

基于 Faster-RCNN 的行人检测方法探讨

李 恒 朱东弼

摘 要

行人检测在计算机视觉领域中有着重要应用。但由于复杂的人体姿态等因素,会影响检测行人的准确率。本文提出基于 Faster-RCNN 的改进型行人检测方法。该方法使用 ResNet50 作为主体网络,提取特征图后,使用改进的区域建议网络(RPN)显示目标区域,最后对目标区域进行识别及分类,既不增加检测时间又提高了检测的准确性。

关键词

行人检测;区域建议网络;深度学习

中图分类号: G06K9

文献标识码: A

DOI: 10.19694/j.cnki.issn2095-2457.2020.26.36

Abstract

Pedestrian detection has an important application in the field of computer vision. However, the accuracy of pedestrian detection will be affected by complex body posture and other factors. This paper proposes an improved pedestrian detection method based on Faster-RCNN. In this method, ResNet50 is used as the main network. After the feature map is extracted, the improved regional recommendation network (RPN) is used to display the target area. At last, the target area is identified and classified.

Key words

Pedestrian detection; Regional recommendation network; Deep learning

李 恒

硕士研究生, 主要研究领域为人工智能, 延边大学。

朱东弼

博士, 教授, 硕士生导师, 主要研究领域为通信网, 人工智能, 延边大学。

0 引言

行人检测(Pedestrian Detection)是检测图像中的目标位置,利用相关技术判定所提供的图像是否存在行人并给出精确的位置标识^[1]。行人检测方法可分两大类,一类是基于背景建模的方法,核心是将前景中的目标进行特征提取然后判断是否存在目标,但背景建模会受到环境因素的影响会导致模型复杂。另一类是基于统计学习的检测方法,它的核心是依据海量样本对图像特征进行提取,然后分类,但是也会受到多种因素干扰而发生漏检或者重复检测的问题,而影响检测效果。

1 基于深度学习的现有行人检测方法分析

2012年后,目标检测进入深度学习时代,2013年,Sermanet等人使用卷积稀疏编码对卷积神经网络进行行人检测的研究^[2]。谢林江等人针对复杂背景因素影响检测效果问题,提出了改进版的神经网络模型,使得行人的特征更加明显^[3]。当前经典的目标检测算法网络模型分为两大类:一类是基于区域建议的 two-stage 算法,如2015年的开创者 R-CNN、改进的 Fast R-CNN 及2016年进阶版的 Faster R-CNN^[4],另一大类是 one-stage 算

法,比如经典的YOLO系列、SSD等算法,相比于two-stage算法,这些算法不包含区域建议模块,不利于网络学习,从而拉低了整体的准确率。

虽然行人检测系统与深度学习相结合使得检测效果有所提升,但是由于人体姿态的复杂程度和背景因素的影响,且行人属于小尺度目标,因此会影响检测出行人的准确率。

2 基于Faster-RCNN的行人检测方法

2.1 行人检测过程

Faster-RCNN行人检测流程首先使用全卷积神经网络(Convolutional Neural Network, CNN)对原始图像进行特征提取,然后对特征图进行候选框推荐,对于更可能是目标物体的给与更高的概率完成候选区域的推荐。同时对特征图进行窗口的选择,由于推荐的窗口大小不同,采用自适应池化层进行统一,最后固定检测框在全连接层实现回归分类任务。

2.2 RPN

RPN用于生成网络建议框,在输入图片时,产生候选框并进行定位显示分类概率。将RPN和Faster-RCNN进行训练,从而得到输出^[4]。对于Faster-RCNN,没有固定的输入原始图片尺寸,但通常将输入的图片短边设置成600。如对于一张1200×1800的图片,会把图片缩放到600×900上。对于所输出的特征图尺寸进行五次成倍数裁剪,这里有两个基本的模块,一是Conv Block,为了改变网络维度;另一个是Identity Block,用于加深网络。为在图片中生成候选区域,要在输出的特征图上进行滑窗选择。针对特征图中的多个窗口,分别要预测出k个目标候选区域,称为“锚”(Anchor)。每个anchor有相对应的尺寸。如在3×3的滑窗的中心处,根据三种不同的长宽比得到九个不同尺寸的矩形三种形状尺寸的anchor。使得精确度有所提升。

2.3 改进RPN

在一张图片中,真正需要检测的目标不会占据太大面积,这会产生过多的负样本,并且固定的anchor尺寸使检测受限,

所以提出使用特征去指导每个矩形框的中心,自行生成长宽比,既可以节约时间,又可以生成合适的长宽比,既可以减少生成anchor的时间,又可以生成合适的长宽比,这样对于多变的人体姿态检测效果更好。在RPN结构中每一层都加入Guided Anchoring模块,作用是对特征图进行建议框预测,其中的一个anchor generation模块作用是矩阵框的获取不同于以往的滑窗,是要在预测特征图的每一个点的像素作为目标置信度,作用是降低负样本的数量,并且高和宽的比例尺寸不是固定不变的,用CNN的方法进行回归这样与行人的尺度可以更好地对齐。另一个模块是Feature adaption模块,将anchor的高和宽转换成一个有两个通道的特征图,之后再次对该特征图进行卷积,即特征重采样。用该方法通过RPN和检测网络共享卷积层的方法缩减图像中区域建议框的计算时间来提高检测精度。

3 结论

为了提高对于行人的检测能力,本文以Faster R-CNN模型为基础,提出对RPN中的模型结构进行改进,可提高行人检测的精准度,并且去除了重复性计算,使得综合性能有不错的效果。

参考文献

- [1]Zhang SS, Benenson R, Schiele B, et al. CityPersons: A diverse dataset for pedestrian detection. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2017, 4457-4465.
- [2]SERMANET P, KAVUKCUOGLU K, CHINTALA S, et al. Pedestrian detection with unsupervised multi-stage feature learning [C]. 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2013: 3626-3633.
- [3]谢林江, 季桂树, 彭清, 等. 改进的卷积神经网络在行人检测中的应用[J]. 智计算机科学与探索, 2018, (05): 708-718.
- [4]REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]. Advances in Neural Information Processing Systems, 2015: 9199.