

# 同济大学生创新/创业项目

## 季度报告

### 一、项目基本信息

项目名称	门禁检测口罩佩戴系统		
项目编号	X2023061	项目级别	校级
起止时间（年月）	2023 年 3 月 至 2024 年 3 月		
项目负责人	梁斯凯	所在院系	电子与信息工程学院
学号	2253540	专业	计算机科学与技术
手机号	13193033800	邮箱	1286387732@qq.com
指导老师	刘春梅	所在院系	电子与信息工程学院

### 二、季度报告内容

1) 项目进展情况			
<input checked="" type="checkbox"/> 按计划进行 <input type="checkbox"/> 进度提前 <input type="checkbox"/> 进度滞后			
2) 项目主要研究			
序号	研究阶段	研究内容	完成情况
1	模型优化和改进、数据库修补和扩充	对于佩戴口罩和未正确佩戴口罩分辨率的提升	已完成
2	模型整体评估	根据图像分析性能	已完成

### 3) 项目研究成果

序号	季度报告成果名称	成果形式
1	优化后的模型	新增 non_standard 的 batch
2	多个图像内容分析	分析报告

### 4) 项目季度报告

项目主要工作已经完成，进行一些总结和优化性工作。

#### 1. 模型优化

在第三季度的项目实施中，我们尝试考虑到人脸可能的姿态变化，在本季度的测试中发现模型在部分情况下对于佩戴口罩和未正确佩戴口罩两者分类效果不叫。经过分析和实验发现是因为我们新增的数据集图片样本数量较小的原因，虽然 bounding-box 的总量足够，但是图片数量不足影响了验证的有效性。因此我们明在第四季度增添了未正确佩戴口罩的图片总量。

调整后模型对于两者分辨率提升，结果更加精确。

#### 2. 结果分析和评估

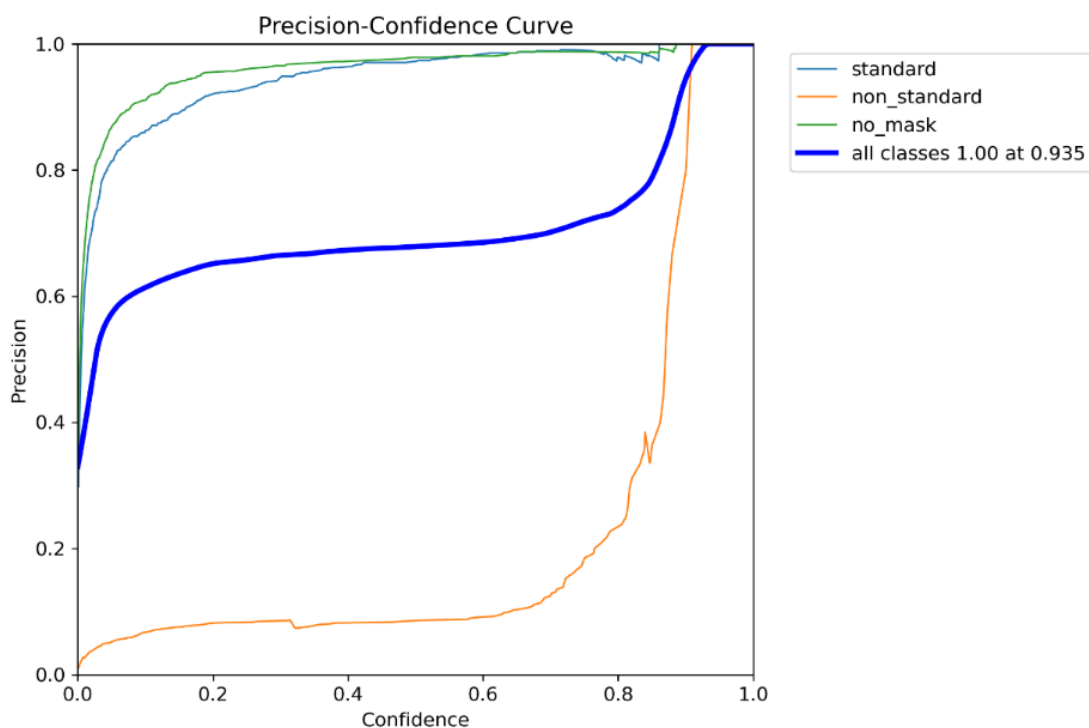
我们对 YOLO 模型在检测是否佩戴口罩的任务中的表现进行一些分析和讨论。本模型被训练来识别三种类别：未佩戴口罩、佩戴口罩、以及佩戴不标准。

在 Precision-Confidence Curve（后简称 P\_curve 图）中，图表显示随着置信度（模型对其预测的自信程度）的增加，精确度（预测正确的正样本占预测为正样本总数的比例）也随之增加。这是符合预期的，因为高置信度通常意味着模型对其预测结果更有信心，因而预测结果的准确性也更高。

同时三种类别在不同置信度阈值下的精确度增长速度存在差异。具体来说：未佩戴口罩和佩戴口罩这两类在置信度超过 0.1 后，其精确度迅速上升，接近 1.0。这表明模型能够以较高的准确度识别出是否佩戴了口罩，且在较低的置信度阈值就能做出准确判断。对于佩戴不标准的情况，模型在置信度达到 0.8 以

后，精确度才开始急剧上升。这说明佩戴不标准的情况在视觉上更为复杂或多变，使得模型在较高的置信度阈值下才能准确判断。

分析后得出模型对于明显的类别（未佩戴口罩和佩戴口罩）识别能力较强：对于这两种情况，模型可以在较低的置信度下就做出高精确度的预测，说明模型对于这两类的特征学习得比较好，能够有效区分。对于复杂情况的识别能力有待提高：对于佩戴不标准这一类，模型需要更高的置信度才能达到较高的精确度，这是因为此类情况的视觉特征更为复杂和多样，例如口罩挂在耳朵上、口罩没有完全覆盖鼻子等，这些都增加了模型识别的难度。

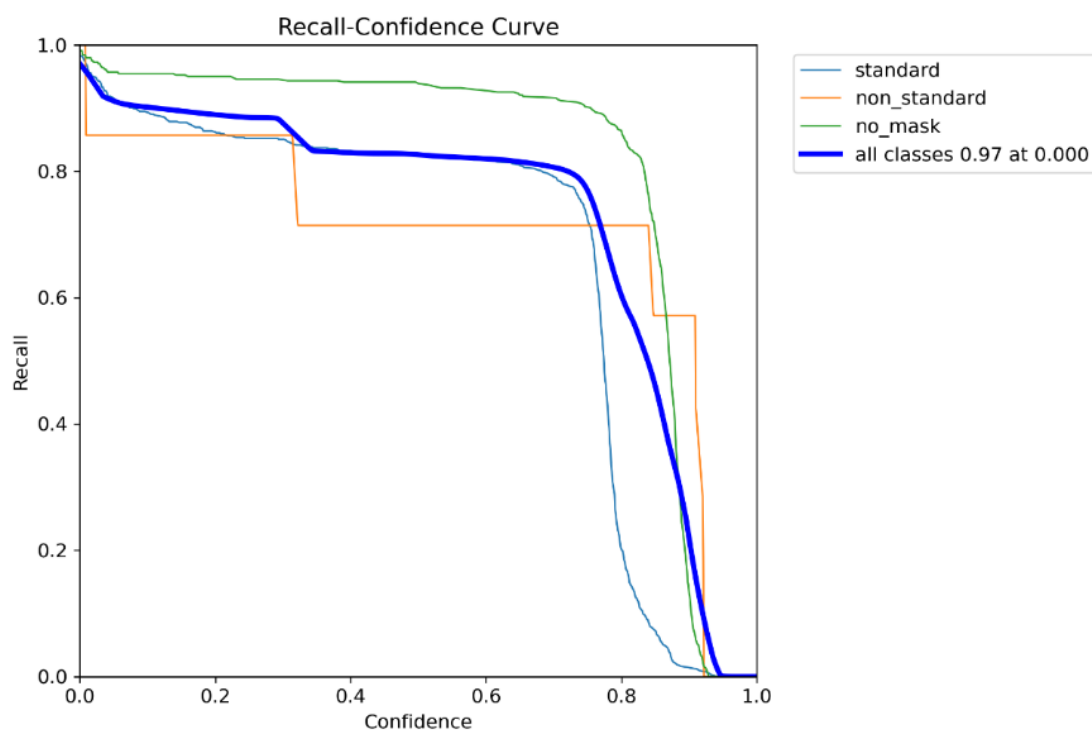


在 Recall-Confidence Curve（后简称 R\_curve 图）中对于所有种类，召回率在置信度低于 0.7 的区间内平缓下降，之后急剧下降。这表明随着置信度的提高，模型越来越挑剔，只有那些它更有信心的预测才会被视为正样本，从而导致召回率下降。

同时不同类别的表现存在差异：佩戴口罩相比未佩戴口罩，佩戴口罩的召回率整体较低，且在置信度 0.7 以上时开始急剧下降。这意味着模型在高置信度下

对佩戴口罩的判断更为严格，导致召回率下降。而未佩戴口罩的召回率在置信度 0.8 以上时急剧下降，这表明在高置信度要求下，模型对未佩戴口罩的检测能力有所下降。最后佩戴不标准类别的召回率呈阶梯状下降，并在置信度 0.8 以后急剧下降。这种阶梯状的下降同样是由于佩戴不标准这一类别的样本在视觉特征上的多样性和复杂性，导致模型在不同置信度阈值下的表现出现明显差异。

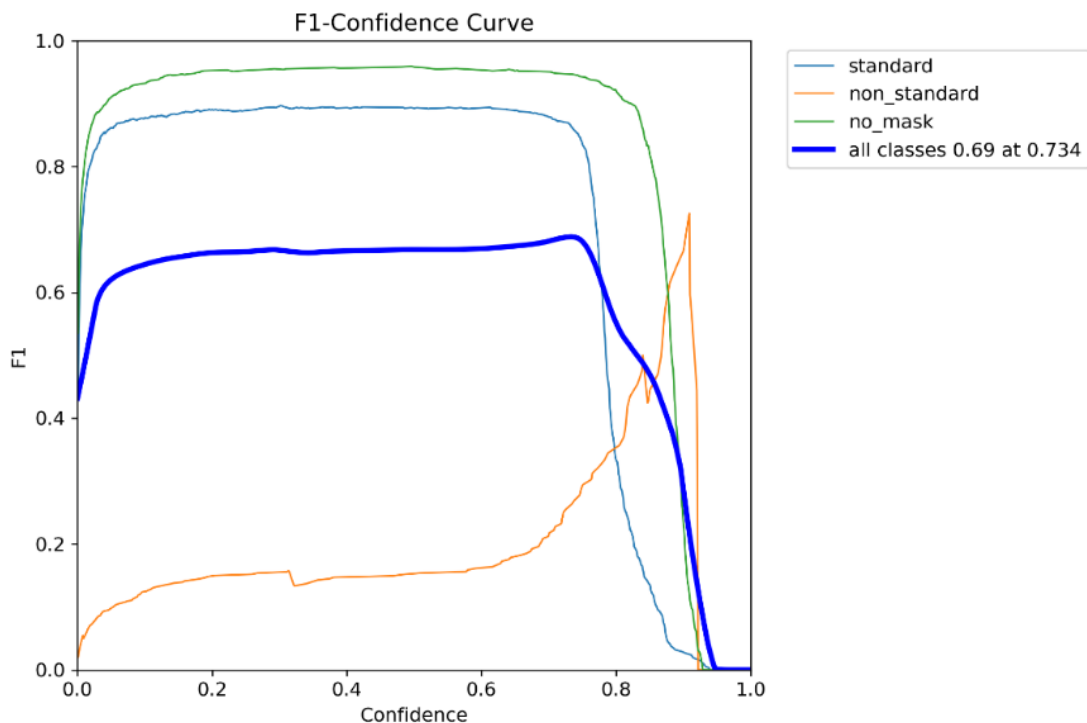
分析后得出：对于明显的类别，模型的召回率较高：在低置信度阈值下，模型对于未佩戴口罩和佩戴口罩的情况有较高的召回率，说明模型能够较好地覆盖这两类样本。而对于复杂情况，模型的召回率较低：对于佩戴不标准的情况，模型的召回率在整个置信度范围内都相对较低，尤其是在高置信度阈值下，这表明模型在识别这类复杂情况时存在一定的困难。



在 F1-Confidence Curve（后简称 F1\_curve 图）中，F1 值是精确率（precision）和召回率（recall）的调和平均，它是评估模型性能的重要指标，特别是在正负样本不平衡的情况下。F1 值能够平衡精确率和召回率，从而提供一个综合的性能评估。

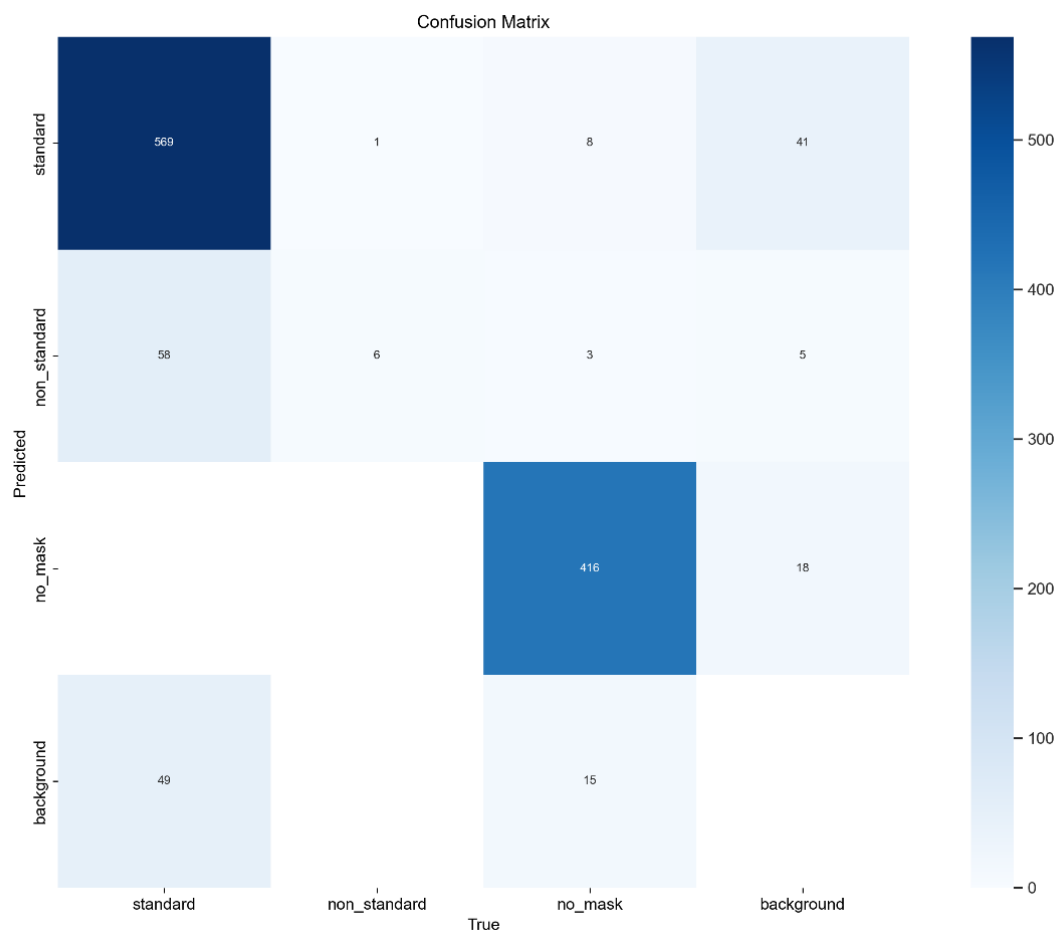
所有类别的总体趋势为在非常低的置信度阈值开始时，所有类别的 F1 值有所上升。这可能是因为是在低置信度阈值下，模型能够捕捉到更多的正样本（虽然这也可能包括一些错误的预测），从而提高了召回率。之后从置信度 0.1 到 0.8 区间，F1 值几乎保持平稳。这表明在这个置信度范围内，精确率和召回率的平衡得到了较好的维持，模型的性能相对稳定。最后置信度超过 0.8 后，F1 值开始急剧下降，这可能是因为高置信度阈值导致模型变得过于保守，虽然精确率可能仍然较高，但召回率的大幅下降导致 F1 值下降。

在不同类别方面，佩戴口罩的总体 F1 值比未佩戴口罩低，但平稳部分的 F1 值均在 0.8 以上，这表明即使佩戴口罩的识别难度可能更大，模型在这个置信度区间内依然能够保持较高的性能。而佩戴不规范口罩这一类别的 F1 值在平稳部分低于 0.2，但在置信度 0.7 到 0.9 区间有大幅上升，F1 值从 0.2 增长到 0.7。这一现象可能表明模型在特定的置信度阈值范围内对佩戴不规范口罩的识别有显著提升，但这种提升不是持续的，超过 0.9 置信度后 F1 值急剧下降。



这种趋势表明，在选择模型的置信度阈值时需要权衡。对于佩戴口罩和未佩戴口罩的情况，模型能够在较宽的置信度范围内保持较高的 F1 值，这表明模型对这两种情况的识别比较稳健。而对于佩戴不规范口罩的情况，模型性能的波动较大，需要更仔细地调整置信度阈值以优化性能。

由于对于不同的类别，其 F1 值的变化显示出不同的特点。特别是佩戴不规范口罩的情况，所以需要特别注意置信度阈值的选择，以免 F1 值出现大幅波动。在实际应用中，如果我们更重视避免错过任何佩戴不规范口罩的情况（即更高的召回率），我们可能会选择一个较低的置信度阈值。相反，如果我们更重视确保预测的准确性（即更高的精确率），我们可能会选择一个较高的置信度阈值。



这张图展示了一个混淆矩阵，它是评估分类模型性能的常用工具。从这张图我们可以得出以下结论：

根据混淆矩阵，口罩检测模型在识别正确佩戴口罩类别方面表现出色，大部分预测都是正确的，漏判和误判很少。然而，在识别未正确佩戴口罩和未佩戴口罩类别时，模型表现不佳，存在将它们误分类为其他类别的情况。特别是未正确佩戴口罩类别中有相当数量的样本被错误地归类为正确佩戴口罩，而未佩戴口罩类别中也有一些被错误地分类。

综合而言，该模型在识别未正确佩戴口罩和未佩戴口罩类别时精度低于正确佩戴口罩的精度。这可能表明模型在处理边界模糊或特征相似的类别时存在困难。为解决这些问题，可能需要进一步的数据采集以增加样本多样性，进行特征工程以突出不同类别的区别，或者进行模型调优以提高分类性能。

### 3. 结题报告和答辩准备

撰写项目结题报告，详细记录整个项目的研究过程、实验结果和结论等内容。准备结题答辩，对项目进行全面的展示和解释。

## 三、指导老师意见

导师签字：

年 月 日

## 四、院系意见

学院（签章）：

年 月 日

## 五、学校意见

学校（签章）：

年        月        日