

doi: 10.3969/j.issn.2095-4565.2021.02.008

一种基于改进 YOLO v3 的小目标检测方法

蔡鸿峰, 吴观茂*

(安徽理工大学 计算机科学与工程学院, 安徽 淮南 232001)

摘 要: 为了提升小目标物体检测的精度, 尽可能避免发生漏检情况, 提出了一种改进的 YOLO v3 算法。选用 Darknet-49 为主干网络, 将原有的 3 个检测尺度扩展为 5 个。同时, 引入 DIoU 函数, 对损失函数进行了改进, 将 PASCAL VOC 2012 数据集作为测试数据集和训练数据集进行实验。研究发现, 改进后的 YOLO v3 算法与原 YOLO v3 算法相比, mAP 值提升了约 2.4%, 且检测速度与原算法接近。实验结果表明, 改进原 YOLO v3 算法网络结构和损失函数, 可以增强算法对小目标物体的检测精度。

关键词: 小目标; 特征融合; 损失函数

中图分类号: TP393 **文献标志码:** A **文章编号:** 2095-4565(2021)02-0033-04

A Small Target Detection Method based on Improved YOLO v3

CAI Hongfeng, WU Guanmao*

(School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001)

Abstract: In order to improve the accuracy of small object detection and reduce the occurrence of missed detection, an improved YOLO v3 algorithm is proposed. The algorithm uses Darknet-49 as the backbone network, and expands the original three detection scales to five detection scales. At the same time, DIoU function is introduced to improve the loss function. PASCAL VOC 2012 data set is used as test data set and training data set for experiments. Compared with the original YOLO v3 algorithm, the map value of the improved YOLO v3 algorithm has been improved by about 2.4%, and the detection speed is close to the original algorithm. The experimental results show that by improving the network structure and loss function of the original YOLO v3 algorithm, the detection accuracy of small objects can be enhanced.

Key words: small target; feature fusion; loss function

目标检测包含目标的定位和分类。早期的目标检测方法是使用手工提取特征, 再在此基础上构造模型。使用该方法设计的模型结构复杂, 提升模型精度困难。卷积神经网络具有较好的学习特征能力, Girshick 等^[1]将 CNN 应用到目标检测中, 提出了 R-CNN 模

型, 使模型的检测性能有了很大提升, 但其网络只能输入固定尺寸的图像, 且 R-CNN 通过 SS 算法生成的候选区大量重叠, 造成计算资源浪费。针对 R-CNN 模型的这类缺点, He 等^[2]提出了 SPP-Net 网络。Girshick^[3]与 Ren 等^[4]先后提出了 Fast R-CNN 与 Faster R

收稿日期: 2020-11-30

基金项目: 安徽省自然科学基金面上项目(项目编号: 1908085MF189)。

作者简介: 蔡鸿峰, 硕士生。Email: 1355356679@qq.com

* 通信作者: 吴观茂, 副教授, 博士, 研究方向: 深度学习。Email: 269919@qq.com

-CNN 模型。从 R-CNN 发展到 Faster R-CNN,模型的检测速度与检测精度均有所提高,但这类两阶段的检测方法在检测速度上与单阶段检测方法仍存在一定差距。典型的单阶段目标检测方法有 YOLO 系列算法和 SSD 算法。Redmon 等^[5]在 2015 年提出了第一个单阶段检测方法 YOLO 算法,Liu 等^[6]借鉴了 YOLO 和 Faster R-CNN 的优点提出了 SSD 算法,检测速度快并且实现了多尺度预测。在 YOLO v1 的基础上,Redmon 等^[7-8]又提出了 YOLO v2 和 YOLO v3 算法。其中,YOLO v2 进行了各种尝试,使用了批标准化技术;YOLO v3 使用了 Darknet-53 和 sigmoid 函数。本文以 YOLO v3 作为研究基础,通过网络结构和损失函数的改进来提升小目标检测的精度。

1 YOLO v3 算法

相较于 YOLO v2,YOLO v3 算法在整体网络结构上有较大改变,使用 sigmoid 函数代替了 softmax 函数,并实现了多尺度预测。YOLO v3 能够在 3 个不同位置的 3 个不同尺寸的特征图上使用大小为 1×1 的卷积核实现检测,其中的 2 个特征图是经过上采样和特征融合以后得到的。Darknet-53 作为主干网络进行特征提取,借鉴了残差网络的结构,实现了跳层连接与残差模块。YOLO v3 网络结构的加深使其在检测速度上较之前的版本有所降低,但在检测准确度上得到了提升。

Darknet-53 是全卷积网络,使用了残差结构,共有 53 个卷积层,主要作用是提取图像特征。对于像素为 256×256 的输入图像,用于检测的 3 个特征图的步长为 8,16,32。而对于像素为 416×416 的输入图像,用于检测的 3 个特征图的步长为 13,26,52。3 个不同尺度的特征图经过 FPN (Feature Pyramid Network) 进行融合后,采用多尺度方法帮助 Darknet-53 网络学习到不同层次的特征信息,然后将融合后的特征信息输入 YOLO 层,实现类别预测和边界框回归。

YOLO v3 中使用了多个尺度的特征图预测,根据不同尺寸的输入图像,得到不同大小

的输出特征图,最终的特征图包含了检测框位置信息、检测置信度和类别信息。YOLO v3 共使用了 9 种 anchor box 作为先验框。这 9 种 anchor box 由 K-means 聚类算法得到,先验框尺寸与特征图尺度大小成反比。 13×13 的特征层用于检测大目标, 26×26 的特征层用于检测中等目标, 52×52 的特征层用于检测小目标。

YOLO v3 检测置信度的计算公式为:

$$\text{Confidence} = \text{Pr}(\text{object}) \times \text{IoU} \quad (1)$$

式(1)中, $\text{Pr}(\text{object})$ 用于判断预测边界框中是否包含待检测目标;IoU 为预测边界框与真值边界框的交并比。

得到全部预测边界框后,通过设置阈值与非极大值抑制等操作得到目标边界框。有了先验框与输出特征图,就可以解码检测框。边界框位置预测示意图如图 1 所示。

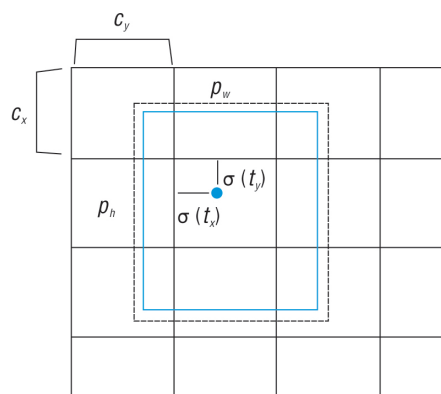


图 1 边界框位置预测示意图

YOLO 的位置预测是指相对于左上角格点坐标预测偏移量,计算每个网格实际预测框的中心位置和尺寸大小。计算公式为:

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = P_w e^{t_w} \\ b_h = P_h e^{t_h} \end{cases} \quad (2)$$

式(2)中, c_x, c_y 为坐标偏移量; b_x, b_y 为实际预测框的中心位置; P_w, P_h 为实际先验框的宽度和长度; b_w, b_h 为实际预测框的尺寸; t_x, t_y 为坐标预测值; t_w, t_h 为边界框宽度和高度的预测值。

2 YOLO v3 算法的改进

2.1 网络结构的改进

YOLO v3 算法使用 Darknet - 53 网络进行特征提取, 利用卷积核实现下采样操作, 步长为 2, 大小为 3×3 。该网络虽然感受野较大, 但是牺牲了空间分辨率, 经过深层卷积提取特征后容易使小目标的信息丢失。针对小目标物体的检测更加依赖浅层特征这一特点, 将网络层数较少的 Darknet 作为检测网络的 backbone, 使用 49 层主干网络, 记为 Darknet - 49。

Darknet - 49 共有 49 个卷积层, 5 个残差块。由于使用 ReLU 激活函数会对低维特征信息造成大量损失, 根据 MobileNet 网络中倒残差结构的最后一个 1×1 的卷积层, 使用线性激活函数来保留低维特征的思想, 在 Darknet - 49 的第 1 个卷积层使用线性激活函数, 减少低维信息的损失。Darknet - 49 网络对输入图像进行 5 次下采样操作, 得到 5 个尺度的特征图。改进后的检测网络将原有的 3 个检测尺度扩展为 5 个检测尺度, 多尺度的特征融合能够提升检测性能^[9]。改进后的检测网络结构如图 2 所示。

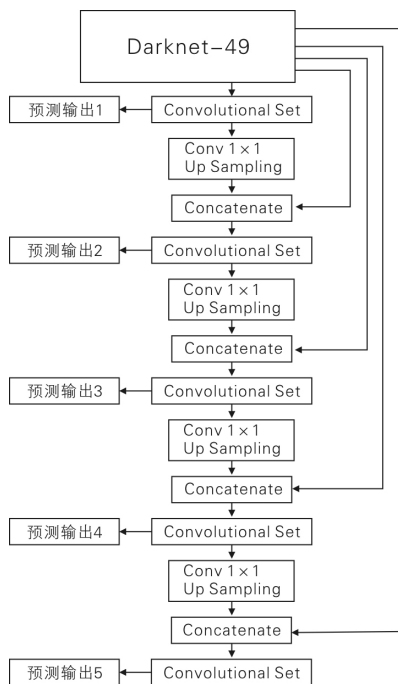


图2 改进后的检测网络结构

2.2 损失函数的改进

YOLO v3 是一种端到端的单阶段目标检测算法, 其损失函数由坐标回归损失、置信度损失和分类损失 3 个部分共同组成, 其中坐标回归损失使用均方误差进行计算, 其余使用交叉熵进行计算。

目标尺度的变化会影响坐标回归的准确性, 使用均方误差不能处理检测目标尺度敏感问题。IoU 为预测目标边界框与真实目标边界框之间的交并比, 其公式为:

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (3)$$

式(3)中, B 为预测目标边界框; B^{gt} 为真实目标边界框。

IoU Loss 具有尺度不变性的特点, 能够更好地反映重合程度。虽然 IoU Loss 比均方误差更有优势, 但是当预测框与 ground-truth 是包含关系时, 其 IoU Loss 为定值, 检测效果会具有较大差异; 当两者不相交时, IoU Loss 的值为 0, 则无法进行优化。

IoU Loss 的计算公式为:

$$L_{IoU} = 1 - IoU \quad (4)$$

针对 IoU Loss 的不足, 本文使用 **DIoU Loss** (Distance IoU Loss)^[10], DIoU 示意图如图 3 所示。

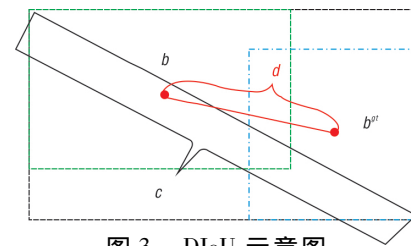


图3 DIoU 示意图

DIoU 公式为:

$$DIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} = IoU - \frac{d^2}{c^2} \quad (5)$$

式(5)中, b 为预测目标边界框的中心坐标; b^{gt} 为真实目标边界框的中心坐标; ρ 为 2 个中心点之间的距离; c 是 2 个目标边界框最小外接矩形的对角线长度。

则 DIoU Loss 表示为:

$$L_{DIoU} = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \times (1 - DIoU_{predict}^{groundtruth}) \quad (6)$$

式(6)中, S^2 为网格总数; B 为边界框总

数; I_{ij}^{obj} 用于判断第 i 个网格中第 j 个边界框是否负责预测该目标; DIOU 能够适应预测目标边界框与真实目标边界框出现包含关系和不相交惯性的情况。

置信度损失 L_{conf} 与分类损失 L_{class} 均用均方误差进行计算, 计算公式为:

$$L_{conf} = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (C_i^j - \hat{C}_i^j)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (C_i^j - \hat{C}_i^j)^2 \quad (7)$$

$$L_{class} = \sum_{i=0}^{S^2} \sum_{c \in classes} I_{ij}^{obj} (P_i^j - \hat{P}_i^j)^2 \quad (8)$$

式(7)~(8)中, C_i^j , \hat{C}_i^j 分别为预测目标边界框和真实目标边界框的置信度; P_i^j , \hat{P}_i^j 分别为预测目标边界框和真实目标边界框的类别概率。

最终损失函数公式为:

$$Loss = L_{DIOU} + L_{conf} + L_{class} \quad (9)$$

3 实验分析

3.1 实验环境

实验使用 ubuntu18.04 操作系统, 使用 Python3.7 与 pytorch 进行编程, 通过 GPU 加速计算。

3.2 数据集

实验使用 PASCAL VOC 2012 数据集, 共有 20 种类别, 背景为第 21 类。PASCAL VOC 2012 数据集见表 1。

表 1 PASCAL VOC 2012 数据集

Title	Sort
Person	person
Animal	bird cat cow dog horse sheep
Vehicle	aeroplane bicycle boat bus car motorbike train
Indoor	bottle chair dining-table potted-plant sofa TV monitor

3.3 实验结果

将 PASCAL VOC 2012 数据集作为测试数据集和训练数据集, 将改进后的 YOLO v3 算法与原 YOLO v3, Faster R-CNN, SSD 进行对比, 实验结果以 mAP 值和 FPS 值作为评估指标。FPS 表示模型每秒钟可以检测的图片数量, FPS 值越大, 模型的检测速度越快。mAP 指标即各类别 AP 的均值, 公式为:

$$mAP = \frac{\sum_i^c AP_i}{C} \quad (10)$$

式(10)中, C 为样本种类数; AP 为样本的平均精度。

4 种算法模型的 mAP 和 FPS 见表 2。由表 2 可以看出, 改进后的 YOLO v3 算法的 mAP 值相较于原 YOLO v3 算法有一定提升, 提升了约 2.4%, 但低于 Faster R-CNN 和 SSD300 算法; 其 FPS 值与原 YOLO v3 算法相比有所降低, 但远超 Faster R-CNN 和 SSD300 算法。这是因为虽然主干网络选取了比原网络层数更少的 Darknet-49 网络, 但同时也增加了网络的预测输出层数, 对检测速度产生一定影响, 致使 FPS 值降低。总体来说, 改进后的 YOLO v3 算法在 mAP 值上有一定提升, 其检测速度与原算法接近。

表 2 4 种算法模型的 mAP 和 FPS

Method	mAP/%	FPS
Faster R-CNN	73.3	7.0
SSD300	74.4	48.0
YOLO v3	65.1	200.0
改进后 YOLO v3	67.5	187.0

改进前后 YOLO v3 算法各类别的 AP 值如图 4 所示。由图 4 可以看出, 改进后的 YOLO v3 的 mAP 值提高了约 2.4%, 对应的各个类别的 AP 值都有一定提升。例如, train 这一类的目标在算法改进前后 AP 值的变化不大, 主要原因是这一类大尺度图片本身已经达到了较高的精度且这类大尺度图片占有较大的图片比例, 不容易因为尺寸的缩放而丢失精度; 而改进后类别为 bottle 的 AP 值增加最多, 上升了约 0.13%, 说明通过增加特征融合层的数量可以提高小目标物体的检测精度。

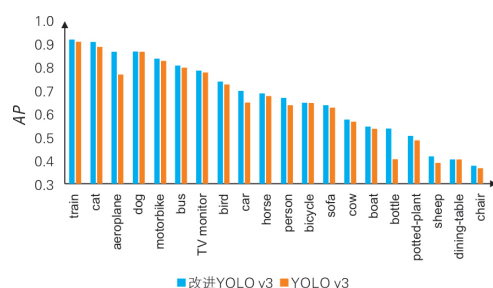


图 4 改进前后 YOLO v3 算法各类别的 AP

(下转第 47 页)

4 结束语

道路开挖会对隧道产生隆起和沉降位移,影响隧道结构安全和稳定,需要进行相关理论计算。在道路开挖施工中,跳仓抽条法对明挖隧道周围地层及结构产生的竖向位移明显小于分层均布法,最大隆起值跳仓抽条法比分层均布法降低了 55.7%,最小沉降量降低约 1/3,有效减小隧道结构的变形,利于轨道交通的安全和稳定。跳仓抽条开挖法在道路施工中的应用,可为相似工程提供借鉴。

参考文献

- [1] 刘尊景,楼永良,邢玉芳,等. 道路施工对下部

(上接第 36 页)

4 结论

通过对原 YOLO v3 算法的网络结构和损失函数进行改进,增强了算法对小目标物体的检测精度。但是,改进后的算法检测速率有所降低,并且在实际检测过程中仍然会存在漏检的情况。

参考文献

- [1] 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. Columbus: IEEE, 2014: 580–587.
- [2] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(9): 1904–1916.
- [3] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440–1448.
- [4] REN S Q, HE K M, GIRSHICK R, et al. Faster R

地铁隧道影响的有限元分析 [J]. 佳木斯大学学报(自然科学版), 2013(6): 842–846.

- [2] 牟建华. 上海市区某大型顶管工程引起道路沉降的有限元分析与监测 [J]. 工程勘察, 2013, 41(2): 9–12, 26.
- [3] 涂慕溪, 陈礼彪, 曾俊铖, 等. 基于有限元的地基-路堤-路面协同变形机理分析 [J]. 公路交通科技, 2019(11): 33–39.
- [4] 魏发达. 明挖公路隧道基坑支护结构分析与施工工艺改进 [D]. 西安: 长安大学, 2014.
- [5] 廖峻, 吴胜忠, 肖博, 等. 城市明挖隧道基坑桩板墙支护设计 [J]. 公路隧道, 2014(3): 47–50.
- [6] 李向阳. 铁路临边明挖隧道深基坑支护施工技术的选择与应用 [J]. 中国标准化, 2016, 31(20): 186–188.

(责任编辑 张银凤)

- CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6): 1137–1149.

- [5] 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. Las Vegas: IEEE, 2016: 779–788.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceedings of the 2016 European Conference on Computer Vision. Cham: Springer, 2016: 21–37.
- [7] REDMON J, FARHADI A. YOLO 9000: better, faster, stronger [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6517–6525.
- [8] REDMON J, FARHADI A. YOLO v3: an incremental improvement [J]. arXiv, 2018: 02767.
- [9] 樊玮, 周末, 黄睿. 多尺度深度特征融合的变化检测 [J]. 中国图象图形学报, 2020, 25(4): 669–678.
- [10] ZHENG Z, WANG P, LIU W, et al. Distance-IoU Loss: faster and better learning for bounding box regression [J]. arXiv, 2019: 08287.

(责任编辑 王 书)