

DOI:10.19651/j.cnki.emt.2106496

基于改进 SSD 算法的行人检测方法

于波 刘畅

(东北石油大学 物理与电子工程学院 大庆 163000)

摘要: 随着计算机视觉的快速发展,行人检测技术越来越成为人们关注的热门话题之一。行人检测算法中 SSD 算法相对于其他行人检测算法更加稳定,针对 SSD 算法做行人检测时产生的检测精度低的问题,提出一种以 ResNet 残差网络作为 SSD 的前馈网络,提取行人特征,并通过特征值生成目标框的行人检测算法。将原始 SSD 算法与改进后的 SSD 算法分别通过 INRIA 行人检测数据集的训练集和测试集做实验进行精度对比。实验结果表明,与 SSD 算法做行人检测时的检测效果相比,改进的行人检测算法更具鲁棒性,算法平均准确率提高了接近 5%。

关键词: 行人检测;SSD 算法;残差网络

中图分类号: TN4;TN91 **文献标识码:** A **国家标准学科分类代码:** 510.1050

Pedestrian detection based on improved SSD

Yu Bo Liu Chang

(School of Physics and Electronic Engineering, Northeast Petroleum University, Daqing 163000, China)

Abstract: With the rapid development of computer vision, pedestrian detection technology has become one of the hot topics. SSD algorithm in pedestrian detection algorithm is more stable than other pedestrian detection algorithms. Aiming at the problem of low detection accuracy when SSD algorithm is used in pedestrian detection, this paper proposes a pedestrian detection algorithm which uses ResNet residual network as the feed-forward network of SSD to extract pedestrian features and generate the target box through the eigenvalues. The original SSD algorithm and the improved SSD algorithm are compared through the training set and the test set of INRIA pedestrian detection data set. The experimental results show that, compared with SSD algorithm, the improved pedestrian detection algorithm is more robust, and the average accuracy of the algorithm is improved by nearly 5%.

Keywords: pedestrian detection;SSD algorithm;residual network

0 引言

随着计算机科学的进步与发展,很多以前需要耗费大量人力或时间才能完成的工作和任务,现在可以由计算机视觉来完成^[1]。行人检测技术是计算机视觉的一个重要研究方向,广泛应用于车辆辅助驾驶、智能视频监控、人机交互等领域^[2]。行人检测技术是依据目标检测算法来完成行人定位及识别的。目标检测算法经历了传统机器学习到深度学习的转变,使得目标检测技术在准确性和执行速度上得到了质的飞跃。深度学习领域的目标检测算法主要包含 R-CNN 系列、YOLO 系列和 SSD 等。R-CNN 方法在检测精度上更好,YOLO 方法在检测速度上表现更优^[3]。行人检测的目的是实时检测出现在图像里的行人,并用矩形框给出行人的位置^[4]。针对行人检测技术中检测精度低的问题,本文提出了基于改进 SSD 算法的行人检测方法。通过

替换 SSD 的前置网络使之成为具有残差学习能力的 ResNet 网络。并将改进内容通过 Python 编程语言和 Tensorflow 框架进行代码实现,通过在 INRIA 行人检测数据集进行测试和训练模型,并将实验结果与改进前的 SSD 算法进行比较。得出最终结论,与 SSD 算法做行人检测时的检测效果相比,改进的行人检测算法更具鲁棒性,算法平均准确率提高了接近 5%。

1 相关模型

1.1 SSD 检测原理

SSD(single shot MultiBox detector)是 Wei Liu 等在 ECCV 2016 上提出的一种目标检测算法,它结合了 YOLO 中的回归思想和 Faster R-CNN 中的锚点机制,使用全图各个位置的多尺度区域特征进行回归^[5]。不仅是回归,还有检测头进行分类。SSD 算法通过前馈 CNN 网络提取图

收稿日期:2021-04-24

像特征值,并对不同尺度的特征层进行分类和定位^[6]。其中,浅层特征图相对尺寸较大,适合对尺度较小的目标进行检测^[7]。相反的深层特征图适合检测尺寸较大的目标。SSD算法比Faster R-CNN算法的运算速度快,比YOLO的检测精度高。SSD300是以VGG16作为前置网络,本文是在此基础上进行改进的网络架构。

SSD网络结构主要分为VGG16基础层、特征提取层、

检测层、非极大值抑制(NMS)、锚点、MultiBox损失这几部分。SSD网络以VGG16为基础的特征提取层,保留了VGG16的5组卷积即Conv5-3之前的网络层,将VGG16的全连接层FC6和FC7更改为卷积层Conv6和Conv7,舍弃了FC8,并且添加了卷积块Conv8,Conv9,Conv10和Conv11用来提取更深层次、更多尺度特征图^[8]。SSD300的详细版网络结构如图1所示。

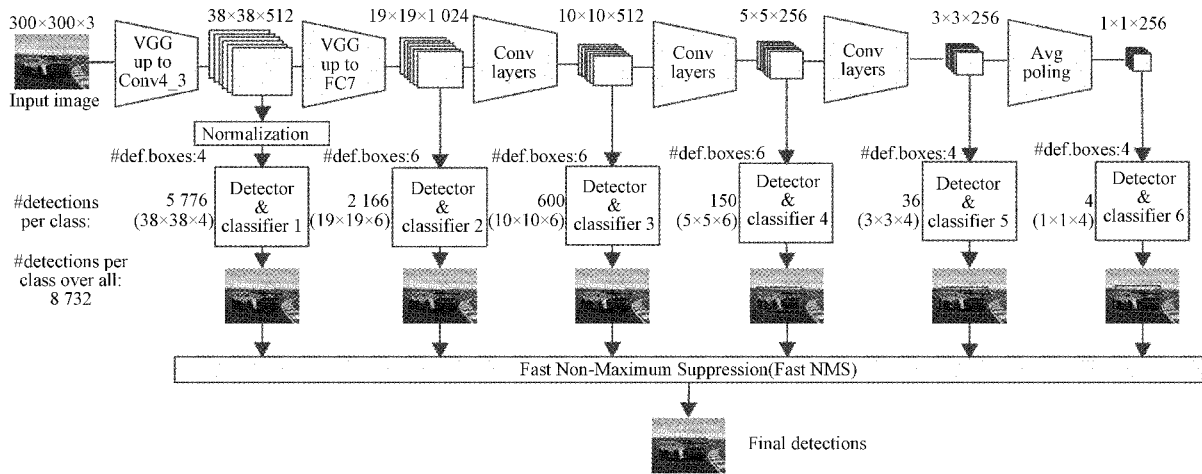


图1 SSD300网络结构

以上是准备好的特征金字塔,用于目标检测。接下来通过锚点机制来生成一些预测框。以生成的框为基本,通过位移和长宽比的变化,慢慢地向真实目标位置靠近。预测框的尺寸大小不一样,最底层的scale=0.2,最高层的scale=0.95,其他层尺寸由式(1)计算。

$$s_k = s_{min} + \frac{s_{max} - s_{min}}{m - 1}(k - 1), k \in [1, m] \quad (1)$$

SSD300中m为6。当在特征层上scale一定时,将设置不同的长宽比ratio=[1, 2, 3, 1/2, 1/3]的预测框,长宽比的计算公式如下:

$$\omega_k^a = s_k \sqrt{a_r}, h_k^a = s_k / \sqrt{a_r} \quad (2)$$

当ratio=1时,长宽比计算公式如下:

$$s'_k = \sqrt{s_k s_{k+1}} \quad (3)$$

以上全部准备就绪,检测层要实现两个功能,类别预测和框预测。根据基础网络生成的边界框和物品类别的置信度,通过调整特征层为相同尺寸,接下来使用非极大值抑制(NMS)和MultiBox损失产生最终检测结果。

1.2 ResNet 模型

CNN网络在发展过程中,大家只知道增加网络的深度可以提升模型性。但在网络超过30层的时候就会出现退化问题。为了解决这一问题,何凯明团队设计了残差学习单元来解决网络退化的问题^[9]。残差网络的使用不仅加速深层网络的训练,提升深度网络的准确率,而且能够避免梯度消失或梯度爆炸的问题^[10]。图2所示为残差学习模块。即在网络层中加入跳跃连接(shortcut connections),

让网络拥有更强的恒等映射(identity mapping)能力,改善深层网络梯度消失问题^[11]。由图2可以看出,输入数据可以通过F(x)处理得到输出,也可以通过跳跃链接直接经过F(x)加入到输出中。

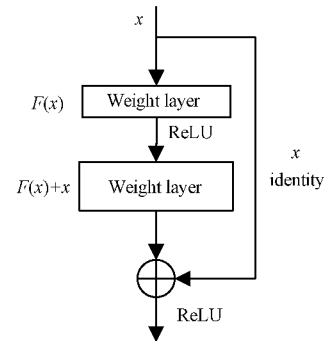


图2 残差学习模块

从数学理论上理解残差学习单元的输入输出,数其学公式可以表示如式(4)所示。

$$y_l = h(x_l) + F(x_l, W_l) \quad (4)$$

$$x_{l+1} = f(y_l)$$

x_l 和 x_{l+1} 是第l个残差单元的输入和输出,F是残差函数,表示学习到的残差,f是ReLU激活函数。ResNet网络是参考了VGG19网络,在其基础上进行了修改,并通过短路机制加入了残差单元。ResNet直接使用步长为2的卷积做下采样,并且用全平均池化层替换了全连接层。其重要的设计原则是当特征图大小降低一半时,特征图的

数量增加一倍,这保持了网络层的复杂度。通过残差学习在网络中的应用,使得网络解决了网络过深产生的退化问题。在 ImageNet 测试集上当采用 ResNet-152 时误差降到了 4.49%,当采用集成模型后,误差可以降低到 3.57%。

2 行人检测算法

2.1 引入残差网络的 SSD 网络结构

基础网络也叫前置神经网络,在 SSD 目标检测算法中从输入的图像中进行特征提取,为后续网络提供输入信息,其中 SSD 使用的前置网络为 VGG16^[12]。

本文用残差网络替换 VGG16 作为 SSD 的基础网络,将图片调整成 300×300 像素,RGB3 通道的形式输入进网

络。与 VGG16 一样,引入残差网络的基础网络使用的也是 3×3 的卷积核,由于 VGG16 采用了 5 段卷积,每一段有一个或多个卷积层^[13]。本文的网络是在 5 段卷积中每两段之间加入残差学习结构,所以形成了 4 个残差学习结构。残差结构卷积层大多是 3×3 的卷积核,由于要输出相同大小的特征图,各层的卷积核数量也是相同的,当特征图大小减半,为了保持每一层的时间复杂度,卷积核的数量将增加一倍。通过步幅为 2 的卷积层进行下采样。以上部分是引入残差结构的基础网络结构说明,图 3 所示为引入残差网络的 SSD 网络结构图。图中可以看出除了引入残差的基础网络部分,其余部分与 SSD 网络结构中对部分一致。具体内容可以了解文献[14]。

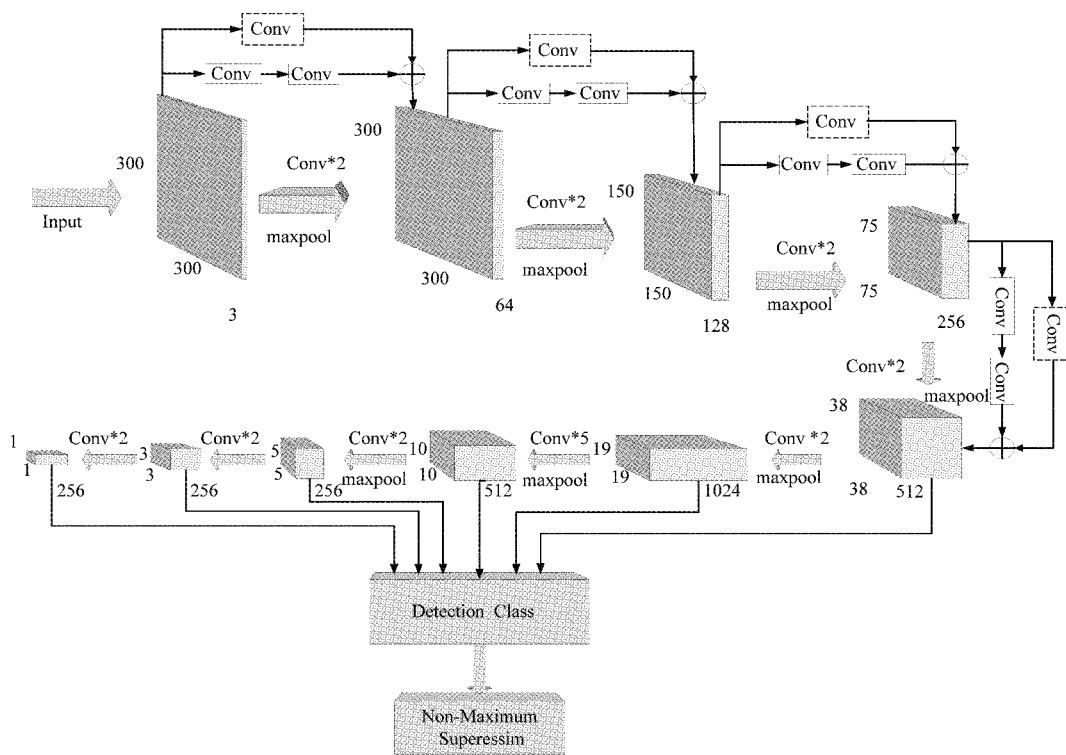


图 3 引入残差网络的 SSD 网络结构

2.2 行人检测算法流程

整个算法流程如图 4 所示,首先将数据集图片调整成 300×300 像素大小输入进待训练网络,数据通过分类网络和矩形框锚定产生特征图和矩形框,输出数据通过反向传播调整网络参数,然后将调整好的数据导出模型,在通过数据集进行测试模型效果。

3 实验与分析

3.1 实验平台

实验环境为:Ubuntu 操作系统,一块 GTX1060 6 GB 的 GPU,三块 E5 系列 CPU。以 Tensorflow 为框架,以 Python 为主要编程语言进行模型训练和测试。

3.2 数据集的选择与模型训练

训练模型选择的是 INRIA 行人检测数据集,该数据库是静态行人检测数据库,提供原始图片及相应的标注文件。

图片中人体大部分为站立姿势且高度大于 100 个像素。本次模型训练用了该数据集中随机选取的 1 000 张图片,630 张图片做训练集,270 张图片做测试集。eval_step=2 500, Batch_size=32,迭代 12 000 次,log_step=10,基础学习率 0.001,第 8 000 次到 10 000 次之间每 1 000 次学习率下降 0.1 倍。损失函数的下降代表着真实值与预测值越来越接近,所以本次实验为了调整模型的精度,先后训练了 4 次,最终确定上述参数的大小。图 5 所示为训练时产生的总

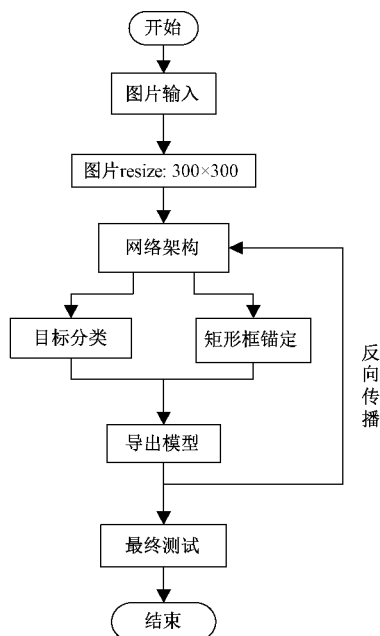


图4 算法流程

损失函数和学习率的图像。从图中曲线的变化可以确定模型在迭代 10 000 次之后更加稳定。

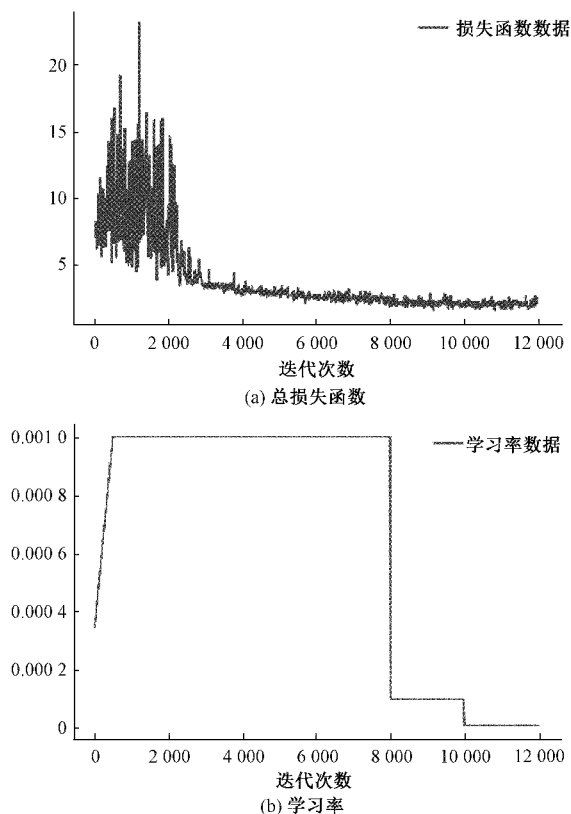


图5 模型损失函数和学习率

3.3 评价指标

评价指标选择的是 AP (average precision)^[15] 即每一类别 P 值的平均值, P 值就是准确率 (precision, P) 公式

如下:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

式中: TP 表示正确检测到行人数量; FP 表示错误地将非行人检测成行人的数量; $TP + FP$ 表示检测的行人总数。计算 AP 值首先要在训练结束后获取从大到小的打分排名, 取前 m 个, 然后获取最大准确率 P_1, \dots, P_m , 最后通过式(6)计算得出 AP 值。

$$AP = \text{sum}(P_1 + \dots + P_m) / m \quad (6)$$

本次实验获得 SSD 模型和引入残差结构的 SSD 模型的 AP 值如表 1 所示。通过对比可以看出, 改进的模型比原始模型性能提升接近 5%。

表1 模型 AP 值对比

	SSD 模型	引入残差的 SSD 模型
AP 值/%	72.94	77.93

3.4 结果分析

本次实验是为了验证引入了残差网络的 SSD 算法与原 SSD 算法在算法性能上和检测精度上哪个具有更好的效果。

图 6 是原始 SSD 模型和引入残差的 SSD 模型进行行人检测最后产生的效果图, 图 6(a) 展示的是在 SSD 模型下进行行人检测的效果图, 从图中可以看出原始 SSD 模型的效果图中有漏检的缺点。图 6(b) 展示的是在引入残差的 SSD 模型中进行行人检测的效果图, 从图中可以看出改进 SSD 模型也同样有漏检的缺点, 但是检测出是行人的数值相比于图 6(a) 的数值高, 既检测精度比原始模型检测精度高。



图6 模型效果图对比

4 结 论

本文以残差网络作为 SSD 模型的基础网络, 使得网络性能提高了接近 5%, 在效果对比中也能明显观察到引入残差网络的 SSD 算法要比原始 SSD 算法更具鲁棒性。但

是改进的算法对于远距离小目标、被遮挡的行人等目标在检测上还是存在检测不出来的情况,所以未来的研究过程中,改善小目标和遮挡问题也将是一个全新的挑战。

参考文献

- [1] 王昊. 基于 YOLO 的行人检测算法研究[D]. 无锡: 江南大学, 2020.
- [2] JIN R, JUN H. A new multi-scale pedestrian detection algorithm in traffic environment[J]. Journal of Electrical Engineering & Technology, 2021, 16(2): 1151-1161.
- [3] 永昌, 单玉刚, 袁杰. 基于改进 SSD 算法的行人检测方法[J]. 计算机工程与设计, 2020, 10(41): 2921-2926.
- [4] 孙锐, 王慧慧, 叶子豪. 融合深度感知特征与核极限学习机的行人检测[J]. 电子测量与仪器学报, 2019, 33(2): 39-47.
- [5] 周晓彦, 王珂, 李凌燕. 基于深度学习的目标检测算法综述[J]. 电子测量技术, 2017, 40(11): 89-93.
- [6] 姚树春, 蔡黎亚, 刘正. 基于级联回归网络的多尺度旋转人脸检测方法[J]. 电子测量与仪器学报, 2019, 33(5): 32-38.
- [7] 李海涛, 曹纯. 基于改进 SSD 的航天发动机试验报警方法[J]. 国外电子测量技术, 2020, 39(4): 149-153.
- [8] 沙俊鹏. 基于深度学习的行人检测与重识别方法研究与实现[D]. 开封: 河南大学, 2020.
- [9] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (Cvpr), 2016: 770-778.
- [10] 林楨哲, 王桂棠, 陈建强, 等. 基于残差网络深度学习的肺部 CT 图像结节良恶性分类模型[J]. 仪器仪表学报, 2020, 41(3): 248-256.
- [11] 吴边. 基于深度学习的行人检测算法研究[D]. 镇江: 江苏大学, 2020.
- [12] 吴天舒, 张志佳, 刘云鹏. 基于改进 SSD 的轻量化小目标检测算法[J]. 红外与激光工程, 2018, 47(7): 47-53.
- [13] KAREN S, ANDREW Z. Very deep convolutional networks for large-scale image recognition [J]. Presented at the Computer Vision and Pattern Recognition 2015.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector [J]. European Conference on Computer Vision (ECCV), 2016 (In press), DOI:10.1007/978-3-319-46448-0_2.
- [15] BOURKE D L. Precision averaging[J]. Pain, 2004, 110(1-2): 502.

作者简介

于波, 工学博士, 教授, 主要研究方向为信号处理、振动故障诊断技术。

E-mail: 347153464@qq.com