计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(10):301−306 [doi: 10.15888/j.cnki.csa.008119]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

面向光学遥感图像典型目标检测的 SSD 模型优化①

薛俊达[1,2], 朱家佳[1,2], 李晓辉[1], 张　静[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 定量遥感信息技术重点实验室, 北京 100094)
[2](中国科学院大学 光电学院, 北京 100049)
通讯作者: 张　静, E-mail: zhangjing@aoe.ac.cn

摘　要: 本文面向光学遥感图像目标检测应用, 针对光学遥感图像中的典型目标—飞机和汽车, 提出一种改进的 SSD 模型: 首先在 SSD (Single Shot multibox Detector) 网络模型基础上引入多尺度特征融合模块, 实现深层特征与浅层特征的融合以获得更多的特征上下文信息, 增强网络对目标特征的提取能力; 其次根据数据集目标样本尺寸分布特征进行聚类分析获得更准确的默认目标框参数, 从而有效提升网络对目标位置信息的提取能力. 将本文模型与 SSD 及 YOLOv3 模型在常用遥感图像目标检测数据集上进行对比, 目标检测精度均有较大提升, 验证了该模型的有效性.

关键词: 光学遥感图像目标检测; 目标框聚类; 多尺度特征融合; SSD

引用格式: 薛俊达,朱家佳,李晓辉,张静,窦帅,米琳,李子扬,苑馨方,李传荣.面向光学遥感图像典型目标检测的 SSD 模型优化.计算机系统应用,2021,30(10):301–306. http://www.c-s-a.org.cn/1003-3254/8119.html

SSD Model Optimization for Typical Object Detection in Optical Remote Sensing Images

XUE Jun-Da[1,2], ZHU Jia-Jia[1,2], LI Xiao-Hui[1], ZHANG Jing[1], DOU Shuai[1], MI Lin[1], LI Zi-Yang[1], YUAN Xin-Fang[1], LI Chuan-Rong[1]

[1](Key Laboratory of Quantitative Remote Sensing Information Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China)
[2](School of Optoelectronics, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Oriented to object detection in optical remote sensing images, this study proposes an improved Single Shot multibox Detector (SSD) model aiming at typical objects, i.e., aircraft and car, in the images. First, a multi-scale feature fusion module is introduced to the SSD network model to fuse deep features and shallow features. As a result, more contextual information of features can be obtained and the network's ability to extract object features is enhanced. Then, cluster analysis is performed according to the size distribution characteristics of target samples in the data set to obtain more accurate default bounding box parameters, thereby effectively improving the network's ability to extract target location information. Finally, the proposed model is compared with SSD and YOLOv3 models on data sets common for object detection in remote sensing images, which demonstrates that the mean Average Precision (mAP) of object detection has been greatly improved and verifies the effectiveness of our model.

Key words: detection of optical remote sensing image objects; bounding box clustering; multi-scale feature fusion; Single Shot multibox Detector (SSD)

① 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)
Foundation item: National Key R & D Program of China (2018YFB050540); Strategic Priority Program of Chinese Academy of Sciences (Category A) (XDA17040303)
收稿时间: 2020-12-31; 修改时间: 2021-01-29; 采用时间: 2021-02-08

随着计算机处理能力的显著提升以及深度学习、卷积神经网络 (Convolutional Neural Network, CNN) 技术在自然图像目标检测应用中取得巨大成功, 为实现高精度、自动、高效的遥感图像目标检测提供了新的技术途径与动力. 2016 年, Liu 等[1] 提出了 SSD (Single Shot multibox Detector) 模型. 作为单阶段目标检测模型的代表之一, SSD 模型在目标定位与分类过程中借鉴了双阶段模型 Faster-R-CNN[2] 的“anchor boxes”以及多尺度特征提取的思想, 使其在保持单阶段目标检测模型的高检测效率的同时, 检测精度有了很大提升. SSD 在自然图像 PASCAL VOC 2007 数据集的平均检测精度 mAP 达到 75.1%, 每秒检测帧率 (Frame Per Second, FPS) 达到 58, 很大程度上实现了检测精度与速度的平衡. SSD 的上述技术特点, 使其在遥感图像目标检测任务中展现出很好的适用性和技术潜力, 受到关注.

相比自然图像, 遥感图像的图幅更大、场景和目标更为复杂, 将 SSD 模型直接应用于遥感图像目标检测中难以获得满意的效果, 必须针对遥感图像特点与目标分布特征对 SSD 模型进行适当的改进与优化.朱敏超等[3] 针对遥感图像目标检测提出了改进的 FD-SSD 网络, 借鉴特征金字塔网络 (Feature Pyramid Networks, FPN)[4] 增强低层特征空间语义信息, 在 DOTA 遥感图像数据集[5] 的检测精度较原 SSD 模型有一定提升, 对飞机、小汽车的检测精度分别为 71.98% 和 41.56%. 史文旭等[6] 提出了改进的 FESSD 模型, 分别利用多分支卷积和双路径网络思想增强网络特征特征提取能力, 在自己构建的遥感图像数据集上的 mAP 达到 79.36%, 对飞机的检测精度为 80.96%. Wang 等[7] 提出了一种特征融合 FMSSD 模型, 通过采用空洞空间特征金字塔模块、区域加权代价函数以及优化 Loss 计算方法等优化措施, 对 DOTA 遥感图像数据集的飞机和小汽车等典型目标的检测精度分别达到 89.11% 和 69.23%.

在光学遥感图像目标检测研究与应用中, 飞机、汽车是最为典型且使用最为普遍的目标. 一方面, 这些目标在日常生活中非常普遍, 包含这些目标的光学遥感图像非常容易获得, 目标样本量大, 对目标检测模型/方法进行训练和验证的可行性非常高.另一方面, 飞机在光学遥感图像中表现为形态特征较明显、目标与背景对比度较高、目标样本尺度分布较宽 (小到几十乘几十像素、大到几百乘几百像素) 等特点; 汽车在光学遥感图像中则表现为与背景对比度较低、目标尺寸小且样本尺度分布窄 (绝大部分目标框尺寸在 50×50 像素以下)、密集分布、数量很大等特点. 在光学遥感图像目标检测研究中, 这两类目标在形态、样本尺度分布、样本数量以及特征提取难度等方面具有非常显著的差异, 对于分析和验证目标检测模型/方法的通用性和适用性具有实际意义.

本文面向光学遥感图像目标检测应用, 以提升 SSD 模型对光学遥感图像中的飞机和汽车类典型目标的检测精度为目标, 提出一种结合多尺度特征融合与目标框聚类分析的 SSD 优化模型: 1) 借鉴特征金字塔多尺度特征融合思想设计并引入多尺度特征融合模块, 实现深层特征与浅层特征的融合以获得更多的特征上下文信息, 增强网络对目标特征的提取能力; 2) 根据数据集目标样本尺寸分布特征进行聚类分析获得更准确的默认目标框参数, 以有效提升网络对目标位置信息的提取能力.

1 模型设计

本文模型框架如图 1 所示.

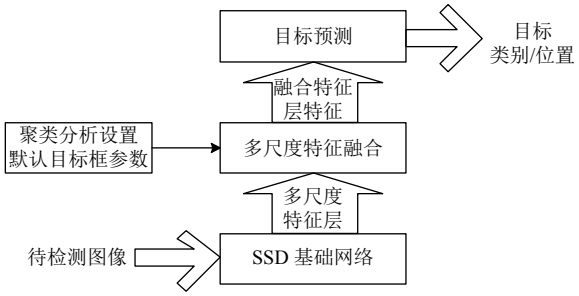


图 1 本文模型框架图

本文模型充分考虑各特征层之间的上下文联系, 在 SSD 模型的基础上引入由效率更高的反池化实现上采样的多尺度特征融合模块, 同时辅以分类别 K-means 聚类分析获得更符合数据集目标框尺寸分布的模型默认目标框参数.在不显著增加模型运算量的情况下, 提升 SSD 模型的特征提取能力以及目标定位精度.

1.1 多尺度特征融合

原始的 SSD 模型由改造后的 VGG-16 网络增加 5 个卷积层 (FC7、Conv8_2、Conv9_2、Conv10_2、Con11_2) 组成. SSD 模型使用网络中的 6 个不同尺度的特征层组对目标进行分类和定位: 浅层特征层

(Conv4_3 和 FC7) 主要用来预测小尺寸目标, 深层特征层 (Conv8_2、Conv9_2、Conv10_2、Con11_2) 主要用来预测大尺寸目标, 从而提升了 SSD 模型对于不同尺度目标检测的适用性.以上设计虽然考虑了多尺度特征的使用, 但没有考虑不同尺度特征之间的关联信息, 目标特征提取并不充分, 特别是对小尺寸目标 (如汽车) 的检测精度提升有限.

为了能更充分提取目标特征, 提升模型对小尺寸目标的检测精度, 本文借鉴 FPN 思想, 设计并引入了多尺度特征融合模块, 采用如图 2 所示的“由深至浅+横向连接”的方式, 提取 SSD 网络中的 7 个不同尺度的特征层 (Conv4_3、Conv5_3、FC7、Conv8_2、Conv9_2、Conv10_2、Con11_2) 的输出特征进行目标特征融合.

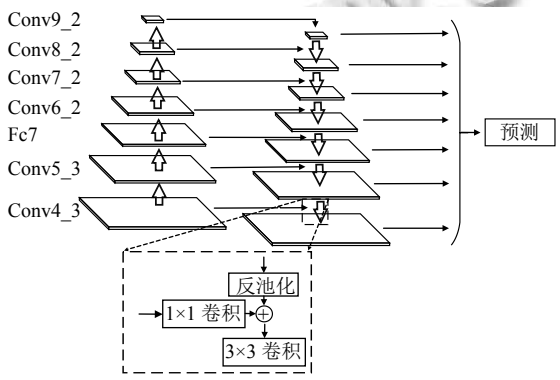


图 2 多尺度特征融合模块结构

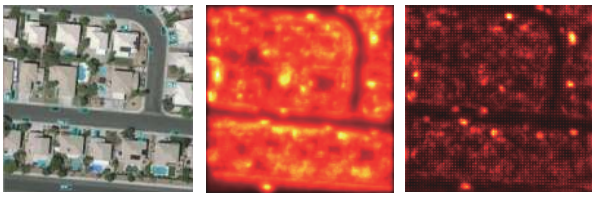
由于反池化操作不需要参数学习, 可以有效减少模型参数、提高学习及检测效率[8]; 并且反池化操作在非感兴趣特征处补 0, 可以更有效区分感兴趣特征与背景、更容易检出目标特征, 本文在“由深至浅”过程中采用反池化操作代替目前普遍采用的反卷积操作来实现融合特征图上采样. 表 1 显示了 SSD 模型在分别引入由反卷积与反池化实现上采样的多尺度特征融合后的目标检测效率对比; 图 3 显示了两种特征融合方式得到的融合特征图的对比.

从表 1 可以看出, SSD 模型在使用反池化实现上采样后, 检测效率没有明显下降, FPS 仅减少了 2, 性能表现优于反卷积. 从图 3 中可以看出, 使用反池化实现上采样得到的融合特征图中的目标纹理更加清晰, 与背景的区别度更高, 更利于提取目标特征.

另外, 为了支持多尺度特征融合模块进行多尺度特征融合, 在 SSD 网络生成多尺度特征层的池化过程中采用最大值池化, 并输出最大值位置索引, 用于多尺度特征融合模块“由深至浅”过程中的反池化操作.

表 1 3 种目标检测模型的目标检测效率对比

模型	FPS
SSD	26
SSD+多尺度特征融合(反卷积)	21
SSD+多尺度特征融合(反池化)	24



(a) 原图 (b) 反卷积 (c) 反池化

图 3 两种特征融合方式所得融合特征图对比

1.2 聚类分析设置默认目标框参数

本文模型利用多尺度特征融合模块输出的 7 个尺度的融合特征图进行目标预测, 需要在模型训练前预设各融合特征图上的默认目标框参数. SSD 模型利用经验公式设置默认目标框参数. 但是, SSD 使用的经验公式源于自然图像, 并不能很好契合遥感图像中的目标分布特征, 这也是 SSD 模型直接应用于遥感图像目标检测难以取得满意效果的原因之一.

如前文所述, 在光学遥感图像中, 飞机与汽车除了在形态上具有显著差异, 在目标尺寸、数量以及尺度分布等方面都具有显著差异. 以当前应用最广、样本最丰富的 DOTA 遥感图像数据集为例, 表 2 列出了数据集中飞机和汽车样本目标框尺寸分布情况.

表 2 DOTA 数据集样本目标框尺寸分布统计表

目标框尺寸 (pixel×pixel)	最小 (2)	1–50	51–200	201–512	最大 (454)	合计
飞机总量	0	2431	17008	564	1	20004
汽车总量	1	361345	50438	0	0	411784
总样本量	1	363776	67446	564	1	431788

从表 2 中可以看出, 飞机样本的目标框尺寸在 51×51–454×454 像素间, 其中, 85.02% 的样本尺寸在 51×51–200×200 像素间; 汽车样本的目标框尺寸在 2×2–200×200 像素之间, 但 87.75% 的样本尺寸小于

50×50 像素, 目标尺寸总体偏小. 飞机比汽车具有更宽的目标样本尺度分布, 同时汽车的样本量远大于飞机的样本量, 占总样本量的 95.37%. 上述差异都为预设目标框参数增加了一定难度.

YOLOv3 模型[9] 使用基于全局目标框 K-means 聚类方法来进行预设目标框大小, 但此方法容易导致聚类结果偏向样本量大的一方. 本文借鉴聚类方法预设目标框参数的思路, 采用分类别 K-means 聚类分析法, 以得到更具代表性的默认目标框参数.

图 4 显示了对 DOTA 数据集的训练数据集中的两类目标样本的目标框进行分类别 K-means 聚类分析得到的聚类个数 k 值与目标平均覆盖度[9] 的变化关系.由图 4 可以看出, 随着聚类个数的增加, 平均覆盖度变化趋于稳定, 且根据函数趋势可以看出, 两类目标的平均覆盖度变化开始趋于稳定的临界点均在聚类个数 k=8 附近, 可以认为聚类个数 k>8 时的聚类结果比较理想.

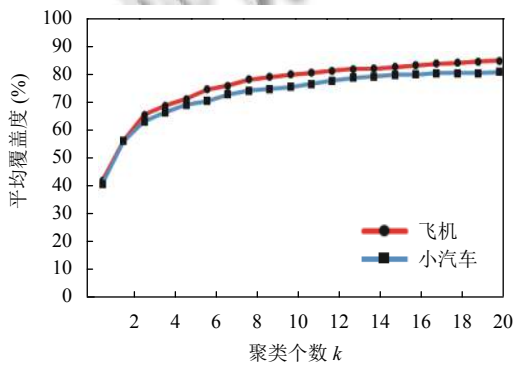


图 4 两类样本目标框 K-means 聚类折线统计

经过进一步的聚类分析, 针对汽车和飞机分别选择 k=8 和 k=15 时的目标框聚类结果进行整合得到最终的 23 个默认目标框参数, 分别为 [5, 10]、[10, 6]、[9, 11]、[9, 18]、[20, 11]、[21, 20]、[14, 21]、[36, 37]、[47, 48]、[28, 29]、[56, 67]、[41, 39]、[18, 18]、[44, 61]、[142, 135]、[68, 67]、[90, 69]、[94, 93]、[77, 84]、[55, 54]、[108, 110]、[178, 174]、[254, 256]. 图 5 给出了采用分类别聚类方法和全局聚类方法获得的目标框尺寸聚类结果对比.表 3 显示了采用 SSD 经验公式、全局聚类和分类别聚类方法获得默认目标框参数的目标尺寸平均覆盖度定量评估结果.

从表 3 可以看出, 采用分类别聚类方法获得默认目标框参数, 对飞机和汽车目标的平均覆盖度均优于 80%, 说明该方法获得的默认目标框参数能够更好反映遥感数集中飞机和汽车这两类目标的尺寸分布. 从图 5 和表 3 可以看出, 基于全局 K-means 聚类方法得到的聚类结果大多集中在 50×50 像素以下, 在 50×50–512×512 像素之间仅有一个聚类中心, 对飞机目标的平均覆盖度较低, 极大限制了模型对于飞机目标的检测; 而分类别 K-means聚类方法在保证平均覆盖度有一定提升的情况下, 避免了聚类结果偏向样本数量大的一方的问题, 聚类结果能够同时很好契合汽车和飞机目标的分布特征, 能够获得更高和更均衡的目标平均覆盖度.

表 3 3 种默认目标框参数设置结果的平均覆盖度对比 (%)

默认目标框参数设置方法	飞机	汽车	平均覆盖度
SSD模型经验公式	80.33	36.29	58.31
基于全局的K-means聚类	74.08	85.01	82.95
分类别K-means聚类	84.11	82.42	83.57

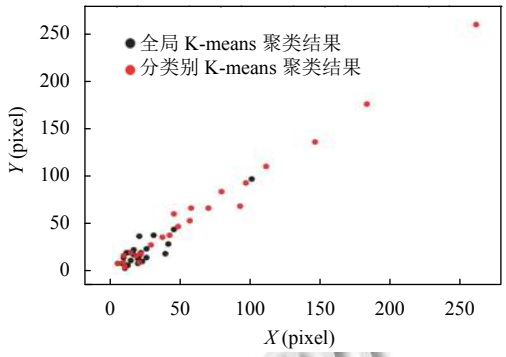


图 5 分类别聚类方法与全局聚类方法的聚类结果分布对比图

本模块最终选用 Conv4_3、Conv5_3、FC7、Conv6_2、Conv7_2、Conv8_2 以及 Conv9_2 七层特征用于后续目标预测. 同时沿用 SSD 模型默认目标框在各特征层上的分布范围, 将分类别聚类结果设置到各特征图上, 分布结果见表 4.

表 4 各特征层默认目标框参数设置

特征层	默认目标框
Conv4_3	[5, 10][9, 11][9, 18][20, 11][21, 20][14, 21][18, 18]
Conv5_3	[28, 29][36, 37][41, 39][47, 48]
FC7	[55, 54][56, 67][44, 61][68, 67][90, 69][94, 93][108, 110]
Conv6_2	[142, 135][178, 174]
Conv7_2	[254, 256]
Conv8_2	[330, 330]
Conv9_2	[425, 425]

注: [x,y]表示同时进行目标框翻转; [x,y]表示人为添加聚类结果以外的目标框.

2 实验结果与分析

2.1 模型训练

本文选择 DOTA 数据集中的飞机和小汽车两个目标类别组成数据集进行模型训练和验证.由于 DOTA 数据集原图较大, 为了便于训练, 将影像按 200 像素重叠裁剪为 512×512, 并将裁剪所得影像 1/2 用于训练, 1/6 用于验证, 1/3 用于测试.

实验配置的显卡为两块 NVIDIA TITAN V, 操作系统为 Ubuntu 18.04, 深度学习框架为 Caffe. 在模型训练阶段, 设置初始学习率=0.001, batch size=16, decay=0.0005, momentum=0.9; 当训练迭代次数在 40 000 及 60 000 次时, 设置学习率衰减率=0.1, 以使模型更快收敛.

图 6 为模型训练过程中损失值与 mAP 随迭代次数的变化.模型收敛在损失值 2.5 左右, mAP 达到 86.67%.

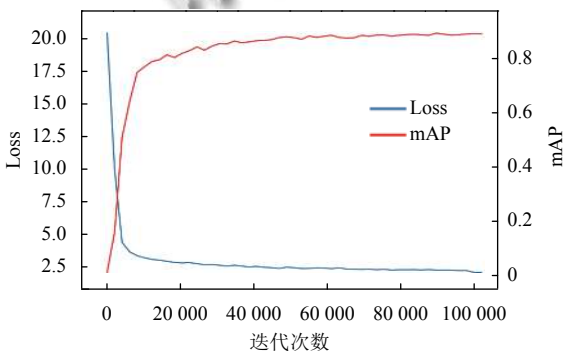


图 6 损失值及 mAP 随迭代次数变化

2.2 目标检测定量评价

在相同训练方式下, 将本文训练好的模型与 SSD 模型、YOLOv3 模型对比, 各模型在 DOTA 遥感数据集的测试数据集与 NWPU VHR-10 遥感数据集[10] 中的检测精度 (mAP) 与效率 (FPS) 如表 5 所示. 其中, SSD*及 YOLOv3*分别表示模型使用本文分类别聚类方法设置默认目标框参数.

从表 5 中可以看出, 针对光学遥感图像中飞机和汽车典型目标检测, 本文模型在 DOTA 数据集中的 mAP 较 SSD 及 YOLOv3 分别提升了 26.78% 和 18.32%, 在 NWPU VHR-10 数据集中的 mAP 较 SSD 及 YOLOv3 分别提升了 20.42% 和 10.19%; 由于本文模型在设计时, 通过引入多尺度特征融合以及设置默认目标框参数等措施, 对尺寸小的汽车目标给予了更多的关注, 使得本文模型相比 SSD 模型, 对于汽车的检测精度提升更为显著, 在 DOTA 数据集和 NWPU VHR-10 数据集中, 分别提升了 34.46% 和 32.63%.在使用本文提出的分类别聚类方法设置默认目标框参数后, SSD*及 YOLOv3*对于两类典型目标的检测精度相比 SSD 和 YOLOv3 也均有一定提升, 尤其是对于小尺寸目标占绝大多数的汽车目标的检测精度有了较大的提升, 验证了分类别聚类方法设置默认目标框参数的有效性.在检测速度方面, 由于默认目标框参数有所增加, 导致模型运算量相应增加, 本文模型的检测速度 (FPS=16) 较 SSD 模型 (FPS=26) 有所下降, 但优于模型更为复杂的 YOLOv3 (FPS=13).

表 5 各模型在 DOTA、NWPU VHR-10 数据集检测精度与速度对比

数据集	模型	AP(飞机)	AP(汽车)	mAP	FPS
DOTA	SSD	70.83	48.94	59.89	26
	SSD*	79.14	82.29	80.72	16
	YOLOv3	74.47	62.23	68.35	13
	YOLOv3*	81.92	71.84	76.88	7
	本文模型	**89.94**	**83.40**	**86.67**	15
NWPU VHR-10	SSD	91.33	56.72	74.03	26
	SSD*	92.70	78.40	85.55	16
	YOLOv3	96.28	72.24	84.26	13
	YOLOv3*	96.60	79.21	87.91	7
	本文模型	**99.54**	**89.35**	**94.45**	15

2.3 目标检测结果目视判别

图 7 为 SSD 模型和本文模型对 DOTA 数据集目标检测结果对比, 蓝色框为汽车类目标的检测结果, 绿色框为飞机类目标的检测结果. 从图 7 中可以看出, SSD 模型针对汽车与飞机目标进行检测时均存在一定程度的漏检, 而本文模型对 SSD 模型的漏检情况有一定的改善, 检测精度大大提升.

3 结论与展望

本文针对光学遥感图像中的飞机和汽车类典型目标检测提出一种改进的 SSD 模型: 通过引入多尺度特征融合模块增强模型对目标特征的提取能力; 采用分类别聚类方法设置更符合目标样本尺寸分布特征的默认目标框参数, 以优化网络对目标位置信息的提取能力.模型既对小尺寸、数量大、密集分布的汽车目标进行了重点关注, 又很好兼顾了尺度分布宽的飞机目标, 从而实现对这两类典型目标检测精度的提升.

通过在当前常用、公开的 DOTA 遥感图像数据集

和 NWPU VHR-10 遥感数据集上的测试结果表明, 相比 SSD、YOLOv3 等当前典型单阶段目标检测模型, 本文提出的 SSD 改进模型在检测速度优于 YOLOv3 的情况下, 对于飞机和汽车目标的检查精度有了很大提升, 验证了上述改进措施的有效性.

(a) SSD 模型

(b) 本文模型

图 7 检测结果对比

参考文献

1 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot Multibox detector. Proceedings of the 14th European Conference on Computer Vision. Cham: Springer, 2016. 21–37.

2 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]

3 朱敏超, 冯涛, 张钰. 基于 FD-SSD 的遥感图像多目标检测方法. 计算机应用与软件, 2019, 36(1): 232–238. [doi: 10.3969/j.issn.1000-386x.2019.01.042]

4 Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2016. 936–944.

5 Xia GS, Bai X, Ding J, *et al*. DOTA: A large-scale dataset for object detection in aerial images. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 3974–3983.

6 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. 光子学报, 2020, 49(1): 148–157.

7 Wang PJ, Sun X, Diao WH, *et al*. FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(5): 3377–3390. [doi: 10.1109/TGRS.2019.2954328]

8 Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for Image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495. [doi: 10.1109/TPAMI.2016.2644615]

9 Redmon J, Farhadi A. YOLOv3: An incremental Improvement. arXiv: 1804.02767, 2018.

10 Cheng G, Zhou PC, Han JW. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12): 7405–7415. [doi: 10.1109/TGRS.2016.2601622]