

基于YOLOv4的课堂姿态检测

侯百川 周丽

四川大学锦城学院

DOI:10.12238/mef.v4i8.4004

[摘要] 本文采用了一种基于YOLOv4网络结构的姿态检测方法,先通过采集学生姿态图像数据建立数据集,再通过YOLOv4网络对数据集进行训练,从而获得权重文件也就是预训练模型,最后调用权重文件实时判断同学们上课的姿态并进行画框操作。实验结果表明,该方法能够有效辨别同学们姿态变换,但还需要继续深入研究对比,从而找出效果最好的姿态检测模型。

[关键词] YOLOv4; 姿态检测; 智慧课堂

中图分类号: TP391.41

文献标识码: A

Classroom Posture Detection Based on YOLOv4

HOU Baichuan, ZHOU Li

Jincheng College, Sichuan University

[Abstract] In this paper, a posture detection method based on YOLOv4 network structure is adopted. Firstly, the data set is established by collecting the students' attitude image data, and then the data set is trained by YOLOv4 network, so as to obtain the weight file, which is also the pre-training model, and then the weight file is called to judge the students' attitude in class in real time and carry out frame operation. The experimental results show that this method can effectively identify the posture transformation of students, but further study and comparison are needed to find out the best posture detection model.

[Key word] YOLOv4; posture detection; wisdom classroom

1 引言

随着当前社会对人才的需求以及对教育的关注度不断提高,年轻人受教育的程度也不断提升。老师如何积极投入课程教学的同时,也能了解到每一位同学的上课情况成为了当前受关注的问题。智慧课堂将人的智慧与信息智能相结合,使整个教学过程有利于学生的个性化成长,以达到引导学生智慧能力充分发展的目的。在目前的智慧课堂领域,大多数智慧课堂应用是针对辅助老师课堂教学,一些仪器如课堂上平板、投影仪等的应用,软件如学习通、钉钉等的应用,而缺乏老师课后了解学生课堂变化状态的辅助软件。因此,需要一种能有效捕获学生课堂变化状态自动检测与识别的有效方法。

深度学习是通过学习找出样本数据的规律,从而获得分析学习能力,能够

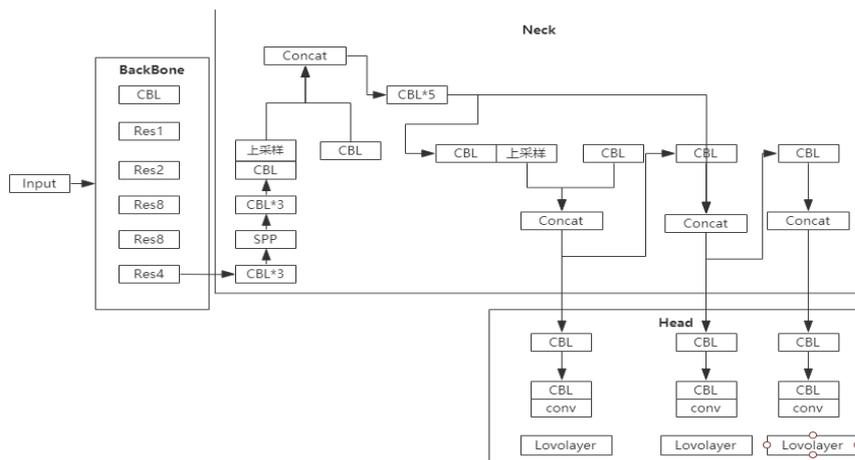


图1 YOLOv4 组成结构

识别偏向样本数据规律的一类目标。现如今主流的目标检测算法可分为两种,一种是以SSD、YOLO为代表的One-Stage算法;另一种是以Faster R-CNN为代表的Two-Stage算法。One-Stage算法具有高检测速度的特点,而Two-Stage具有高

精度的特点。YOLO系列算法是目前使用最多的目标检测算法,其中YOLOv4因具有高检测速度、较高精度的特点而被广泛应用,所以本文使用YOLOv4网络结构来训练模型。

2 YOLOv4算法

2.1 YOLOv4概述

YOLOv4模型由Alexey Bochkovskiy于2020年提出,算法处理速度快,精确度较高,常用于多目标实时检测系统。YOLO将物体检测作为回归问题来求解,是一个独立的end-to-end网络,完成原始图像的输入经过预训练模型的处理过程到输出结果图像的网络结构。YOLOv4的主干特征提取网络、数据增强、特征金字塔、激活函数等方面在YOLOv3上进行了改进和优化,在检测速度和精度方面都有了显著的提升。

2.2 YOLOv4组成结构

YOLOv4算法由输入端Input、主干特征提取网络层BackBone、特征金字塔中间层Neck以及输出层Head四个部门组成, YOLOv4组成结构如图1所示:

输入端Input: 输入原始图片,通过图像预处理阶段,将输入的图片缩放到网络的图像大小608*608*3并进行归一化等操作。在网络训练阶段,使用Mosaic算法对图像进行数据增强操作从而提高模型的训练速度与精度, Mosaic算法具有丰富数据集和减少GPU的优点。mosaic数据增强是在原有Cut Mix数据增强方法的基础上提出的新的数据增强方法,理论上和Cut Mix类似。即将四张图片使用随机缩放、随机裁剪、随机排布的方式拼接。

主干特征提取网络层BackBone: 采用Mish函数+CSPDarknet53+Dropblock的组合提取一些通用的特征表示。由输入端进入主干特征提取网络的第一个组件CBM, CBM由Conv+Bn+Mish组成, Mish函数相对于其他激活函数更加平滑,提升了模型的精度。通过CBM后进入由5个CSP模块组成的CSPDarknet53网络结构,经CSPDarknet53网络结构特征图变化规律是608->304->152->76->38->19, CSPDarknet53网络结构具有保持准确性、降低计算瓶颈、降低内存成本的优点。在其中加入Dropblock缓解过拟合。

特征金字塔中间层Neck: 通过主干特征提取网络层BackBone后进入第一个特征金字塔中间层组件CBL, CBL由Conv+Bn+Leaky relu组成。后进入SPP模块,

SPP模块由三个最大化池层平行组成,在对经池化后的特征图进行Concat操作,通过SPP模块通过融合不同大小的最大化池层以更有效的增加主干特征的接受范围。后通过FPN+PAN结构, FPN+PAN组合模块通过自顶向下来捕获强语义特征和自底向上传达强定位特征来完成目标定位功能。组成在主干特征提取网络层BackBone与输出层Head之间添加SPP模块和FPN+PAN2种方式以获得更健壮的特征表示。

输出层Head: 然后通过CIOU_Loss损失函数与DIOU_NMS来进一步提升算法的检测精度,从而得到输出结果。

3 系统实现

3.1 数据采集

本文以班级拍摄的多姿态录像通过每隔5帧截取一次图片为基础建立数据集,由于根据老师教学方式的多样性在课堂出现的情况也是复杂多样,所以所选研究对象姿态为sleep、practice、usephone、study这些常见的四个大类, sleep以个人姿态睡觉为目标; practice以个人姿态练习为目标; usephone以个人姿态玩手机为目标; study以个人姿态认真听讲为目标。具体类别还需根据年纪的不同来分析,不同年纪会出现不同的姿态变换,如小学为不认真听讲、讲话、认真听讲等,初中过后便会出现玩手机、看小说等情况。视频数据总时长为1h30min,为了使类别标记数量均衡,所以每种姿态都专门拍摄了不同场景的录像。截取图片大小为1920*1080,数据采集如图2所示。



图2 数据采集

3.2 数据标签

根据实验要求我们对图像进行预处理,首先由于截取图像名称规律变换

不清晰,为方便使用YOLOv4做预训练模型,对图片名称按顺序进行更名,这样方便通过遍历文件名的方式调用文件,其次我们将所截取的图像大小调整为608*608,然后通过图像标注工具Labelme对图像进行人工打标签操作,以个人姿态为目标进行打标签,标签名称为sleep、practice、usephone、study四类。将所打标签的图片转换为json格式。由于在打标签的过程种发现study类的样本数据过低防止出现欠拟合的情况,我们对study类样本进行了补录。最终建立的课堂姿态数据集共6875个样本。Labelme如图3所示。

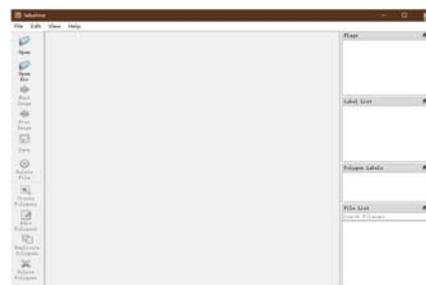


图3 Labelme

3.3 训练

实验硬件配置: 处理器Intel (R) Core(TM) i7 10870H @2.90GHZ 2.90GHZ, 独立显卡NVIDIA GEFORCE RTX 2060。实验软件环境: Python3.6, pytorch1.6, Cuda11.1。本实验使用的是YOLOv4源代码来训练预训练模型。修改配置文件cfg.py, 设置为不适用darknet、批量batch设为1、分部设为1。修改配置文件,将filters设为24,将三层yolo的类别设为3。接着我们开始训练,预训练模型进行60次迭代,将学习率设为0.001、预训练的主干网络yolov4_conv_137,开始训练,从而得到Epoch(被训练次数)60的预训练模型。

3.4 实验结果与分析

为了验证YOLOv4训练模型的效果,我们调用一张图片进行测试,先创建class.names文件,里面的类别写入为自己数据集的类别,顺序与预训练模型时的classes顺序一致。使用我们训练所得到的预训练模型。得到的YOLOv4预训练模型检测效果图如图4所示。



图4 YOLOv4预训练模型检测效果图

从图4可以看出,在迭代60次下的预训练模型能清晰的框出每个人的大致姿态,但对于姿态的准确性还有待提升。所以为提升准确性,我们可以适当增加迭代次数,调整学习率,或增加数据集来防止欠拟合。当准确性过高时,该功能效果会有显著提升。且为保证准确率,我们可通过迁移学习的方式,使用多种方法,如Faster R-CNN、SSD等网络模型在相同参数下的训练准确率,从而使效果更好。Faster R-CNN是一种基于区域建议深度学习模型的目标检测方法。SSD网络是一个基于卷积神经网络的目标检测算法。选择哪种网络模型在保证速度的同时又能达到我们预期的精度要求仍是未来有待研究比较的问题。我们将训练效果最佳的模型运用到摄像头,通过每分钟调用一次课程姿态检测来检

测学生们的姿态,再将姿态变化绘制成折线图,使变化情况更加直观。且类别只分类,对同学们上课时的姿态变化类别不完善,可以增加更多的类别。

4 应用前景

与传统高等数学教学相比,混合式教学更加注重自学能力的培养,教师作为引导者,引导学生进行有效的自主学习。现如今传统高等教学方式已经逐渐被淘汰,混合式教学方式越来越受大家的追捧。混合式教学包含课前、课中、课后三个阶段,本文所提到的功能主要运用于课后,我们通过YOLOv4所训练出来的姿态检测功能,将一堂课中学生的姿态变换情况绘制成折线图。这样老师下课后能直观的了解学生课堂上的表现情况,可以使老师发现没有认真听讲的同学,单独约谈进行交流沟通;也可以改变教学方式,在同学们注意力最集中的时间段进行重点知识的讲解,在注意力最涣散的时间段调动学生的注意力;还可以把相同注意力曲线变换的同学分为一个班,进行一班一方法的因地制宜的教学方式。现代化的辅助教学方式能够促使学生学习更加有效。

[参考文献]

[1]吴婕.信息化环境下智慧课堂的构

建策略[J].文学教育(下),2021(06):59-61.

[2]周宸,高伟,郭谋发.基于YOLOv4模型的玻璃绝缘子自爆缺陷识别方法[J].电气技术,2021,22(05):38-42+49.

[3]黄海新,金鑫.基于YOLOv4的小目标缺陷检测[J].电子世界,2021(05):146-147.

[4]Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2017, 39(06): 1137-1149.

[5]刘顺民,马致明,陈炳才.基于SSD算法的交通指示牌目标检测[J].现代电子技术,2021,44(13):144-147.

[6]陈宏健.基于智慧教育背景下的高等数学混合式教学研究[J].福建轻纺,2021(07):33-35+45.

作者简介:

侯百川(1999--),男,汉族,四川巴中人,四川大学锦城学院2018级计算机与软件专业本科在读。

周丽(1984--),女,汉族,四川成都人,讲师,硕士,研究方向:图像处理、深度学习。

