引用格式:严开忠,马国梁,许立松,等. 基于改进YOLOv3的机载平台目标检测算法[J]. 电光与控制,2021,28(5):70-74. YAN K Z, MA G L, XU L S, et al. Improved YOLOv3 based target detection algorithm for airborne platform[J]. Electronics Optics & Control, 2021, 28(5):70-74.

基于改进YOLOv3的机载平台目标检测算法

严开忠, 马国梁, 许立松, 尚海鹏, 于 睿
(南京理工大学,南京 210094)

摘 要: 针对小型智能侦察无人机机载平台存在的计算力受限、检测速度较慢的问题,提出了一种基于YOLOv3改进的目标检测算法。首先引入深度可分离卷积改进YOLOv3的骨干网络,降低网络的参数和计算量,提高算法的检测速度,再根据机载视角下目标形状的特点,预置K-means产生先验框的初始聚类中心,并在边框回归中引入CIoU损失函数,将DIoU与NMS结合,改善YOLOv3对密集目标的漏检问题,最后再通过TensorRT优化加速后部署到英伟达Jetson TX2机载计算平台。实验结果表明,所改进的算法在验证集上的平均精度均值(MAP)达到了82%,检测速度从3.4帧/s提升到16帧/s,满足实时性要求。

关键词: 目标检测; 侦察无人机; YOLOv3; 深度可分离卷积; DIoU; TensorRT
中图分类号: V247 **文献标志码**: A **doi**:10.3969/j.issn.1671-637X.2021.05.016

Improved YOLOv3 Based Target Detection Algorithm for Airborne Platform

YAN Kaizhong, MA Guoliang, XU Lisong, SHANG Haipeng, YU Rui
(Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: To overcome the problems of limited calculation power and slow detection speed of the small intelligent reconnaissance UAV platforms, an improved target detection algorithm based on YOLOv3 is proposed. First of all, depthwise separable convolution is introduced to improve the backbone network of YOLOv3, which greatly reduces the quantity of parameters and calculation cost of the network, and improves the detection speed of the algorithm. Then, according to the characteristics of the target shape under the perspective of the airborne platform, the initial clustering center of K-means is preset when generating prior box, and CIoU loss function is introduced in the box regression. DIoU is combined with NMS to reduce the missed detections for dense targets. Finally, the improved model is optimized and speeded up by TensorRT, and deployed to the NVIDIA Jetson TX2 airborne computing platform. The experimental results show that the Mean Average Precision (MAP) of the improved algorithm on the verification set reaches 82%, and the detection speed is increased from 3.4 to 16 frames, which can meet the real-time requirements.

Key words: target detection; reconnaissance UAV; YOLOv3; depthwise separable convolution; DIoU; TensorRT

0 引言

随着人工智能技术的快速发展,武器装备系统的智能化成为军事领域研究的热点。由于小型无人机具有成本低廉、易于部署等诸多优点,其在军事侦察、治安监控、交通指挥等各个领域均有广泛应用,而针对小型无人机载平台目标检测技术的研究也受到越来越多的关注。传统的目标检测算法通常是采用滑动窗口对图像遍历得到候选区域,然后用基于手工设计的特征提取器(如Haar[1],SIFT[2],HOG[3]等)提取候选区域的特征,最后再用SVM[4],AdaBoost[5]将特征分类。由于无人机平台计算力有限,机载视角下目标的特征信息较少,传统的检测方法效果并不理想。基于深度学习的目标检测算法主要分为two stage和one stage两类:two stage类算法是通过算法生成一系列候选区域,再用卷积神经网络提取特征进行分类识别,比较典型的算法有R-CNN[6],以及改进后的Fast R-CNN[7]和Faster R-CNN[8],这类算

收稿日期:2020-05-15 修回日期:2021-04-20
基金项目:国家自然科学基金青年基金(11302106)
作者简介:严开忠(1995—),男,贵州毕节人,硕士生,研究方向为图像处理,目标检测与目标跟踪,1935458275@qq.com。

法检测速度太慢，无法满足实时性的要求；one stage 类算法则不需要先提取候选区域，速度得到极大提升，典型的算法有 YOLO[9] 和 SSD[10] 等。YOLOv3 是 YOLO 系列效果最好的一个版本，精度高、速度快，但是在无人机这种计算力受限的嵌入式平台上，原始的 YOLOv3 无法做到实时运行。

针对以上问题，本文在 YOLOv3 的基础上引入深度可分离卷积[11] 来降低网络的参数和计算量，预置 K-means 产生 anchor 的初始聚类中心，并在边框回归中引入 CIoU[12] 损失函数，将 DIoU[12] 与 NMS[13] 结合，最后再通过 TensorRT 优化加速后部署。实验结果表明，改进算法满足实用性要求。

1　YOLOv3 算法的框架结构

YOLO 直接把输入图像分成 $N \times N$ 个网格（grid），每个网格完成检测识别后，再利用边框回归得出更加逼近于真实框的预测框。YOLOv3 的结构如图 1 所示，其中，n 表示 DBL 和 Res_unit 组件的个数。

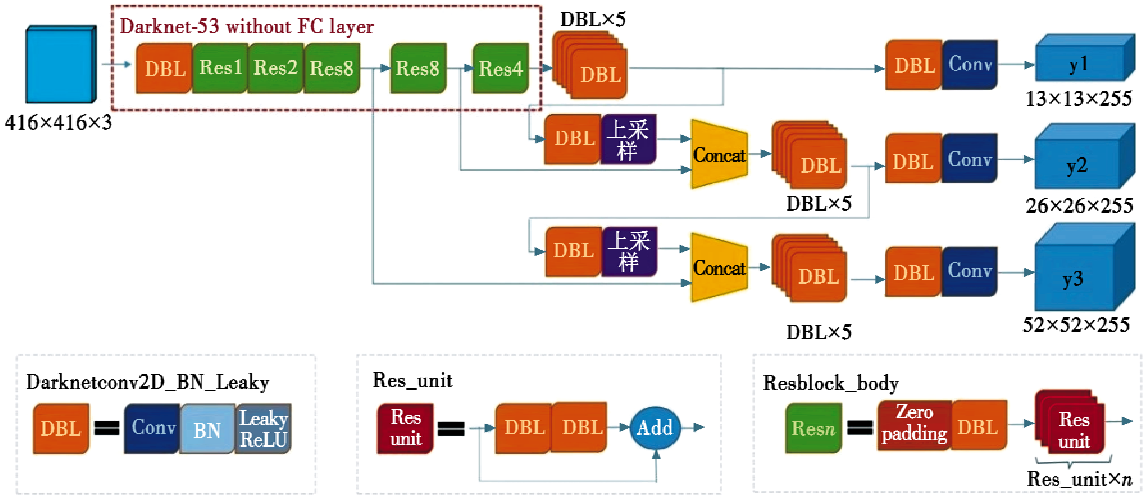


图 1　YOLOv3 检测算法结构图

Fig. 1　Structure of YOLOv3 algorithm

YOLOv3 借鉴了 Faster R-CNN 中的锚点机制（anchor），并引入了特征金字塔网络（FPN）结构，分别在 13×13，26×26，52×52 这 3 个特征尺度上做检测，大幅改善了 YOLO 系列算法对小目标的检测效果。

由图 1 可知，YOLOv3 采用了性能更强的特征提取网络 Darknet-53，由于引入了残差结构，Darknet-53 将网络加深到了 53 层，特征提取能力进一步提升。但是，由于 Darknet-53 的网络结构过于复杂，使得 YOLOv3 的检测速度进一步下降，在无人机等计算力受限的边缘设备上无法实时运行。

2　机载平台目标检测算法的改进

2.1　深度可分离卷积

2017 年，Google 在 MobileNet 中首次提出了深度可分离卷积（Depthwise Separable Convolution，DSC）的概念。深度可分离卷积在几乎不影响精度的情况下大幅降低了网络的计算量。在常规卷积中，卷积核的通道数和卷积图像的通道总是保持一致，而深度可分离卷积是将常规卷积分为深度卷积（Depthwise Convolution）和逐点卷积（Pointwise Convolution）两步，结构如图 2 所示。

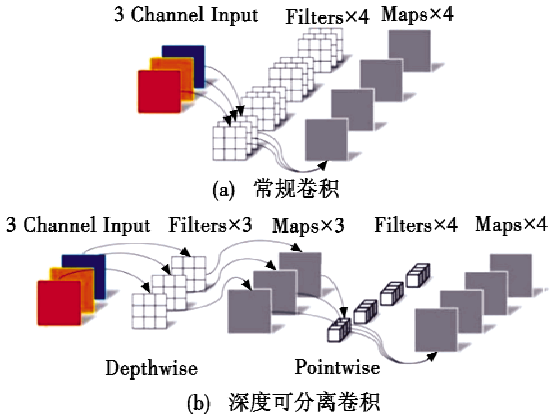


图 2　常规卷积和深度可分离卷积的结构

Fig. 2　Structure of standard convolution and DSC

由图 2 可知，对于一个输入尺寸为 $[H, W, C]$ 的特征图，如果用 kernel size 为 Z（Z 为奇数）的卷积核，stride 为 1，padding 为 0，输出尺寸为 $[N, M, K]$，则

$$N = H - Z + 1 \tag{1}$$

$$M = W - Z + 1 \tag{2}$$

常规卷积需要的乘法次数为

$$N_{sc} = K \times Z \times Z \times C \times N \times M \tag{3}$$

深度卷积需要的乘法次数为

$$N_{DC} = C \times Z \times Z \times 1 \times N \times M \tag{4}$$

逐点卷积需要的乘法次数为

$$N_{PC} = K \times 1 \times 1 \times C \times N \times M \tag{5}$$

所以深度分离卷积总共需要的乘法次数为

$$N_{DSC} = (Z \times Z + K) \times C \times N \times M \tag{6}$$

代价比为

$$\frac{N_{DSC}}{N_{SC}} = \frac{1}{K} + \frac{1}{Z^2} \tag{7}$$

通常情况下,K 值比较大,当 Z 取常用的 3 * 3 卷积时,常规卷积的计算量是深度可分离卷积的 8 ~ 9 倍。

本文中,将 YOLOv3 骨干网络中的常规卷积替换为深度可分离卷积,如图 3 所示。

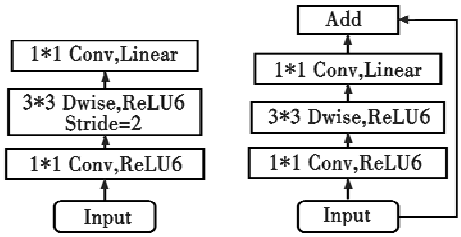


图 3 改进网络中的深度可分离卷积模块

Fig. 3 DSC module in the improved network

与 YOLOv3 网络中的常规卷积相比,深度可分离卷积的优势更明显,在改进的 YOLOv3 中,网络参数大幅减少,计算量大幅降低,检测速度进一步加快。

2.2 预置先验框初始聚类中心

YOLOv3 中的先验框(anchor)是通过对训练集进行 K-means 聚类得到的,K-means 聚类的效果会受到初始聚类中心的影响,K-means ++[14] 对初始值的选取进行了一定的限制,但是,由于无人机视角的特殊性,上述聚类方法产生的 anchor 效果不够明显。本文提出了预置初始聚类中心的方法,通过对机载平台使用场景下目标的特点进行分析,手动选定 K-means 的 9 个初始聚类框。测试结果显示,改进的聚类算法效果更好。表 1 所示为改进的 K-means 聚类算法在本文数据集上的平均 IoU(Avg IoU)大小对比。

表 1 不同聚类算法的 Avg IoU 对比

Table 1 Avg IoU of different clustering methods

聚类算法	聚类中心数量	Avg IoU
K-means	9	0.72
K-means ++	9	0.74
改进的 K-means	9	0.75

由表 1 可知,改进的聚类算法 Avg IoU 更高、效果更好。

2.3 更稳定的边框损失

在 YOLOv3 中采用均方误差(Mean Square Error, MSE)损失函数进行边框回归,边框损失函数为

$$L_{box} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} (2 - w_i \times h_i) [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2] \tag{8}$$

式中:λ_{coord} 表示坐标损失权重;I_{ij}^{obj} 表示第 i 个单元格的第 j 个 box 是否负责预测该目标(1 或 0);$2 - w_i \times h_i$ 是 YOLOv3 为了提升小目标检测乘上的比例系数;(x_i, y_i), w_i, l_i 分别表示预测框的中心点坐标和宽、高;$(\hat{x}_i, \hat{y}_i), \hat{w}_i, \hat{l}_i$ 分别表示真实目标框的中心点坐标和宽、高。

MSE 损失函数无法准确描述边框之间的交并比(IoU)关系,而且对目标框尺度比较敏感,不具有尺度不变性,为了更好地衡量预测框和目标框的重叠关系,常使用 P_{IoU} 损失来进行边框回归,P_{IoU} 的定义为

$$P_{IoU} = \frac{B \cap B^{gt}}{B \cup B^{gt}} \tag{9}$$

式中:B 为预测框;B^{gt} 为真实框。用 L_{IoU} 表示 IoU 损失函数,定义为

$$L_{IoU} = 1 - \frac{B \cap B^{gt}}{B \cup B^{gt}} \tag{10}$$

但是,L_{IoU} 只有在预测框和真实框之间有重叠时才起作用,无重叠的边框根本无法回归。文献[15]中改进后的 Generalized IoU(GIoU)仍然严重依赖 IoU,误差比较大,很不稳定,GIoU 的损失定义为

$$L_{GIoU} = 1 - P_{IoU} + \frac{|C - B \cup B^{gt}|}{|C|} \tag{11}$$

式中,C 表示能同时包含 B 和 B^{gt} 的最小矩形。而 DIoU 将目标框与预测框之间的距离、重叠率以及尺度都考虑进去,使得边框回归变得更加稳定,在与目标框不重叠时仍然可以提供有效的收敛方向。L_{DIoU} 定义为

$$L_{DIoU} = 1 - P_{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} \tag{12}$$

式中:b, b^{gt} 分别代表预测框和目标框的中心点;ρ 表示两个中心点之间的欧氏距离;c 表示能同时覆盖 anchor 和目标框的最小矩形的对角线。

为使与目标框有重叠时边框回归更有效,本文在 YOLOv3 的边框回归的损失中引入 Complete-IoU 损失(CIoU Loss),CIoU 在 DIoU 的基础上再把边界框纵横比考虑进去,CIoU Loss 定义为

$$L_{CIoU} = 1 - P_{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\nu \tag{13}$$

式中:α 为用于平衡比例的参数;ν 用来衡量 anchor 框和目标框之间的比例一致性,即

$$\nu = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{14}$$

$$\alpha = \frac{\nu}{(1 - P_{IoU}) + \nu} \tag{15}$$

由式(13) ~ 式(15)可以看到,L_{CIoU} 更加倾向于往重叠区

域增多的方向优化，有利于目标的定位。

2.4 非极大值抑制的改进

YOLOv3 中使用非极大值抑制(NMS)算法来消除多余的边框，NMS 算法的具体步骤如下所述。

1）选取当前类别中得分(scores)最大 s_i 的边框，记为 M；

2）计算其他边框 B_i 和 M 的 $IoU(M,B_i)$，若 $IoU(M,B_i)$ 大于设定的阈值 ε，就舍弃这些边框；

3）从剩余的边框中再找出 scores 最大的一个，重复步骤 2)；

4）如此循环，直到完成所有类别的 NMS 计算。

NMS 将 $IoU(M,B_i)$ 大于阈值的边界框 scores 置为零，即

$$s_i=\begin{cases}s_i & IoU(M,B_i)<\varepsilon\\ 0 & IoU(M,B_i)\geqslant\varepsilon\end{cases} \tag{16}$$

NMS 算法对冗余的边界框有显著的抑制作用，但会粗略地将目标重叠率比较高的边界框删除，使得该算法对密集目标的检测效果并不理想。针对此问题，本文在 NMS 计算中引入 DIoU，DIoU-NMS 定义为

$$s_i=\begin{cases}s_i & P_{\mathrm{IoU}}-R_{\mathrm{DIoU}}(M,B_i)<\varepsilon\\ 0 & P_{\mathrm{IoU}}-R_{\mathrm{DIoU}}(M,B_i)\geqslant\varepsilon\end{cases} \tag{17}$$

$$R_{\mathrm{DIoU}}(M,B_i)=\frac{\rho^2(M,B_i)}{c^2} \tag{18}$$

式中：ρ 表示 M,B_i 两个中心点之间的欧氏距离；c 表示能同时覆盖 M,B_i 的最小矩形的对角线。改进后的 DIoU-NMS 能充分判断重叠率较高的两个边界框是否属于同一个目标，从而有效地进行边框抑制，此外，由于机载平台视角下目标重叠率比自然视角下要低，所以本文在改进的 NMS 中使用更小的阈值 ε，进一步降低算法的漏检率。

3 实验及结果分析

3.1 模型训练与部署

为验证本文所改进目标检测算法的效果，收集小型四旋翼无人机所采集的图片共 5000 幅，通过裁剪、翻转、变化颜色通道等操作将数据集扩张到 7600 幅，并选取其中的 15% 作为测试集。网络采用分批次训练的策略，初始学习率为 0.001，衰减系数为 0.0005，开启多尺度训练，网络迭代了 40000 次以后趋于稳定，最终 Loss 值稳定在 0.8 左右。

TensorRT 是一个高性能的深度学习推理(Inference)优化器，可以为深度学习应用提供低延迟、高吞吐率的部署推理，广泛应用于嵌入式平台或自动驾驶平台。TensorRT 能够支持 TensorFlow，Caffe，Mxnet，Pytorch 等几乎所有的深度学习框架进行快速和高效的部署推理。TensorRT 主要是对训练好的模型进行优化，用于推理阶段的加速。将训练好的模型转换为 ONNX 格式，并使用 TensorRT 中的 ONNX 解析器解析模型并构建 TensorRT 引擎。

3.2 算法实验结果分析

表 2 所示为本文改进后的网络模型和原始的 YOLOv3 在参数量、模型大小、浮点计算量 3 个指标的对比。

表 2　改进的网络模型参数对比

Table 2　Parameters of the improved network model

网络模型	参数量/M	模型大小/MiB	浮点计算量
YOLOv3 tiny	8.7	33.1	5.4
YOLOv3	61.6	246.5	19.1
改进的模型	18	60.8	5.7

由表 2 可以看到，在 YOLOv3 中引入深度可分离卷积后，网络的计算量和参数都大幅降低，与原始的 YOLOv3 相比，浮点计算量下降了 70.2%，参数降低了 70.8%，模型也只有原来的 1/4 大小，提升效果非常明显。

为了进一步评价本文改进 YOLOv3 算法的实验结果，本文采用平均精度均值(Mean Average Precision，MAP)、帧频(Frames Per Second，FPS)2 个指标作为评判标准。改进的算法在本文测试集上的结果如表 3 所示，IoU 阈值为 0.5，测试平台为 NVIDIA Jetson TX2，本文改进的 YOLOv3 算法记为 M-YOLOv3 算法。

表 3　算法的检测结果对比

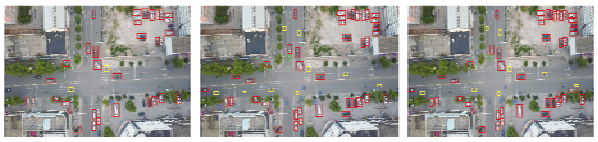
Table 3　Comparison of detection results of the algorithms

算法	平均精度均值/%	帧频
YOLOv3 tiny	68	14
YOLOv3	84	3.4
M-YOLOv3	82	6.7
M-YOLOv3 + TensorRT	82	16

由表 3 可知，原始的 YOLOv3 算法在 NVIDIA Jetson TX2 计算平台上的帧频只有 3.4 帧/s，根本无法实时运行，引入深度可分离卷积后，YOLOv3 算法在该平台上的帧频提升到 6.7 帧/s，并保持了较高的精度，达到 82%，远高于 YOLOv3 tiny 算法的 68%，改进的 YOLOv3 算法经过 TensorRT 加速后在 TX2 平台可以达到 16 帧/s，说明本文改进后的 YOLOv3 目标检测算法在机载平台下能够满足实时性的要求。图 4 所示为 YOLOv3-tiny，YOLOv3，M-YOLOv3 这 3 种不同算法对某街道路口的实际检测结果对比。

由图 4 可以看到，YOLOv3-tiny 算法的漏检率和误检率都比较高，检测结果较差，而 YOLOv3 算法对小目标的检测效果更好，但是对密集场景下的目标存在漏检的情况。由于改进的 YOLOv3 算法在边框回归中引入 CIoU Loss，并使用 DIoU-NMS 进行边框抑制，所以定位更加精确。图 5 所示为密集场景目标检测结果，图

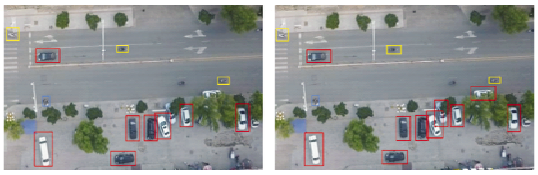
5(a)、图 5(b)分别为图 4(b)、图 4(c)的局部放大。



(a) YOLOv3-tiny (b) YOLOv3 (c) M-YOLOv3

图 4 不同算法检测结果

Fig. 4 Detection results of different algorithms



(a) YOLOv3 (b) M-YOLOv3

图 5 密集场景目标检测结果

Fig. 5 Detection results of targets in dense scene

由图 5 可以看到,在目标比较密集的场景中,原始的 YOLOv3 算法使用的 NMS 算法容易粗略地把重叠率比较高的目标边框消除,造成漏检的情况,而改进后的 DIoU-NMS 能充分判断重叠率较高的两个边界框是否属于同一个目标,一定程度上避免了重叠目标边框被过度消除的问题,在密集场景下效果更好。

4 结束语

本文深度分析了机载平台下目标检测任务中的关键问题,提出了一种基于 YOLOv3 改进的目标检测算法。通过在 YOLOv3 的骨干网络中引入深度分离卷积,降低网络的计算量和参数,再根据无人机视角下目标形状的特点,在生成先验框时预置 K-means 算法的初始聚类中心,并在边框回归中引入 CIoU 损失函数,同时将 DIoU 与 NMS 结合,改善密集场景目标的边框过度抑制问题,最后再通过 TensorRT 将 M-YOLOv3 优化加速部署到 TX2 计算平台上。实验结果表明,所改进的算法在本文数据集上的平均精度均值达到了 82%,与原始 YOLOv3 相比,改进算法网络的参数量和模型尺寸都大幅降低,在 TX2 计算平台上达到 16 帧/s,满足实时性的要求,有较高的实用价值。

参 考 文 献

[1] VIOLA P A, JONES M J, SNOW D. Detecting pedestrians using patterns of motion and appearance[C]//Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003:734-741.

[2] LOWE D G. Distinctive image features from scale-invariant key points[J]. International Journal of Computer Vision, 2004, 60(2):91-110.

[3] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005:886-893.

[4] VAPNIK V. An overview of statistical learning theory[J]. IEEE Transactions on Neural Networks, 1999, 10(5):988-999.

[5] VIOLA P A, JONES M J. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2):137-154.

[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014:580-587.

[7] GIRSHICK R. Fast R-CNN[C]//IEEE International Conference on Computer Vision (ICCV), 2015:1440-1448.

[8] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. Transactions on Pattern Analysis and Machine Intelligence, IEEE, 2016, 39(6):1137-1149.

[9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:779-788.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision, 2016:21-37.

[11] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:4510-4520.

[12] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020:12993-13000.

[13] SALSCHEIDER N O. FeatureNMS: non-maximum suppression by learning feature embeddings[J]. arXiv preprint arXiv:2002.07662, 2020.

[14] VASSILVITSKII S, ARTHUR D. K-means++: the advantages of careful seeding[C]//Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2006:1027-1035.

[15] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019:658-666.