1*
R4	93.8	93.5	91.2	97.1	94.2	93.5	94.9	99.5*
mAP/% ↑	88.3	88.5	89.4	93.2	94.2	93.1	93.2	96.9*
FPS ↑	15.9	17.5	61.3	27.6	—	56.3	61.1	58.8
FLOPs ↓	38.7	38.4G	11.4G	25.2G	—	12.8G	12.5G	12.7G

注:表中粗体加*表示最佳检测效果

定性分析:本文与 Baseline 在 RSOD 数据集的检测结果对比如图 12 所示,从图 12(a),图 12(b)中可看到本文方法改善了对遥感场景中常见的小目标飞机及油桶的检测效果,可适应其密集排列的方式,不仅提高了整体检测精度,还同时修正了一些误检问题;从图 12(e),图 12(f)可进一步看到常见的天气、薄雾、昏暗等外界环境在遥感检测任务带来的影响,在不同的程度的环境干扰下,所提方法对这类小尺度或存在形变的目标依然有着稳定的检测效果;同时,由图 12(c)、图 12(g)中可看到在昏暗光线及复杂地形下,单

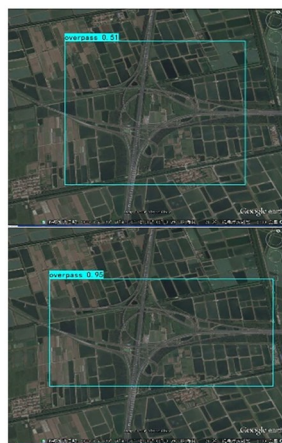
个及多个立交桥目标较难分辨,其所处的环境从耕地到居民区等都十分复杂检测难度较高,但从本文结果输出的检测框不难看出所提方法的有效性,同时标签优化后的检测框对此类纵横比较大目标的适应性较好,使得检测框可以更加贴合目标物体的尺度大小,这对提升精度无疑是有一定助益的;对图 12(d),图 12(h)中地形复杂和存在阴影遮挡等情况,所提方法也能有效克服干扰,很大程度上改善了漏检问题。各方面结果均表明,所提方法对目标类内及类间尺寸差异大等特性有着较强的适应性,还能有效避免各类外界



(a) 小尺度目标
(a) Small-scale objects



(b) 尺度不一
(b) Multi-scale objects



(c) 背景昏暗
(c) A dim background



(d) 存在植被遮挡
(d) Shaded by vegetation

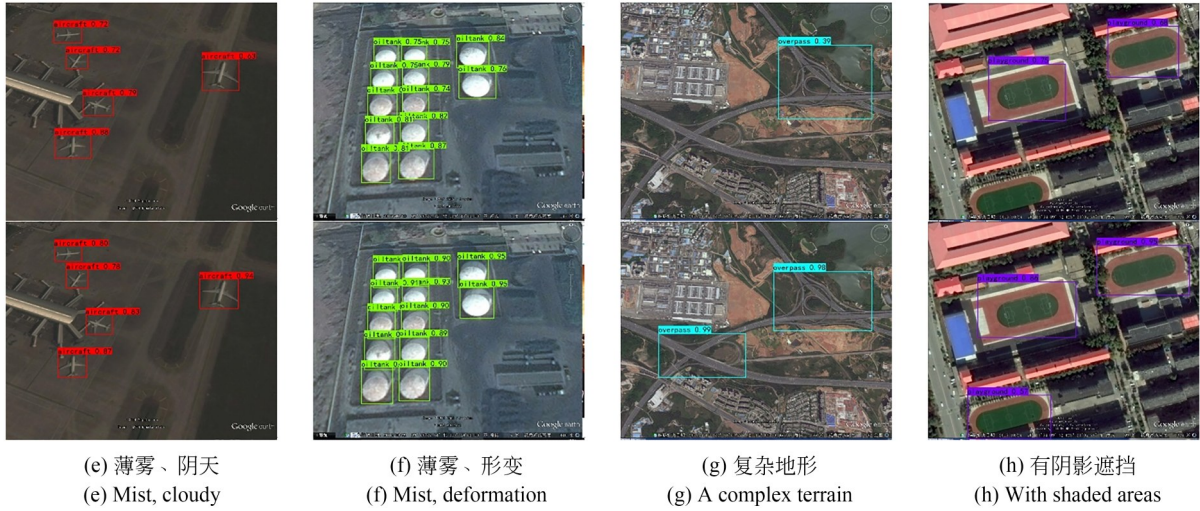


图 12 本文方法与 Baseline 在 RSOD 数据集的实验结果(每组图中上图为 Baseline, 下图为本文方法)
 Fig. 12 Experimental results of our proposed method(shown on the down) and Baseline(shown on the up) on RSOD dataset

环境干扰,重点关注多尺度目标位置信息等表征并使其充分表达,减少目标漏检、误检现象,整体检测精度显著提高。

3.6 消融实验与分析

本文消融实验主要围绕所提多元特征提取模块 MFE 及表征优化策略 COS 两部分进行。

首先,在 MFE 模块中主要关注跨组连接、上下文模块 CM 及通道注意力 SE,在 RSOD 数据集上,基于 Baseline 依次加入以进行消融实验设置,如表 4 所示。可以看出,在模块内部进行跨组连接操作带来了一定的计算复杂性,影响检测效率,但提升 1.7% 的精度表现也十分突出;注意力 SE 增加了所需通道层面网络的非线性,有助于特征图的高质量输出,也有着相应的精度提升;加入 CM 模块精度增长了 0.5%,说明对小尺度目标的有效性,结果表明所提方法没有以牺牲过多检测效率来换取精度,整体检测效果十分显著。

其次,对于表征优化策略 COS 而言,主要涉

及标签优化操作和坐标像素注意力模块 CPA,在 RSOD 数据集上的消融实验设置为,在 Baseline 基础上依次加入以进行验证,如表 5 所示,可看出标签优化对大纵横比目标的针对性改善,整体精度提升了 0.3%,计算复杂性小幅增加;加入 CPA 进一步关注遥感多尺度目标的有效信息,精度增长 0.6%;整体效果较为明显,计算复杂性的增长仅轻微影响了检测速度,说明了 COS 的有效性。

此外,为说明 MFE 中通道注意力 SE^[24]及 COS 中坐标注意力 CA^[26]对本文方法的适应性,在 RSOD 数据集上分别与卷积注意力机制 CBAM^[35]以及文献[27]中的通道注意力 CA 进行比较,如表 6 及表 7 所示。可以看到,在 MFE 中 SE 和 CBAM 各类表现基本持平,但本文采用的 SE 可关注所需通道信息的同时计算复杂性相对较小;在 COS 中,从计算复杂性和检测速度上本文略逊于文献[27],但其精度提升甚微且仅关注通道信息,而本文采用的坐标注意力 CA 可同时重点捕获遥感目标所需关键位置信息及通道信息,适应性更强。

最后,从整上为验证多元特征提取模块 MFE 及表征优化策略 COS 两部分内容的有效性,分别在 DIOR, HRRSD 及 RSOD 数据集上进行了测试,结果如表 9 所示,在 Baseline 基础上先后融入 MFE 模块及 COS 策略进行消融实验设置。所提方法在 DIOR 数据集上的精度依次提

表 4 在 RSOD 数据集上的消融实验(MFE)

Tab. 4 Ablation experiments on RSOD datasets(MFE)

Method	mAP/% ↑	FPS ↑	FLOPs ↓
Baseline	93.2	61.5	12.48 G
+跨组连接	94.9	59.9	12.64 G
+跨组连接+SE	95.2	59.7	12.69 G
+跨组连接+SE+CM	95.7	59.4	12.71 G

表 5 在 RSOD 数据集上的消融实验 (COS)

Tab. 5 Ablation experiments on RSOD datasets (COS)

Method	mAP/% ↑	FPS ↑	FLOPs ↓
Baseline	93.2	61.5	12.48 G
+标签优化	93.5	61.4	12.49 G
+标签优化+CPA	94.1	60.1	12.52 G

升了 3.1% 和 1.6%，检测速度降低了 2.4 FPS；在 HRRSD 数据集上的精度依次提升了 2.3% 和 0.8%，检测速度降低了 1.8 FPS；在 RSOD 数据集上的精度依次提升了 2.5% 和 1.2%，检测速度降低了 2.7 FPS。消融实验结果充分表明，本文所提多元特征提取模块 MFE 和表征优化策略 COS 可在保证一定检测效率且不增加过多计算复杂性的前提下，对算法的整体检测精度提升

表 6 在 RSOD 数据集上对比实验 (SE、CA)

Tab. 6 Contrast experiments on RSOD datasets (SE、CA)

Method	mAP/% ↑	FPS ↑	FLOPs ↓
Baseline	93.2	61.5	12.48 G
+MFE(with SE)	95.7	59.4	12.71 G
+MFE(with CBAM)	95.8	59.4	12.72 G
+COS(with CA ^[26])	94.1	60.1	12.52 G
+COS(with CA ^[27])	93.9	60.2	12.50 G

较为显著，在不同数据集上的检测效果均十分优越，对遥感多尺度目标以及大纵横比目标等针对性更强，说明了本文方法在遥感领域检测难点不尽相同的复杂任务中均具有一定的适应性和鲁棒性，整体性能良好。

表 7 在 DIOR、HRRSD 及 RSOD 数据集上的消融实验

Tab. 7 Ablation experiments on DIOR、HRRSD and RSOD datasets

消融实验	DIOR				HRRSD				RSOD				FLOPs
	MFE	COS	mAP/%	FPS	MFE	COS	mAP/%	FPS	MFE	COS	mAP/%	FPS	
Baseline	—	—	66.2	30.5	—	—	87.1	45.5	—	—	93.2	61.5	12.5 G
✓	✓	—	69.3	29.2	✓	—	89.4	43.1	✓	—	95.7	59.4	12.7 G
✓	✓	✓	70.9	28.1	✓	✓	90.2	43.7	✓	✓	96.9	58.8	12.7 G

4 结 论

针对复杂背景下遥感多尺度目标存在细粒度级别多尺度特征提取困难，预测部分有效表征不足的问题，本文基于无锚框思想，提出一种多元特征提取与表征优化的遥感多尺度目标检测方法 MFC。在特征提取部分，通过设计联合上下文信息的多元特征提取模块 MFE 扩大感受野使网络最大程度适应不同尺度的遥感目标，并重点关注小目标周围背景信息进行辅助检测；利用深度聚合结构加强特征的充分融合，提高特征图质量。在预测部分，提出表征优化策略 COS，通

过标签优化和坐标像素注意力组合关注纵横比较大目标及通道、位置及像素等有效信息并使其突出表征，降低复杂背景干扰，提高检测准确性。在地域跨度广泛、数据来源及质量参差不齐和天气条件恶劣等复杂遥感场景下的实验也表明，本文方法对遥感多尺度目标具有较为突出的检测效果，适应性和鲁棒性也较强，在 DIOR、HRRSD 及 RSOD 数据集上 mAP 分别达到了 70.9%，90.2% 和 96.9%，说明了所提模型的有效性。后续将考虑对遥感鸟瞰角度下以任意方向且密集排列的遥感旋转目标作为进一步检测精度优化的方向。

参考文献：

[1] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* Feature Pyramid Networks for Object Detection [C]. 2017 IEEE Conference on Computer Vision and Pat-

tern Recognition (CVPR). 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 936-944.

[2] HE K M, GKIOXARI G, DOLLÁR P, *et al.* Mask R-CNN [C]. 2017 IEEE International Conference on Computer Vision (ICCV). 22-29, 2017,

- Venice, Italy. IEEE, 2017: 2980-2988.
- [3] GE Z, LIU S, WANG F, *et al.* YOLOX: Exceeding YOLO Series in 2021[EB/OL]. 2021: *arXiv*: 2107.08430. <https://arxiv.org/abs/2107.08430>
- [4] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal Loss for Dense Object Detection[C]. 2017 IEEE International Conference on Computer Vision (ICCV). 22-29, 2017, Venice, Italy. IEEE, 2017: 2999-3007.
- [5] HOU J Y, MA H B, WANG S J. Parallel Cascade R-CNN for object detection in remote sensing imagery [J]. *Journal of Physics: Conference Series*, 2020, 1544(1): 012124.
- [6] 沙苗苗, 李宇, 李安. 改进 Faster R-CNN 的遥感图像多尺度飞机目标检测[J]. *遥感学报*, 2022, 26(8): 1624-1635.
- SHA M M, LI Y, LI A. Multiscale aircraft detection in optical remote sensing imagery based on advanced Faster R-CNN[J]. *Journal of Remote Sensing*, 2022, 26(8): 1624-1635. (in Chinese)
- [7] CHENG G, SI Y J, HONG H L, *et al.* Cross-scale feature fusion for object detection in optical remote sensing images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(3): 431-435.
- [8] 程德强, 赵佳敏, 寇旗旗, 等. 多尺度密集特征融合的图像超分辨率重建[J]. *光学精密工程*, 2022, 30(20): 2489-2500.
- CHENG D Q, ZHAO J M, KOU Q Q, *et al.* Multi-scale dense feature fusion network for image super-resolution [J]. *Opt. Precision Eng.*, 2022, 30(20): 2489-2500. (in Chinese)
- [9] 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): 130-140.
- 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): 130-140. (in Chinese)
- [10] XU D Q, WU Y Q. FE-YOLO: a feature enhancement network for remote sensing target detection[J]. *Remote Sensing*, 2021, 13(7): 1311.
- [11] 陈欣, 万敏杰, 马超, 等. 采用多尺度特征融合 SSD 的遥感图像小目标检测[J]. *光学精密工程*, 2021, 29(11): 2672-2682.
- CHEN X, WAN M J, MA C, *et al.* Recognition of small targets in remote sensing image using multi-scale feature fusion-based shot multi-box detector[J]. *Opt. Precision Eng.*, 2021, 29(11): 2672-2682. (in Chinese)
- [12] 朱威, 王立凯, 靳作宝, 等. 引入注意力机制的轻量级小目标检测网络[J]. *光学精密工程*, 2022, 30(8): 998-1010.
- ZHU W, WANG L K, JIN Z B, *et al.* Lightweight small object detection network with attention mechanism[J]. *Opt. Precision Eng.*, 2022, 30(8): 998-1010. (in Chinese)
- [13] 赵浩光, 王平, 董超, 等. 结合多尺度视觉显著性的舰船目标检测[J]. *光学精密工程*, 2020, 28(6): 1395-1403.
- ZHAO H G, WANG P, DONG C, *et al.* Ship detection based on the multi-scale visual saliency model[J]. *Opt. Precision Eng.*, 2020, 28(6): 1395-1403. (in Chinese)
- [14] LAW H, DENG J. CornerNet: detecting objects as paired keypoints [J]. *International Journal of Computer Vision*, 2020, 128(3): 642-656.
- [15] LAW H, TENG Y, RUSSAKOVSKY O, *et al.* CornerNet-lite: Efficient Keypoint Based Object Detection [EB/OL]. 2019: *arXiv*: 1904.08900. <https://arxiv.org/abs/1904.08900>
- [16] 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). 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 850-859.
- [17] DUAN K W, BAI S, XIE L X, *et al.* CenterNet: Keypoint Triplets for Object Detection[C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). October 27 - November 2, 2019, Seoul, Korea (South). IEEE, 2020: 6568-6577.
- [18] TIAN Z, SHEN C H, CHEN H, *et al.* FCOS: fully convolutional one-stage object detection[C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). October 27 - November 2, 2019, Seoul, Korea (South). IEEE, 2020: 9626-9635.
- [19] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as Points [EB/OL]. 2019: *arXiv*: 1904.07850. <https://arxiv.org/abs/1904.07850>
- [20] LIU Y J, YANG J L, CUI W Q. Simple, Fast, Accurate Object Detection Based on Anchor-Free Method for High Resolution Remote Sensing Images[C]. IGARSS 2020 - 2020 IEEE International

- Geoscience and Remote Sensing Symposium. September 26 - October 2, 2020, Waikoloa, HI, USA. IEEE, 2021: 2443-2446.
- [21] ZHANG T, WANG G Q, ZHUANG Y, *et al.* Feature Enhanced Centernet for Object Detection in Remote Sensing Images [C]. IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. September 26 - October 2, 2020, Waikoloa, HI, USA. IEEE, 2021: 1639-1642.
- [22] PAN X J, REN Y Q, SHENG K K, *et al.* Dynamic Refinement Network for Oriented and Densely Packed Object Detection [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 11204-11213.
- [23] YU F, WANG D Q, SHELHAMER E, *et al.* Deep Layer Aggregation [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 2403-2412.
- [24] GAO S H, CHENG M M, ZHAO K, *et al.* Res2Net: a new multi-scale backbone architecture [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(2): 652-662.
- [25] NAJIBI M, SAMANGUEI P, CHELLAPPAR, *et al.* SSH: Single Stage Headless Face Detector [C]. 2017 IEEE International Conference on Computer Vision (ICCV). 22-29, 2017, Venice, Italy. IEEE, 2017: 4885-4894.
- [26] HOU Q B, ZHOU D Q, FENG J S. Coordinate Attention for Efficient Mobile Network Design [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 13708-13717.
- [27] QIN X, WANG Z L, BAI Y C, *et al.* FFA-net: feature fusion attention network for single image dehazing [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 11908-11915.
- [28] LI K, WAN G, CHENG G, *et al.* Object detection in optical remote sensing images: a survey and a new benchmark [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 159: 296-307.
- [29] LU X Q, ZHANG Y L, YUAN Y, *et al.* Gated and axis-concentrated localization network for remote sensing object detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(1): 179-192.
- [30] LONG Y, GONG Y P, XIAO Z F, *et al.* Accurate object localization in remote sensing images based on convolutional neural networks [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(5): 2486-2498.
- [31] YANG J F, FU X Y, HU Y W, *et al.* PanNet: A Deep Network Architecture for Pan-Sharpening [C]. 2017 IEEE International Conference on Computer Vision (ICCV). 22-29, 2017, Venice, Italy. IEEE, 2017: 1753-1761.
- [32] ZHANG J, XIE C M, XU X, *et al.* A contextual bidirectional enhancement method for remote sensing image object detection [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 4518-4531.
- [33] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional Block Attention Module [M]. Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.

作者简介:



孟月波(1979—),女,陕西西安人,教授,西安建筑科技大学信息与控制工程学院博士生导师,2014年于西安交通大学获得工学博士学位,主要从事计算机视觉理解、建筑环境智能感知与调控、建筑智能化技术方面的研究。E-mail: mengyuebo@163.com

通讯作者:



刘光辉(1976—),男,陕西西安人,副教授,西安建筑科技大学信息与控制工程学院硕士生导师,2016年于西安建筑科技大学获得工学博士学位,主要从事计算机视觉理解、建筑环境智能感知与调控、建筑智能化技术方面的研究。E-mail: guanghuil@163.com