

基于 YOLO 框架的口罩类型及合理佩戴的识别检测系统开发

袁帅祥

计算机与人工智能学院阿里云大数据学院软件学院

常州大学

指导教师：倪彤光（副教授），周忠良（高级工程师）

2023 年 5 月

摘要—随着新冠疫情的流行，人们普遍认识到佩戴口罩的重要性。如何快速准确地判断口罩类型及其是否合理佩戴对于预防疾病传播具有重要意义。随着现代成像设备的不断升级，图像数据的获取变得越来越容易。同时，深度学习等新兴技术的发展，为口罩检测提供了高效、准确和快速的解决方案。目前，常用的目标检测算法包括一阶段和二阶段的算法，一阶段目标检测算法主要是基于深度学习的卷积神经网络模型，如 YOLO、SSD、RetinaNet 等。这类算法具有检测速度快、部署简单、精度高等特点。在口罩检测方面，使用一阶段算法可以快速地实现对口罩的检测，对于检测时间要求较为严格的场景具有一定的优势。二阶段目标检测算法则主要是基于区域提取和分类的方法，包括 R-CNN、Fast R-CNN、Faster R-CNN 等。这类算法通常具有更高的检测精度，同时也具有更复杂的结构和计算量。在口罩检测方面，使用二阶段算法可以实现更加精确的口罩检测，并且能够在一定程度上避免误判或漏检。在实际应用中，我们需要根据实际的应用场景来进行模型的选择，本文使用基于 YOLO 框架的目标检测方法，对口罩进行快速准确的识别和佩戴状态的检测。首先，建立了一个口罩种类图像数据集，包括医用口罩、n95 口罩、普通口罩等不同类型的口罩。然后，使用 YOLOv5 算法对图像进行训练，得出了一个高精度的口罩分类检测模型。随后将模型部署到 QT 界面交互的系统中，并且通过多种算法训练了相同的数据集，进行了不同模型的比较，本文进行了大量实验，对实验结果进行了分析。以证明模型在各方面都有较高的优越性，适合于在实际场景下进行部署和应用。本文设计了一个可交互的口罩类型及合理佩戴的识别检测系统。

Index Terms—口罩识别, YOLO, 深度学习, 目标检测

I. 引言

随着新冠疫情的全球爆发，口罩成为了人们防疫的重要利器。然而，由于人员众多，口罩佩戴不当或佩戴不合适的情况时有发生，这些情况会增加疫情传播风险，因此训练一个口罩是否合理佩戴以及佩戴的口罩类

型是否安全的识别系统，将对疫情防控工作产生重要的意义。

A. 课题的意义

此课题研究可以提高口罩佩戴的准确性，在疫情防控期间，正确佩戴口罩是最基本的防护措施之一，但是口罩佩戴不当、佩戴不合适等问题时有发生，这些问题会增加疫情传播风险。训练一个口罩是否合理佩戴以及佩戴的口罩类型是否安全的识别系统，可以通过人脸识别技术，准确判断口罩是否佩戴在正确的位置上，提醒佩戴者正确佩戴口罩。同时，可以判断佩戴的口罩是否为防护等级较高的医用口罩或 N95 口罩。这样可以有效降低口罩佩戴不当或佩戴不合适的情况发生，提高口罩佩戴的准确性，从而减少疫情传播风险。

B. 国内外发展状况和发展趋势

YOLOv5 是一种基于深度学习的目标检测算法，由美国加州大学伯克利分校的 Alexey Bochkovskiy、Chien-Yao Wang 和 Hong-Yuan Mark Liao 共同开发。相比于 YOLOv4，YOLOv5 在精度、速度和模型大小等方面都有了显著的提升，成为当前最先进的目标检测算法之一。

II. YOLOv5 算法原理

YOLOv5 是一种基于深度学习的目标检测算法，它将目标检测问题转化为回归问题，通过单次前向传播即可得到目标的位置和类别信息。YOLOv5 的核心包括以下几个部分：

A. 模型结构

YOLOv5 的模型结构可以分为三个部分：Backbone、Neck 和 Head。Backbone 用于提取特征，常用的是 CSPDarknet；Neck 用于融合不同尺度的特征，常用的是 FPN+PAN 结构；Head 用于预测目标的位置和类别。Focus 网络结构是 YOLOv5 中使用的一种特殊的卷积网络架构，主要用于检测小目标。它通过将原始输入图像进行逐层缩小和堆叠，得到高分辨率、高维度的特征图，从而有效提高检测精度。具体地说，在 Focus 网络结构中，首先对输入图像进行降采样，按照一定的步长间隔取样，得到四张不同的子图。然后将这四张子图堆叠起来，作为下一层的输入，再进行降采样，重复上述步骤，直到得到目标尺寸的特征图。通过这种方式，Focus 网络结构可以在保证尺寸不变的情况下，得到更多的特征信息，并且可以对小目标进行更加准确的定位和检测。在堆叠过程中，每个子图的通道数量都为 1，因此第一层输入的通道数是原始输入图像通道数的 4 倍。同时，经过若干层堆叠后，通道数也会相应增加。在 YOLOv5 中，使用了两个 Focus 网络结构，分别用于检测较小的物体和较大的物体，以提高检测的效率和准确性 [7]。在图 4.2.1 中，假如输入了一张 640 行 640 列像素，3 个通道的图片，通过降采样的方式就可以获得更多的特征，特征层相对于原来的 3 通道，变成了原来的四倍，高和宽得到了一定的压缩。[1] 之后会不断地进行卷积、标准化和激活函数进行特征提取，可以提取出不同尺度和语义的特征信息在进行了一系列操作后，CspLayer 的主要作用是来自不同分支的特征信息进行融合，使得最终的特征图具有更丰富的语义信息。在 CspLayer 中，首先进行输入特征的维度变换，以匹配来自分支的不同特征图。随后，使用一个相接操作将不同特征图拼接在一起，然后通过卷积层、标准化层和激活函数层进行后续处理。最后，再通过一个残差连接将融合前后的特征进行加和，以增强特征表示的稳定性。从图 4.2.1 中我们可以看到，最终 Backbone 部分得到了三个有效特征层，然后将三个有效特征层传入 FPN(特征金字塔) 中进行一个加强特征提取。在 FPN 中主要是进行特征融合，因为虽然经过 CspLayer 进行了特征的融合，但是最终得到的三个特征层仍然有可能存在冗余或者重叠，所以需要进一步地进行特征融合以便更好地利用不同语义的特征。首先从最小的有效特征层开始，进行一次卷积操作，以增强其特征表达能力。接着，将该特征层进行上采样，使得其特征图的分辨率

与上一层的特征图相同。这样就可以进行特征的堆叠后使用 CspLayer 进行特征融合，在融合后，再对得到的特征图进行一次卷积操作和上采样操作，得到更高分辨率和更丰富的语义信息的特征图。最终重复上采样融合后，再进行下采样自上而下地进行堆叠和融合，每次融合还会输出到 YoloHead 中进行分类和回归的预测，在头部模块中，每个特征层都会进行一次卷积操作和上采样操作，以增强其特征表达能力和分辨率。最终，这些特征图会被堆叠在一起并进行非极大值抑制后，得到检测结果。

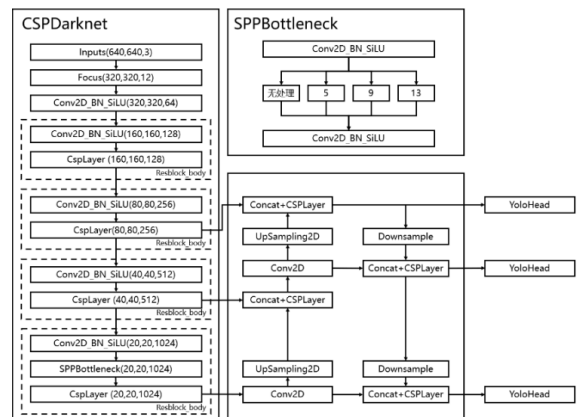


图 1. YOLOv5 整体网络结构

B. 损失函数

YOLOv5 的损失函数包括三部分：定位误差、分类误差和置信度误差。定位误差使用的是 CIoU Loss，分类误差和置信度误差使用的是二元交叉熵损失。YOLOv5 中使用的激活函数是 SiLU 激活函数，SiLU 是 Sigmoid 和 ReLU 的改进版。SiLU 具备无上界有下界、平滑、非单调的特性。SiLU 在深层模型上的效果优于 ReLU。可以看做是平滑的 ReLU 激活函数。[1] YOLOv5 模型的训练，本质上就是一个迭代式的优化问题，通过不断地迭代更新模型参数以最小化损失函数，进而实现模型的不断优化和提高检测性能。在具体实现过程中，需要定义损失函数以及选择合适的优化算法和调参策略。YOLOv5 使用了多个损失函数来综合考虑各个方面的训练目标。其中，最主要的三个损失函数分别是 confidence loss、classification loss 和 box regression loss。confidence loss 是针对每个框的置信度 (confidence) 进行优化的损失函数，其作用是鼓励模型对框内物体进行正确的检测，同时惩罚模型对背景误判为物体的情况。它可以用二元交叉熵 (binary cross-entropy) 作为度量

指标。classification loss 是针对每个框的类别预测值进行优化的损失函数，其作用是指导模型对图像中的不同类别物体进行准确的分类，并惩罚模型对类别的错误预测。它可以使用交叉熵（cross-entropy）作为度量指标。box regression loss 是针对每个框的位置坐标进行优化的损失函数，其作用是指导模型对图像中不同目标的边界框位置进行准确预测，并惩罚模型对边界框位置的错误预测。它可以使用平方误差（square error）作为度量指标。

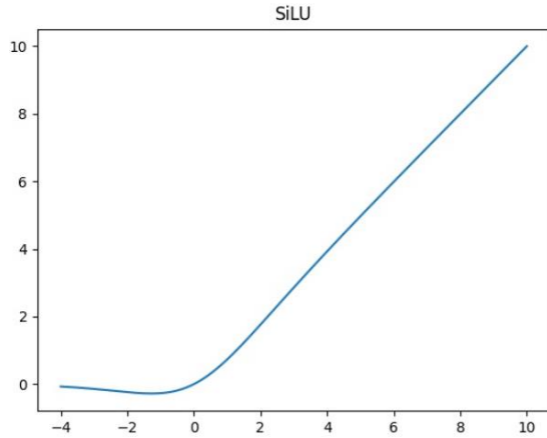


图 2. SiLU 激活函数曲线

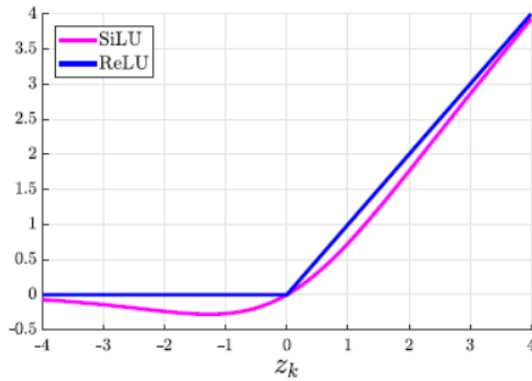


图 3. SiLU 激活函数与 ReLU 激活函数曲线

C. Anchor 机制

YOLOv5 使用 Anchor 机制来预测目标框，每个 Anchor 根据特征图的不同尺度进行预测，从而能够检测出不同大小的目标物体。

III. 系统的总体设计

A. 系统可行性分析

系统功能划分为以下几个模块：

- 数据集采集与预处理模块
- 模型训练模块
- 口罩类型检测与合理佩戴检测模块
- 界面设计模块

B. 系统功能需求分析

本系统是基于 YOLOv5 算法开发的口罩是否合理佩戴以及口罩类型的检测系统，旨在通过计算机视觉技术，对人脸进行检测和识别，实现对口罩的佩戴情况和类型的自动识别和记录，提高口罩的佩戴率和口罩的使用效果，从而有效预防疾病传播。系统功能划分为以下几个模块：

1、数据集采集与预处理模块

包括口罩图像数据的采集、标签化处理，划分训练集和测试集、数据预处理、数据增强等操作。

2、模型训练模块

包括五种目标检测算法对相同数据集进行训练。

3、口罩类型检测与合理佩戴检测模块

包括使用 YOLOv5 算法和其他四种目标检测算法对口罩检测并分类到防护能力普通的口罩、防护安全的口罩、无口罩、佩戴不正确、正确佩戴五个类型。

4、界面设计模块

包括 QT 界面的设计、实现以及功能实现，例如图片检测功能、视频检测功能以及摄像头检测的功能。算法模型选择，包括算法模型的选择、加载、调用等操作。

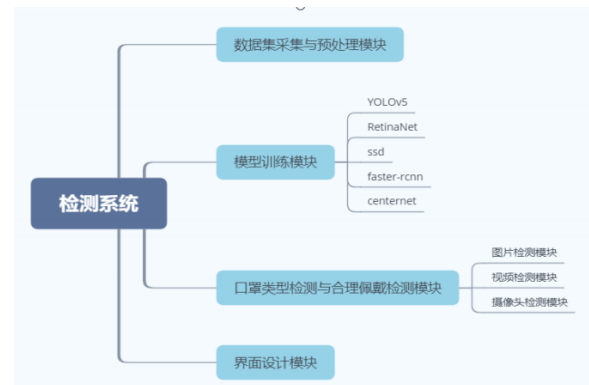


图 4. 系统模块图

IV. 系统前端的设计

A. QTdesigner

QTdesigner 是由 Qt 提供的一款可视化图形界面设计工具，可以帮助开发人员快速地创建各种 GUI 应

用程序,包括窗口、对话框、面板、菜单、工具栏、按钮、文本框、下拉列表等常见的 GUI 元素。

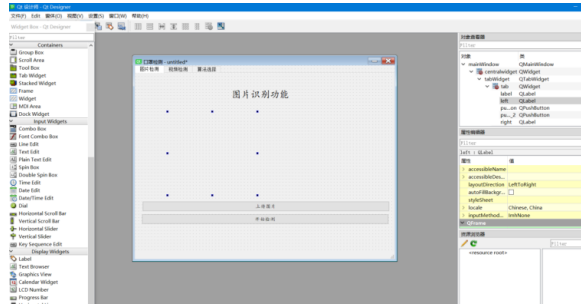


图 5. 使用 QtDesigner 设计检测系统 QT 界面

V. 系统实现

A. 数据集采集与预处理模块的实现

通过爬取百度图片和下载部分生成数据集,人工选取了约两千张佩戴口罩的图片,并通过 labeling 软件对图片进行逐张标注,框出图片中的检测目标并选择类别。



图 6. 使用 labeling 标注软件为图片进行标注

VI. 总结与展望

A. 设计目标与意义

基于 YOLO 算法的口罩类型识别与是否合理佩戴的检测系统,这是一个有实际应用价值的研究方向,因为在新冠疫情背景下,口罩成为了人们生活中的必需品,目前疫情仍然有卷土重来的趋势,所以仍然需要重视防护。而通过对口罩佩戴情况的识别和分析,可以有效地帮助相关部门监管和防控疫情,也能够对公众健康教育产生重要的作用。

B. 选择算法与实验设计

在研究过程中,选取了 YOLOv5 算法,并通过实验进行对比研究。通过实验,发现 YOLOv5 算法不仅训练速度快,而且准确率较高,在目标检测的领域具有很大的优势。然而,其在小物体检测上的表现相对较差,还需要进一步优化。在实验结果方面,YOLOv5 算法模型对于口罩类型的识别和合理佩戴的检测都取得了良好的效果,并与其他算法进行了对比。在置信度和检测性能上,YOLOv5 综合比所有参考算法都更具有该场景下的优势,并且处理速度也很快,但是在小目标的检测上还有一定的提升空间,仍然可以通过进一步优化网络结构,或者选取更好的数据集,更多、更平衡的数据样本等来提升模型的检测精度。

参考文献

- [1] Benjumea A, Teeti I, Cuzzolin F, et al. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles[J]. arXiv preprint arXiv:2112.11798, 2021.
- [2] 魏贞. 机器学习:Python 实践. 北京: 电子工业出版社,2018.
- [3] 张德军, 吴亦奇, 徐战亚. 基于深度学习的姿态估计与行为识别方法. 武汉: 武汉大学出版社,2022.
- [4] 谭志. 基于深度学习的目标检测与识别技术. 北京: 化学工业出版社,2021.08.
- [5] 王硕, 孙洋洋. PyQt 5 快速开发与实战. 北京: 电子工业出版社,2017.
- [6] 陈亦新. 深度学习理论与实战:Pytorch 案例详解. 北京: 清华大学出版社,2021.02.
- [7] 孟婧, 江平, 王凯, 蒋鑫宇. 改进的 SSD-ResNet 算法 [J]. 合肥工业大学学报 (自然科学版),2023, 第 46 卷 (3): 326-332.
- [8] 刘晋川, 黎向锋, 叶磊, 刘安旭, 赵康, 左敦稳. 基于改进 RetinaNet 的行人检测算法 [J]. 科学技术与工程,2022, 第 22 卷 (10): 4019-4025.
- [9] 李悦言, 程培涛, 杜淑幸. 一种改进 CenterNet 的轻量化目标检测算法 [J]. 西安电子科技大学学报,2022, 第 49 卷 (5): 137-144.
- [10] Yang Sun; Jiankun Song; Yong Li; Yi Li; Song Li; Zehao Duan. IVP-YOLOv5: an intelligent vehicle-pedestrian detection method based on YOLOv5s[J]. Connection Science, 2023; 1-20.
- [11] DFCED. 图解目标检测之 Faster-RCNN 原理详解.CSDN,2020(3-27). <https://blog.csdn.net/DFCED/article/details/105138996>.

致谢

感谢指导教师倪彤光副教授和周忠良高级工程师的悉心指导,以及常州大学计算机与人工智能学院、阿里云大数据学院、软件学院的大力支持。