

湖州师范学院

2023 届毕业设计(论文)

课 题 名 称: 基于 YOLOv7 模型的口罩佩戴检测技术研究

课 题 名 称(英文): Mask Wearing Detection Technology
based on model YOLOv7

学 生 姓 名: 严嘉铭 学 号: 2019082429

专 业 名 称: 计算机科学与技术

指 导 教 师: 高强 职 称: 讲师

所 在 学 院: 信息工程院校

完 成 日 期: 2023 年 3 月 27 日

基于 YOLOv7 模型的口罩佩戴检测技术研究

摘要: 为了保证疫情环境下人民身体健康, 在一些场合下佩戴口罩是非常必要的。传统的口罩佩戴检测方法需要人工干预, 效率较低, 而基于深度学习技术的口罩佩戴检测技术可以实现自动化, 大大提高检测效率。

早期的基于深度学习的检测算法主要基于卷积神经网络, 由于这些算法虽然检测精度较高但速度较慢, 难以满足实时检测的需求。YOLO (You Only Look Once) 模型具有检测精度高, 运算量小等优点, 非常适合实时检测。因此本文研究基于 YOLOv7模型的口罩佩戴检测技术, 通过选择目标检测算法 YOLO 算法进行口罩佩戴检测技术的研究, 在 Pytorch 框架上构建算法模型, 并搜集了8000多幅人脸面部口罩佩戴图像对模型进行训练和测试。并统计该模型的平均检测精度 (mean Average Precision, mAP) 以及召回率 (Recall Rate) 等指标。

最后分析了算法的检测结果和性能, 并对实验进行了总结。实验表明, YOLOv7模型对于所有目标的整体 mAP 达到了83%, 召回率达到了81%, 说明该模型具有较高的检测精度。

关键词: 深度学习; 人脸目标检测; 口罩佩戴

Mask Wearing Detection Technology based on model YOLOv7

Abstract: In order to ensure the health of people in the epidemic environment, it is necessary to wear a mask on some occasions. Traditional mask wearing detection methods require manual intervention and are inefficient. However, mask wearing detection technology based on deep learning technology can achieve automation and greatly improve detection efficiency.

Early detection algorithms based on deep learning were mainly based on convolutional neural networks, which were difficult to meet the needs of real-time detection due to their high detection accuracy but slow speed. YOLO (You Only Look Once) model has the advantages of high detection accuracy and low computational complexity, making it very suitable for real-time detection. Therefore, this paper studies the mask wearing detection technology based on the YOLOv7 model. By selecting the target detection algorithm YOLO algorithm, the mask wearing detection technology is studied. An algorithm model is constructed on the Python framework, and more than 8000 facial mask wearing images are collected for training and testing the model. The mean Average Precision (mAP) and Recall Rate of the model were calculated.

Finally, the detection results and performance of the algorithm are analyzed, and the experiments are summarized. The experiment shows that the overall mAP of YOLOv7 model for all targets reaches 83%, and the recall rate reaches 81%, indicating that the model has high detection accuracy.

Keywords: Deep Learning, Face Target Detection, Wearing Mask

目录

第一章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状及发展趋势	1
1.2.1 国内外研究现状	1
1.3 研究内容	3
1.4 本章总结	4
第二章 基本原理介绍	5
2.1 传统目标检测技术介绍	5
2.2 YOLO 系列模型原理	5
2.2.1 基本思想	5
2.2.2 交并比	5
2.2.5 非极大值抑制	6
2.2.6 边界框回归	6
2.3 YOLOv7 模型基本原理	8
2.3.1 网络结构	8
2.3.2 模型结构重参数化	9
2.3.3 正样本分配策略	11
2.3.4 AUX 辅助输出	13
2.4 本章小结	13
第三章 仿真实验	14
3.1 样本数据集的选择和构建	14
3.2 实验方案	15
3.2.1 机器配置	15
3.2.2 实验步骤	15
3.2.3 参数设置	16
3.2.4 实验指标	17
3.3 实验验证与分析	17
3.3.1 实验训练结果	17
3.3.2 口罩佩戴检测识别结果	19
3.3.3 结果分析	20
3.4 本章小结	20
第四章 总结与展望	22
4.1 总结	22
4.2 展望	22
参考文献	23
致谢	24

第一章 绪论

1.1 研究背景及意义

2019 年底 2020 年初,新冠病毒的爆发对全球人民都是一次巨大的灾难^[1],作为应对疫情传播的重要防护作用之一,在公共场所佩戴口罩是每个人民都需做到的责任^[2]。除此之外,在医院、诊所、实验室、无尘车间等特殊场合都规定必须佩戴口罩。口罩佩戴检测技术是目前疫情防控中的一个重大工作项目。

对于口罩检测技术来说,国内外对于口罩佩戴检测识别算法都是基于人脸检测识别系统的开发进行的。目前,针对口罩检测技术面临着众多的难题,例如,单个个体的口罩佩戴情况、多人群体的全部口罩佩戴情况,正面佩戴口罩能否成功检测以及非正面佩戴口罩能否成功检测,另外还有口罩种类的多样性以及非口罩遮挡面部等问题所导致的丢失目标、产生错检漏检等问题给民众带来的影响,同时还伴有公共场所需要的实时监测要求等问题。

传统的人力检测既费力又费时,不仅消耗大量的人力资源,还容易产生一定误差,将其他遮挡物错认成口罩,出现漏检错检的问题,存在较大的安全隐患。对既需高精度又需实时性的口罩佩戴检测的带来巨大的挑战。因此研究基于深度学习的口罩佩戴检测技术具有较高的实用价值。针对这些问题,随着 YOLO (You Only Look Once) 系列模型提出,作为具有能进行实时检测的模型,基于 YOLOv7 模型的口罩佩戴检测的目标检测算法的应用能很好的进行解决,且具有便捷性,高效性,实时性和系统性,具有广阔的发展前景。

1.2 国内外研究现状及发展趋势

近些年,由于 YOLO 系列模型的发展和改进,使得最新的 YOLOv7 系列的全类平均正确率 (mean Average Precision, mAP) 以及每秒传输帧数 (Frame Per Second, FPS) 相比前几代模型都得到提高,对于口罩佩戴检测识别领域,国内外众多学者相继提出了多种目标检测方法。目前,口罩佩戴检测的研究算法已经有了巨大的成果。

1.2.1 国内外研究现状

在众多的目标检测识别技术中,YOLO 目标检测识别算法相较于其他网络,具有实时性高而且精度比其他实时检测网络高 1 倍的特点。同时 YOLO 能够通过检测整张图片,避免对背景的错误识别,以及能够学习到高度的泛化特征的特点。能够很好的解决高精度和实时性的需要。

近些年,随着深度学习的不断发展和优化,基于卷积神经网络 (Convolutional Neural Network, CNN) 的 YOLO 目标检测识别技术进一步得到发展。YOLOv7 模型作为 YOLO 系列最新的模型,速度和精度相比之前模型分别提高了 509% 和 2%,凭借其更准确的检测精度,更快更强的网络结构以及更精确的损失函数被大量应用于各种目标检测问题。

目前,在计算机视觉研究中,人脸检测技术既要追求准确性又要兼顾实时性一直是一个较为困难的问题,精度较高就代表着需要大量的计算,传统的人脸检测识别很难同时做到。2017 年人脸检测模型 FaceBoxes 的出现作为一种具有高精度的 CPU 实时人脸检测器^[3],在确保实时性的同时保持着较高的精度,并且不需要关注场景中人脸的数量。在这之后,2020 年 CenterFace 网络的出现^[4],可以对人脸检测和关键点检测进行同时实现,并且其精度和检测速度均高于之前主流的 FaceBoxes 模型。

目前,基于深度学习的目标检测算法逐步取代了传统的人脸检测识别算法,不再通过人工提取特征

的方式,成为了目标检测技术的主流。2016年,Redmon等提出了 You Only Look Once 目标检测模型^[5],简称 YOLO。YOLO 算法是一种对目标进行实时检测的卷积神经网络,是一种基于回归问题的算法。将图像划分为 $N \times N$ 的多个区域,并预测每个区域的边界框,选择输出具有最大概率的类,然后对选择的边界框进行非极大值抑制(Non-Maximum Suppression, NMS)操作,根据置信度阈值过滤置信度较低的边界框,保留与目标的真实框(Ground Truth)最接近的那一项,即为最终的检测结果。虽然 YOLO 模型的最大优点在于对目标检测的检测速度较高,能够达到 45fps,但是因为输出的是 NMS 操作消除后,置信度最高的边界框,所以对于场景中那些小目标的检测精度并不高,mAP 值只有 66.4%。伴随着 YOLO 系列的提出,越来越多的 One-Stage 算法被提出,2016年,Liu等提出的 Single Shot MultiBox Detector 目标检测模型,简称 SSD^[6]。可以直接产生物体为目标种类的概率和坐标信息,只需经过单次检测就能达到最终的检测结果,相比 YOLO,检测精度和检测速度都更加优越。2017年,因为 YOLO 算法的检测精度不足,Redmon等又在 YOLO 的基础上进行改进提出了 YOLOv2 模型^[7]。在每个卷积层后加入批量归一化层(Batch Normalization, BN),防止模型过拟合,使得模型的收敛速度得到提升;又通过高分辨率预训练分类网络(High Resolution Classifier)将图片分辨率提高为 448×448 等操作,相较于之前 YOLO 算法的 mAP 值获得了不小的提升。2018年 Redmon 等又进而提出了 YOLOv3 模型^[8],调整了网络结构,使用新的网络结构 Darknet53,对于距离位置相近的多个目标之间的检测具有更好的精度。2020年,发布的 YOLOv4 模型在 YOLOv3 的基础上,骨干网络采用了 CSPDarknet53,引入 spatial pyramid pooling network 网络结构简称 SPP-Net,不再只能输入固定尺寸的图片,而是通过提取输入的图像的特征图,再从特征图中提取特征,最终使得检测速度和检测精度都发展到了较高水平,mAP 相比 YOLOv3 提高了 10%,检测速度提升了约 12%^[9]。2022年6月,由于 YOLOv5 等方法在效率和速度上仍然有很大的提升空间,所以美团提出了 YOLOv6 模型^[10]。通过训练策略、测试策略、量化和参数优化方法等,以便适用于不同场合和不同尺度部署网络,在检测精度和检测速度上都超越了之前的检测模型。2022年7月,YOLOv7 模型的发布相比之前的 YOLO 系列模型精度更高^[11],检测速度快了 120%。设计了几种可训练的 bag-of-freebies 简称 BoF,使实时检测器在不增加模型的复杂度和推理的计算量的情况下大大提升模型的精度和检测速度。解决了目标检测中的重要优化方向:模型结构重参化和动态标签分配。

2020年2月,牛作东等引入自注意力机制,发表改进 RetinaFace 的自然场景口罩佩戴检测算法^[12]。相较之前提高了算法的 mAP 达到 87.7%,FPS 为 18.3。但是对于场景中具有的多目标和小目标人群以及面部遮挡等问题,并没有达到很高的适用性与泛用性。

2020年11月,基于 YOLO 改进算法的安全帽和口罩佩戴自动同时检测^[13],孙世丹等针对工地等危险区域的安全帽与口罩检测佩戴通过优化 K-means 聚类算法,改进 YOLOv3 模型处理分析数据集,不仅提高了检测的 mAP 和 FPS 使其分别达到 86.9%和 45fps,能够较为精确的处理原本画面中遇到多目标和小目标检测精度较低的情况,体现了其实用性,具有商业价值。

2020年12月,李国进等在 YOLOv3 的基础上通过替换主干网络为 ResNet50 残差网络,并且引入可变形卷积 DCN 和通道注意力机制 SENet,提出基于 DCN-SERes-YOLOv3 的人脸佩戴口罩检测算法^[14]。较好的应对了面部遮挡和小目标检测的误检和漏检问题,精度和检测速度也达到了 91.38%和 59.6fps,具有更好的应用前景。

2021年3月,任小康等的基于改进的 YOLOv3 口罩佩戴检测和识别^[15],虽能相比之前检测出被严重遮挡的人脸目标,但对于遮挡率超过 90%的被严重遮挡的口罩目标依然存在漏检行为,其 mAP 为 58.05%,FPS 为 13.2,略显不足。同时模型的复杂度较高,其单次卷积运算需要的十亿次浮点运算数 BFLOPs 均高于之前的 YOLOv2, YOLOv3 模型。

2021年,刘子龙等的基于改进后的 SSD 人脸口罩佩戴识别方法研究与典型应用的发布^[16],解决了

传统 SSD 网络的鲁棒性较差,精度和实时性不足等缺点。将主干网络替换为 DenseNet 网络,并引入特征金字塔模块,使得改进后的 SSD 网络的 mAP 达到 92.25%,FPS 达到 41.2,具有较好的精确性和实时性。

2021 年 3 月朱杰等提出的基于 YOLOv4-tiny 改进的轻量级口罩检测算法^[17],通过改进网络结构,在网络结构中引入金字塔池化结构 SPP 模块实现不同尺度的特征融合,使得对口罩目标的 mAP 达到了 94.7%,FPS 达到了 76.8。在精度和实时性方面都达到了较高水平。但是难以处理复杂多变的情景,对于复杂的场景情况下的人脸识别识别的精度较为不足,mAP 只有 85.7%。

2021 年 8 月,薄景文等基于 YOLOv3 的轻量化口罩佩戴检测算法的提出^[18],把原来 YOLOv3 的 Darknet-53 网络结构替换为 ShuffleNetv2 网络,并且引入 SKNet 注意力机制到特征融合网络,同时采用 CIou 作为边界框回归损失函数。在保持 mAP 为 93.38%的同时 FPS 提高为了 59.09。比传统的 YOLOv3 提高了近 34fps。但是对于复杂背景下的目标检测能力不足。

2021 年 12 月,曹小喜等提出的基于改进 YOLOv4 的密集人群口罩佩戴检测算法^[19]。针对密集人群的面部遮挡,遇到高曝光和小目标检测时的漏检误检等难题,通过引入通道注意力机制 SEnet 和复用空间金字塔池化,分别增强了算法的对有用特征的利用能力和算法在不同感受野提取并融合空间特征的能力。使得算法相较于 YOLOv4 而言,mAP 达到了 94.81%得到了不小的提升,但是 FPS 为 39 略低于 YOLOv4 算法,实时性相比略显不足。

2022 年 3 月,项融融等的基于改进 YOLOv5s 的口罩佩戴检测算法的提出^[20],通过引入自注意力机制和改变 Neck 层的卷积结构,提高模型的显著特征,采用基于双尺度的特征融合目标检测,更好地进行特征提取。增强了对遮挡目标的检测能力,对于遮挡目标和小目标检测更具有高效性,精确性和快速性。但是对于光线不足的场景下的目标检测效果不够理想。

2022 年 6 月杨震山等发表的口罩佩戴智能检测算法研究^[21],通过构建具有多样性的口罩佩戴检测数据集解决了对口罩多样性的数据匮乏以及没有公开带有标注的数据集这类问题,并且设计了两种检测方法。一种是基于 YOLOv4 的 RFBNet-s 模型,其 mAP 达到 99.3%和 FPS 值达到 37,均达到了较高水平,相比于之前的目标检测算法既能够保持更高的精度,又能达到更高的检测速度。另一种 YOLO-Mask 模型虽然 mAP 值在 95.6%相较于 RFBNet-s 模型较低,但是 FPS 值达到了 67.6 是 RFBNet-s 模型的 1 倍以上,用更为高效的实时性来换取精度的较为降低。两种设计模型分别满足更高精度和更实时性的要求,很好的避免了之前检测过程中的漏检误检和实时性较低的问题,为疫情防控带来了显著的帮助。但是对于口罩佩戴是否完全合规,即完全遮挡口鼻而非露出鼻子或未完全遮挡嘴巴等问题,以及如何兼顾更高精度的同时大幅度提升实时性还未能很好地进行解决。

基于 YOLOv7 模型的口罩佩戴检测技术是不仅仅是疫情防控条件下的一个重要组成部分,也是医院,诊所,实验室等地点的重要组成部分。对于保障人民在疫情环境下的安全,提高人民生活质量,减轻口罩佩戴检测的人力资源负担,改善人民生活环境等方面处于关键性的地位,起到十分重要的作用。

1.3 研究内容

本课题基于 YOLOv7 模型的口罩佩戴检测识别算法,先介绍了研究所需要的基本原理以及 YOLOv7 模型的基本原理。然后使用 Pytorch 深度学习框架和 YOLO 开源的口罩佩戴数据集对算法模型进行训练和测试,得到最优的权重模型。接着通过口罩佩戴数据集对训练好的模型进行验证。最后通过分析平均检测精度和召回率等指标对算法模型的性能进行评价分析。本课题的研究方案图如图 1-1 研究方案所示。

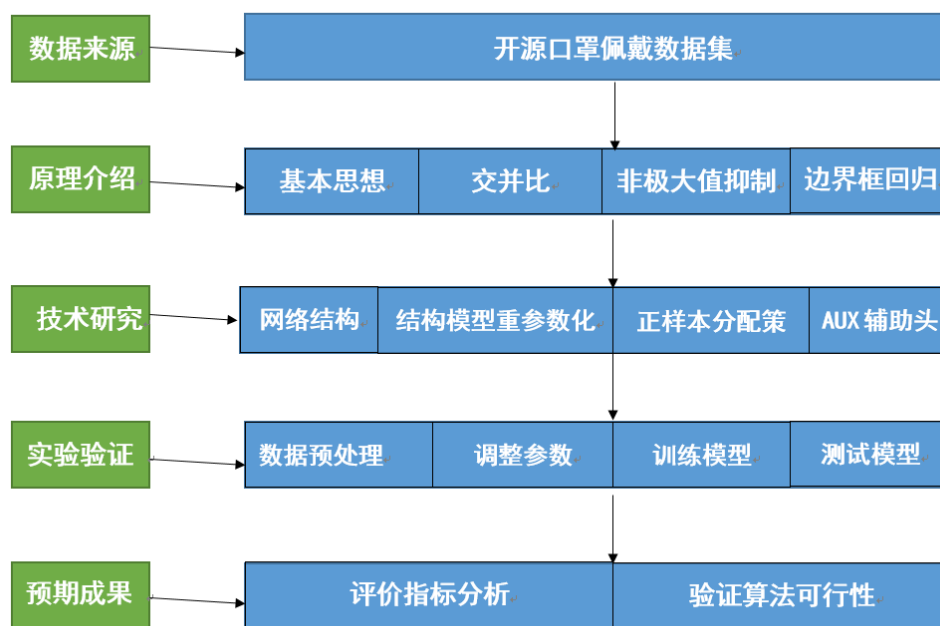


图 1-1 研究方案

1.4 本章总结

本章介绍了本课题的选题意义，同时介绍了对于 YOLO 模型和目标检测识别技术的国内外研究现状，还分析了基于 YOLOv7 模型的目标检测识别技术在当前环境下的发展趋势，对于本课题关于基于 YOLOv7 模型的口罩佩戴检测技术的研究有着重要的作用。

第二章 基本原理介绍

2.1 传统目标检测技术介绍

目标检测技术是通过给定一张图像，能够识别出图像中的目标是否需要预测的目标，并且能够标记出目标的精确位置信息。在基于深度学习的目标检测技术还未出现的时候，传统的目标检测技术基于传统的手动设计特征，并结合滑动窗口的方式来进行目标检测和定位，大致分为三个部分：区域选择，特征提取，分类器分类。但传统的目标检测技术具有特征设计困难，效率低，对于某些特定的条件不适应，鲁棒性差；通过滑动窗口的方式来提取候选框，对候选框进行分类判定，导致时间复杂度高和窗口冗余度高等问题。传统目标检测技术基本流程图如图 2-1 所示。

2012 年，卷积神经网络（Convolutional Neural Network，CNN）兴起之后，利用卷积神经网络进行特征提取，基于深度学习的目标检测技术逐渐替代了传统目标检测技术。但是并没有从本质上改变候选框提取目标区域的策略，因此时间复杂度依然较高，2015 年，提出通过区域生成网络（Region Proposal Network）来代替原先的滑动窗口的策略。自此，基于深度学习的目标检测技术彻底完成了一个端到端的过程，不仅在检测精度上，还包括检测速度上，均达到了一个非常大的提升。

直到后来，基于回归方法的目标检测识别，直接回归候选框位置的方法来完成目标的检测和定位。在确保检测精度的同时，又进一步提升了目标检测技术的检测速度。

至此，基于深度学习的目标检测方法也主要分为了，一阶段目标检测算法（One-Stage）和双阶段目标检测算法（Two-Stage）两条线路。双阶段目标检测算法（Two-Stage）即基于候选框选取的深度学习目标检测技术，一阶段目标检测算法（One-Stage）即基于回归方法的深度学习目标检测技术。

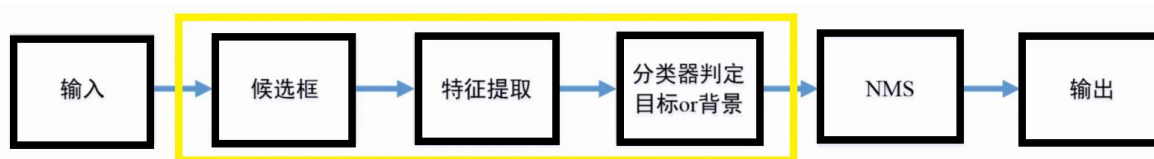


图 2-1 传统目标检测基本流程

2.2 YOLO 系列模型原理

2.2.1 基本思想

YOLO 是典型的一阶段目标检测算法（One-Stage），其基本思想就是用卷积神经网络（Convolutional Neural Network，CNN）结构来预测输入图像的边界框（bounding box）和物体类别概率，将物体检测问题转化为回归问题，采用回归的方法进行目标检测识别。

YOLO 算法的基本流程为：

- 1) 先调整输入图像的大小。
- 2) 再运行卷积神经网络 CNN。
- 3) 最后通过模型的置信度值对检测结果设置阈值。

2.2.2 交并比

交并比（Intersection over Union, IoU）是目标检测中的一个重要概念，是衡量定位精确度的一种方式。用来测量目标检测过程中预测框与真实框的重合度，即预测框和真实框的交集和并集的比值。即：

$$IoU = \frac{A \cap B}{A \cup B} \quad (1)$$

IoU 的比值越高,说明预测框与真实框的重合度越高,也就代表预测定位更精确。如果 IoU=1 就代表完全重合,图 2-2 为交并比。

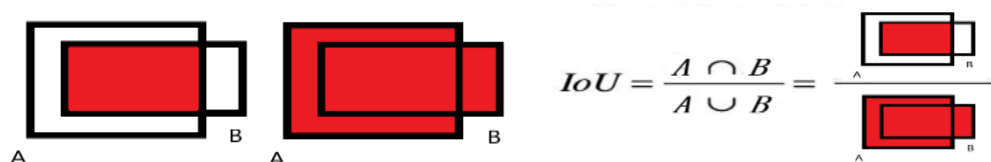


图 2-2 交并比

2.2.5 非极大值抑制

非极大值抑制 (Non-Maximum Suppression, NMS), 是用来抑制极大值之外的其他元素, 是一种局部最大搜索的策略。在目标检测识别过程中, 会出现对于一个物体的识别存在多个预测框, 并且可能会出现交叠的概率, 然后会对每个预测框计算置信度值并进行排序, 这时就需要运用非极大值抑制策略 NMS 来选取 IoU 的计算结果最优的预测框, 删除其余冗余的预测框, 即选择是待识别物体的概率最大的那一项。图 2-3 为非极大值抑制效果图。

非极大值抑制的流程为:

- 1) 先将所有的预测框按照置信度值进行排序。
- 2) 选择置信度值最大的那一类及其对应的预测框。
- 3) 遍历剩下的预测框, 计算置信度值最大的那一个预测框与其他预测框的交并比 IoU。
- 4) 如果交并比 IoU 大于一定的阈值, 就说明该预测框与置信度值最大的那一个预测框识别的是同一物体, 则需要删除该预测框, 目的是仅保留置信度值最大的预测框。
- 5) 回到 1) 进行迭代, 直至处理完所有预测框。

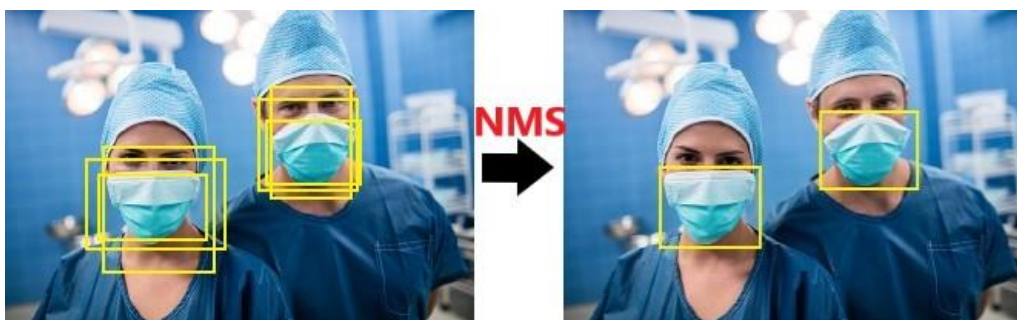


图 2-3 非极大值抑制效果图

2.2.6 边界框回归

边界框回归 (Bounding-Box Regression) 是目标检测识别中一个重要操作。在目标检测识别过程中, 不仅要识别目标的类别, 还要通过获取目标在原始图像中的位置信息来给出目标的精确定位。对于一个

待检测目标来说，有真实框（Groud Truth）和预测框（Anchor Box），即使预测框检测出目标的类别，但是真实框与预测框的交并比 IoU 的值低于阈值，则表示预测框的定位不够精确，没有正确检测出目标的位置。因此，我们需要运用边界框回归来调整目标的预测框，使得调整后的预测框的位置更加接近真实框，提高两者的交并比 IoU，以进一步确认预测框识别的物体是否是待识别的物体，同时提高检测的精度。

对于预测框和真实框，一般通过 (x, y) 来表示框的中心点坐标， (w, h) 来表示框的宽度和高度。如图 2-4 所示， G 为目标的真实框， P 为目标原始的预测框， \hat{G} 为使用边界框回归算法预测后的预测框。运用边界框回归的目的是通过某种函数关系映射，使得目标原始的预测框 P 经过映射得到一个更接近于目标的真实框 G 的一个预测框 \hat{G} 。即给定一组预测框 $P = (P_x, P_y, P_w, P_h)$ ，通过一个映射 f ，使得：

$$f(P_x, P_y, P_w, P_h) = (\hat{G}_x, \hat{G}_y, \hat{G}_w, \hat{G}_h) \approx (G_x, G_y, G_w, G_h) \quad (2)$$

边界框回归的基本流程是：

1) 先进行平移变换，即：

$$\begin{aligned} \hat{G}_x &= P_w d_x(P) + P_x \\ \hat{G}_y &= P_h d_y(P) + P_y \end{aligned} \quad (3)$$

2) 再进行尺度缩放，即：

$$\begin{aligned} \hat{G}_w &= P_w \exp(d_w(P)) \\ \hat{G}_h &= P_h \exp(d_h(P)) \end{aligned} \quad (4)$$

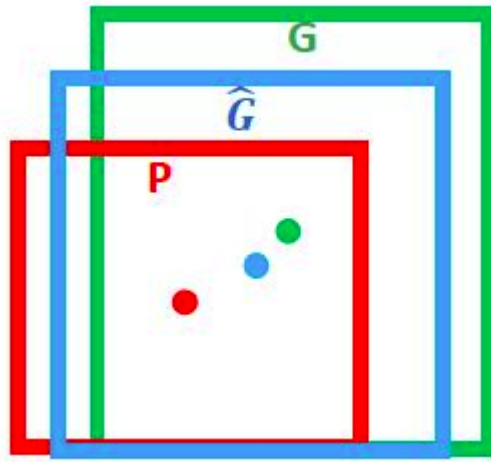


图 2-4 边界框回归

2.3 YOLOv7 模型基本原理

2.3.1 网络结构

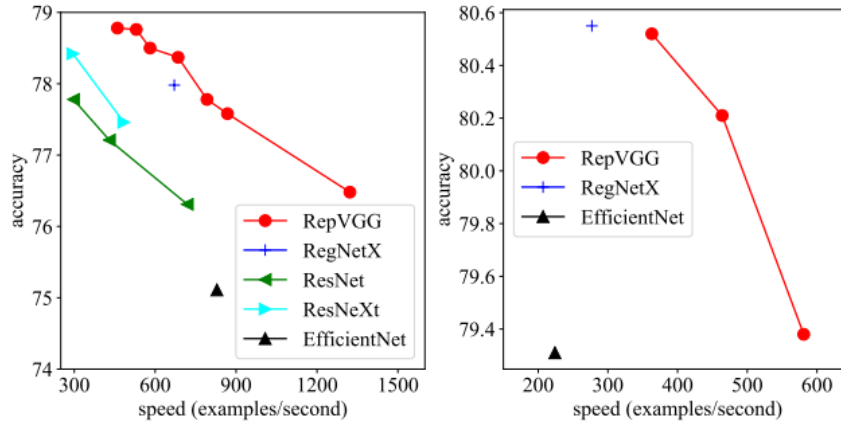


图 2-5 RepVGG 网络与其他网络的对比^[22]

YOLOv7 采用了 RepVGG 的网络结构。RepVGG 网络是一种类似于 VGG（Visual Geometry Group Network）的网络结构，只由卷积来组成，但具有一种 3×3 的卷积结构和线性整流函数（Rectified Linear Unit, ReLU）。与其他模型相比最大的特点是没有分支结构。具有高精度的同时实时性也高。ResNet 网络在残差结构中使用分支结构，但是使得显存占用率较大。而 RepVGG 网络沿用了 VGG 网络的顺序结构使得显存占用率小，同时作为一个可融合的网络，在模型进行推理的时候能够将多路的网络结构转变为单路的网络结构，提升了模型推理的速度。体现在 ImageNet 比赛中在实时速度下排名前三的精确度，其精确率达到 80% 多，比 ResNet-50 网络结构和 ResNet-101 网络结构分别快了 80% 多和 100% 多。同时 RepVGG 网络在相同的精确率下推理速度分别是 ResNet 网络的两倍多，是 EfficientNet 网络的三倍多。如图 2-5 RepVGG 网络与其他网络的对比所示。

YOLOv7 在网络架构引入了拼接，增加了 E-ELAN(Extended-ELAN)模块，MPConv 模块，SPPCSPC 模块和路径聚合网络(Path-Aggregation Network, PANet)。

- 1) E-ELAN (Extended-ELAN) 模块。YOLOv7 通过将网络结构中之前的几层得到的特征图作为输入，通过四个不同的链路，分别经过一个卷积层，一个卷积层，三个卷积层和五个卷积层，得到不同的特征图输出，并对最终结果进行拼接，将四个不同层级得到的特征图拼接在一起，得到一个新的特征图。使得经过 E-ELAN 模块后，对比于输入，特征图的通道数加倍。E-ELAN 模块的结构如图 2-6 所示。

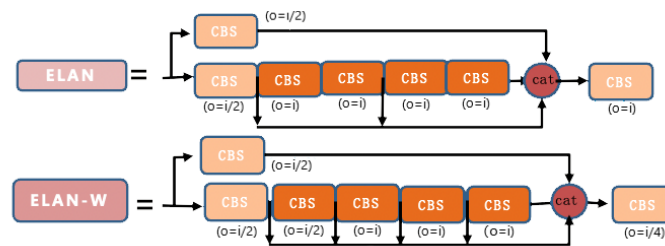


图 2-6 E-ELAN 模块的结构

- 2) **MPCov** 模块。之前的 YOLO 模型下采样的方法都是通过最大池化层 **Maxpooling** 或者通过卷积层 **Conv**，步长为 2 进行，得到的特征图大小均类似。YOLOv7 模型则是这两种方法都采用，综合考虑最大池化层和卷积层的下采样结果。一种结构是将最大池化层 **Maxpooling** 和卷积层 **Conv** 结合进行下采样，另一种结构是通过两种不同的卷积层 **Conv** 进行下采样，之后对这两种结构得到的特征图进行拼接。**MPCov** 模块的结构如图 2-7 所示。

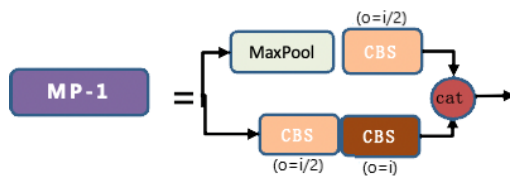


图 2-7 MPCov 模块的结构

- 3) **SPPCSPC** 模块。YOLOv7 模型采用了 **SPP** 的思想。之前的 YOLOv5 模型是通过三个不同的最大池化层 **Maxpooling** 得到不同尺度的特征，再进行拼接。而 YOLOv7 模型则是首先对特征图经过三次卷积层，之后同样分别通过 5×5 ， 9×9 ， 13×13 这三个不同的最大池化层 **Maxpooling** 得到不同尺度的特征外，再和原特征图这四者进行拼接，实现对不同感受野的特征融合。之后再进行一次卷积，再与进行一次卷积的原特征图进行拼接。不仅减少了一半的计算量，使得检测速度得到了提升，同时使得检测精度也得到提升。**SPPCSPC** 模块的结构如图 2-8 所示。

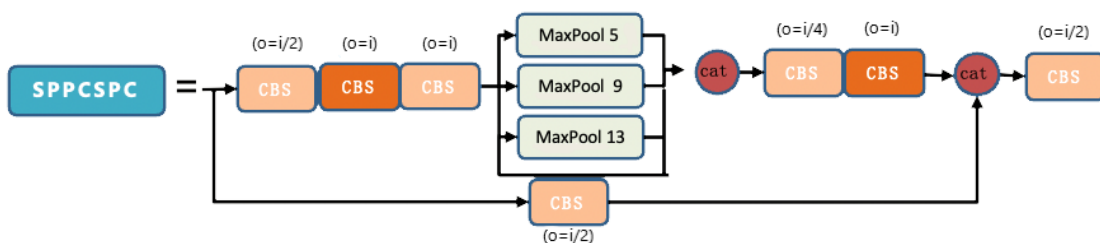


图 2-8 SPPCSPC 模块的结构

- 4) 路径聚合网络 (Path-Aggregation Network, PANet)。YOLOv7 模型继续采用了 YOLOv5 模型的 PAN，将经过 **SPPCSPC** 模块的特征图不断进行上采样，并与低层信息进行融合，实现低层信息与高层信息的特征融合。之后再进行一次下采样，与低层特征进行融合，实现高层信息与低层信息的特征融合。

2.3.2 模型结构重参数化

YOLOv7 模型允许训练网络结构与推理网络不相同。通过模型结构重参数化思想，将训练网络时的多路结构转化为推理网络时的单路结果，不仅提升了网络的推理速度，同时节省了空间冗余度。测试网络时将多路结构中的 3×3 的卷积， 1×1 的卷积和同等映射全部合并成 3×3 的卷积，之后将卷积和批处理归一化 (Batch Normalization, BN) 进行合并。网络结构在训练时不进行变化，只在训练结束之后，推理网络时将分支中的 3×3 卷积， 1×1 卷积和同等映射全部合并成 3×3 的卷积。模型结构重参数化的过程如下：

- 1) 卷积层与批处理归一化层的合并。先将 BN 层的计算公式拆解为形如卷积层的计算公式：

$$Conv_{(x)} = W_{(x)} + b \quad (5)$$

转变为：

$$f_j^i = W_{BN} * (W_{Conv} * f_j^i + b_{Conv}) + b_{BN} \quad (6)$$

使得：

$$\begin{aligned} W &= W_{BN} * W_{Conv} \\ b &= W_{BN} * b_{Conv} + b_{BN} \end{aligned} \quad (7)$$

使每一层卷积层都与紧接着的批处理归一化层进行合并，用一个卷积层代替原来的卷积层+批处理归一化层，因为只有训练过的批处理归一化层才能进行计算。卷积层与批处理归一化层合并之后本质上还是一个卷积层，但是权重考虑了批处理归一化层的参数。批处理归一化层使得模型便于收敛，加快了模型的收敛速度，获得 0-1 分布。同时引入了两个可学习的参数 γ 和 β ，参数 γ 使得数值不会过大或过小，特征不会过于集中或分散，参数 β 使得数值不会过左或过右，偏移 0-1 分布。这两个可学习参数使得数据可以分布的更好一些，更有利于机器进行学习。

- 2) 3*3 的卷积层和 1*1 的卷积层进行合并。把 1*1 的卷积核全部转换成 3*3 的卷积核，不仅可以提高运算效率，还可以与 3*3 的卷积层进行合并。为了确保计算结果不会发生变化，通过给卷积核填充，在 1*1 的卷积核周围加入一圈 0 来进行填充，变为 3*3 的卷积核的形式，然后与 3*3 的卷积核进行相加合并。同样也要对原始图像周围也通过加一圈 0 来进行填充相互对应。
- 3) **ResNet 残差转换**。将输入层直接等于输出层，即将输入的每个通道的每个元素直接输出到对应的通道中。通过将卷积核设定为：

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (8)$$

的形式，使得得到的结果不变，保证每次进行卷积时都不改变输入。模型结构重参数化的流程如图 2-9 所示。

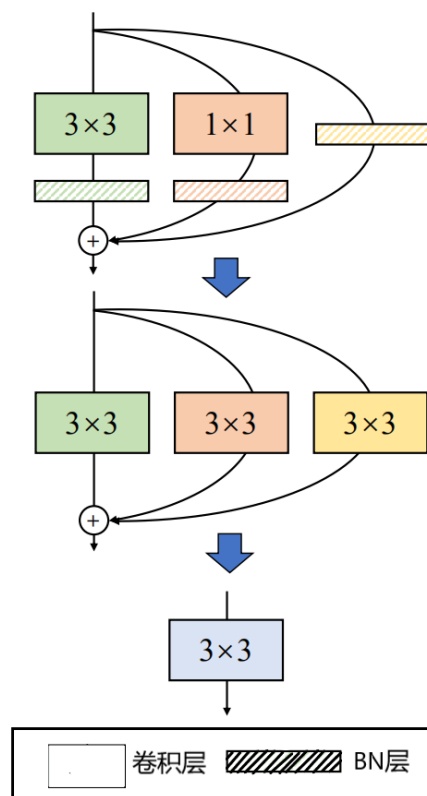


图 2-9 模型结构重参数化的流程

2.3.3 正样本分配策略

YOLOv7 模型将 YOLOv5 模型与 YOLOx 模型的特点结合了起来，之前的 YOLO 模型是将真实框 Ground-Truth 与预测框的交并比 IoU 值较高的作为正样本。YOLOv7 模型中为了使正样本更容易匹配，可能性更多一些，YOLOv7 模型是将真实框的中心点往四周进行偏移 0.5 个单位，使相邻两个网格所对应的预测框也作为正样本，也就是说使偏移后的中心点所在的网格对应的预测框也作为正样本，提升了 YOLOv7 模型的召回率，防止漏检问题。YOLOv7 模型相比之前对于预测框的选择进行了更改，做了两次筛选，被称为初筛和复筛。

- 1) 初筛：真实框与被作为正样本的预测框的长宽比例要在 0.25 到 4 之间，也就是说，即 $1/4 < \text{真实框长宽比例} < 4$ 。
- 2) 计算 IoU：被作为正样本的每个预测框都要与每个真实框计算交并比 IoU 值。之后把计算好的交并比 IoU 值对应每个真实框从大到小进行排序，每个真实框选择 IoU 最大的 10 个被作为正样本的先验框保留。准备再从这 10 个被作为正样本的先验框中进行复筛。
- 3) 把每个真实框对应的这 10 个先验框的交并比 IoU 值进行累加和，并进行截断，通过 IoU 的大小来判断选取的最优个数，并非是固定数目。最后进行动态匹配，保留与截断后的累加和数值相等的先验框被作为正样本的先验框，最少必须保留 1 个。但是有时候，一个先验框可能会对应多个真实框，复筛不一定能筛查出去。YOLO 规定，每一个预测框只能作为正样本对应最多一个真实框，也就是说，如果遇到一个被作为正样本的预测框与多个真实框的交并比 IoU 都很大，只能选择最大的那个进行学习。

- 4) 计算类别预测损失：为了确保检测物体的类别一致，对每一个预测框的长宽比与真实框之间的差异，交并比 IoU 与真实框之间的差异，预测的类别与真实框之间的差异，这三点进行加权（其中 IoU 的权重更大一些），计算一个损失总值。

正样本分配策略的具体计算如下，如图 2-10 所示：

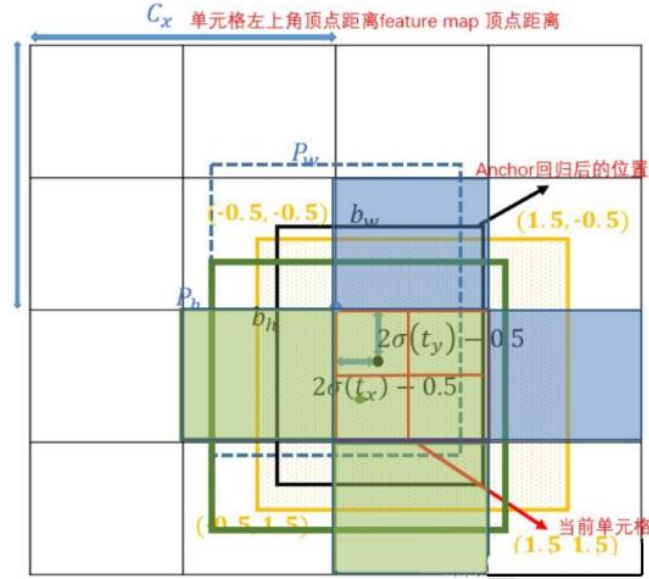


图 2-10 正样本分配策略的具体计算

其中：

- 1) b_x, b_y, b_w, b_h 是预测值，代表预测的预测框的中心点坐标 宽 高。
- 2) t_x, t_y, t_w, t_h 是卷积得到的特征图在梯度方向的值。
- 3) $\sigma(t_x)$ 和 $\sigma(t_y)$ 是 sigmoid 函数，在原先的 YOLOv3v4 模型中， $\sigma(t_x)$ 和 $\sigma(t_y)$ 的值直接是当前网格的左上角坐标距离真实框的中心的位移，要确保它的值在 0~1 之间，这样才能确保预测出来的坐标坐落在当前网格内。现在因为 YOLOv7 模型将真实框的中心点往周围进行偏移 0.5 个单位，因此在 YOLOv7 模型中，将 $\sigma(t_x)$ 和 $\sigma(t_y)$ 的值扩大了 2 倍再减去 0.5，才能确保当前网格的左上角坐标距离真实框的中心的位移值在 -0.5~1.5 之间。
- 4) C_x, C_y 是当前网格的左上角坐标，我们假设当前网格是它负责预测的物体的真实框的中心点。
- 5) P_w, P_h 是事先聚类得到的预测框值。
- 6) YOLOv7 模型对 P_w 和 P_h 采用 sigmoid 函数 σ 是为了使梯度曲线较平缓，更适合收敛。 $2 * \sigma(t_w)^2$ 和 $2 * \sigma(t_h)^2$ 的取值范围在 0~4 之间，是为了初筛的时候要确保真实框与被作为正样本的预测框的长款比例小于 4。

这些参数具有如下关系：

$$\begin{aligned} b_x &= 2\sigma(t_x) - 0.5 + c_x \\ b_y &= 2\sigma(t_y) - 0.5 + c_y \end{aligned} \quad (9)$$

$$\begin{aligned} b_w &= P_w * 2\sigma(t_w)^2 \\ b_h &= P_h * 2\sigma(t_h)^2 \end{aligned} \quad (10)$$

2.3.4 AUX 辅助输出

YOLOv7 模型拥有两个头，除了主头（lead_head）之外，还额外拥有一个辅助头（aux_head）。其网络结构的输入由原先的 640×640 改为了 1280×1280 ，上来第一层就对图片进行下采样操作，采用间隔下采样的方法，得到 4 个 640×640 的实际输入，再进行卷积，使得 3 个维度通道变为 12 个维度通道。YOLOv7 模型选择在维度上进行增加而不是最大池化 maxpooling，是因为 maxpooling 太绝对了，会把数值小的特征统统去除，而 YOLOv7 模型则是将这些数值小的特征放到不同的输入层中。辅助头 aux_head 主要是为了增加召回率，防止漏检的问题，主头 lead_head 再基于辅助头 aux_head 做进一步筛选。

对于辅助头来说，辅助头不影响输入层，只影响输出层。输出层变为了 4 个输出层，因为带了辅助头一共得到了 $4 \times 2 = 8$ 个输出层，每层还是 3 个预测框。每个输出层都会选择一个预测框对应的正样本，按理来说，无论是主头还是辅助头都要匹配正样本，要拿预测结果与真实框进行对比，然后更新预测结果。但是，辅助头却只有在训练时运用，实际测试应用时并没有辅助头，因此在选择正样本时仍然使用主头的预测结果。值得注意的是，主头和辅助头的层级特征是不同的，可能匹配到的正样本也不同。意思就是说额外引入 4 层让主头进行训练，让主头预测的更加精确。另外，主头和辅助头的偏移量略有区别，主头的偏移量是 0.5，辅助头的偏移量是 1。辅助头偏移量为 1 是基于正样本分配策略下，为了让周围 4 个网格对应的预测框都能作为正样本进行训练，因此有五个网格，提高召回率。在损失函数上，辅助头的损失计算用 0.25 进行缩放。

2.4 本章小结

本章对本课题研究所需要了解的基本原理进行了介绍与简述，通过对传统的目标检测识别技术的介绍，引入了以卷积神经网络为基础的基于深度学习的目标检测识别技术，进而引入基于深度学习的目标检测识别的两大分类，一阶段目标检测算法和二阶段目标检测算法。基于 YOLO 系统的目标检测识别技术就是一阶段目标检测算法，以回归问题为主导。之后介绍了 YOLO 系列的模型的原理，介绍了其中的基本思想，交并比，非极大值抑制和边界框回归四个内容，还介绍了本课题研究的 YOLOv7 模型的基本原理，介绍了其运用的 RepVGG 的网络结构，模型结构重参数，正样本分配策略和 AUX 辅助头输出这四个内容。为接下来的搭建模型，进行仿真实现提供了大量的理论基础。

第三章 仿真实验

3.1 样本数据集的选择和构建

本次实验的数据集是来自 YOLOv7 中开源的口罩佩戴检测数据集，由 4,935 个 256 像素的人脸有佩戴口罩和 3,023 个 256 像素的人脸无佩戴口罩的图像组成。该数据集口罩种类丰富，背景多样性大，适用于各种复杂场景下对人脸佩戴口罩实时检测的条件。数据集样例图如图 3-1 所示。

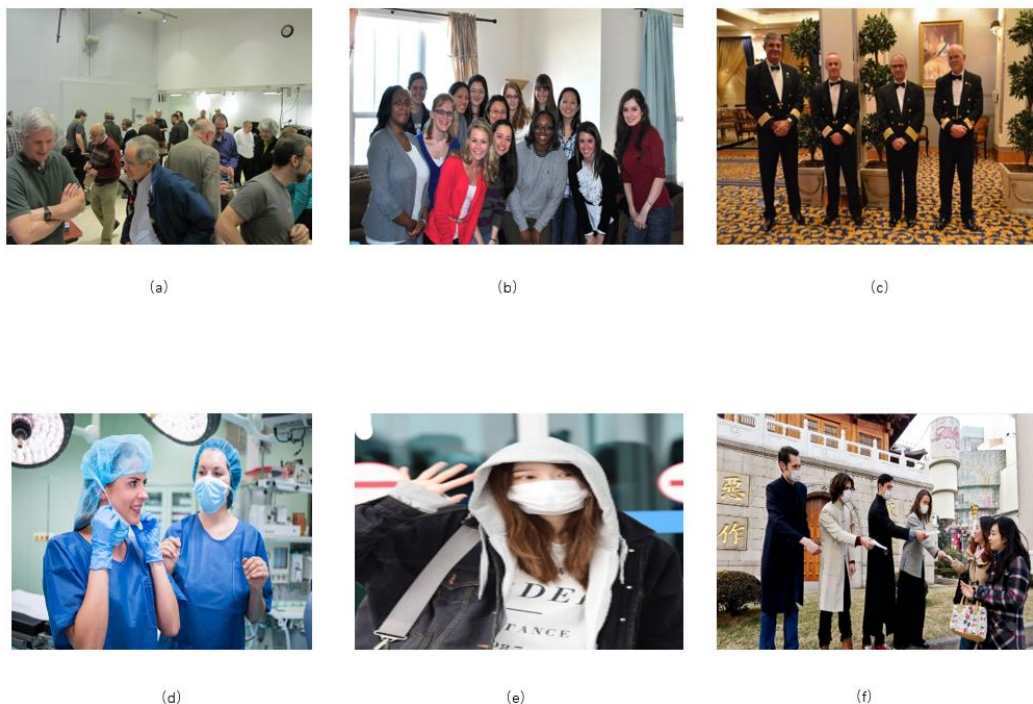


图 3-1 数据集样例图

在本次目标检测的研究中，将 7,958 张图片全部用于实验训练，其中 7,162 张图片用于训练，796 张图片用于测试。

在训练模型前，先对数据集进行了预处理。从每张图片中提取了图片的类别，大小，边框的位置信息以及每张图片的名称，其中边框的位置信息的标签文件是以目标的真实框的中心点的坐标以及边框的宽度和高度的形式（center_x, center_y, w, h），文件名与每张图片的名称相同，文件后缀格式为 xml 的 VOC 格式的标签文件进行存储。之后将数据集的 VOC 格式的标签文件转换为适用于 YOLOv7 模型的 TXT 格式的标签文件，将其中的（center_x, center_y, w, h）坐标形式转换为（xmin, ymin, xmax, ymax）的坐标形式进行存储，即目标真实框的左上角坐标和右下角坐标。其中，第一列代表目标的类别，第二列代表中心归一化横坐标 $x = (xmin + xmax) / 2 * dw$ ，第三列代表中心归一化纵坐标 $y = (ymin + ymax) / 2 * dh$ ，第四列代表归一化宽度 $w = (xmax - xmin) / 2 * dw$ ，第五列代表归一化高度 $h = (ymax - ymin) / 2 * dh$ 。同时生成与每张图片名称相同的文件后缀格式为 txt 的标签文件。具体的 xml 标签文件信息和 txt 标签文件信息的样式与对比如图 3-2 所示。

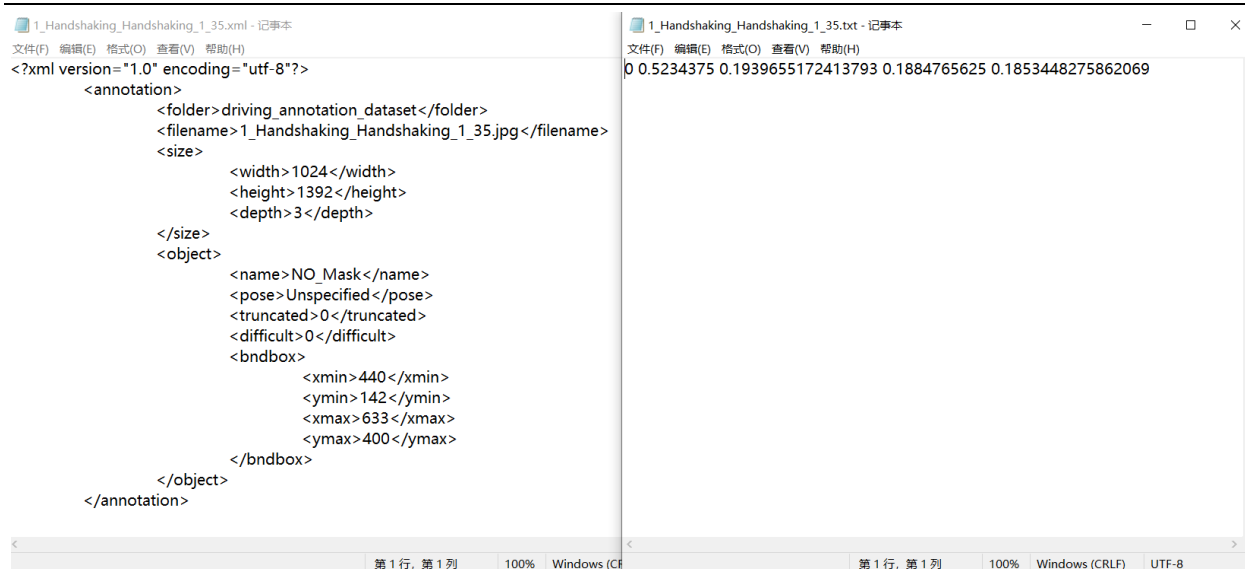


图 3-2 xml 标签文件信息和 txt 标签文件信息的样式与对比

3.2 实验方案

3.2.1 机器配置

运行深度学习算法需要强大的计算力，因此需要合适的机器配置作为支撑。具体实验的相关配置表格如表 3-1 所示：

表 3-1 机器配置

硬件设备名称	配置
CPU	Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz 2.60 GHz
内存	64G
GPU	GeForce RTX 1650
GPU 加速库	CUDA 10.2
操作系统	Windows 11.0
深度学习框架	Pytorch 1.10

3.2.2 实验步骤

在准备完数据集之后，开始进行口罩佩戴检测识别的具体实验。安装好研究需要的软件，搭建好相关的配置环境，网络模型，设置好损失函数后，接下来对网络模型进行训练与测试。具体的实验步骤如下：

- 1) 开始训练前先将数据集进行划分, 划分为训练用数据集和测试用数据集。再将划分好的 VOC 格式的标签文件的数据集转换为 YOLO 可用的 txt 格式的标签文件, 并各自生成训练集和测试集的路径列表文件。
- 2) 选择适合的 YOLOv7 权重模型, 本课题选择 yolov7.pt 作为预训练的权重模型。
- 3) 根据自己的数据集修改存储模型结构的配置文件 cfg, 本课题目标的类别只有 Mask 和 NO_Mask, 因此将类别数 nc 改为 2, 目标类别的名称 num_classes 修改为自己的类别 Mask 和 NO_Mask。
- 4) 根据自己的数据集修改存储训练集和测试集的文件 data。data 里的内容为测试集和训练集的绝对路径名, 类别数量 nc 设定为 2 以及类别名称 num_classes 设定为自己的 Mask 和 NO_Mask。
- 5) 根据机器配置修改合适的 epoch, batch-size, img-size, devices 和 workers。本课题采用的 epoch 为 200, batch-size 为 4, img-size 为 640, devices 为 0, workers 为 0。
- 6) 通过运用 forward_once 函数进行前向传播, 将网络结构整体分为 backbone 和 head 两个模块。前 50 层网络结构为 backbone 模块, 用来提取图片的特征。通过遍历每一层网络, 将进行卷积后得到的特征图作为输入。第 51-105 层是 head 层, 是输出的模块。通过遍历 3 个输出层, 对输出层的每个通道加上一个偏置进行微调, 但是这个偏置并非是固定值, 而是一个均值为 0, 方差较小的可训练参数, 会不断进行更新。
- 7) 在 SPPCSPC 模块中, 对 3 个不同的最大池化层 Maxpooling 进行填充, 为后续的拼接做准备, 因为这 3 个不同的最大池化层的卷积核不一样, 因此进行填充时也不一样。为了让不同的最大池化层运行完的结果形式相同, 卷积核的大小越大, 填充就需要越大。将这 3 个与进行过 3 次卷积的原特征图得到的特征图, 4 者进行拼接。拼接完再进行两次卷积。
- 8) 通过 self.act(self.rbr_dense(inputs) + self.rbr_1x1(inputs) + id_out) 方法将 3*3 的卷积, 1*1 的卷积和同等映射全部合并成 3*3 的卷积, 之后将卷积和批处理归一化 BN 行合并。
- 9) 使用训练集对模型进行训练, 在每一轮 epoch 训练完之后, 使用验证集对模型进行验证, 之后调整模型的参数。
- 10) 在模型训练完之后, 将模型的最优解和最后解进行保存, 保存为.pt 文件。通过将权重模型替换为训练完的最优解模型, 运用训练后的最优解模型对测试集进行测试, 评估模型的性能。
- 11) 计算模型的平均检测精度 mAP 和模型的召回率, 对算法的性能进行研究分析。

3.2.3 参数设置

本实验具体的参数设置如表 3-2 所示:

表 3-2 实验参数设置

参数名	参数设置
weights	yolov7.pt
cfg	yolov7.yaml
data	mydata.yaml
epoch	200
batch-size	4
img-size	640
devices	0
workers	0

3.2.4 实验指标

通过精确率和召回率（Recall Rate）对实验结果进行评价分析。召回率是指在所有实际样本为正样本中检测为正样本的概率，即正确预测为正样本的数量占有所有实际为正样本数量的比例。即：

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

其中，TP 为实际样本为正样本的情况下，模型正确预测为正样本的数量，FP 实际样本为负样本的情况下，模型错误预测为正样本的数量，TN 为实际样本为负样本的情况下，模型正确预测为负样本的数量，FN 为实际样本为正样本的情况下，模型错误预测为负样本的数量。

3.3 实验验证与分析

3.3.1 实验训练结果

在初次实验过程中，将实验用参数设置为 epoch 为 200，batch-size 设置为 16，img-size 设置为 640，device 设置为 0，workers 设置为 8。虽然程序成功开始运行，但是通过该次实验结果可知，模型在训练过程中经常出现程序意外中断导致训练停止的情况。程序报错内容为运行超时，超出显存容量。经过后续排查程序报错原因以及对程序不断进行调整，重新设置新的实验参数，成功解决该次实验遇到的运行超时，超出显存容量的问题。

在之后的实验过程中，将实验用参数调整为 epoch 为 100，batch-size 调整为 8，img-size 继续设置为 640，device 继续设置为 0，workers 调整为 0。虽然成功解决上述初次实验时遇到的运行超时，超出显存容量的实验问题，也成功得到了该次实验的实验结果和召回率变化曲线图。但是通过该次实验结果可知，该次实验得到的实验数据并不理想，用该次训练得到的最优解权重模型对测试集进行测试时，没有成功识别出目标的口罩佩戴情况。同时召回率变化曲线图也同样表明几乎没有准确识别出目标，实验结果表示该次训练模型的检测性能并不理想。经过后续排查模型训练后实验数据不理想，检测性能低的原因以及对程序不断进行调整，重新设置新的实验参数，成功解决该次实验遇到的模型训练后实验数据不理想，检测性能低的问题。图 3-3 为不理想的实验召回率变化曲线图。

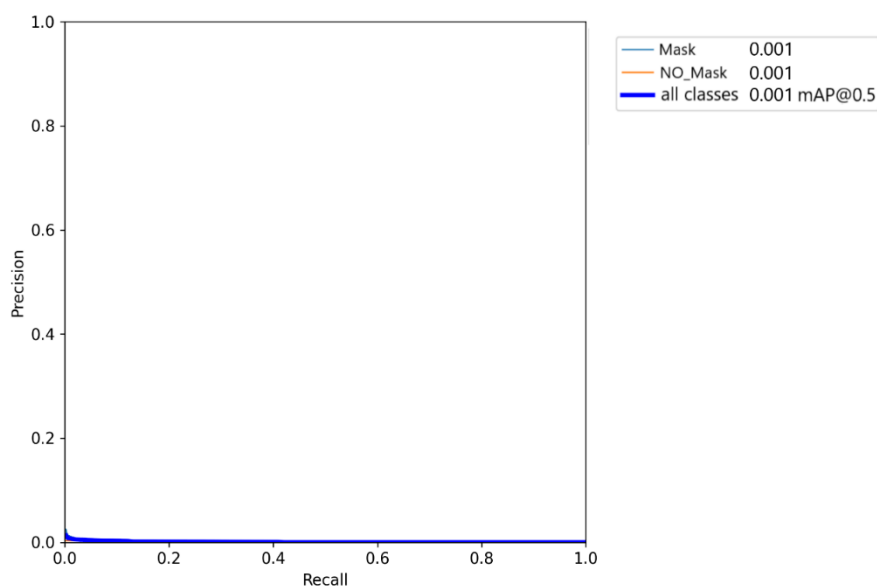


图 3-3 不理想的实验召回率变化曲线图

图 3-4 为最终实验得到的召回率变化曲线图。在最终实验时设置的实验参数为 epoch 为 200, batch-size 设置为 4, img-size 设置为 640, device 设置为 0, workers 设置为 0。通过最终实验得到的模型训练的召回率变化曲线图可知,该次实验成功得到理想的实验数据。YOLOv7 模型在训练之后,平均检测精度 mAP 能达到 83%,佩戴口罩的目标平均检测精度 mAP 为 78%,未佩戴口罩的目标平均检测精度 mAP 为 86%,对未佩戴口罩目标的检测识别的 mAP 值略高于对佩戴口罩目标的检测识别。在检测精度较高的同时也具有较高的召回率,说明对图片进行口罩佩戴检测识别的效果理想,表示模型的总体学习能力较好。

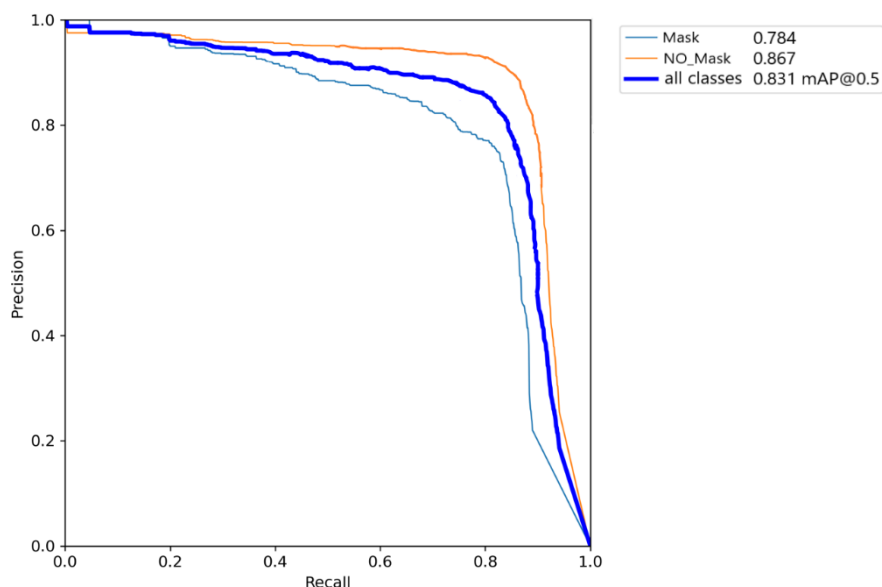


图 3-4 最终实验的召回率变化曲线图

最终基于 YOLOv7 模型的口罩佩戴检测识别结果的分析如表 3-3 所示。

表 3-3

类别	Recall/%	mAP/%
Mask	77.4	78.4
NO_Mask	85.6	86.7

3.3.2 口罩佩戴检测识别结果

通过保存训练完之后的模型的最优解权重模型 `best.pt`，运用于测试集进行检测，测试集的检测结果如图 3-4 复杂场景下小目标的口罩佩戴检测识别效果图以及图 3-5 出现面部遮挡的口罩佩戴检测识别效果图所示。其中，橙色框表示检测到目标佩戴了口罩，蓝色框表示检测到目标并未佩戴口罩。

根据图 3-4 进行分析，该场景为复杂场景下小目标的口罩佩戴检测识别。基于 YOLOv7 模型的口罩佩戴检测识别表现出较好的目标检测效果，能够应对复杂场景下的小目标检测识别，正确识别出目标的佩戴口罩情况，具有较高的检测精度。



图 3-4 复杂场景下小目标的口罩佩戴检测识别效果图

根据图 3-5 进行分析，基于 YOLOv7 模型的口罩佩戴检测识别能够应对面部出现遮挡物时的检测识别，对出现面部遮挡的口罩佩戴检测识别也能识别出目标未进行佩戴口罩的情况，具有较高的检测精度。



图 3-5 出现面部遮挡的口罩佩戴检测识别效果图

3.3.3 结果分析

在模型的训练过程中如果参数设置不合适，可能会出现由于显卡内存不足会导致 batch-size 等参数设定的阈值受限，设定过大会导致出现显卡内存不足的报错，进而导致模型无法继续训练下去。同时 epoch 参数设定的实验轮次较低会导致最终实验得到的训练好后的权重模型无法准确识别出待检测目标以及召回率变化曲线图表现出的检测精度过低等问题。设置合适的实验参数才能使得模型成功进行训练并得到理想的实验数据。

以最终实验得到的理想的检测模型来进行实验结果分析，通过精确率计算在所有预测结果为正样本的结果中，正确预测的结果的概率，其值越大代表误检的目标就越少。通过召回率计算所有实际样本为正样本中检测为正样本的概率。其值越大代表漏检的目标就越少。

对于精确率来说，如果阈值越小，更多的样本就会被检测为正类别，导致整体上精确率会降低。如果阈值越大，在某些情况下预测出的结果可能都是正样本，导致整体上精确率提高。对于召回率来说，所有真实正样本的数量是一个定值，如果降低阈值，会导致整体上召回率提高。

该实验模型从实验结果可知，平均检测精度 mAP 能达到 83%，佩戴口罩的目标平均检测精度 mAP 为 78%，未佩戴口罩的目标平均检测精度 mAP 为 86%，对未佩戴口罩目标的检测识别的 mAP 值略高于对佩戴口罩目标的检测识别。从对测试集进行检测识别的结果可知，该模型能够应对复杂场景下的小目标检测识别，正确识别出目标的佩戴口罩情况；也能够应对面部出现遮挡物时的检测识别，对出现面部遮挡的口罩佩戴检测识别能识别出目标未进行佩戴口罩的情况。通过评价结果可知，YOLOv7 模型在检测精度较高的同时也具有较高的召回率，说明对图片进行口罩佩戴检测识别的效果理想，表示模型的总体学习能力较好。模型检测的精确率比召回率略高，说明在预测过程中依然存在部分误检漏检的问题，模型的整体检测性能并没有达到最为理想的状态。造成的这一结果的原因经由多次实验分析得出，可能是 IoU 阈值的设置不够准确以及模型的实验参数还有待继续进行更新调整，在后续的研究中可以通过从这方面入手对模型加以改进和优化。

3.4 本章小结

本章对课题的研究内容进行仿真实验，首先介绍了实验的数据集的选择和来源以及构建。然后搭建了算法运行的虚拟环境，接着对数据集进行预处理，划分为训练集和测试集，并且将 VOC 格式的标签文件转换为 YOLO 模型可用的 txt 格式的标签文件输入到 YOLOv7 网络模型中。再按照自己的机器配置调整修改相应的参数进行 YOLOv7 模型的训练，在多次实验的过程中逐步解决遇到的各种实验问题，

如程序运行超时，显卡内存不足以及训练好后的模型检测性能差，无法正确检测出目标，召回率变化曲线图表现出低精确度等问题，最终找到最为合适的实验参数进行模型训练。模型训练结束后使用训练好的最优解权重模型，再对测试集进行测试，检测模型能否成功识别待检测目标。最后，通过平均检测精度和召回率等指标对 YOLOv7 模型的检测性能进行评价分析，得到模型在检测精度较高的同时也具有较高的召回率，对图片进行口罩佩戴检测识别的效果理想，模型的总体学习能力较好。但是在预测过程中依然存在部分误检漏检的问题，模型的整体检测性能并没有达到最为理想的状态。

第四章 总结与展望

4.1 总结

口罩佩戴检测识别技术对于维护疫情防控有着重要的作用，YOLO 系列模型作为具有高精度的实时检测器，已经被广泛的运用于多种实时检测系统中。因此基于 YOLOv7 模型的口罩佩戴检测识别系统能够在人流量密集，背景复杂的公共场合下快速进行实时检测，获取公众口罩佩戴信息，实现精确定位，保障防疫安全。同时还减少了大量的人力资源的消耗，更好的解决了人力可能会出现漏检，误检等以及难以全天候实时检测等问题。除此之外，传统的目标检测技术基于传统的手动设计特征，并结合滑动窗口的方式来进行目标检测和定位，具有特征设计困难，效率低，对于某些特定的条件不适应，鲁棒性差；通过滑动窗口的方式来提取候选框，对候选框进行分类判定，导致时间复杂度高和窗口冗余度高等一系列问题。因此，随着深度学习的发展，基于 YOLOv7 模型的口罩佩戴检测识别技术取得了较好的检测效果和算法性能。

本课题先从传统的目标检测识别技术出发，分析了传统的目标检测识别技术的劣势和基于深度学习的目标检测识别技术的优势。分析国内外目标检测识别技术的研究，提出了研究的内容和实验的目标。然后对实验所涉及到的基本原理以及 YOLOv7 模型的基本原理进行了介绍。然后通过仿真实验进行实验训练并测试模型的性能：

- 1) 安装需要的软件，搭建算法运行的虚拟环境
- 2) 接着对数据集进行预处理，划分为训练集和测试集，并且将 VOC 格式的标签文件转换为 YOLO 模型可用的 txt 格式的标签文件输入到网络模型中
- 3) 再按照自己的机器配置调整修改相应的参数进行模型的训练
- 4) 训练完结束后使用训练好的最优解权重模型，对测试集进行测试。

最后对实验结果进行分析。

4.2 展望

在本课题的研究中，因为对深度学习相关理论知识的学习和了解不足，以及时间上的限制，导致模型训练的过程中遇到了多种问题，网络模型也没有得到更好的调参和训练测试，所以对于训练结果以及算法的性能有待优化和提升。希望在此后的研究中，能够对深度学习有着更好的研究和理解，掌握更多知识，对模型进行优化和改进，提高模型的检测性能，得到更好的检测结果。

参考文献

- [1]新冠病毒感染肺炎暂定名“新冠肺炎”，英文简称 NCP！[J].中国医学计算机成像杂志,2020,26(01):44.
- [2]刘延波,郝铭,刘玲玲,刘垚,蔡秉燚,陈志军.口罩对新冠病毒的防护作用及灭菌后重复使用方法[J].国际纺织导报,2020,48(02):42-46+52.
- [3]Shifeng Zhang,Xiaobo Wang,Zhen Lei,Stan Z. Li. Faceboxes: A CPU real-time and accurate unconstrained face detector[J]. Neurocomputing,2019,364(C).
- [4]Xu Yuanyuan,Yan Wan,Yang Genke,Luo Jiliang,Li Tao,He Jianan. CenterFace: Joint Face Detection and Alignment Using Face as Point[J]. Scientific Programming,2020,2020.
- [5]Joseph Redmon,Santosh Kumar Divvala,Ross B. Girshick,Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection.[J]. CoRR,2015,abs/1506.02640.
- [6]Wei Liu,Dragomir Anguelov,Dumitru Erhan,Christian Szegedy,Scott E. Reed,Cheng-Yang Fu,Alexander C. Berg. SSD: Single Shot MultiBox Detector.[J]. CoRR,2015,abs/1512.02325.
- [7]Loey Mohamed,Manogaran Gunasekaran,Taha Mohamed Hamed N.,Khalifa Nour Eldeen M.. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection[J]. Sustainable Cities and Society,2020(prepublish).
- [8]Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [9]Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
- [10]Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.
- [11]Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv preprint arXiv:2207.02696, 2022.
- [12]牛作东,覃涛,李捍东,陈进军.改进 RetinaFace 的自然场景口罩佩戴检测算法[J].计算机工程与应用,2020,56(12):1-7.
- [13]孙世丹,郑佳春,赵世佳,黄一琦.基于 YOLO 改进算法的安全帽和口罩佩戴自动同时检测[J].集美大学学报(自然科学版),2021,26(04):379-384.DOI:10.19715/j.jmuzzr.2021.04.12.
- [14]李国进,荣誉.基于 DCN-SERes-YOLOv3 的人脸佩戴口罩检测算法[J].计算机与现代化,2021(09):12-20+30.
- [15]任小康,刘行行.基于改进的 YOLOv3 口罩佩戴检测和识别[J].计算机工程与科学,2022,44(10):1812-1821.
- [16]刘子龙,周志鹏,万森,许瑞.基于改进后的 SSD 人脸口罩佩戴识别方法研究与典型应用[J].中国安全防范技术与应用,2021(02):29-34.
- [17]朱杰,王建立,王斌.基于 YOLOv4-tiny 改进的轻量级口罩检测算法[J].液晶与显示,2021,36(11):1525-1534.
- [18]薄景文,张春堂.基于 YOLOv3 的轻量化口罩佩戴检测算法[J].电子测量技术,2021,44(23):105-110.DOI:10.19651/j.cnki.emt.2107568.
- [19]曹小喜,程凡永,王飞州,张明艳.基于改进 YOLOv4 的密集人群口罩佩戴检测算法[J].安徽工程大学学报,2022,37(03):49-57+69.
- [20]项融融,李博,赵桥.基于改进 YOLOv5s 的口罩佩戴检测算法[J].国外电子测量技术,2022,41(07):39-44.DOI:10.19652/j.cnki.femt.2203765.
- [21]杨震山. 口罩佩戴智能检测算法研究[D].山东工商学院,2022.DOI:10.27903/d.cnki.gsdsg.2022.000114.
- [22] Ding X, Zhang X, Ma N, et al. Repvgg: Making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.

致谢

经过这段时间的研究和学习，我最终完成了我本科毕业设计和论文。在一过程中，从课题的选题到文献的查阅，到代码的研读，编写和修改，再到论文的编排和撰写，这其中的经历都让我收获颇。

在这里，我真挚地感谢我毕业设计的指导老师高强老师。高强老师他那严谨的科研态度和悉心且不厌其烦的帮助指导在各个方面都给予了我极大的帮助与鼓励。从课题的选题建议和指导到论文的整体框架编排，再到文章内容的各处细节要素和格式上的规范要求，都是高强老师一点一滴指导来的，没有高强老师的细细教导，也就没有我完成的毕业设计和论文。同时，也要感谢本科四年里所有教过我，传授我知识的老师们，是他们的循循善诱，才让我学到了扎实的专业知识与技能。其次，我还要感谢在四年里所有给予过我帮助和鼓励的朋友，是他们给足了我克服困难的动力和勇气，让我一次次坚定了自己的目标。最后，感谢师院四年的悉心栽培，让我能够成为有能力报答社会的有志青年。