

基于 FFC-SSD 模型的光学遥感图像目标检测

薛俊达^{1,2}, 朱家佳^{1,2*}, 张静^{1,2*}, 李晓辉^{1***}, 窦帅¹, 米琳¹, 李子扬¹,
苑馨方¹, 李传荣¹

¹中国科学院空天信息创新研究院, 中国科学院定量遥感信息技术重点实验室, 北京 100094;

²中国科学院大学, 北京 100049

摘要 面向高效、高精度光学遥感图像目标检测应用, 重点针对提升 SSD(single shot multibox detector)模型对图像中聚集分布的小尺寸目标检测精度的难点, 提出一种 FFC-SSD(multi-scale feature fusion & clustering SSD)改进模型; 设计目标框分组聚类(BGC)模块, 采用分组聚类的方法获得更符合目标样本尺寸分布的默认目标框参数并给予小尺寸目标更多关注, 以有效提升网络对目标位置信息的提取能力; 设计反池化高效多尺度特征融合(MSFF)模块, 以在增强模型目标特征提取能力的同时有效减小模型效率损耗。实验结果显示了所提模型对光学遥感图像目标检测的有效性 with 适用性, 较好地实现了精度与效率的平衡, 对小尺寸目标具有较高的检测精度。

关键词 图像处理; 目标检测; 光学遥感图像; 多尺度特征融合; 聚类

中图分类号 TP753

文献标志码 A

DOI: 10.3788/AOS202242.1210002

Object Detection in Optical Remote Sensing Images Based on FFC-SSD Model

Xue Junda^{1,2}, Zhu Jiajia^{1,2*}, Zhang Jing^{1,2*}, Li Xiaohui^{1***}, Dou Shuai¹,
Mi Lin¹, Li Ziyang¹, Yuan Xinfang¹, Li Chuanrong¹

¹*Aerospace Information Research Institute, Key Laboratory of Quantitative Remote Sensing Information Technology, Chinese Academy of Sciences, Beijing 100094, China;*

²*University of Chinese Academy of Sciences, Beijing 100049, China*

Abstract For the applications of efficient high-precision object detection in optical remote sensing (RS) images, this paper focuses on the difficulty of improving the detection accuracy of the SSD (single shot multibox detector) model on small and densely distributed objects in such images. An improved model FFC-SSD (multi-scale feature fusion & clustering SSD) is thereby proposed. For this purpose, a bounding-box group clustering (BGC) module is designed. Group clustering is implemented to obtain default object frame parameters that are more consistent with the size distribution of object samples and gives more attention to small objects. This module effectively improves the network's ability to extract object locations. Then, an efficient de-pooling multi-scale feature fusion (MSFF) module is designed to enhance the ability of the model to extract object features and effectively reduce the efficiency loss of the model at the same time. The experimental results demonstrate the effectiveness and applicability of the FFC-SSD model for object detection in optical remote sensing images. The proposed model achieves a favorable balance between precision and efficiency and has high detection accuracy on small objects.

Key words image processing; object detection; optical remote sensing image; multi-scale feature fusion; clustering

收稿日期: 2021-09-24; 修回日期: 2021-11-03; 录用日期: 2021-11-25

基金项目: 国家重点研发计划(2018YFB050540)、中国科学院战略性先导科技专项(A类)(XDA17040303)

通信作者: *zhangjing@aoe.ac.cn; **jjzhu@aoe.ac.cn; ***xhli@aoe.ac.cn

1 引言

高精度、快速、自动化的遥感图像目标检测在军事、灾害应急、资源勘测、城市规划等领域具有广阔的应用前景,一直以来都是研究的热点。计算机处理能力的显著提升以及深度学习技术在自然图像目标检测应用中取得的巨大成功,为遥感图像目标检测提供了新的动力与技术途径。目前基于深度学习技术的目标检测主要分为双阶段方法和单阶段方法。其中,单阶段方法直接在特征图上同时输出目标的位置及类别信息,检测效率高但精度并不理想。2016 年提出的 SSD (single shot multibox detector)^[1]作为单阶段方法的代表之一,借鉴了双阶段方法 Faster R-CNN^[2]的锚框设计以及多尺度特征提取思想,在保持高检测效率的同时,检测精度也取得一定提升,对于大数据时代下具有较高时效性需求的应用显示出良好的适用性。

由于成像方式和拍摄角度等的不同,光学遥感图像相比自然图像具有图幅更大、背景更复杂、目标尺度分布宽、小尺寸目标数量大且聚集分布等检测难点,直接沿用自然图像目标的检测模型往往难以获得满意的检测效果。一些研究^[3-4]尝试对 SSD 进行适应性改进,将其应用于光学遥感图像中车辆目标检测和多类别目标的检测。SSD 对小尺寸目标的检测具有局限性,其检测效果并未有明显改善。为获得更好的特征融合效果,Lin 等^[5]结合单一特征映射、金字塔特征层次和综合特征的优点,提出了特征金字塔网络(FPN)。FPN 有效提升了模型对小尺寸目标检测的适应性,成为目前最流行的多尺度网络^[6-7],得到广泛应用^[8-11]。针对光学遥感图像中的小尺寸目标检测问题,一些结合 FPN 的 SSD 改进模型被相继提出,朱敏超等^[12]提出改进的 FD-SSD (Feature Fusion and Dilated Convolution SSD),实现对 DOTA 数据集 15 类目标的检测精度的提升;史文旭等^[13]提出的特征增强的 FESSD (Feature Enhancement SSD)对其自建数据集中以小目标为主的飞机、油罐、船的检测效果有明显改善;黄国新等^[14]采用 ResNet-50 结合 FPN 并调整 SSD 默认目标框尺度,实现了小尺寸目标的检测精度提升;Wang 等^[15]提出一种特征融合 SSD(FMSSD)模型,设计空洞空间特征金字塔、区域加权代价函数并优化 Loss 函数以给予小尺寸目标更多关注,FMSSD 对 DOTA 数据集 15 类目标的检测 mAP(mean Average Precision)为 72.43%。总的来说,FMSSD 适应光学

遥感图像多类别、宽尺度、小尺寸目标数量大且密集分布的特点,解决小尺寸目标检测难点,在保持 SSD 高检测效率的同时提升模型检测精度,是现阶段 SSD 模型改进优化的一个主要研究方向。

本文面向高效、高精度光学遥感图像的目标检测应用,重点针对提升 SSD 对聚集分布的小尺寸目标检测精度的难点,提出一种基于目标框分组聚类与高效多尺度特征融合的 FFC-SSD 模型;分析光学遥感图像中目标的分布特征,设计目标框分组聚类 BGC 模块,采用目标框分组聚类方法获得更契合目标样本尺寸分布的默认目标框参数,给予小尺寸目标更多关注,从而有效提升目标检测精度;设计高效多尺度特征融合反池化高效多尺度特征融合(MSFF)模块,采用反池化进行特征图上采样,在提升模型特征提取能力的同时有效减小模型效率损耗。

2 模型设计方法

SSD 在光学遥感图像目标检测应用中难以达到满意效果,问题可能在于:1)光学遥感图像中不同目标的尺寸、数量差异大,且小尺寸目标(如汽车、舰船等)呈现数量多、密集分布等特点,SSD 采用基于自然图像数据的经验公式设置默认目标框参数,不能很好地契合光学遥感图像中的目标尺寸与尺度分布,对数量众多、密集分布的小尺寸目标关注不够,导致模型无法回归得到更准确的目标位置信息;2)SSD 没有利用不同尺度特征层间的关联信息,特征提取不充分。现有改进模型虽然引入了 FPN,但没有很好地解决特征融合过程中计算效率降低的问题。本文重点针对以上问题进行模型优化设计,提出了一种基于分组聚类与高效多尺度特征融合的 SSD 改进模型——FFC-SSD,模型框架如图 1 所示。

如图 1 所示,FFC-SSD 在微调 SSD 网络基础上,引入目标框分组聚类 BGC 模块和高效多尺度特征融合 MSFF 模块,以实现目标特征提取能力的有效提升。

2.1 目标框分组聚类 BGC 模块设计

对 SSD 进行优化,使其默认目标框参数适应光学遥感图像目标分布特征,首先需选择能正确反映光学遥感图像目标特征表征和尺度分布特征的数据集。数据集应符合以下要求:1)取自高分辨率光学遥感图像;2)图像尺寸至少为 512 pixel×512 pixel;3)包含较大规模的目标类别和样本量;4)场景、目标样本尺寸和形状具有多样性;5)包含密集分布的小尺寸目标。

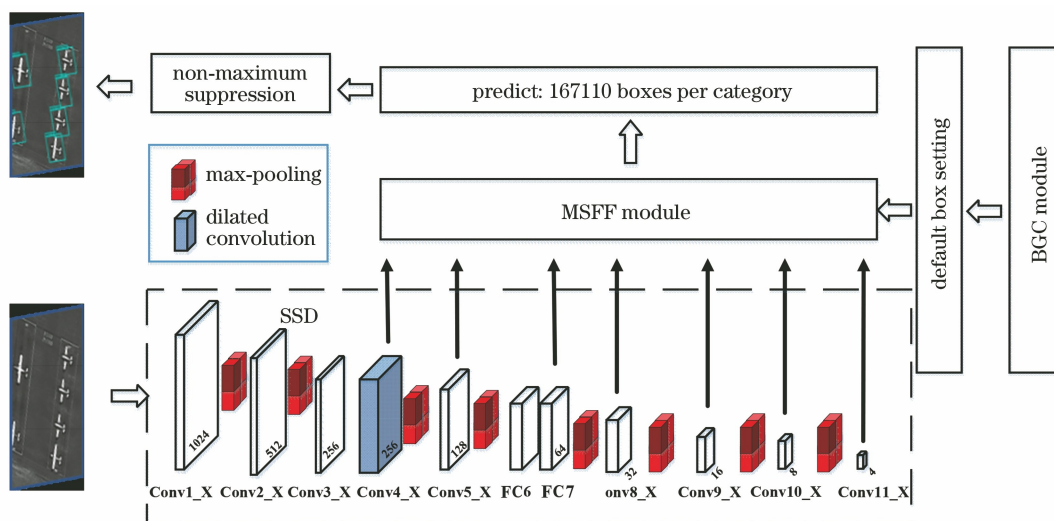


图 1 FFC-SSD 模型框架图

Fig. 1 Framework of FFC-SSD model

本文选择当前应用最广、目标类别最多、样本最丰富的 DOTA 数据集^[16]进行 SSD 默认目标框参数设置方法优化和 BGC 模块设计。DOTA 数据集由武汉大学发布,包含多种光学遥感器获取的 2806 幅高分辨率遥感影像(尺寸从 800 pixel×800 pixel 到 4000 pixel×4000 pixel),共 15 个目标类别、超过 340000 个样本。DOTA 中不同类别间的样本数量和样本尺寸存在巨大差异:样本量最大的小型车辆包含 126501 个样本,田径场、棒球场、篮球场等类别的样本量不足 1000;样本的目标框尺寸从不足 10 pixel²(小型车辆)到超过 10⁶ pixel²(田径场),尺度分布非常宽。DOTA 的小、中、大目标样本的占

比分别为 57%、41% 和 2%,包含大量小型车辆、大型车辆、船等密集分布的小目标,有 9 景影像的样本量大于 2000(最多 1 景有 10206 个样本);小型车辆的小目标样本占比超过 99%,46.47% 的样本尺寸小于 400 pixel²。图 2 显示了 DOTA 数据集各类目标的样本量分布及尺寸分布情况。图中及后文的小、中、大目标采用 DOTA 数据集的定义^[16]:目标框水平高度 $h < 50$ pixel 的目标为小目标; $h > 300$ pixel 的目标为大目标; $50 \text{ pixel} \leq h \leq 300 \text{ pixel}$ 的目标为中目标。图 2(b)中的大圆点为利用 SSD 经验公式计算得到的默认目标框尺寸,该默认目标框尺寸分布对小目标和样本量大的目标的关注不够。

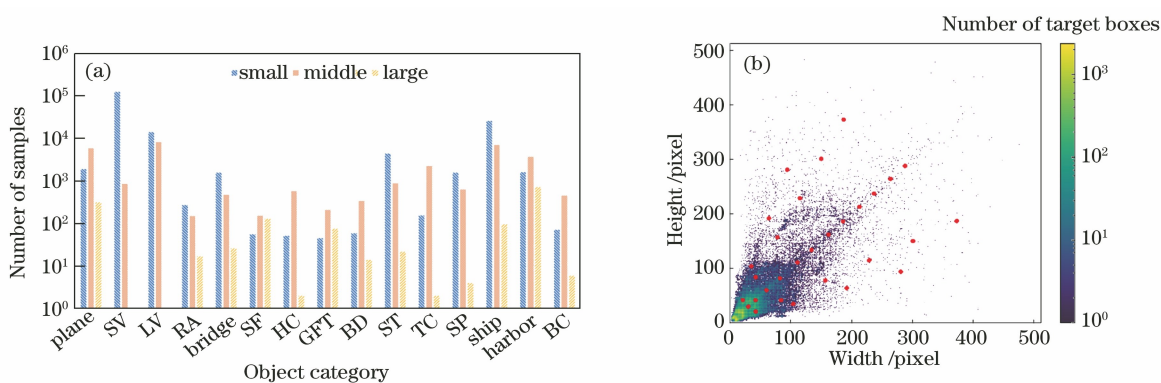


图 2 DOTA 数据集各类目标样本量分布及尺寸分布情况。(a)样本量分布;(b)尺寸分布

Fig. 2 Number of samples and size distribution of each category in DOTA dataset.

(a) Number of samples; (b) size distribution

本文使用平均覆盖度^[17,9]定量分析默认目标框参数与数据集目标样本尺寸的符合程度。平均覆盖度 A_{IoU} 定义为全体目标框与各聚类中心的最大交并比的均值^[17]:

$$A_{IoU} = \frac{\sum_{i=0}^N \max[R_{IoU1}(i), \dots, R_{IoUj}(i), \dots, R_{IoUN}(i)]}{N}, \quad (1)$$

式中: N 表示目标框个数; $R_{i \cap j}(i)$ 表示第 i 个目标框与第 j 个默认目标框的交并比。

文献[15]对 SSD 默认目标框的参数设置公式进行了优化,通过增加参数个数、扩大尺度分布范围,使默认目标框参数对 DOTA 数据集目标样本的平均覆盖度由 43.74% 提升到 50.22%, mAP 也提升 3%, 这表明优化默认目标框参数是提升精度的一个有效途径。但是,该方法设置的默认目标框的尺寸分布相对均匀,对遥感图像中数量巨大、尺寸更小(小于 20 pixel × 20 pixel)的目标关注有限,无法对此类目标位置信息进行准确提取。另一方面,单纯增加默认目标框数量会增加计算量、降低检测效率。例如,当输入图像的尺寸为 1024 pixel × 1024 pixel 时,在 Conv4_3 层特征图(128 pixel × 128 pixel)上每增加一个默认目标框,会增加 16384 次计算、约 491.52 MB 显存占用,当默认目标框数量大于 30 时,SSD 引入 FPN 或 MSFF 模块后甚至难以在常规配置的计算机(12G 显卡)上正常训练。因此,为实现检测精度提升并保持效率优势,应在有限数量的默认目标框参数下实现其与目标样本尺寸

分布的更好契合。综合考虑光学遥感应需求、计算效率和计算机硬件条件限制,本文将默认目标框的最大数量设定为 30。

为使有限数量下的默认目标框参数更好地契合光学遥感图像目标尺度分布,使其既适应宽尺度目标分布又重点关注数量巨大的小目标,本文根据目标样本尺寸分布聚类的方法设置默认目标框参数。但是,直接对所有目标样本进行聚类^[9] 容易使结果偏向样本量大的类别而难以顾及小样本量类别,这对目标样本量差异大、尺度分布宽的遥感图像并不适用。故本文提出目标框分组聚类方法:首先根据各类别目标样本量将目标框按类别分成若干组,使每组内的各类别样本量不存在显著差异;然后分别对每组内样本尺寸进行 k -means 聚类。

折衷考虑检测效率与精度的影响,按照分组数尽可能少且组内各类别间样本数量差异尽可能小的原则,对 DOTA 数据集中每类目标的样本量进行 k -means 聚类,得到 5 个聚类中心(平均覆盖度为 92.89%),由此将 15 个目标类别划分为 5 组,如表 1 所示。

表 1 DOTA 数据集中目标类别分组描述

Table 1 Grouping description of target categories in DOTA datasets

Group	Object category	Sample number per category
T0	Small vehicle (SV)	>100000
T1	Large vehicle (LV), ship	20000~40000
T2	Plane, storage tank (ST), harbor	5000~10000
T3	Bridge, tennis court (TC), swimming pool (SP)	2000~5000
T4	Roundabout (RA), soccer field(SF), ground field track (GFT), Baseball diamond (BD), basketball court (BC), helicopter (HC)	<2000

分别对 T0~T4 组内所有样本尺寸进行 k -means 聚类,如图 3 所示。图 3 中,随着聚类个数 k 的增加,各组样本的平均覆盖度随之增加,当 $k \geq 8$ 时平均覆盖度变化趋于平缓,故取 $k \geq 8$ 的聚类结果会比较理想。进一步进行聚类分析并折衷考虑默

认目标框参数数量、精度以及效率,T0~T4 组分别选择 k 为 8,9,8,8,9 的聚类结果进行整合并剔除冗余,最终得到 30 个默认目标框参数。

FFC-SSD 将 7 个不同尺度的特征层的特征输入 MSFF_U 模块进行特征融合,各融合特征图的默认目标框参数可表示为

$$\left\{ \begin{array}{l} K_{1-\min} = 0.0097, n = 1 \\ K_{2-\min} = 0.05, n = 2 \\ K_{n-\min} = S_{\min} + \frac{S_{\max} - S_{\min}}{3}(n - 3), n \in [3, 7] \\ K_{1-\max} = 0.05, n = 1 \\ K_{2-\max} = 0.1, n = 2 \\ K_{n-\max} = K_{n-\min} + \frac{S_{\max} - S_{\min}}{3}, n \in [3, 7] \end{array} \right. \quad (2)$$

式中: $K_{n-\min}$ 、 $K_{n-\max}$ 分别表示第 n ($n = 1, 2, \dots, 7$) 层

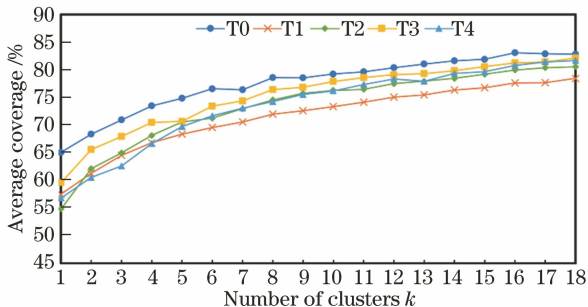


图 3 5 组样本平均覆盖度随聚类个数 k 的变化趋势

Fig. 3 Average coverage of five groups varying with number of clusters k

融合特征图默认目标框在原图中占比的最小、最大值; S_{min} 、 S_{max} 分别表示默认目标框在原图中占比的最小、最大值。

图 4 显示了分组聚类法获得的 30 个默认目标框参数的尺寸分布(图中的大圆点)。表 2 列出了这 30 个默认目标框参数在 7 个参与多尺度融合的特征层上的分布。对比图 4 与图 2(b)可以看到, 分组聚类法在目标尺寸小于 50 pixel×50 pixel 的区域设置了更多的默认目标框参数, 给予小目标更多的关注。表 2 中 Conv4_3 层包含了 8 个默认目标框参数, 尺寸均小于 40 pixel×40 pixel, 最小为 5 pixel×10 pixel, 这些小尺寸的默认目标框参数有利于网络获取数量众多的小目标和微小目标更准确的目标位置信息, 进而提取更准确的目标特征。

表 2 各融合特征图上设置的默认目标框尺寸

Table 2 Default target box size on each fusion feature map

n	Layer	Feature map size / (pixel×pixel)	Size of default box $w \times h$ / (pixel×pixel)
1	Conv4_3	256×256	5×10, 10×6, 12×21, 20×11, 14×13, 22×20, 18×27, 36×17
2	Conv5_3	128×128	27×39, 75×45, 48×29, 42×69, 40×39, 84×23, 24×66
3	FC7	64×64	72×80, 92×89, 56×72, 48×91, 105×60, 73×100
4	Conv8_2	32×32	149×96, 159×152, 40×133, 130×120
5	Conv9_2	16×16	167×201, 97×187, 59×210
6	Conv10_2	8×8	246×248
7	Conv11_2	4×4	290×323

表 3 为利用分组聚类法和 SSD 公式法获取的默认目标框参数对 DOTA 数据集各类别目标样本的平均覆盖度对比。由表 3 可知, 分组聚类法得到的默认目标框参数对各类别目标的平均覆盖度均高于 78%, 小目标占 99% 的小型车辆的平均覆盖度更是由 35.65% 提升到 80.99%, 且标准差 Std 仅为

表 3 两种方法设置的默认目标框参数对于 DOTA 数据集中各类别的平均覆盖度对比

Table 3 Comparison of average coverage for each category in DOTA of default object frame parameters set by two methods

Object category	SV	LV	Ship	Plane	ST	Harbor	Bridge	TC
SSD	35.65	66.11	66.29	80.12	54.04	76.20	59.25	77.27
BGC	80.99	78.09	80.10	87.24	87.31	78.77	80.21	86.94
Object category	RA	SF	HP	GFT	BD	SP	BC	
SSD	77.36	67.46	66.84	78.96	85.21	73.63	77.70	
BGC	88.87	79.46	79.20	80.51	88.06	84.56	82.48	

2.2 高效多尺度特征融合 MSFF 模块设计

为进一步提升模型的特征提取能力, 本文同样采用 FPN 方式构建多尺度特征融合 MSFF 模块, 结构如图 5 所示。

为引入 MSFF 模块、更好地适应小目标检测, 首先对 SSD 网络进行微调: 1) 去除 Conv3_X 与 Conv4_

表 2 中 w 为宽度, h 为高度。

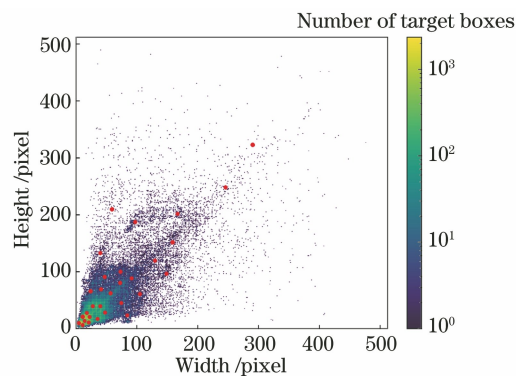


图 4 DOTA 数据集样本目标框尺寸分布与分组聚类法获得的默认目标框尺寸分布

Fig. 4 Distribution of sample target box dimensions in DOTA and default target box dimensions set by group clustering

3.74%, 这说明 BGC 模块设置的默认目标框参数给予了小目标更多关注, 同时又顾及了中、大目标, 对各类样本的覆盖度更均衡, 与光学遥感图像中目标样本尺度分布的契合度更高, 使网络能在后续预测过程中对目标进行更为准确的定位和特征提取, 进而提升目标检测精度。

X 之间的池化层, 使 Conv4_3 的分辨率提高 1 倍; 2) 在 Conv4_2、Conv4_3 层采用空洞卷积(卷积核为 3×3)以增大感受野。以上微调增强了浅层特征层对小目标位置、细节等空间特征信息的提取能力, 这些增强信息进入 MSFF 模块后与深层语义信息融合, 将进一步增强网络对小目标特征的提取能力。

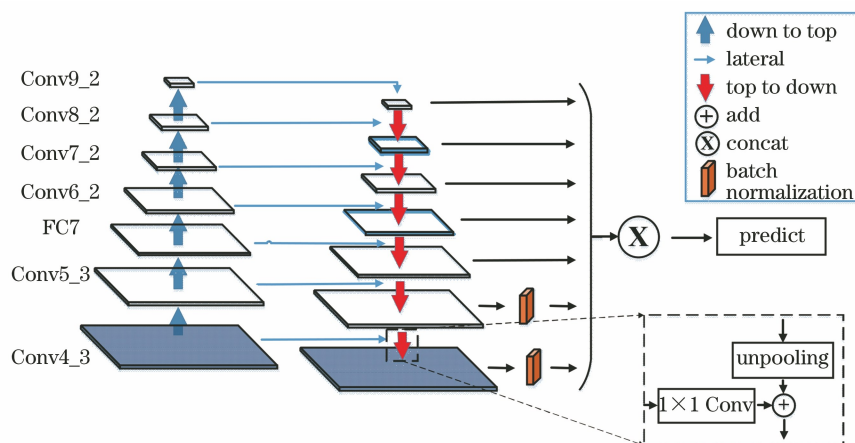


图 5 MSFF 模块示意图

Fig. 5 Diagram of MSFF module

小目标由于包含的像素少,其特征与位置信息会随网络的加深逐渐丢失,浅层特征层包含的目标信息对小目标检测更为重要,这些特征在 FPN 中能通过高层语义信息的融入进一步强化小目标的特征,改善对小目标的检测效果。本文在以上网络微调基础上选择 Conv4_3、Conv5_3、FC7、Conv8_2、Conv9_2、Conv10_2、Conv11_2 共 7 个特征层参与特征融合,通过增加浅层特征层、增强浅层特征信息,提升小目标检测效果。

现有文献设计的 MSFF 模块普遍采用反卷积操作实现特征图上采样^[8-13],计算量大导致检测效

率下降。反池化无需参数学习,可有效减少模型参数、提高学习及检测效率^[18],故本文设计的 MSFF 模块(以下简称 MSFF_U 模块)采用反池化操作实现融合特征图上采样;反池化过程中,通过在非感兴趣特征处进行补 0 操作,使特征图中的感兴趣特征与背景得到更有效的区分,使特征更具辨识度,这进一步降低了目标检出难度。图 6 显示了反卷积上采样的 MSFF 模块(以下简称 MSFF_D 模块)和本文设计的反池化上采样的 MSFF_U 模块的输出特征图对比。图 6 中 MSFF_U 模块输出特征图的前景与背景的分度度更高,纹理更加清晰。

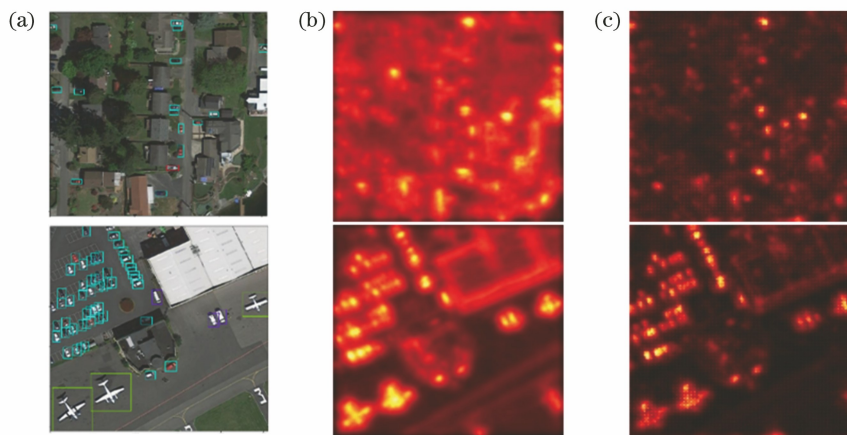


图 6 MSFF_D 与 MSFF_U 模块输出特征图对比。(a)原图;(b) MSFF_D 模块输出特征图;(c) MSFF_U 模块输出特征图

Fig. 6 Output feature maps of MSFF_D and MSFF_U modules. (a) Original images; (b) output feature maps of MSFF_D module; (c) output feature maps of MSFF_U module

MSFF_U 模块采用反池化操作实现上采样,显著减少了模型参数、降低了计算复杂度,也使模型能够选取更多浅层特征层参与特征融合,这有利于增强小目标特征。将 SSD 分别引入 MSFF_D、MSFF_U 模块,在 DOTA 数据集上测算检测精度

和效率,可以发现:将 SSD 引入 MSFF_U 模块 [mAP 为 64.6%,检测帧率(FPS)下降 2]相比引入 MSFF_D(mAP 为 60.5%,FPS 下降 6),在提升检测精度的同时也降低了效率损耗。此外,使用 Adam 优化器代替 SSD 的 SGDM,可使模型收敛更

快、更充分,检测精度也随之得到提升。

3 模型验证与结果分析

3.1 实验环境与参数配置

本文实验环境:CPU 为 Intel Xeon(R) Gold 5118 2.30 GHz;GPU 为 NVIDIA TITAN V (12 GB 内存) \times 2;操作系统为 Ubuntu 18.04;深度学习框架为 Caffe。

数据集:本文选用 DOTA 数据集进行消融实验,采用 DOTA 和 NWPU VHR-10 数据集^[19]进行不同模型间的性能对比分析。分别随机抽取 2 个数据集中 1/2、1/6、1/3 的影像形成训练集、验证集和测试集^[15-16]。为便于 FFC-SSD 与其他模型的性能对比,本文选用与相关文献相同的实验参数与数据裁剪方式^[15,20]:将影像按相互间重叠 200 pixel 的方式裁剪为大小为 1024 pixel \times 1024 pixel 的图像;检测帧率 FPS 指输入图像(1024 pixel \times 1024 pixel)在一块显存为 12G 的显卡上的处理速度。

模型训练参数:初始学习率为 0.001, batch size 为 8, decay 为 0.0005, momentum 为 0.9, momentum2 为 0.999;迭代次数在 30000 及 60000 次时,设置学习率衰减率为 0.1。测试阶段设置非极大阈值为 0.45。图 7 为模型在 DOTA 数据集的

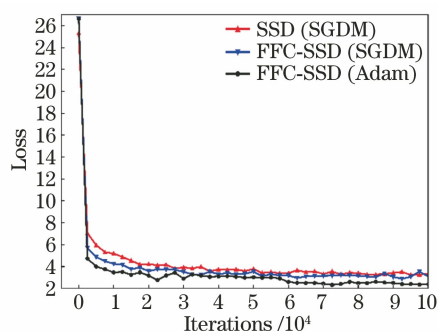


图 7 损失函数收敛曲线

Fig. 7 Convergence curves of loss function

表 4 各模块对 DOTA 数据集中目标检测的精度及速度的影响

Table 4 Influence of each module on mAP and FPS of object detection

Experiment No.	Model	Aps / %	APless / %	mAP / %	FPS
1	SSD	33.5	52.5	55.9	26
2	SSD+MSFF_U	44.3	63.5	64.6	24
3	SSD+BGC	49.5	62.4	63.6	16
4	SSD+BGC+MSFF_U (FFC-SSD)	69.3	69.9	74.9	15
5	SSD+BGC+MSFF_D	63.4	64.4	70.0	12

图 8 为实验 1~5 对 DOTA 数据集中每类样本的检测精度对比。从图 8 可见,FFC-SSD 对各类别

损失函数收敛曲线对比,从图中可见 FFC-SSD 采用 Adam 时收敛更充分,更有助于性能提升。

3.2 消融实验

本文基于 DOTA 数据集设计 5 组实验以分析 FFC-SSD 及各优化模块(BGC、MSFF_U)对光学遥感图像目标检测性能的影响,如表 4 所示。表中,APs、APless 分别表示小目标和小样本量目标(样本量 $<$ 2000,表 1 中的 T4 组目标)的检测精度。

表 4 中实验 1,SSD 的 APs、APless 及 mAP 均处于较低水平,APs 比 mAP 低 20%。实验 2 为 SSD 引入多尺度特征融合 MSFF_U 模块,mAP 提升 8.7%,APs 及 APless 也均有约 10%的提升,MSFF_U 的引入使模型对小目标以及所有目标的检测效果得到明显改善;由于采用反池化高效多尺度融合,效率并没有显著降低,FPS 仅下降 2。实验 3 为 SSD 引入 BGC 模块,由于默认目标框参数契合目标尺度分布,实验 3 的整体精度提升仅略低于实验 2;实验 3 给予了小目标更多关注,增强了小目标的位置信息,小目标的检测精度 APs 提升了 16%,比实验 2 提升更明显,这显示了默认目标框参数设置对于尺度分布宽、小目标数量大、密集分布的遥感图像目标检测的重要性;BGC 模块在浅层特征层设置了更多默认目标框参数,增加了计算量,FPS 下降至 16。实验 4 为 SSD 同时增加 MSFF_U 和 BGC 模块,即 FFC-SSD,增强的目标浅层特征信息在 MSFF_U 模块中与高层语义相融合,目标特征得到进一步增强,mAP 进一步显著提升,APs 提升尤为显著,APs、APless 及 mAP 差异缩小至 5%,FFC-SSD 对小目标、小样本量目标的检测效果接近整体水平,显示出很好的适用性。实验 5 将实验 4 的 MSFF_U 模块替换为 MSFF_D 模块,即多尺度特征融合采用传统反卷积上采样,其检测精度提升略低于 FFC-SSD,但反卷积增加了计算量,使检测效率下降更明显,这验证了 FFC-SSD 的多尺度特征融合性能优势。

的检测精度较 SSD 均提升显著。

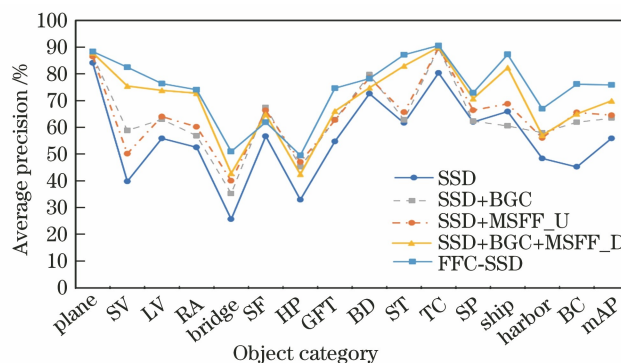


图 8 各实验对 DOTA 数据集中每类样本的检测精度

Fig. 8 Average precision (AP) for each category in DOTA testing dataset for each experiment

以上实验表明,光学遥感图像中各类目标的特征通过 BGC 和 MSFF_U 两个优化模块得到增强,FFC_SSD 在减小效率损耗的同时,使小目标及所有目标的检测精度得到显著提升。

3.3 模型性能对比实验

表 5 为 FFC-SSD 与典型目标检测模型在 DOTA 数据集的性能对比,表中 FRCNN(Faster R-CNN)、DSSD(Deconvolutional SSD)、FMSSD 对应的数据引自文献[15,20]。

表 5 中,双阶段模型 FRCNN 的 FPS 较低,且由于仅使用网络最末层进行预测,没有利用目标的多尺度特征,小目标特征在网络深层损失,难以被检测,故 FRCNN 对小目标居多的类别(小型车辆、大型车辆、储藏罐、游泳池和船)的检测精度均较低。

单阶段模型 YOLOv3 和 DSSD 都使用了多尺度特征融合策略,mAP 有一定提升,但两个模型均源于自然图像,没有针对遥感图像的目标分布特征进行优化,对 DOTA 数据集各类别目标的检测精度差异很大,标准差 stdAP 偏高,对光学遥感图像多类别目标检测的适应性和均衡性不足;两个模型的复杂度均较 FMSSD 和 FFC-SSD 更高,故检测效率偏低,DSSD 的 FPS 甚至小于 10。FMSSD 和 FFC-SSD 都针对遥感图像目标检测任务进行了优化,分别设计了 MSFF 模块,获得了更高 mAP(>70%)、更低 stdAP 和超过 15 的 FPS,较好地实现了精度与效率的平衡。但是,FMSSD 采用公式法设置默认目标框参数,对小目标的位置等浅层特征进行提取的准确性低于 FFC-SSD,且由于 FFC-SSD 的

表 5 FFC-SSD 与其他典型模型对 DOTA 数据集的检测性能

Table 5 Detection performance of FFC-SSD and other models on DOTA dataset

Model	SSD ^[1]	YOLOv3 ^[9]	FRCNN ^[2,20]	DSSD ^[15,21]	FMSSD ^[15]	FFC-SSD
Plane	84.2	91.0	80.3	91.1	89.1	88.4
Small vehicle	39.9	40.3	53.6	79.0	69.2	82.5
Large vehicle	55.9	76.9	52.5	77.2	73.6	76.4
Roundabout	52.6	58.5	49.8	72.6	67.5	74.1
Bridge	25.7	50.0	32.9	54.6	48.2	51.0
Soccer field	56.7	18.0	57.0	38.0	52.7	62.0
Helicopter	33.0	85.2	41.9	28.9	60.2	54.3
AP						
Ground field track	54.8	30.2	68.1	66.4	68.0	74.7
Baseball diamond	72.7	68.3	77.6	71.8	81.5	78.3
Storage tank	61.7	82.1	59.6	69.7	73.3	87.2
Tennis court	80.4	92.0	90.4	87.6	90.7	90.6
Swimming pool	62.0	80.2	56.5	59.4	80.6	73.0
Ship	65.9	89.2	50.0	87.5	76.9	87.4
Harbor	48.4	69.3	61.7	75.4	72.4	67.2
Basketball court	45.3	62.4	75.1	52.1	82.7	76.2
mAP	55.9	66.2	60.6	67.4	72.4	74.9
stdAP	15.8	22.2	14.9	17.4	11.7	11.6
FPS	26	13	7	9	16	15

MSFF_U 模块设计增强了对密集分布目标的区分能力,故 FFC-SSD 对 DOTA 数据集中小目标占比高、密集分布的小型车辆、储藏罐和船的检测精度较 FMSSD 更高,对小目标居多的类别(小型车辆、大型车辆、储藏罐、游泳池和船)的检测精度均处在较高水平。FFC-SSD 较好地解决了小目标检测精度低的问题,且保持了较高的检测效率,对光学遥感图像中的多类别目标和小尺寸目标的检测显示出很好的适用性。

表 6 列出了 FFC-SSD 与典型模型在 NWPU VHR-10 数据集的性能对比,表中 FRCNN、FMSSD 数据引自文献[15]。对于以中目标为主(占比 83%)、尺度差异较小、检测难度较低的 NWPU

表 6 FFC-SSD 与其他典型模型对 NWPU VHR-10 数据集的检测性能

Table 6 Detection performance of FFC-SSD and other models on NWPU VHR-10 dataset

Model	SSD ^[1]	YOLOv3 ^[9]	FRCNN ^[2,15]	FMSSD ^[15]	FFC-SSD	
AP	Plane	98.2	95.6	94.6	99.7	99.7
	Ship	83.9	88.6	82.3	89.9	96.3
	Storage tank	75.9	77.9	65.3	90.3	88.1
	Baseball diamond	90.2	91.7	95.5	98.2	99.4
	Tennis court	85.6	89.1	81.9	86.0	90.3
	Basketball court	79.6	89.8	89.7	96.8	99.4
	Ground track field	92.2	84.8	92.4	99.6	99.9
	Harbor	77.1	81.2	72.4	75.6	96.1
	Bridge	67.8	70.8	57.5	80.1	98.5
	Vehicle	75.6	87.8	77.8	88.2	89.0
	mAP	82.6	85.7	80.9	90.4	95.7
	stdAP	8.70	6.92	12.19	7.89	4.49

3.4 目视判别与应用效果实验

图 9 为 SSD 与 FFC-SSD 对 DOTA 数据集检测结果的目视效果对比。从图 9 中可见,FFC-SSD 极大减少了目标误检和漏检的情况,且目标定位更准确[图 9(d)中港口及棒球场]。

由于 DOTA 数据集在目标类别、样本数量、目标尺度/空间分布等方面具有很好的光学遥感图像目标特征表征代表性,本文将在 DOTA 数据集上训练得到的 SSD 与 FFC-SSD 模型直接应用于 NWPU VHR-10、RSOD^[22]、UCAS-AOD^[23] 光学遥感图像数据集,以测试 SSD 和 FFC-SSD 对光学遥感图像目标检测的应用效果,结果如表 7 所示。

表 7 不同模型在 3 个光学遥感图像数据集上的 mAP 对比
Table 7 mAP of different models on three optical remote sensing datasets

Model	NWPU VHR-10	RSOD	UCAS-AOD
SSD	61.7	45.1	48.9
FFC-SSD	76.5	60.2	69.6

VHR-10 数据集,各典型模型的检测精度均高于 80%。由于对遥感图像目标特征进行了针对性优化,FMSSD 和 FFC-SSD 的 mAP 均优于 90%,比其他通用模型提升更为显著。其中,FFC-SSD 的 BGC 模块对默认目标框的设置参数与数据集目标样本的尺度分布更为契合,位置信息提取更准确,对 MSFF_U 模块的特征融合与提取进行了增强,使 mAP 提升较 FMSSD 更大、stdAP 更小,检测效果更好。

以上实验从模型和数据对比的角度验证了 FFC-SSD 对光学遥感图像目标检测任务的适用性,也显示了合理设置默认目标框参数对于提升 SSD 模型精度的重要性。

从表 7 中可见,FFC-SSD 在 3 个数据集上的 mAP 均优于 60%,较 SSD 均有显著提升。对比表 6 和表 7 可以发现:FFC-SSD 即使不经过 NWPU VHR-10 数据集的训练,其检测精度(76.5%)也仅略低于经过训练的 SSD、FRCNN 模型。实验结果表明,数据集是提升模型泛化能力的一个重要因素,基于有很好光学遥感图像目标特征表征代表性的数据集(如 DOTA)训练得到的 FFC-SSD,具有良好的应用通用性和普适性,能够适应常规光学遥感图像目标检测任务。

4 结 论

面向高效、高精度光学遥感图像目标检测应用,重点针对提升 SSD 对聚集分布的小尺寸目标检测精度的技术难点,提出了一种基于目标框分组聚类 BGC 模块结合高效多尺度特征融合 MSFF 模块的改进模型——FFC-SSD。BGC 模块使模型的默认

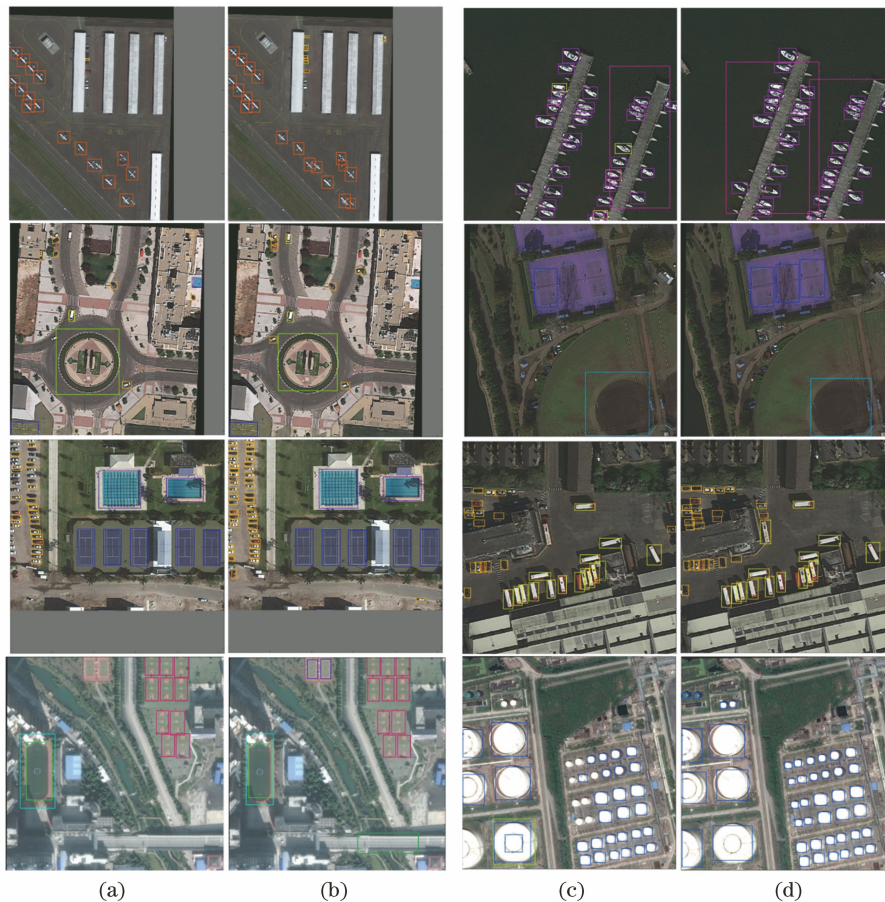


图 9 SSD 与 FFC-SSD 模型检测结果对比。(a)(c) SSD;(b)(d) FFC-SSD

Fig. 9 Test results of SSD and FFC-SSD models. (a)(c) SSD; (b)(d) FFC-SSD

目标框参数设置更契合遥感图像目标尺度分布,在平衡中、大目标的同时给予了小目标更多关注,增强了目标位置信息;MSFF 模块利用其反池化高效多尺度特征融合策略,在有效减少参数和计算量的同时,使被增强的目标浅层特征(位置、纹理等)在其与高层语义融合的过程中得到进一步增强,使各融合特征图输出更为准确、丰富和有辨识度,使小目标以及所有目标的检测效果得到显著提升。

多个实验结果表明,FFC-SSD 较好地实现了精度与效率的平衡,相比同类 SSD 改进模型,FFC-SSD 对图像中密集分布的小尺寸目标具有更好的检测效果,对光学遥感图像目标检测应用显示出较好的有效性与适用性。后续将结合实际应用需求,在损失函数、效率提升等方面开展优化与验证,以进一步提升模型性能与适用性。

参 考 文 献

- [1] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [2] 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.
- [3] 李国进, 胡洁, 艾娇燕. 基于改进 SSD 算法的车辆检测[J]. 计算机工程, 2022, 48(1): 266-274.
Li G J, Hu J, Ai J Y. Vehicle detection based on improved SSD algorithm[J]. Computer Engineering, 2022, 48(1): 266-274.
- [4] 成喆, 吕京国, 白颖奇, 等. 结合 RPN 网络与 SSD 算法的遥感影像目标检测算法[J]. 测绘科学, 2021, 46(4): 75-82, 99.
Cheng Z, Lü J G, Bai Y Q, et al. High-resolution remote sensing image object detection algorithm combining RPN network and SSD algorithm [J]. Science of Surveying and Mapping, 2021, 46(4): 75-82, 99.
- [5] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern

- Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [6] 高新波, 莫梦竟成, 汪海涛, 等. 小目标检测研究进展[J]. 数据采集与处理, 2021, 36(3): 391-417.
Gao X B, Mo M, Wang H T, et al. Recent advances in small object detection [J]. Journal of Data Acquisition and Processing, 2021, 36(3): 391-417.
- [7] 陈科圻, 朱志亮, 邓小明, 等. 多尺度目标检测的深度学习研究综述[J]. 软件学报, 2021, 32(4): 1201-1227.
Chen K Q, Zhu Z L, Deng X M, et al. Deep learning for multi-scale object detection: a survey[J]. Journal of Software, 2021, 32(4): 1201-1227.
- [8] Cao G, Xie X, Yang W, et al. Small object detection using deep feature pyramid networks[J]. Proceedings of SPIE, 2018, 10615: 106151E.
- [9] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-05-08]. <https://arxiv.org/abs/1804.02767>.
- [10] Li X H, He M Z, Li H F, et al. A combined loss-based multiscale fully convolutional network for high-resolution remote sensing image change detection[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 8017505.
- [11] Zhao L, Zhang X F. Object detector based on enhanced multi-scale feature fusion pyramid network [C] // 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference, March 12-14, 2021, Chongqing, China. New York: IEEE Press, 2021: 289-293.
- [12] 朱敏超, 冯涛, 张钰. 基于 FD-SSD 的遥感图像多目标检测方法[J]. 计算机应用与软件, 2019, 36(1): 232-238.
Zhu M C, Feng T, Zhang Y. Remote sensing image multi-target detection method based on FD-SSD[J]. Computer Applications and Software, 2019, 36(1): 232-238.
- [13] 史文旭, 谭代伦, 鲍胜利. 特征增强 SSD 算法及其在遥感目标检测中的应用[J]. 光子学报, 2020, 49(1): 0128002.
Shi W X, Tan D L, Bao S L. Feature enhancement SSD algorithm and its application in remote sensing images target detection [J]. Acta Photonica Sinica, 2020, 49(1): 0128002.
- [14] 黄国新, 李炜, 张比浩, 等. 改进 SSD 的机场场面多尺度目标检测算法[J]. 计算机工程与应用, 2022, 58(5): 264-270.
Huang G X, Li W, Zhang B H, et al. Improved SSD-based multi-scale object detection algorithm in airport surface [J]. Computer Engineering and Applications, 2022, 58(5): 264-270.
- [15] Wang P J, Sun X, Diao W H, et al. FMSSD: feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(5): 3377-3390.
- [16] Xia G S, Bai X, Ding J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3974-3983.
- [17] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [18] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [19] Cheng G, Zhou P C, Han J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12): 7405-7415.
- [20] Yan J Q, Wang H Q, Yan M L, et al. IoU-adaptive deformable R-CNN: make full use of IoU for multi-class object detection in remote sensing imagery[J]. Remote Sensing, 2019, 11(3): 286-301.
- [21] Fu C Y, Liu W, Ranga A, et al. DSSD: deconvolutional single shot detector[EB/OL]. (2017-01-23) [2021-05-08]. <https://arxiv.org/abs/1701.06659v1>.
- [22] 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.
- [23] Zhu H G, Chen X G, Dai W Q, et al. Orientation robust object detection in aerial images using deep convolutional neural network [C] // 2015 IEEE International Conference on Image Processing, September 27-30, 2015, Quebec City, QC, Canada. New York: IEEE Press, 2015: 3735-3739.