

基于 SSD 的改进目标精定位检测算法

陈传华¹, 侯志强^{1,2}✉, 余旺盛¹, 李 军¹, 廖秀峰¹, 王皎尧¹

(1. 空军工程大学信息与导航学院, 西安, 710077; 2. 西安邮电大学, 西安, 710121)

摘要 目标检测问题是计算机视觉中的热门问题, 如何提高目标检测定位精度是检测过程中面临的一个难题。在 SSD 算法的基础上, 通过结合选择性搜索算法, 提出了一种提高检测定位精度的方法。该算法首先通过 SSD 算法框架对图像进行目标初始检测, 获得目标粗略位置和类别, 然后采用一种改进的选择性搜索算法对目标所在区域进行选择性搜索, 生成目标边界候选框, 最后采用文中提出的边界判断方法得到目标精确位置, 完成由粗到精(Coarse-to-Fine)的目标定位检测。文中算法对 PASCAL VOC2012 数据集中的 22 531 张图像进行了测试, 实验结果显示文中算法检测目标中心位置误差 7.6, 平均覆盖率 90.8%, 相比于其他算法, 中心位置误差更低, 覆盖率更高, 能提高目标检测定位精度 20%~30%。

关键词 目标检测; 目标定位; 定位精度; 选择性搜索

DOI 10.3969/j.issn.1009-3516.2018.06.012

中图分类号 TP391.4 **文献标志码** A **文章编号** 1009-3516(2018)06-0073-06

An Improved Object Precision Positioning Detection Algorithm Based on SSD

CHEN Chuanhua¹, HOU Zhiqing^{1,2}✉, YU Wangsheng¹, LI Jun¹, LIAO Xiufeng¹, WANG Jiaoyao¹

(1. Information and Navigation College, Air Force Engineering University, Xi'an 710077, China;

2. Xi'an University of Posts & Telecommunications, Xi'an 710121, China)

Abstract: Detection is one of the important tasks in image processing. How to improve the accuracy of object location is a difficult problem in the process of detection. On the basis of the SSD algorithm, this paper proposes a rough location and object class combined with selective search algorithm. Then, the paper searches the area of object by using an improved selective search algorithm to generate an object candidate frame. Finally, the paper adopts a boundary judgment method to obtain the accurate position of object and complete the detection from coarse to fine. In this paper, a lot of experiments are carried out based on 22531 images of PASCAL VOC data sets. The results show that the algorithm has a location error of 7.6 centers and a mean coverage rate of 90.8%. Compared with the other algorithms, this algorithm has lower center location error and higher coverage rate, improving the location accuracy of object detection by 20%~30%.

Key words: object detection; object location; positioning accuracy; selective search

收稿日期: 2017-12-01

基金项目: 国家自然科学基金(61473309;61703423;41601436);陕西省自然科学基金(2016JM6050)

作者简介: 陈传华(1994—),男,福建仙游人,硕士生,主要从事深度学习目标检测。E-mail:1327067049@qq.com

通信作者: 侯志强(1973—),男,陕西眉县人,教授,博士生导师,主要从事模式识别,计算机视觉,信息融合研究。E-mail:h-zhq@sohu.com

引用格式: 陈传华,侯志强,余旺盛,等.基于 SSD 的改进目标精定位检测算法[J].空军工程大学学报(自然科学版),2018,19(6):73-78.
CHEN Chuanhua, HOU Zhiqing, YU Wangsheng, et al. An Improved Object Precision Positioning Detection Algorithm Based on SSD[J].
Journal of Air Force Engineering University (Natural Science Edition), 2018, 19(6): 73-78.

随着计算机技术的不断发展,为更好地进行图像目标检测结果的后续处理,如目标分割,图像识别和视觉跟踪等^[1-3],目标检测定位精度的要求越来越高,如何提高目标检测定位的精度受到了广泛的关注^[4]。

近年来,基于深度学习的算法比传统算法在性能上有了极大的提高,深度学习方法在图像分类、自然语言处理、人脸识别和目标跟踪等方面得到了广泛应用。在目标检测领域,基于深度学习的目标检测框架主要可以分成2类方法^[5]:基于区域(Region-based)的方法和非基于区域(Region-free)的方法。

基于区域的检测方法采用候选区域方法生成区域候选框,用深度学习网络对这些候选框进行特征提取,然后,设计一个区域性的子网(Region-wise Sub-network)对这些候选框分类。R-CNN^[6],SPP-Net^[7]和Fast R-CNN^[8]等方法,在应用中都获得了良好的检测效果。由于Fast R-CNN需要大量提取候选框,处理一副图像需要大量时间,因此,Ren等^[9]又提出Faster R-CNN算法。该算法提出采用RPN网络来提取候选区域进行分类,实现端到端的目标检测。

非基于区域的方法将目标检测当作单次检测(Single Shot)问题,直接使用全卷积网络的全局特征预测每个位置存在目标类别生成分数(Score)并对目标框的形状进行回归预测调整。这类方法在目标检测问题上具有速度快、精度高等优点,目前主要方法有YOLO^[10]和SSD^[11]等。

与基于区域的方法相比,非基于区域方法检测速度快,但检测精度略低,对目标的定位框调整采取

的回归策略,存在目标定位精度不高、定位框偏大和定位偏移等问题,这些问题给图像的后续处理带来不便。

在非基于区域的方法中,SSD算法相比YOLO等算法检测速度快,精度更高,所以本文采用SSD算法作为基础检测框架,主要针对非基于区域方法存在的目标定位精度不够高的问题,提出了一种改进的检测定位方法,通过结合SSD和选择性搜索算法^[12],有效提高了目标检测定位精度。

1 SSD算法

1.1 SSD模型

SSD是基于前馈神经网络,产生一系列固定大小的边界框和目标类别的分数,最后进行非极大值抑制产生最终检测。SSD算法将VGG-16 Net^[13]作为基础网络(在分类层之前截断),在基础网络之后添加尺寸逐层递减的卷积特征层,以产生多个尺度检测的预测值。

由图1可以看出SSD采用特征金字塔结构进行检测,检测时利用不同特征层的特征图,在多个特征图上同时进行softmax^[14]分类和位置回归。

在SSD中引入默认框,类似于Faster R-CNN中使用的锚点,卷积处理时,在具有不同尺度的若干特征图上每个点的中点为中心,生成一条到同心的不同长宽比的默认框,对于每个默认框,预测对所有目标类别的形状偏移和置信度,如图2所示。

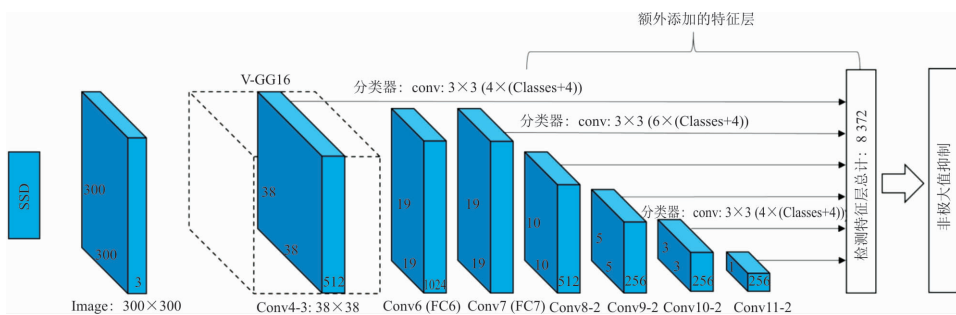


图1 SSD网络模型

Fig. 1 SSD detection model

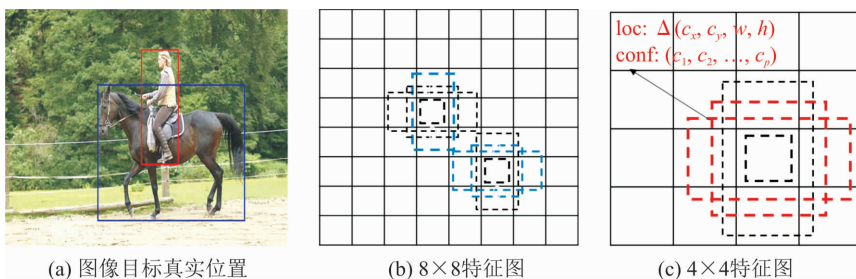


图2 SSD架构

Fig. 2 SSD framework

1.2 SSD 模型训练

对 SSD 模型进行训练,首先将默认框匹配到真实标签框。模型损失是位置损失 L_{loc} 和置信度损失 L_{conf} 的加权求和。

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (1)$$

式中: N 为检测过程中与真实标签框匹配到的默认框数量; L_{loc} 为预测框 l 和真实标签框 g 参数之间的平滑 L_1 损失; 置信度损失 L_{conf} 为 softmax 损失对多类别置信度 c 和权重 α 的交叉验证。

SSD 使用低层特征图检测小目标,使用高层特征图检测大目标。经过匹配后,正负样本严重不平衡。所以,SSD 根据置信度大小将所得到的负样本排序,挑出排序在前的负样本,使得负样本和正样本保持 3:1。

2 基于 SSD 的改进目标检测算法

在输入端输入带有目标的图像,首先通过 SSD 算法对图像进行初始检测,得到图像目标类别和粗略位置;然后采用选择性搜索算法对目标所在区域进行选择搜索,得到目标位置的可能所在区域并生成目标边界候选框;最后采用本文提出的边界判断方法得到目标精确位置,完成由粗到精的目标定位检测。并采用选择性搜索算法对目标所在区域进行精确定位,输出得到目标的精确位置。

选择性搜索是一种结合了穷搜索(Exhaustive Search)^[15]和分割^[16]的方法,寻找所有潜在的物体真实位置来进行目标检测。选择性搜索算法在进行区域分割和合并过程中采取了多种策略,相对于穷搜索大量减少了搜索时间和空间。

本文采用选择性搜索算法只对目标所在区域进行搜索,这有效减少了目标候选框的生成数量。在针对最后一步生成目标候选框时提出了新的目标候选框算法,这大大简化并加快了算法的运行速度。针对大量的目标候选框,本文提出一种目标边界判断方法。设相对阈值 α 为:

$$\alpha = A_n / L \quad (2)$$

式中: $A_n = \{A_1, A_2, \dots, A_n\}$ 是由图像边缘向目标区域中心点扫描到的候选框的长度; L 是目标所在区域边缘长度。当候选框的长度值 A_n 大于等于相对阈值 α 时,就认定此候选框的边界为真实目标位置的一个边界,以此来确定目标的真实位置。但是存在一种特殊情况:

$$A_n < \alpha \quad (3)$$

即所生成的目标候选框普遍偏小,且小于相对阈值

时,就默认离目标图像区域边缘最近的候选框 A_i 为图像目标真实边框。

2.1 改进的选择性搜索算法整体流程

1) 使用基于图表示的图像分割算法^[17]对输入图像进行初始分割,获取原始分割区域 $R = \{r_1, r_2, \dots, r_n\}$;

2) 设置相似集 $S = \emptyset$;

3) 计算生成区域的相似度,将结果放入相似集 S ;

4) 通过计算区域相似度来进行区域合并。在相似集 S 中选择相似度最大的区域 r_i 和 r_j ,并将其融合成新区域 r_i ,从相似集 S 中剔除 r_i 和 r_j 的相似度,计算合并的新区域 r_i 与相邻区域(原先 r_i 和 r_j 相邻的区域)的相似度,将得到的相似度加入相似集 S 中,并将 r_i 添加到区域集 R 中;

5) 得到每个区域的边界框,这个边界框就是目标潜在的真实位置;

6) 根据得到的目标区域的边界框,采用目标边界判断方法确定图像目标的真实边界,获得目标的精确位置定位。

2.2 相似度计算

在对生成区域进行区域合并时,必须得考虑到物体的多样性,包括物体颜色,纹理等等。选择性搜索算法中采用了 4 种特征用于计算分割区域之间的相似度。

1) 颜色相似度。

将色彩空间由 RGB 空间转换到 HSV 空间,HSV 的每个颜色通道都归一化为 25 bins 的颜色直方图,然后每个区域都能得到一个 75 维的向量 $C_i = \{c_i^1, c_i^2, \dots, c_i^n\}$,计算区域之间的颜色相似度:

$$S_{color}(r_i, r_j) = \sum_{k=1}^n \min(c_i^k, c_j^k) \quad (5)$$

2) 大小相似度。

区域里有多少像素可以用来表示区域之间大小相似度^[18]。计算大小相似度可以让包含较少像素的区域优先合并,避免被大的区域吞并。

$$S_{size}(r_i, r_j) = 1 - \frac{size(r_i) + size(r_j)}{size(im)} \quad (6)$$

3) 吻合相似度。

吻合相似度是衡量 2 个区域是否相似的准则,其指标是合并后的区域的边界框(能够框住区域的最小矩形)越小,其吻合度越高。其计算方式如下:

$$fill(r_i, r_j) = 1 - \frac{size(BB_{ij}) - size(r_i) - size(r_j)}{size(im)} \quad (7)$$

4) 纹理相似度。

文献^[19]采用 SIFT-Like 特征作为纹理相似度的计算。为了计算简便,本文采用 LBP 纹理特征作

为纹理相似度的计算。区域之间纹理相似度计算方式和颜色相似度计算方式类似,通过累加每个直方图区间最小值计算相似度:

$$S_{\text{texture}}(r_i, r_j) = \sum_{k=1}^n \min(t_i^k, t_j^k) \quad (8)$$

5) 将前文的计算公式组合写为式(19),其中 $a_i \in \{0, 1\}$:

$$S(r_i, r_j) = a_1 S_{\text{color}}(r_i, r_j) + a_2 S_{\text{texture}}(r_i, r_j) + a_3 S_{\text{size}}(r_i, r_j) + a_4 \text{fill}(r_i, r_j) \quad (9)$$

3 实验结果与分析

采用 Python 程序对本文算法进行了测试,实验在处理器为 Intel(R)Core i7 2.8 GHz 的计算机上进行。算法中,相对阈值取值 $\alpha = 0.32$ 。为验证本文算法的目标定位性能,实验选用 PASCAL VOC2012 测试集作为实验数据库,本文将测试集分成 4 类,其中单目标简单背景图 6 645 张,单目标复杂背景图 6 052 张,多目标简单背景图 5 787 张,多目标复杂背景图 4 760 张,共 22 531 张图像。并将本文算法定位结果与真实位置(Ground Truth)进行实验对比,同时和传统 SSD 算法结果进行比较,从定性和定量 2 个层面分析本文算法的定位性能。

3.1 定性分析

下面给出部分实验结果,分为 4 种情况对本文算法进行定性分析:

3.1.1 单目标简单背景

以图 3(a)为例,在原图中,目标背景较简单,传统 SSD 算法能够顺利将目标检测出来,但是没能将目标本体的全部部分框住,存在目标定位框偏大,遗漏等问题。采用本文算法能够完全的将目标包含在定位框之内,定位十分准确,从直观视觉上可判断,采用本文算法所得到的定位框与真实位置相差不大。

3.1.2 单目标复杂背景

在目标检测中,目标背景较为复杂,如果不进行调整,对目标的定位会更加不准确,不符合人眼的直观感受。从图 3(b)中我们可以看出,当目标背景较为复杂时,原算法对目标的定位与原图目标的真实位置相差较大。而本文算法可以很好地解决这一问题,在背景复杂的情况下也能准确地将目标定位出来。

3.1.3 多目标简单背景

以图 4(a)为例,传统 SSD 算法能够顺利将目标检测出来并用定位框将其包含在内,但是存在定位框漂移、偏小等问题。采用本文算法对单个目标进行处理,能够很好地解决这个问题。

3.1.4 多目标复杂背景

以图 4(b)为例,传统 SSD 算法虽然能够顺利将其检测出来并定位,但是定位框与原图目标的真实位置相差较大,定位框定位的区域偏小。采用本文算法对多目标进行拆分为单个目标进行处理,能够很好地解决这个问题,符合人眼的直观视觉。



图 3 单目标检测的实验结果

Fig. 3 The results of single object detection



图 4 多目标检测实验结果

Fig. 4 The results of multi-objects detection

3.2 定量分析

为对实验结果进行定量分析,本文采用中心位置误差和覆盖率 2 个指标来评价本文算法的定位性能。其中,中心位置误差(Center Location Error, CLE)是指图像定位候选框的中心位置与原图像目标的真实位置中心位置之间的欧氏距离。本文计算了与原图像真实位置中心位置误差,并与 SSD 算法的中心位置误差进行了比较,见表 1。覆盖率是指目标定位框和目标真实位置的重叠部分所占的比重,覆盖率 $(|S_T \cap S_G|)/(|S_T \cup S_G|)$ 中 S_G 和 S_T 分别表示目标定位框和原图目标真实位置, \cap 和 \cup 分别表示 2 个区域的交集和并集, $|\cdot|$ 表示区域的面积。表 2 给出了本文算法目标定位覆盖率和 SSD 算法覆盖率的对比结果。表 1 和表 2 中的结果是对 PASCAL VOC2012 数据库中的 22 531 幅图像的统计结果。

表 1 中心位置误差值
Tab. 1 Center location error 像素

图像	SSD 算法	本文算法
单目标简单背景	7.4	2.3
单目标复杂背景	22.3	4.6
多目标简单背景	46.7	10.8
多目标复杂背景	68.5	13.5

注:中心位置误差值越小,目标定位精度越高

表 2 覆盖率
Tab. 2 Overlap %

图像	SSD 算法	本文算法
单目标简单背景	87.2	95.1
单目标复杂背景	83.4	92.3
多目标简单背景	78.3	88.9
多目标复杂背景	73.6	85.2

注:覆盖率越大,目标定位精度越高

分析实验结果可知,本文算法在大多数的图像目标定位中保持了较低的中心误差和较高的覆盖率,表现出良好的定位精度。

3.3 本文算法检测速度和均值平均精度

本文算法在处理器为 Intel(R)Core i7 2.8 GHz 的计算机上运行,在 PASCAL VOC2012 数据集上检测速度达到 25 fps,传统 SSD 算法在本计算机上运行速度为 31 fps,本文算法比传统 SSD 算法稍慢,但比 YOLO 的 22 fps 要快,符合实时检测的要求。

本文算法在 PASAL VOC 2012 测试集上进行测试,表 3 显示了本文算法和其他算法模型的检测结果。通过对比可以看到本文算法优于 Fast RCNN 算法,与 Faster RCNN 算法性能相当,比传统 SSD 算法高 0.1%。与 YOLO 相比,本文算法性能

明显更好。

表 3 PASCAL VOC 2012 测试集检测结果

Tab. 3 The results of detection on PASCAL VOC 2012

Method	均值平均精度
Fast RCNN	68.4
Faster RCNN	70.4
YOLO	57.9
SSD	70.3
本文算法	70.4

4 相关工作

目前基于 SSD 的改进目标检测方法有以下 2 种性能较好的算法,文献[20]采取融合中间特征层的方法,将特征层 Conv4_3 和 Conv5_3 采取不同的 2 个方式融合,在 PASCAL VOC 2012 数据集上获得均值平均精度为 71.2,检测速度达到 23 fps。文献[21]基于 SSD 提出 CSSD 算法,该算法直接在 SSD 特征层上添加 2 个加宽的多尺度卷积层,经过重新训练后,在 PASCAL VOC 2012 数据集上获得均值平均精度为 71.4,检测速度达到 22 fps。虽然以上算法在性能上超过了本文算法,但检测速度仍略低于本文算法。表 4 给出了本文算法目标定位中心位置误差和覆盖率与上述 2 种算法的对比结果。表 4 中的结果是对 PASCAL VOC 2012 数据库中的 22 531 幅图像的统计结果。

表 4 中心位置误差值和覆盖率

Tab. 4 Center location error and overlap %

算法	中心位置误差	覆盖率
本文算法	7.6	90.8
特征融合	12.8	80.3
CSSD	13.5	81.6

注:中心位置误差值越小,覆盖率越大,目标定位精度越高

由实验结果我们可以看出本文算法在目标检测定位精度上远优于这 2 种算法,表现出良好的目标定位精度。

5 结语

SSD 算法相对于其他深度学习算法优势明显,在速度上比 YOLO 更快,在精度上也可以和 Faster R-CNN 比较。针对 SSD 算法目标检测定位精度不高的问题,本文提出的基于 SSD 算法目标检测精确定位,大幅提高了目标定位精度。在实验中针对各种类型的图像,本文算法的结果明显比真实位置的覆盖率更高,中心误差更低。

参考文献(References):

- [1] VIOLA P, JONES M J. Robust Real-Time Object Detection[J]. International Journal of Computer Vision, 2002, 57(2): 137-154.
- [2] GIRSHICK R, DONAHUE J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C]// Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2014:580-587.
- [3] 金泽芬, 侯志强, 余旺盛. 多特征博弈的目标跟踪算法[J]. 空军工程大学学报(自然科学版), 2017, 18(1):50-56.
- JIN Z F F, HOU Z Q, YU W S. Object Tracking Algorithm for Multi-Feature Game [J]. Journal of Air Force Engineering University (Natural Science Edition), 2017, 18(1):50-56. (in Chinese)
- [4] HARZALLAH H, JURIE F, SCHMID C. Combining Efficient Object Localization and Image Classification [C]// IEEE 12th International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009:237-244.
- [5] KONG T, SUN F C, YAO A B. RON: Reverse Connection with Objectness Prior Networks for Object Detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA: IEEE, 2017: 5244-5252.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Image Net Classification with Deep Convolutional Neural Networks[J]. International Conference on Neural Information Processing Systems, 2012, 60(2): 1097-1105.
- [7] HE K M, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9):1904-1916.
- [8] GIRSHICK R. Fast R-CNN[C]// IEEE International Conference on Computer Vision. [S. l.]: IEEE, 2015: 1440-1448.
- [9] REN S, 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 & Machine Intelligence, 2015, 39(6):1137-1149.
- [10] REDMON J, DIVVALA S, et al. You Only Look Once: Unified, Real-Time Object Detection [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV: IEEE, 2016: 779-788.
- [11] LIU W, ANGELOV D, ERHAW D, et al. SSD: Single Shot Multibox Detector[C]// European Conference on Computer Vision. [S. l.]: Springer, 2015: 21-37.
- [12] UIJLINGS J R, SANDE K E, et al. Selective Search for Object Recognition [J]. International Journal of Computer Vision, 2013, 104(2):154-171.
- [13] SIMONVAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Vision and Pattern Recognition, 2014(1): 1556.
- [14] MEMISEVIC R, ZACH C, et al. Gated Softmax Classification [C]//International Conference on Neural Information Processing Systems. Vancouver, British Columbia: Curran Associates Inc, 2010:1603-1611.
- [15] FELZENSZWALB P F. Efficient Graph-Based Image Segmentation[J]. International Journal of Computer Vision, 2004, 59(2):167-181.
- [16] ARBELAEZ P, MAIRE M, et al. Contour Detection and Hierarchical Image Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 33(5):898-916.
- [17] SASAKI Y. Exhaustive Search[C]//Encyclopedia of Genetics Genomics Proteomics & Informatics. [S. l.]: Springer, 2008:653-653.
- [18] 刘建华, 毕笃彦, 叶广强. 基于目标模型的红外弱小目标预检测[J]. 空军工程大学学报(自然科学版), 2006, 7(5):36-38.
- LIU J H, BI D Y, YE G Q. Infrared Dim and Small Object Predetection Based on Object Model[J]. Journal of Air Force Engineering University(Natural Science Edition), 2006, 7(5):36-38. (in Chinese)
- [19] XU Y, HUANG S B, JI H, et al. Scale-Space Texture Description on SIFT-Like Textons[J]. Computer Vision & Image Understanding, 2012, 116(9): 999-1013.
- [20] CAO G M, XIE X M, YANG W Z, et al. Feature-Fused SSD:Fast Detection for Small Objects[C]// The 9th International Conference on Graphic and Image Processing. [S. l.]:SPIE, 2017: 1-8.
- [21] XIANG W, ZHANG D Q. Context-Aware Single-Shot Detector[C]//2018 IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe, NV: IEEE, 2018: 1784-1793.

(编辑:徐楠楠)