第 58 卷 第 16 期/2021 年 8 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于 YOLOv3 网络训练优化的高分辨率遥感影像目标检测

杨耘，李龙威*，高思岩，柏晗，江万成

长安大学地质工程与测绘学院，陕西 西安 710054

摘要 传统的 YOLOv3 模型通常利用 ImageNet、COCO 等与测试集目标场景特征差异较大的数据集进行训练，存在对高分辨率遥感影像中复杂场景目标检测精度不高的问题。为解决这一问题，提出了一种对 YOLOv3 网络训练过程进行优化的方法。该方法基于迁移学习的思想，在 YOLOv3 网络训练中，通过生成与目标域更相似的增广数据集对模型进行预训练，实现了训练过程的优化，提高了目标初始预测的精度；利用目标域训练数据对预训练模型参数进行微调，完成了对网络的训练。利用公开的 RSOD 和 DIOR 遥感图像目标检测数据集的子集对飞机、运动场、立交桥三大类目标进行模型训练和检测实验，结果表明：本文提出的训练优化后的 YOLOv3 模型有效地提高了复杂城区场景中上述三类目标的检测精度。与传统的 YOLOv3 模型相比，三类目标的平均精度均值(mAP)提高了 2%以上。

关键词 遥感；目标检测；高分辨率遥感影像；YOLOv3；迁移学习

中图分类号 P237 **文献标志码** A **doi**：10.3788/LOP202158.1601002

Objects Detection from High-Resolution Remote Sensing Imagery Using Training-Optimized YOLOv3 Network

Yang Yun, Li Longwei*, Gao Siyan, Bai Han, Jiang Wancheng

School of Geological Engineering and Surveying and Mapping, Chang'an University, Xi'an, Shaanxi 710054, *China*

Abstract The traditional YOLOv3 model uses ImageNet and COCO datasets for training, in which the scene target characteristics are significantly different from those in test datasets, and leads to low detection accuracy of complex scene targets in high-resolution remote-sensing images. This paper optimizes the training process of the traditional YOLOv3 network using the idea of transfer learning. During the training of the YOLOv3 network, the model is pre-trained by generating an augmented dataset similar to the target domain. The training-optimized method improves the accuracy of the object boundary of target prediction. Also, the parameters of the pre-training model are fine-tuned using a training dataset from the target domain, thus, completing the whole training process of the network. The experiment on the detection of three types of object, including aircraft, playground, overpass, was carried out based on a subset of RSOD & DIOR dataset for remote sensing image object detection. The results show that the proposed YOLOv3 model effectively improves the detection accuracy of the three types of targets in complex urban scenes. The mean average precision of object detection using our model improved by 2% or more, compared with the traditional YOLOv3 model.

Key words remote sensing; object detection; high-resolution remote-sensing image; YOLOv3 network; transfer learning

OCIS codes 010.0280; 040.1880

收稿日期：2020-09-25；**修回日期**：2020-11-16；**录用日期**：2020-12-14

基金项目：长安大学中央高校基本科研业务费(300102269205，300102269304)、国家重点研发计划(2018YFC1504805，2019YFC1509201)

通信作者：*1049730716@qq.com

1 引　　言

自动目标检测是对图像中的感兴趣目标进行自动定位，它是图像识别任务的重要研究内容，对于遥感图像的语义检索、军事目标识别等具有重要意义[1]。早期的目标检测方法[2-4]通常需要事先人工提取特征，再利用提取的特征对图像中的目标进行检测和分类。这类方法将特征提取与目标检测或分类逐步实现，检测精度不高且智能化程度低。

卷积神经网络（CNN）等深度学习理论为智能目标检测及分类等任务提供了有效的解决思路。代表性的目标检测深度网络如 RCNN[5]、SPP-Net[6]、YOLO[7]，其实质是利用已构建的深度网络对输入的图片数据进行自动特征提取，并输出各类目标的位置及其类别，实现端到端（end-to-end）的目标检测[8]。与传统方法相比，这类方法极大地提高了目标检测的精度和智能化水平。其中，YOLOv3 网络[9]凭借其检测速度快以及多尺度的目标检测，得到广泛应用，如文献[10]。但不同目标及场景，其效果不尽相同，因此，近年来，鞠默然等[11]针对不同类目标检测任务对传统的 YOLOv3 网络进行了改进和优化。

高空间分辨率遥感影像中常存在“同物异谱、异物同谱”现象，其中车、行人、树等城市背景会对感兴趣目标的检测产生较大干扰，在这样复杂场景下，空间尺度差异大、结构复杂、形状变化多样的目标检测难度加大。因此，李珣等[12]基于 YOLOv2 网络根据实际道路上的环境变化提出对多种车型进行多目标识别的方法。将 YOLOv3 网络推广应用至高分辨率遥感影像的多目标检测任务具有重要意义。候笑晗等[13]将深度学习的方法用于合成孔径雷达（SAR）图像舰船目标的检测；张家强等[14]基于深度残差全卷积网络（即将 UNet 和 ResNet 网络进行结合），对 Landsat 8 遥感影像进行云的检测。但是，一方面，航空航天遥感影像分辨率相对于自然图像低，且目标背景复杂，干扰较多，导致使用原始的 YOLOv3 网络目标检测精度不高[15]；另一方面，传统的 YOLOv3 模型通常利用 COCO、ImageNet 图像数据集进行预训练，这些数据集中的目标与测试集中的目标场景及特征差异较大，这也降低了 YOLOv3 网络目标检测的精度。因此，急需改进或优化传统的 YOLOv3 网络，以改善其在复杂场景下高分辨率遥感影像多目标检测的精度。在此背景下，本文针对传统 YOLOv3 网络[9]训练过程的优化方法展开了研究。

2 本文方法的理论基础

2.1 YOLO 模型目标检测基本原理

YOLO 模型目标检测的基本原理是将输入图像划分成一定大小的格网，然后检测每个格网中所含的目标，并预测目标边界框位置、大小、定位置信度以及所有类别概率向量；对每个边界框的类别进行预测，计算每个预测边界框与其实际边界框的交并比（IoU），通过非极大值抑制，选出置信度最优的预选框，抑制冗余预选框，最后检测各个目标的边界框，如图 1 所示。

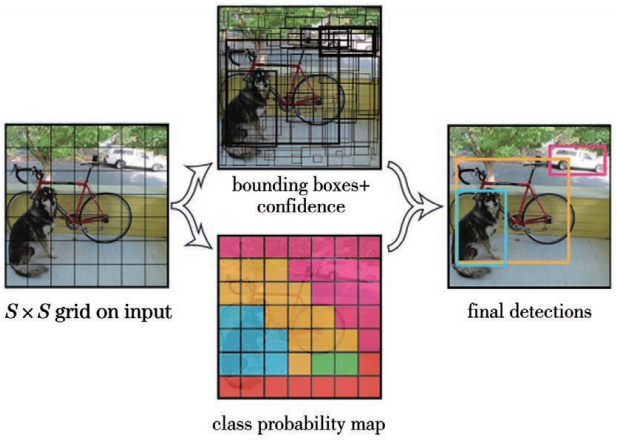

图 1　YOLO 模型目标检测的原理

Fig. 1　Principle of YOLO model for object detection

在多尺度目标检测中，YOLOv2 网络借鉴了快速的基于区域的 CNN 网络（Faster-RCNN）中锚点框（Anchor boxes）的思想，使用 K 均值（K-means）聚类算法对训练集的边界框聚类，找到更加合适的锚点框以实现更好的目标定位，而且可以检测到不同尺度的目标。为了进一步提高检测、分类精度和速度，国内外学者又对 YOLOv2 网络进行进一步的改进，提出了 YOLOv3 模型。该模型采用 Darknet-53 后端网络，类别预测改为 Logistic 函数，可实现多标签的目标检测。基于 YOLOv3 模型进行目标检测时，每一种尺度预测 3 个框，锚点框的设计仍使用聚类方式[16]，通常选取 9 个聚类中心，并按照大小均分给 3 种基本尺度。

2.2 训练优化的 YOLOv3 模型

在传统的 YOLOv3 模型[9]目标检测中，通过对图像数据集的 K-means 聚类来确定锚点框，以预测目标边界框位置。因此，锚点坐标对目标边界框准确性有较大的影响。传统的 YOLOv3 模型是对

COCO、ImageNet 数据集进行 K-means 聚类来确定锚点，但是，这些数据集庞大，目标类别过多，且这些数据集中的目标与待检测目标场景及特征差异较大，影响了模型训练的准确性，同时由于干扰图像数据多，有效图像数据少，降低了模型的泛化能力和训练速度。而现有的与目标域目标场景相似的遥感图像目标检测数据集中，可用的训练数据较少，因此，本文基于迁移学习的思想（源域信息向目标域迁移时，源域信息与目标域特征更相似，迁移效果越好，越有利于目标检测精度的提高）以及前人研究成果[17]，提出了一种 YOLOv3 网络训练过程优化的方法。具体如下：

在对模型训练之前，首先从公开的目标检测数据集中选取包括 3 类感兴趣目标的图像子集作为初始数据集，并对图像进行旋转、随机平移、图像翻转、改变图像的色调和饱和度等增广操作，构建了本文用于预训练的增广数据集（MD 数据集）；进而，通过对 MD 数据集进行 K-means 聚类获得每个目标的锚点位置，预测出更接近于目标的真实边界框，以用于后续的目标检测。

上述 YOLOv3 模型训练过程优化的流程如图 2 所示。

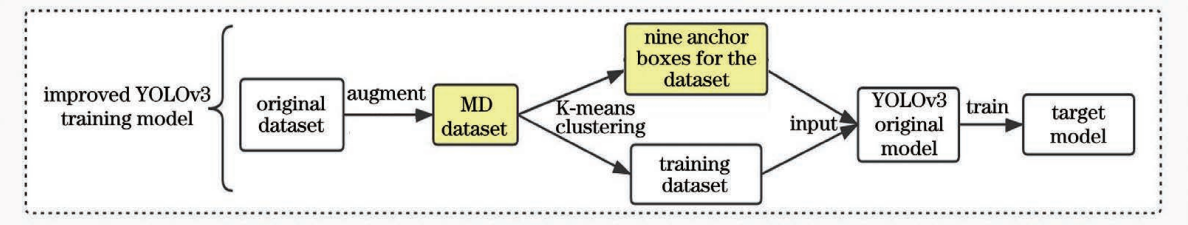

图 2　优化后的 YOLOv3 模型训练流程图

Fig. 2　Our optimized YOLOv3 model training flowchart

3　实验验证

3.1　实验数据准备

从公开的遥感图像目标检测的 RSOD 数据集（参见网址 https://github.com/RSIA-LIESMARS-WHU/RSOD-Dataset）中选取包含飞机（Aircraft）、运动场（Playground）、立交桥（Overpass）3 类目标（如图 3 所示）的图像子集。其中，飞机图像子集有 466 张，共包含 4993 架飞机，运动场的图像子集有 189 张，包含 191 个运动场，立交桥的图像子集有 176 张，包含 180 座立交桥。这些目标图像大都是从多角度拍摄得到的，其色调及空间分辨率有较大差异。通过各类目标特征学习，可加强模型的区分能力，提高模型的泛化能力，有利于提高目标检测及分类的精度。

图 3　RSOD 数据集所包含的 3 种目标图像

Fig. 3　Three objects contained in the RSOD dataset

对选取的样本图像进行了图像旋转、随机平移、图像翻转、改变图像的色调和饱和度等操作，得到增广数据集，如图 4 所示。经过上述策略对数据集进行增广，数据集更加复杂，训练出来的模型鲁棒性更好，能够解决遥感影像中目标复杂多变的情况，改善了检测的实际效果。

使用图像标注软件（Labeling）对增广数据集进行标注，记录图像的类标记以及空间位置，即矩形框中心坐标（x，y），矩形框的长 h 和宽 w 以及矩形框内物体的类别 c。图像标记样例如图 5 所示。

3.2　实验验证与结果分析

基于 Tensorflow 平台、python3.6 语言环境，以及 GPU 为 Tesla V100 GPU，CPU 为 Intel Xeon

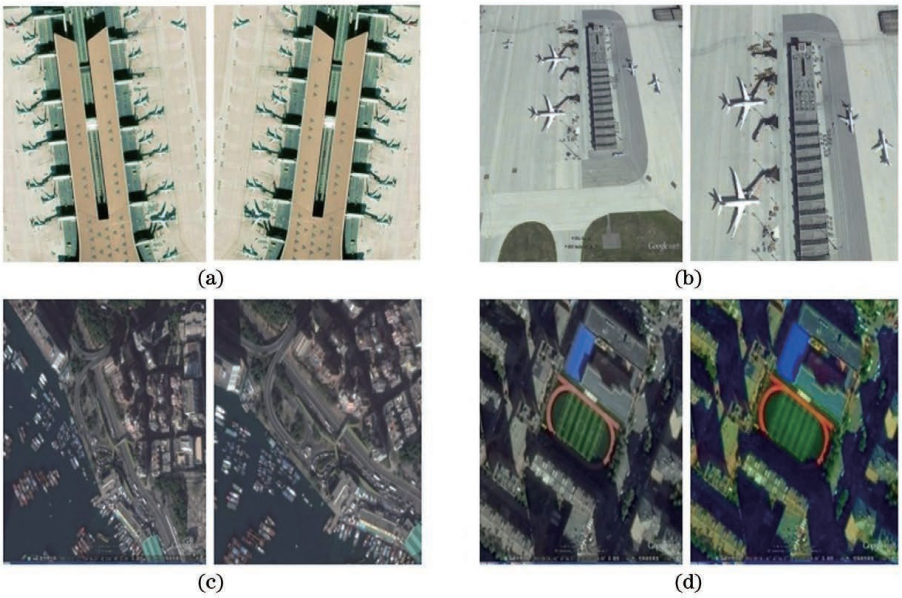

图 4　RSOD 数据集的增广处理。(a)沿 X 轴翻转;(b)图像裁剪;(c)图片旋转;(d)图像饱和度调整

Fig. 4　RSOD dataset augmentation processing. (a)Image flipping along X axis; (b)image cropping; (c) image rotation; (d) image saturation adjustment

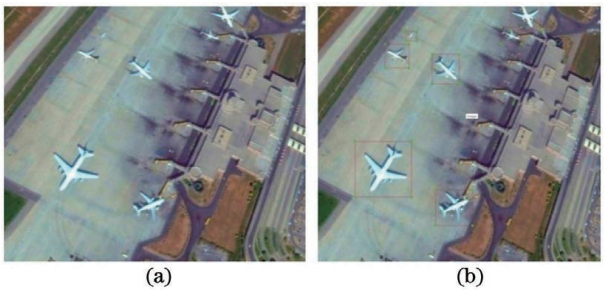

图 5　基于 Labeling 软件的数据集的标记样例。(a)原始图像;(b)标记图像(目标真实边界框)

Fig. 5　Labeling sample of dataset using Labeling software. (a) Original image; (b) labeled image (true object bounding box)

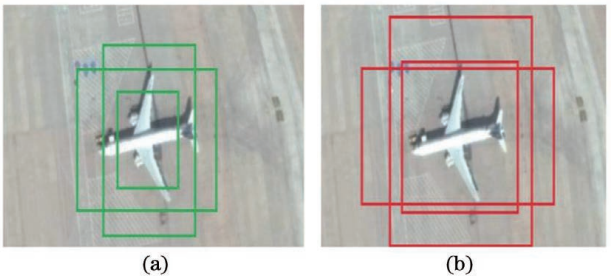

图 6　优化前后生成的锚点对应的三种基本尺寸的预选框对比。(a)传统 YOLOv3 模型预设锚点框;(b)基于聚类增广数据集预测的锚点框

Fig. 6　Comparison of the three basic sizes of anchor boxes before and after our training optimization. (a) Anchor boxes from traditional YOLOv3 model; (b) anchor boxes from our clustered augmented dataset

Gold 5215 的硬件环境开展本文实验。实验所用的训练数据集中飞机(Aircraft)样本 475 个,立交桥(Overpass)样本 230 个,运动场(Playground)样本 210 个,测试集中每一类目标图像有 200 张。

为了验证本文优化 YOLOv3 模型的有效性,开展了优化前、后生成的锚点框对比分析,如图 6 所示。从图中可以看出:基于 COCO 数据集聚类的原始预设的锚点框与目标实际边界框偏差较大,会导致标记的目标不完整或不能正确识别感兴趣目标的问题,而基于本文自定义的增广数据集聚类优化后的锚点框与目标实际边界框更接近,包含了整个被检测物体,这表明本文预测的锚点框位置更接近于目标真实边界。分析其原因:本文通过增广数据产生了与目标域中目标特征更相近且更具代表性的数据集,对该数据集进行网络预训练,优化了预训练模型的预测效果。

使用上述方法实现了数据增广,并经过 K-means 聚类得到 9 个锚点框,为(21,18)、(29,29)、(40,34)、(48,44)、(61,54)、(78,70)、(106,95)、(263,390)、(443,406),其与利用 COCO 数据集预设的锚点框相比有较大改善。其次,对飞机、立交桥和运动场 3 类目标预测框的面积与通常实际目标大小进行了定性对比,结果表明:上述利用本文产生的增广数据集进行训练过程优化的方法的确有利于改善目标锚点框的预测结果。

进而，分别基于原始预设锚点框和基于增广数据集聚类优化后的锚点框对模型进行预训练，在迭代 120 次，交并比（IoU）设为 0.7 时，优化前、后的 YOLOv3 模型部分检测结果如图 7 所示。

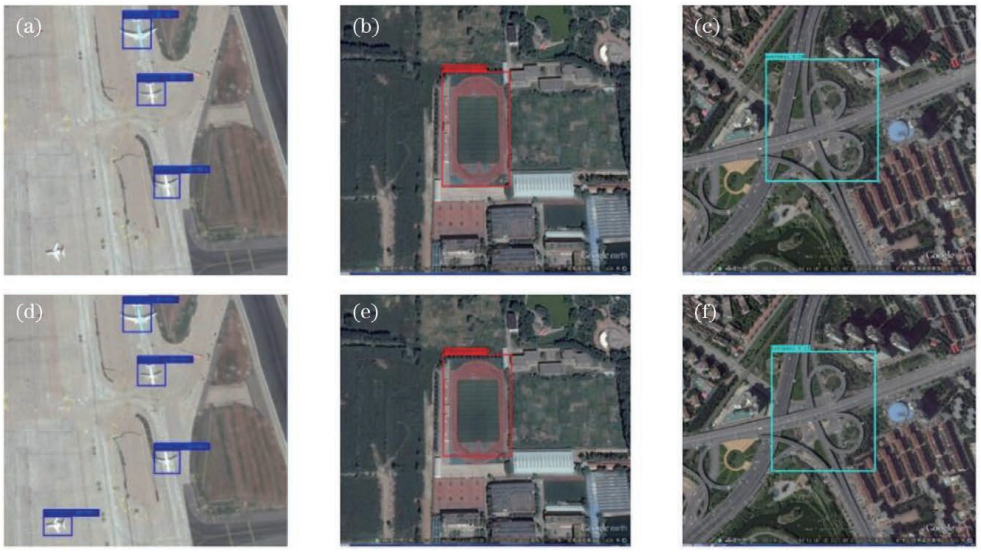

图 7　传统 YOLOv3 模型与本文优化的 YOLOv3 模型对飞机（左）、运动场（中）、立交桥（右）检测结果的对比。（a）～（c）传统 YOLOv3 模型；（d）～（f）本文优化 YOLOv3 模型

Fig. 7　Comparison of the detection results between the traditional and our optimized YOLOv3 models. (a)-(c) Traditional YOLOv3 model; (d)-(f) our optimized YOLOv3 model

上述实验结果表明：利用本文优化后的 YOLOv3 网络模型检测出了传统 YOLOv3 网络模型未检测出来的飞机目标，而运动场和立交桥这两类目标的检测结果差异不大，这是由于预选框的位置和置信度仅有微小变化。表 1 给出优化前、后 YOLOv3 模型目标检测的结果。

表 1　优化前、后 YOLOv3 模型目标检测正确/错误样本

Table 1　YOLOv3 model object detection result before and after optimization

Scene	Detection number of traditional model			Detection number of optimized model		
	Aircraft	Overpass	Playground	Aircraft	Overpass	Playground
Aircraft	350	7	5	390	7	6
Overpass	16	170	4	18	172	3
Playground	14	10	160	15	11	162

从表 1 可以看出，飞机的检测正确样本数增加较多，而其他两种目标检测正确样本数增加较少，分析其原因：飞机训练样本较多，而运动场及立交桥这两类目标训练样本较少，且后者属于复合目标，结构相对复杂，特别是立交桥。

为了进一步评价本文训练优化后模型的检测精度，选取常用的精确率（Precision）、召回率（Recall）、平均精度（Average Precision，AP）、平均精度均值（mean Average Precision，mAP）指标来评价。其中，精确率和召回率的表达式为

$$P=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}},\tag{1}$$

$$R=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}},\tag{2}$$

式中：N_{TP} 表示是正类而被预测成为正类的实例个数；N_{FP} 表示是负类但是被预测成为正类的实例个数；N_{FN} 表示是正类但是被预测成为负类的实例个数。第 c 个类别的平均精度 $P_{\mathrm{A},c}$ 是以该类目标的精确率和召回率作为纵、横坐标，与 Precision-Recall 曲线 $P_c(R)$ 围成的面积，实际计算时，取不同 Recall 值下最高 Precision 值的均值，表达式为

$$P_{\mathrm{A},c}=\int_0^1 P_c(R)\mathrm{d}R。\tag{3}$$

平均精度均值 mAP 即为所有类别 AP 值的均值，用来衡量模型在所有类别上训练效果的好坏，其计算公式为

$$P_{\mathrm{mA}}=\frac{\sum_{c=1}^{C}P_{\mathrm{A},c}}{C},\tag{4}$$

式中，C 表示总的类别数。根据(1)～(4)式，分别算出优化前后的 YOLOv3 模型目标分类及检测的精度，如表 2 所示。

表 2　优化前、后 YOLOv3 模型目标检测结果评价

Table 2　Evaluation results using the traditional and optimized YOLOv3 models

Parameter	Traditional YOLOv3 model			Our optimized YOLOv3 model		
	Aircraft	Overpass	Playground	Aircraft	Overpass	Playground
Precision	0.921	0.909	0.947	**0.924**	**0.907**	**0.947**
Recall	0.737	0.739	0.762	**0.821**	**0.748**	**0.771**
$P_{\mathrm{A},c}$	0.819	0.815	0.844	**0.869**	**0.820**	**0.850**
mAP		0.826			**0.846**	

由表 2 可知，由于立交桥和运动场目标空间尺度较大，使用经过聚类后的锚点框对其检测结果的影响不大，因此，Recall 值和 Precision 值没有明显增加。但飞机类图像所含的细小目标较多，使用预设锚点框可能存在漏检的情况，所以 Recall 值较低。而使用经过聚类的锚点框后，Recall 显著增加，从 73.7% 增加到 82.1%，而 mAP 提高了 2 个百分点。

此外，从表 2 可以看出，飞机和运动场的检测精度相对较高，而立交桥这类较大的复合目标，结构复杂，本文方法对其检测精度改善不明显。因此，本文从 DIOR 数据集[18]中增选了部分立交桥图像数据参与训练。结果表明：立交桥图像训练数据增选后，本文方法对 3 类目标的检测平均精度均值高于 2%。

4　结　　论

传统的 YOLOv3 模型通常利用 COCO、ImageNet 等与测试集目标场景特征差异较大的数据集进行训练，易导致对高分辨率遥感影像中复杂场景目标检测精度不高的问题。为解决这一问题，本文以 YOLOv3 网络为基础，在迁移学习的思想框架下，利用与目标域更相似的增广数据集对传统 YOLOv3 网络的训练过程进行了优化，从而使预测的目标锚点框位置与真实位置更接近。最后，利用其他的 RSOD 训练集对 YOLOv3 预训练模型进行再训练。对源于 RSOD、DIOR 数据集中的飞机、立交桥及运动场 3 类目标的图像数据集进行测试，结果表明：与传统的 YOLOv3 模型相比，优化后锚点框的 YOLOv3 模型用于目标检测时的召回率都有所提高，其中飞机目标的召回率提升显著，从 73.7%提高到 82.1%，3 类目标的 mAP 值也增加了 2 个百分点以上。这表明：本文训练过程的优化方法改善了 YOLOv3 模型的训练效果，提高了模型的泛化能力。

参　考　文　献

[1] Cheng G, Han J W. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117: 11-28.

[2] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 20-25, 2005, San Diego, CA, USA. New York: IEEE Press, 2005: 886-893.

[3] Viola P, Jones M. Robust real-time object detection[J]. International Journal of Computer Vision, 2001, 57(2): 34-47.

[4] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.

[5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.

[6] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[M]// Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8691:

346-361.

[7] 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), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[8] Ward I R, Laga H, Bennamoun M. RGB-D image-based object detection: from traditional methods to deep learning techniques[M]//Rosin P L, Lai Y K, Shao L, et al. RGB-D image analysis and processing. Advances in computer vision and pattern recognition. Cham: Springer, 2019: 169-201.

[9] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2020-09-20]. https://arxiv.org/abs/1804.02767.

[10] Wang D W, He Y H, Li D X, et al. An improved YOLOv3 infrared video image pedestrian detection algorithm[J]. Journal of Xi'an University of Posts and Telecommunications, 2018, 23(4): 48-52, 67.
王殿伟, 何衍辉, 李大湘, 等. 改进的 YOLOv3 红外视频图像行人检测算法[J]. 西安邮电大学学报, 2018, 23(4): 48-52, 67.

[11] Ju M R, Luo H B, Wang Z B, et al. Improved YOLO V3 algorithm and its application in small target detection[J]. Acta Optica Sinica, 2019, 39(7): 0715004.
鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(7): 0715004.

[12] Li X, Shi B B, Liu Y, et al. Multi-target recognition method based on improved YOLOv2 model[J]. Laser & Optoelectronics Progress, 2020, 57(10): 101010.
李珣, 时斌斌, 刘洋, 等. 基于改进 YOLOv2 模型的多目标识别方法[J]. 激光与光电子学进展, 2020, 57(10): 101010.

[13] Hou X H, Jin G D, Tan L N. Survey of ship target detection in SAR images based on deep learning[J]. Laser & Optoelectronics Progress, 2021, 58(4): 0400005.
侯笑晗, 金国栋, 谭力宁. 基于深度学习的 SAR 图像舰船目标检测综述[J]. 激光与光电子学进展, 2021, 58(4): 0400005.

[14] Zhang J Q, Li X Y, Li L Y, et al. Landsat 8 remote sensing image based on deep residual fully convolutional network[J]. Laser & Optoelectronics Progress, 2020, 57(10): 102801.
张家强, 李潇雁, 李丽圆, 等. 基于深度残差全卷积网络的 Landsat 8 遥感影像云检测方法[J]. 激光与光电子学进展, 2020, 57(10): 102801.

[15] Cheng G, Han J W. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117: 11-28.

[16] Zhang S J, Zhao H C. Algorithm research of optimal cluster number and initial cluster center[J]. Application Research of Computers, 2017, 34(6): 1617-1620.
张素洁, 赵怀慈. 最优聚类个数和初始聚类中心点选取算法研究[J]. 计算机应用研究, 2017, 34(6): 1617-1620.

[17] Zhang L, Zhang Y S, Yu Y, et al. Research on data augmentation for object detection of remote sensing image[J]. Journal of Geomatics Science and Technology, 2019, 36(5): 505-510.
张磊, 张永生, 于英, 等. 遥感图像目标检测的数据增广研究[J]. 测绘科学技术学报, 2019, 36(5): 505-510.

[18] Li K, Wan G, Cheng G, et al. Object detection in optical remote sensing images: a survey and a new benchmark[EB/OL]. (2019-08-31)[2020-09-20]. https://arxiv.org/abs/1909.00133v2.