



深圳技术大学
SHENZHEN TECHNOLOGY UNIVERSITY

本科毕业论文（设计）

题目：基于内窥镜图像的 YOLO 建模
与优化研究

姓 名	何锐生
学 院	大数据与互联网学院
专 业	计算机科学与技术
学 号	202100202016
指 导 教 师	樊小毛
职 称	副教授
提 交 日 期	2025 年 5 月 8 日

深圳技术大学本科毕业论文（设计）诚信声明

本人郑重声明：所呈交的毕业论文（设计），题目《基于内窥镜图像的 YOLO 建模与优化研究》 是本人在指导教师的指导下，独立进行研究工作所取得的成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式注明。除此之外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。本人完全意识到本声明的法律结果。

毕业论文（设计）作者签名：何锐生

日期： 2025 年 5 月 8 日

目 录

摘要.....	I
Abstract.....	II
1. 引言.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.3 研究目标与内容.....	3
1.4 论文结构安排.....	4
2. 相关技术与理论基础.....	5
2.1 内窥镜图像处理技术.....	5
2.2 YOLO 模型的基本原理.....	5
2.2.1 YOLO 算法的网络结构.....	6
2.2.2 YOLO 的算法流程.....	6
2.3 目标检测算法概述.....	7
2.4 评价指标.....	8
2.4.1 混淆矩阵 (Confusion Matrix).....	8
2.4.2 准确性 (Accuracy).....	9
2.4.3 召回率 (Recall).....	9
2.4.4 精确度 (Precision).....	9
2.4.5 PR 曲线.....	9
2.4.6 平均精度和均值平均精度 (AP 和 mAP).....	10

2.4.7 F1-score 与 F 值 (F-Measure)	11
2.4.8 IoU 值	12
2.5 小结	13
3. 基于内窥镜图像的 YOLO 建模与优化方法	14
3.1 数据集介绍与预处理	14
3.2 YOLO 模型的选择与构建	15
3.3 模型训练与验证	16
3.4 模型优化方法	18
3.4.1 更换 IoU 之 MPDIoU	18
3.4.2 更换损失函数之 NWD	22
3.4.3 增加小目标检测层	27
3.5 小结	31
4. 实验结果分析与讨论	33
4.1 实验环境与设置	33
4.1.1 硬件环境	33
4.1.2 软件环境	33
4.1.3 模型训练设置	33
4.2 实验结果与讨论	34
4.3 模型性能评估	35
4.4 小结	36
5. 喉部肿瘤辅助诊断系统设计与实现	38
5.1 功能模块实现	38

5.1.1 图像上传模块.....	38
5.1.2 肿瘤检测模块.....	38
5.1.3 结果可视化模块.....	39
5.2 页面展示.....	39
5.3 小结.....	40
6. 总结与展望.....	41
6.1 研究总结.....	41
6.2 研究不足与未来展望.....	42
参考文献.....	44
致谢.....	47

基于内窥镜图像的 YOLO 建模与优化研究

【摘要】 随着人工智能在医学图像分析中的深入发展，基于深度学习的目标检测方法为喉部肿瘤早期筛查提供了新思路。针对内窥镜图像中肿瘤目标尺寸小、边界模糊等问题，本文基于 YOLOv5 模型提出改进方法，并进行多策略优化。

本文的主要研究内容如下：

1. 引入 MPDIoU 与 NWD 损失函数及其组合策略：在传统 CIoU 基础上，尝试将 MPDIoU 和 NWD 引入作为边界框回归损失函数，并通过多组实验探索不同加权组合。结果表明，采用 $0.3\text{NWD}+0.7\text{CIoU}$ 的混合策略在保持高召回率的同时，显著提升了检测精度与 mAP 指标，有效缓解了单一损失函数对复杂目标边界拟合能力不足的问题，为肿瘤定位提供了更稳健的特征对齐能力。

2. 设计小目标检测层以提升小肿瘤识别能力：针对内窥镜图像中小目标易被漏检的问题，在模型结构中引入额外的小目标检测层，增强浅层特征的表达能力，并通过多尺度特征融合提升对尺寸多样目标的识别精度。该改进有效提升了模型对边界模糊、低对比度小肿瘤的检测性能，特别适用于早期病变的识别场景，具有良好的临床实用性。

3. 构建辅助诊断系统：在最优模型基础上开发了辅助诊断系统，集成图像上传、自动检测、分类判断和结果可视化展示等功能，整体界面简洁友好，支持用户交互与诊断结果导出。该系统不仅提升了模型的可访问性与可用性，也为医生提供了一种快速、直观的辅助诊断工具，具备良好的推广前景和临床应用价值。

实验结果表明，优化后模型在 $\text{mAP}@0.5$ 由 0.940 提升至 0.949, $\text{mAP}@0.5:0.95$ 由 0.578 增至 0.619, Precision 提升 2.1%，整体性能优于基线模型。

本文工作为喉部肿瘤的智能检测提供了可行技术路径，也为深度学习在医学图像领域的应用提供参考。

【关键词】 YOLOv5；喉部肿瘤；内窥镜图像；目标检测；损失函数优化；小目标检测；

Research on YOLO Modeling and Optimization Based on Endoscopic Images

【Abstract】 With the in-depth development of artificial intelligence in medical image analysis, the object detection method based on deep learning provides new ideas for the early screening of laryngeal tumors. Aiming at the problems such as small tumor target size and blurred boundaries in endoscopic images, this paper proposes an improved method based on the YOLOv5 model and conducts multi-strategy optimization.

The main research contents of this article are as follows:

1. Introducing the MPDIoU and NWD loss functions and their combination strategies: Based on the traditional CIOU, attempts are made to introduce MPDIoU and NWD as bounding box regression loss functions, and different weighted combinations are explored through multiple sets of experiments. The results show that the hybrid strategy of $0.3\text{NWD}+0.7\text{CIOU}$, while maintaining a high recall rate, significantly improves the detection accuracy and mAP index, effectively alleviates the problem of insufficient fitting ability of a single loss function for complex target boundaries, and provides a more robust feature alignment ability for tumor localization.

2. Design a small target detection layer to enhance the recognition ability of small tumors: Aiming at the problem that small targets in endoscopic images are prone to be missed, an additional small target detection layer is introduced into the model structure to enhance the expression ability of shallow features and improve the recognition accuracy of targets of various sizes through multi-scale feature fusion. This improvement effectively enhances the detection performance of the model for small tumors with blurred boundaries and low contrast. It is particularly suitable for the identification scenarios of early lesions and has good clinical practicability.

3. Construction of an auxiliary diagnosis system: An auxiliary diagnosis system was developed based on the optimal model, integrating functions such as image upload,

automatic detection, classification judgment, and visual display of results. The overall interface is simple and friendly, supporting user interaction and export of diagnosis results. This system not only enhances the accessibility and usability of the model, but also provides doctors with a rapid and intuitive auxiliary diagnostic tool, and has good promotion prospects and clinical application value.

The results show that the optimized model improves from 0.940 to 0.949 at $mAP@0.5$, from 0.578 to 0.619 at $mAP@0.5:0.95$, and improves by 2.1% in Precision. Overall, its performance is superior to that of the baseline model.

This paper provides a feasible technical path for the intelligent detection of laryngeal tumors and also offers a reference for the application of deep learning in the field of medical images.

【Keywords】 YOLOv5; laryngeal tumor; endoscopic images; object detection; loss function optimization; small object detection;

1. 引言

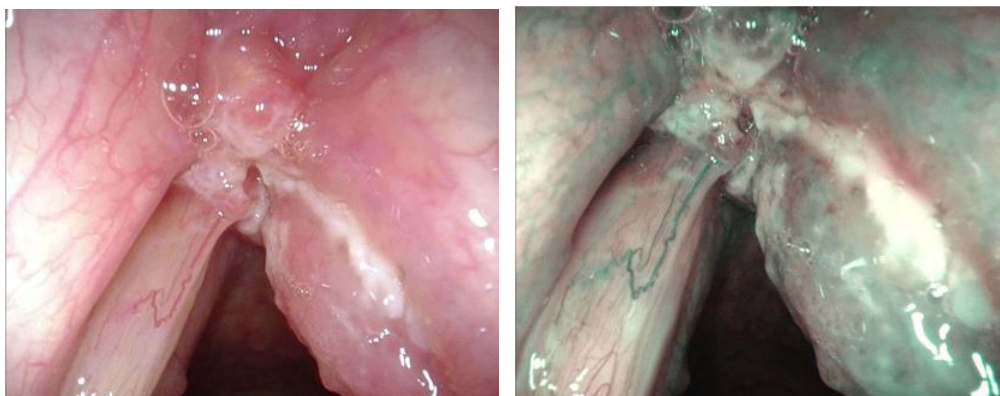
1.1 研究背景与意义

喉癌是上呼吸道中常见的恶性肿瘤之一，主要发生于喉部区域，如声带、喉室及喉返神经等部位。其临床表现通常包括声音嘶哑、持续咳嗽、吞咽或呼吸困难以及颈部淋巴结肿大等症状。由于早期表现不典型，许多患者在确诊时已处于疾病进展阶段，可能导致发声与呼吸功能受损，严重者甚至危及生命^[1]。当前，喉镜活检依然被视为确诊喉癌的权威手段。经鼻电子内窥镜作为临床常用的检查工具，能够清晰观察咽喉区域，在喉癌早期筛查中具有重要价值^[3]。

虽然内窥镜技术是消化道、呼吸道等腔道疾病诊断的“金标准”，但该方法仍存在一定局限性。例如，当病灶较小或位于表浅部位时，其与周围正常黏膜组织在外观上难以区分，容易造成病变区域的遗漏检测^[2]。此外，由于白光内窥镜在分辨率及对比度方面的限制，即便是经验丰富的内窥镜医生，也可能在面对浅层黏膜癌时发生误判或漏检。此外，其影像分析高度依赖医师经验。世界卫生组织统计显示，全球每年新增消化道肿瘤病例超 500 万例，其中早期检出率不足 30%。传统的人工阅片主要存在三大痛点：（1）主观偏差：不同医师对微小病变（<5mm）的判读一致性仅 65%~78%（2）效率瓶颈：单次胃镜检查平均生成 2000+ 帧图像，医师分析耗时约 30 分钟（3）漏诊风险：常规内镜的早期肿瘤漏诊率达 15%~25。因此，探索更先进的检测技术对于提升喉癌的早期筛查效率、降低误诊漏诊率具有重要意义。

随着电子内镜成像技术的进步，窄带成像（narrow band imaging）已成为早期识别恶性肿瘤，尤其是头颈部病变的关键工具之一。与传统白光成像（white light imaging）相比，NBI 具有更高的病灶可视化能力。如图 1-1 所示，NBI 利用特定波长的光增强血管形态结构的对比度，可清晰展现黏膜血管的微小变化及病变区域与健康组织之间的边界^[4]。然而，NBI 的临床使用对医师的专业经验要求较高，其操作和判读存在较高学习门槛，这也在一定程度上限制了其在医疗资源不足地区的普及。同时，传统依赖人工经验的图像判读不仅耗时较长，且具有一定的主观性。

因而，在欠发达地区，结合 WLI 与 NBI 图像实现喉癌的智能诊断与辅助分析，具有实际应用价值和深远意义。



(a) 白光成像

(b) 窄带成像

图 1-1 喉癌示意图：(a) 白光成像；(b) 窄带成像。

1.2 国内外研究现状

最近几年，随着深度学习技术的发展，医学图像分析领域得到了可观的突破。内窥镜图像作为一种重要的医学图像类型，被广泛应用于消化系统、呼吸系统等疾病的辅助诊断中。其中，喉癌的早期发现与诊断对于提高患者生存率具有重要意义。传统内窥镜检查依赖医生的经验，存在主观性强、误诊率高、诊断效率低等问题。为此，利用深度学习模型对喉癌内窥图像进行自动检测与识别成为研究热点。

国内学者在 YOLOv5 算法改进及其医学图像检测应用方面取得了显著的进展，特别是在网络结构优化和多尺度特征融合方面表现突出。成都数之联科技股份有限公司最新申请的“基于改进 YOLOv5 的深层多尺度目标检测方法”专利具有代表性。该专利在 YOLOv5 的 backbone 网络中增加了多个不同深度的卷积层，并在不同层之间建立了复杂连接结构(包括直接连接和跳跃连接)，显著增强了模型对不同尺度目标的特征提取能力。通常情况下，喉癌早期病变只占据图像的很小一部分，用传统算法进行检测容易产生漏检。因此，这种方法特别适用于医疗图像中常见的小目标检测。

在荧光图像细胞检测领域，国内研究者提出了一种融合空域图像增强的改进

YOLOv5 算法。该算法在原始模型的基础上引入了空域图像增强模块，通过对增强系数的精细化优化，有效拓宽了模型对于荧光图像对比度变化的适应能力，显著提升了在复杂图像场景下的鲁棒性^[5]。实验结果显示，经过改进后的模型的 mAP 值达到了 0.983，细胞识别率达到了 91.65%。比原始 YOLOv5 模型提升了 9.19 个百分点。这项研究为喉癌内窥镜图像检测中的对比度增强和特征提取提供了重要参考。

国外学者在喉癌的自动检测方面也取得了显著的进展。Nuzaiha 等人在 24 年提出了一种带有深度学习的矮猫鼬优化算法，它可以对喉癌进行自动检测^[6]。实验结果表明，他们提出的模型是一种准确可靠的喉癌自动检测方法。

在内窥镜检测领域，宾夕法尼亚大学的 Ali 等人在 21 年使用 YOLOv3 对胃肠道内窥镜图像中的息肉进行检测^[7]。这项工作成功展现了 YOLO 模型在医疗图像目标检测方面的潜力。

1.3 研究目标与内容

本研究旨在构建基于 YOLO 模型的喉癌内窥镜图像自动检测方法，并通过对模型结构重新建模与优化，提升其在实际医学图像分析中的准确性与鲁棒性。针对喉癌在早期阶段存在不明显、组织边界模糊、图像质量不稳定等特点，本文力求设计一套高效、实用的图像检测方案，以辅助医生实现更快速、准确的初步筛查与诊断，减轻人工工作负担，提高医疗资源利用效率。

本研究主要包括以下几个方面的内容：首先，整理和分析已有的喉癌内窥镜图像数据，完成图像标注与数据集构建，并在此基础上进行数据清洗和增强操作，提升训练样本的多样性与代表性。其次，选用 YOLOv5 作为基础目标检测框架，构建喉癌检测模型。接着，围绕模型在医学图像场景中的表现特点，开展模型优化研究，包括改进损失函数、通过增加检测层进行重新建模等策略，以进一步提高模型在小目标识别等方面的表现。随后，基于优化后的模型进行实验验证，并通过准确率、召回率和 mAP 等指标对模型性能进行评估，确保其在实际应用中具备良好的检测能力与泛化能力。最后，尝试设计一个简洁的检测系统界面，初步实现内窥镜图像的上传、检测与结果输出功能，探索模型在临床辅助诊断中的实

际可行性。

通过上述研究内容的开展，本文不仅能够验证 YOLOv5 在内窥镜图像喉癌检测任务中的有效性，同时也为医学图像中小目标识别问题提供一种可推广的解决方案，具有一定的理论研究意义与实际应用价值。

1.4 论文结构安排

本文围绕基于 YOLO 模型的内窥镜图像喉癌检测及优化研究，结合模型构建、算法优化与系统应用等多个方面展开，共分为七章，本文的整体结构安排如下：

第一章为引言部分，主要阐述选题背景、研究意义，并对国内外相关工作的研究现状进行综述，进一步明确研究目标与内容，最后对文章结构布局进行概述。

第二章介绍本课题所依托的相关技术与理论基础，重点讲解实现研究目标过程中所涉及的关键技术原理，包括内窥镜图像处理的基本方法、YOLO 模型的基本原理与工作流程、目标检测算法的概述，以及模型评价所采用的各类性能指标。

第三章为基于内窥镜图像的 YOLO 建模，主要包括数据集的来源与预处理过程，YOLO 模型的选择依据与搭建方式，以及模型的训练与验证过程，为后续优化工作提供基础。

第四章为 YOLO 模型的优化研究，从网络结构调整、损失函数优化等角度出发，对原始模型进行针对性改进，并设计多组实验进行验证，通过数据分析对比优化前后的模型性能差异。

第五章为实验与结果分析，详细介绍实验所使用的硬件和软件环境，对模型在不同实验条件下的检测结果进行展示与讨论，进一步从精度、召回率、mAP 等角度评估模型性能。

第六章为喉部肿瘤辅助诊断系统，将模型应用于具体系统开发，实现图像上传、喉癌检测、检测结果可视化等功能模块，展示系统页面与功能逻辑，并对系统进行简要总结。

第七章为总结与展望，总结本文的主要研究成果，归纳研究中存在的不足之处，并对未来在算法精度、临床应用和系统拓展方面的研究方向进行展望。

2. 相关技术与理论基础

2.1 内窥镜图像处理技术

内窥镜图像处理技术是现代医学图像分析领域中的重要研究方向之一，其核心目的是对来自内窥镜设备的图像数据进行处理与分析，以达到增强图像质量、突出病变区域、降低噪声干扰的目的，并为后续的计算机辅助诊断系统提供可靠的输入。与普通自然图像相比，内窥镜图像具有背景复杂、光照变化剧烈、组织结构相似度高、图像畸变明显等特点，这些因素使得内窥镜图像的处理任务更具挑战性。

在实际应用中，内窥镜图像往往由于镜头距离、光源方向、体腔反光、液体遮挡等原因，造成图像部分区域出现过曝或阴影等问题，从而导致图像亮度不均。同时，由于在图像获取过程中，内窥镜设备会频繁移动，这可能会引入模糊、运动伪影等干扰。因此，在将图像输入至目标检测或分类模型前，必须通过一系列图像处理操作对图像质量进行提升。

常见的图像处理流程通常包括图像预处理、图像增强与颜色空间转换等步骤。图像预处理阶段一般包括去噪、锐化、亮度均衡等操作，目的在于提升图像整体的清晰度和可辨识度。图像增强部分常采用对比度受限自适应直方图均衡(CLAHE)来增强局部细节表现力，并通过 Gamma 变换、Retinex 算法等手段调整图像的光照表现，使暗部细节更加清晰可见。在颜色空间处理方面，内窥镜图像常从 RGB 空间转换至 HSV、Lab 或 YCrCb 空间进行处理，以便更加精确地分离亮度和颜色信息，从而提升病灶区域识别效果。

2.2 YOLO 模型的基本原理

YOLO (You Only Look Once) 作为一种高效的端到端目标检测模型，创新性地将目标的分类与定位问题统一为一个回归任务。该方法可在图像输入后，直接输出目标的边界框坐标及其对应类别，大幅提高了检测速度与实时性。与传统的两阶段检测方法（如 R-CNN 系列）不同，YOLO 系列方法通过单个神经网络即可完成从图像输入到目标检测输出的全流程，具有检测速度快、结构简洁、实时性

强等优点，尤其适用于对处理效率要求较高的实际应用场景^[8]。

2.2.1 YOLO 算法的网络结构

YOLO 的网络结构借鉴了 GoogLeNet（图 2-1），拥有 24 个卷积层和 2 个全连接层。

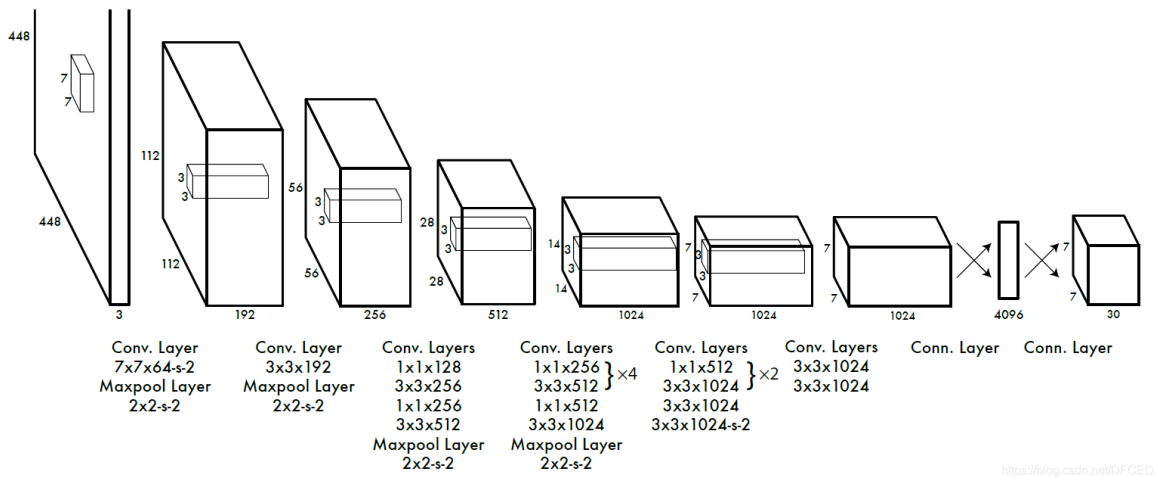


图 2-1 YOLO 的网络结构

YOLO 网络的结构可概括为以下四个关键模块：

- 输入部分：模型接受尺寸为 $448 \times 448 \times 3$ 的 RGB 图像作为输入数据。
- 特征提取层：由多组卷积层与最大池化层堆叠构成，负责提取图像中的深层次特征信息。
- 预测层：包含两个全连接层，用于回归边界框的位置参数并计算各类别的概率值。
- 输出部分： $7 \times 7 \times (2 \times 5 + 20) = 1470$ 的预测结果。其中， 7×7 表示作者将输入图像划分为一个 7×7 的网格；2 表示每个网格预测 2 个边界框；5 表示每个边界框包含 4 个坐标信息（x, y, w, h）和 1 个置信度分数；20 表示 20 个类别的概率。

2.2.2 YOLO 的算法流程

YOLO 算法流程大致如下：

- 1) 图像预处理：将输入图像缩放至固定尺寸（如 448×448 像素），以适应

网络结构的输入要求，并统一后续处理流程。

- 2) 特征提取：将预处理后的图像输入卷积神经网络，提取多层次的图像特征信息，生成特征图。
- 3) 划分网格：将图像划分为 $S \times S$ 个网格单元（如 7×7 ）。每个网格单元负责检测其中心落在该单元内的目标。
- 4) 边界框与置信度预测：每个网格单元预测 B 个边界框和及其对应的置信度分数。分数表示该框内包含目标的概率以及边界框的准确性。
- 5) 类别预测：每个网格单元还预测 C 个类别的条件概率，用于表征该区域包含目标属于某类别的可能性。
- 6) 计算最终分数：将每个边界框的置信度与类别概率相乘，得到该边界框对各类别目标的最终检测分数。
- 7) 非极大值抑制（NMS）：为了减少冗余检测，应用非极大值抑制，保留得分最高的边界框并移除重叠度较高的其他框。
- 8) 输出结果：最终输出检测到的目标类别及其对应的边界框。

2.3 目标检测算法概述

目标检测是计算机视觉中的核心任务之一，旨在从图像中识别出所有感兴趣的物体并精确预测其在图像中的位置与类别。与图像分类任务相比，目标检测要识别图像中是否存在某一类物体，还要精确定位物体在图像中的位置并标注出其边界框。因此，目标检测任务具有更高的复杂性和挑战性。

从发展历程来看，目标检测算法大致可分为两类：基于候选区域的两阶段方法（Two-Stage）和基于回归的单阶段方法（One-Stage）。前者以 R-CNN 系列为代表，典型的如 R-CNN、Fast R-CNN 和 Faster R-CNN。其主要流程是首先生成若干候选区域，然后对这些区域进行分类与回归操作。这类方法通常具有较高的检测精度，尤其适合对精度要求较高的场景。但是，其计算代价较大，检测速度也相对较慢。

相较之下，单阶段方法则直接将检测任务转化为一个回归问题，采用端到端的结构，从图像输入到目标框和类别的输出一次性完成，无需候选区域生成。YOLO（You Only Look Once）系列和 SSD（Single Shot MultiBox Detector）是典

型的代表。YOLO 系列方法通过将整张图像划分为多个网格，并在每个网格内直接预测目标的位置信息与类别概率，实现了在保持较高精度的同时大幅提升检测速度，适合实时性要求较高的场景。SSD 则采用多尺度特征图进行检测，兼顾检测精度与多尺寸目标适应能力。

2.4 评价指标

目标检测任务的效果评估是衡量模型性能的重要环节，通常采用一系列定量指标来反映检测模型在准确性、召回能力以及整体检测效果等方面的综合表现。

2.4.1 混淆矩阵 (Confusion Matrix)

混淆矩阵是一种二维表格结构，用于对比模型的预测输出与实际类别，帮助评估分类性能。对于二分类问题，混淆矩阵如下：

表 2-1 混淆矩阵

Confusion Matrix	Actual Positive	Actual Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

各单词解释如下：

T(True)：最终预测结果正确。 P(Positive)：模型预测其是正样本。

F(False)：最终预测结果错误。 N(Negative)：模型预测其是负样本。

TP：模型正确识别为正类别的样本数。

TN：模型正确识别为负类别的样本数。

FP：模型错误将负类别预测为正类别的样本数。

FN：模型错误将正类别预测为负类别的样本数。

2.4.2 准确性 (Accuracy)

准确性是模型在所有样本中正确预测的比例，计算公式为：

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

准确性是一个总体性能指标，但其在类别不平衡的情况下可能不够敏感。通常来说，正确率越高模型越好。

2.4.3 召回率 (Recall)

召回率是指在实际为正类别的样本中，模型正确预测为正类别的比例。计算公式为：

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.2)$$

召回率衡量模型对正类别的敏感性。召回率越高，实际为正样本被预测出来的概率越高，类似于“宁可错杀一千，绝不放过一个”。

2.4.4 精确度 (Precision)

精确度是指在模型预测为正类别的样本中，有多少是真正的正类别。计算公式为：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.3)$$

精确度衡量了模型的预测准确性。

2.4.5 PR 曲线

在检测过程中，我们当然希望检索结果的 Precision 值是越高越好。同时，

如果 Recall 值也越高的话便再好不过了。但是，事实上这两者是矛盾的。举个例子：30 个良性样本和 30 个恶性样本，我们的模型检测出了 1 个良性且是正确的，此时的 Precision 值是 100%，但是 Recall 值却非常低；但是，如果我们为了找全良性样本把全部样本都认为是良性，此时的 Recall 值是 100%，但是 Precision 值却非常低。因此，在不同的场合中，我们都希望这两个值都能尽量的高，两者之间做一个权衡，这便引入了我们的 Precision-Recall (PR) 曲线。

一般情况下，将 Recall 设置为横坐标，Precision 设置为纵坐标：

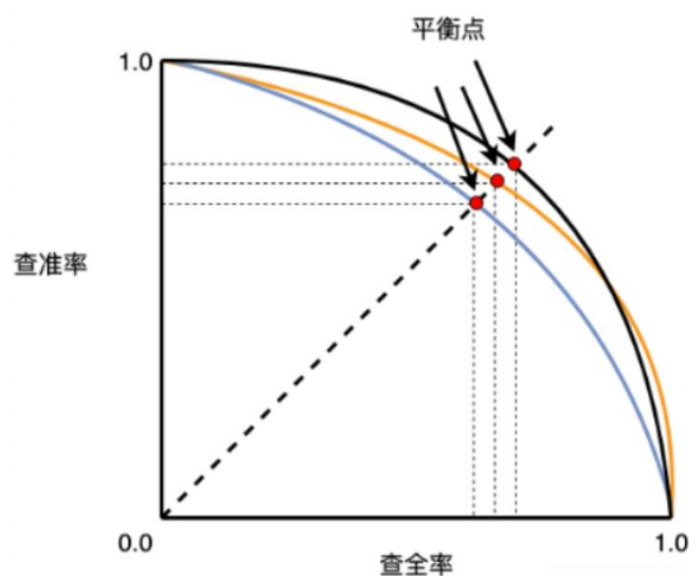


图 2-2 PR 曲线

在众多模型对数据进行学习后，当模型 A 的 PR 曲线完全包含模型 B 的 PR 曲线时，通常可以认为 A 的性能优于 B。然而，也可能出现 A 和 B 的 PR 曲线交叉的情况。在这种情况下，我们常通过计算曲线下方的面积来进行评估，但更常见的是使用平衡点 F1 来比较。平衡点对应的是精度和召回率相等时的取值（斜率为 1），F1 值越高，表示该学习器的性能越优。F1 值的计算我会在后面给出。

2.4.6 平均精度和均值平均精度（AP 和 mAP）

Average Precision (AP)，即平均精度，它是衡量模型在某一个类别下的检测准确性的综合性指标。它通过绘制 Precision-Recall 曲线，计算该曲线下方的面积来表示。通俗来讲，AP 反映了在不同置信度阈值下，模型的精确性和召回能

力之间的整体权衡表现。AP 值越高，说明模型在该类别的检测效果越好。

mean Average Precision (mAP)，即均值平均精度，它是对所有类别的 AP 值进行平均得到的指标，表示模型在所有类别上的综合检测性能。在多类别检测任务中，mAP 是评估模型性能最权威、最常用的指标之一。

假设模型检测的目标共有 N 个类别，且每个类别都计算得到了一个 AP 值，则：

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (2.4)$$

根据 IoU 阈值的不同，mAP 通常可细分为：

- mAP@0.5：表示当 $IoU \geq 0.5$ 时计算得到的 mAP；
- mAP@[0.5:0.95]：表示在 $IoU = 0.5$ 到 0.95 （步长为 0.05 ）之间取多个阈值分别计算 mAP，再取平均，评价更严格。

2.4.7 F1-score 与 F 值 (F-Measure)

F1 分数是精确度和召回率的调和平均值，它在处理类别不平衡问题时特别有用。计算公式为：

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.5)$$

在某些特定情况下，精确度和召回率的重要性并不相同。例如在医疗领域，为了避免遗漏任何患者，召回率显得尤为重要。我们引入一个参数 β 来衡量两者之间的关系，进而通过加权调和均值来综合考虑 P 和 R：

$$F_{\beta} = \frac{(1 + \beta^2) \times P \times R}{(\beta^2 \times P) + R} \quad (2.6)$$

2.4.8 IoU 值

IoU 是目标检测中用于衡量预测框与真实框重合程度的核心指标。其定义是预测框与真实框的交集面积与并集面积之比，其数学表达如下：

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{A \cap B}{A \cup B} \quad (2.7)$$

其中，A 表示预测框的面积，B 表示真实框的面积， $A \cap B$ 表示两者的交集面积， $A \cup B$ 表示两者的并集面积。

IoU 是判断模型预测结果是否正确的基础。在目标检测任务中，通常设定一个 IoU 阈值（例如 0.5），当预测框与真实框的 $IoU \geq$ 阈值时，认为该预测是一个 True Positive（真正例），否则认为是 False Positive（假正例）。不同 IoU 阈值会对检测性能评价产生显著影响：

- $IoU \geq 0.5$ ：检测容忍度较高，常用于一般目标检测任务的评估。
- $IoU \geq 0.75$ 或更高：对重叠精度要求更严格，常用于精细检测或小目标检测。

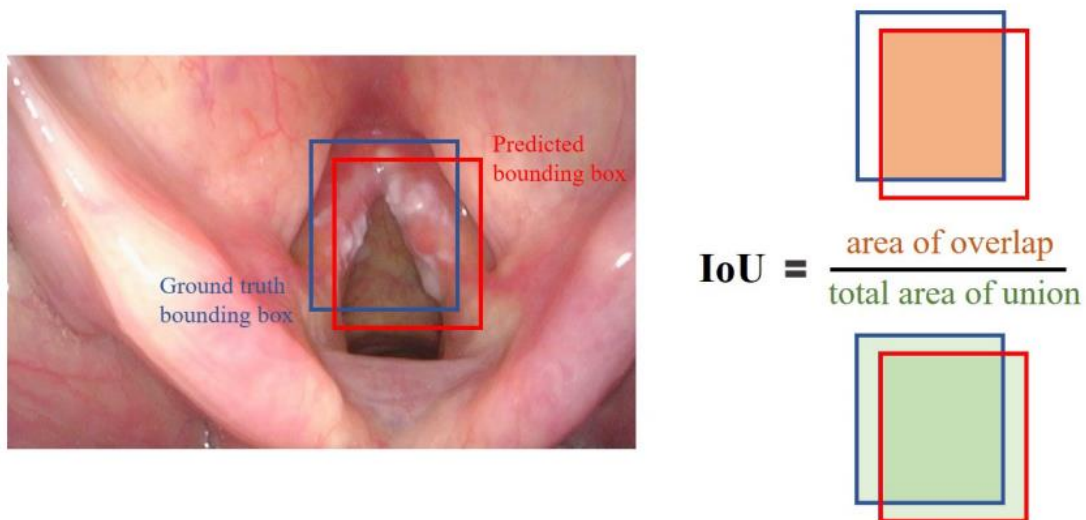


图 2-3 交并比计算的图形表示。蓝色框表示标注框，红色框表示预测框。

2.5 小结

本章围绕本研究所涉及的核心技术与理论基础展开了系统梳理，主要包括内窥镜图像处理技术、YOLO 模型原理、目标检测算法发展概况以及相关评价指标体系。

首先，介绍了内窥镜图像在医学图像处理中的特点及其面临的技术挑战，如图像噪声、光照不均、目标模糊等问题，并强调了其在辅助诊断中的重要应用价值。随后，系统阐述了 YOLO 模型的基本原理与结构，重点解析了 YOLOv5 在网络设计、特征提取与输出层处理等方面的改进，展示了该模型在速度与精度之间所取得的良好平衡。

在目标检测算法综述部分，对比了两阶段（如 Faster R-CNN）与单阶段（如 YOLO 系列）检测方法的差异，进一步突出了 YOLO 模型在实际场景下的优势。最后，详细介绍了目标检测任务中常用的评估指标，包括混淆矩阵、准确率、召回率、精确度、PR 曲线、AP/mAP 值、F1-score 以及 IoU 等，这些指标为后续模型训练与性能评估提供了量化依据。

综上，本章为后文模型设计、实验对比及系统开发提供了坚实的理论基础，确保研究工作的科学性与可行性。

3. 基于内窥镜图像的 YOLO 建模与优化方法

3.1 数据集介绍与预处理

本文章的研究主要基于中山大学第一附属医院提供的数据集展开。在本工作中，数据集被分为带标注数据和仅类别注释数据这两部分。其中，有带标注数据是指具有类别标签和目标框信息的数据，而仅类别注释数据只有类别标签。数据集中标注的类别包括良性肿瘤（benign）和恶性肿瘤（tumor），而未标注的数据则包含正常（normal）类别。根据表 3-1，带标签的内部数据集包含 NBI 和 WLI 两种模态的数据，训练集与验证集的比例大约为 9:1，测试集用于评估目标检测算法的性能。值得强调的是，为了避免不同数据集之间的交叉污染，测试集的数据来自于不同的患者。

表 3-1 数据集的详细分布（带标注数据）

	NBI			WLI			
	训练	验证	测试	训练	验证	测试	Total
Benign	974	107	407	2286	275	771	4820
Tumor	3674	389	1000	4156	462	1000	10681
Total	4648	496	1407	6442	737	1771	15501

由于内窥镜设备拍摄的视频帧有特定的特点，本研究对原始数据进行了必要的预处理。如图 3-1 所示，利用灰度分布对图像进行裁剪，去除冗余信息，以减小其对后续模型训练的干扰。此外，还对错误标签的数据进行了纠正。

此外，本研究还对图像进行了数据增强处理。数据增强能够通过增加样本数量来有效抑制过拟合并提升模型的泛化能力。在本研究中，根据模型的特性，我应用了以下几种数据增强技术：随机仿射变换（Random Affine Transformations）、HSV 数据增强、水平翻转等。

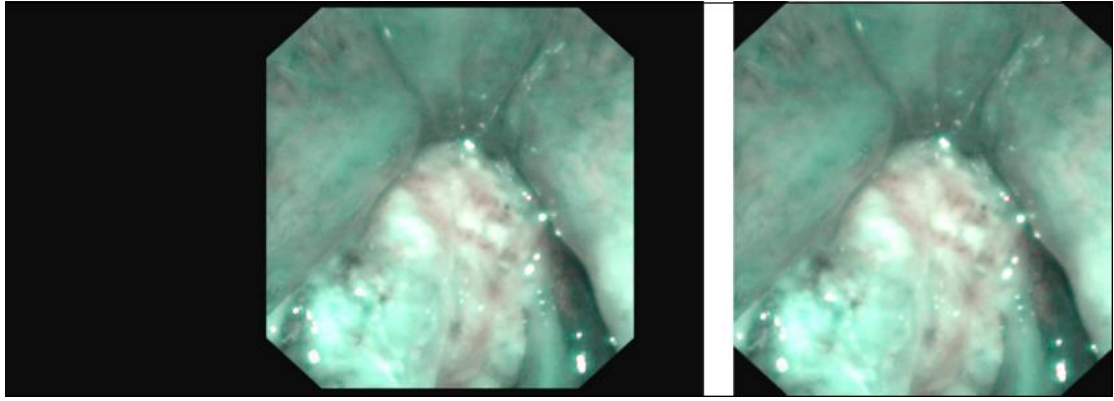


图 3-1 示例图像：展示了裁剪冗余信息前后的对比效果。

3.2 YOLO 模型的选择与构建

本研究选用 YOLOv5 作为喉癌内窥镜图像的目标检测模型，主要考虑到其在精度与速度之间的优越平衡能力，以及在医学图像领域中已有的广泛应用基础。YOLO（You Only Look Once）系列模型作为单阶段目标检测方法的代表，自首次提出以来在多个视觉任务中均展现出极高的效率和可靠性。相较于传统的两阶段检测算法，YOLOv5 在保持较高精度的同时大大提升了检测速度，更适合于处理高分辨率的内窥镜图像及后续的实时推理部署需求。

YOLOv5 在网络结构上进行了深度优化，其主要由四个模块组成：Backbone（特征提取主干）、Neck（特征融合模块）、Head（检测头）和后处理部分。Backbone 采用 CSP（Cross Stage Partial）结构对输入图像进行多层次特征提取，提高特征复用能力并降低模型冗余；Neck 部分则采用 FPN（Feature Pyramid Network）和 PAN（Path Aggregation Network）结构，对不同尺度的特征进行融合，有效提升模型对多尺度目标的检测能力；Head 模块则输出不同层级的预测结果，包括目标类别、置信度和边界框坐标。

在模型构建过程中，本文使用了 PyTorch 深度学习框架，基于开源的 YOLOv5 官方仓库进行二次开发和训练。首先对内窥镜图像数据进行预处理，包括尺寸归一化、图像增强和标签格式转换等。随后在配置文件中调整模型结构与训练参数，如类别数量、损失函数权重、学习率策略等。训练过程中采用随机仿射变换等数据增强方式，以提升模型的泛化能力；并使用 CIoU 作为边界框回归的损失函数，进一步提高定位精度。

考虑到本课题的内窥镜图像数据集具有目标尺寸较小、边界模糊、样本数量有限等特点，YOLOv5s 在保持较高检测效率的同时，具备良好的泛化能力和训练稳定性。此外，该模型对硬件资源要求较低，能够快速完成训练和部署，极大提高了实验效率。因此，YOLOv5s 是在本研究任务中兼顾性能、效率和实现难度的最佳选择。

3.3 模型训练与验证

在完成模型结构的构建与数据集的预处理后，本文对 YOLOv5s 模型进行了训练与验证。训练过程中采用默认配置中的 CIOU^[11] (Complete Intersection over Union) 作为边界框回归的损失函数，作为后续优化的 baseline，以评估模型在原始条件下的性能表现。

在训练过程中，设置如下超参数：初始学习率为 0.01，batch size 设为 16，训练轮数为 200 轮。优化器采用 SGD（带动量），动量系数为 0.937，权重衰减参数设为 0.0005。训练在实验室配备的 AMD EPYC 7282 CPU 与 NVIDIA RTX A6000 GPU 上完成，模型训练时间约为 10 小时。

表 3-2 CIOU 在验证集上的表现

Model	Class	Precision	Recall	mAP50	mAP50-95
YOLOV5-CIOU	Benign	0.839	0.901	0.928	0.531
	Tumor	0.901	0.898	0.952	0.625
	Total	0.870	0.9	0.940	0.578

在验证阶段，评估指标包括 Precision、Recall、F1-score 以及 mean Average Precision (mAP)。在初始训练完成后，模型在验证集上的性能如表 3-2 所示。可以看到，YOLOv5s 在使用 CIOU Loss 作为回归基础的情况下，已经具备较高的检测准确性与定位能力，为后续的模型优化提供了坚实的基线支持。

此外，为了更直观地展示模型的整体检测性能，本文绘制了基于验证集的 Precision-Recall 曲线，如图 3-2 所示。PR 曲线能够反映模型在不同召回率（Recall）下的精确度（Precision）变化情况，从而评估模型在不同检测阈值下的稳定性与鲁棒性。

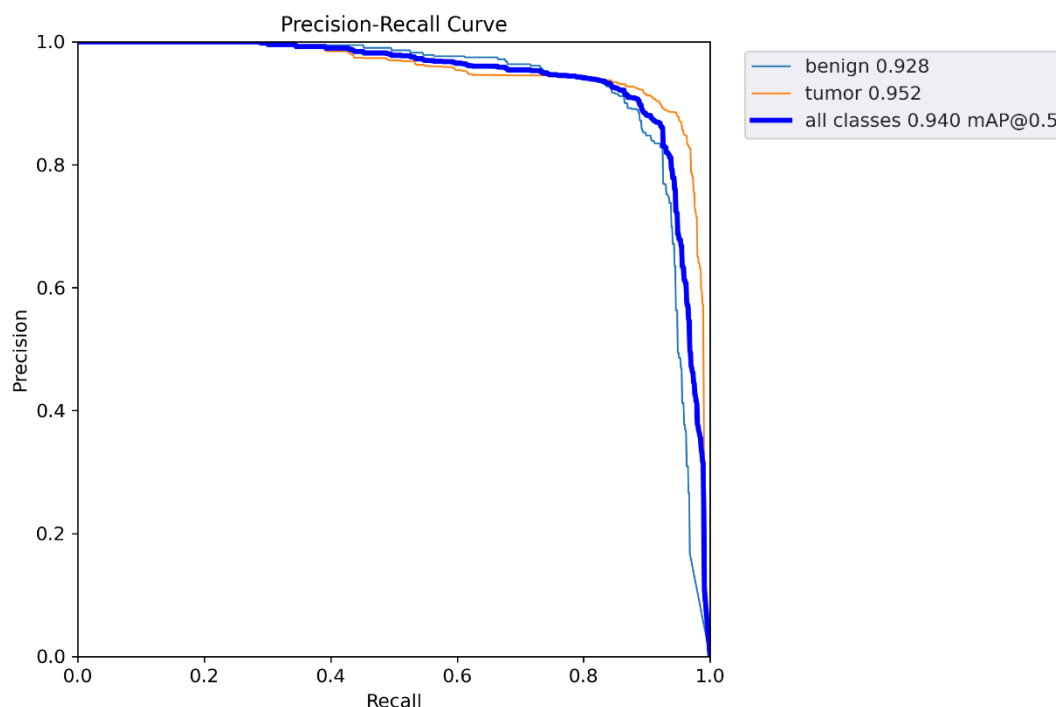


图 3-2 CIoU 训练下的 PR 曲线

从图中可以看出：

benign 类别的 PR 曲线较为平滑，AUC（曲线下的面积）较大，说明模型对良性肿瘤的识别具有较高的准确性，mAP@0.5 达到 0.928；

tumor 类别的 PR 曲线同样表现良好，mAP@0.5 达到 0.952，模型对恶性肿瘤的定位能力较强；

综合两类的表现，模型的整体 mAP@0.5 为 0.940，验证了 YOLOv5s 在该任务下良好的检测能力。

该 PR 曲线展示了模型在多类别检测中的性能表现，为下一步优化提供了参考依据。

3.4 模型优化方法

mAP 是评估深度学习模型性能的关键指标。在图像分类任务中，mAP 用于衡量模型在真实图像数据上的表现，帮助选择最适合任务的模型。

影响 mAP 指标的因素主要有：

① 模型复杂度直接影响 mAP。复杂度较高的模型能够更好地拟合数据，提高测试集上的性能。然而，过于复杂的模型会增加训练时间，降低模型的整体表现。

② 数据集的规模也是影响 mAP 的重要因素。数据集越大，模型性能越好。然而，数据集过大时，训练时间也会延长，可能影响模型效果。

③ 超参数的选择对于模型的性能至关重要。合理的超参数能够改善模型的拟合能力，提高测试集的性能，不当的选择则可能导致性能下降。

基于这些因素，以下是一些优化 mAP 的方法：

① 通过旋转、平移等手段增加训练数据样本，提升模型的鲁棒性和泛化能力。

② 采用先进的模型结构，如 ResNet，或对现有模型进行调整，如增加网络层数、调整学习率等。

③ 选择合适的损失函数，如 IoU Loss，以提高模型对难识别样本的敏感度，从而提升精度。

④ 通过不同尺寸的输入数据进行训练，帮助模型适应不同大小的目标，提高检测能力。

⑤ 将不同的目标检测网络进行融合，如 Faster R-CNN 与 SSD 的结合，提升模型性能。

本研究主要根据第二和第三点优化思路对我们的模型进行有效的性能提升。

3.4.1 更换 IoU 之 MPDIoU

在目标检测任务中，损失函数经常用于度量神经网络预测信息与期望标签的距离，预测信息越接近期望信息，损失函数值越小。在 YOLO 系列的默认设置中通常使用 CIoU，它主要衡量的是长宽比的差异而非长宽位置与置信度之间的实

际偏差，这在检测过程中可能导致一定问题。如图 3-3(a)所示，图中的外框代表实际检测框，内框表示预测框，当两个框的中心点重合且长宽比相同时，CIoU 会失效。另外如图 3-3(b)所示，在多个预测框大面积重叠的情况下，CIoU 不能反映出实际情况。

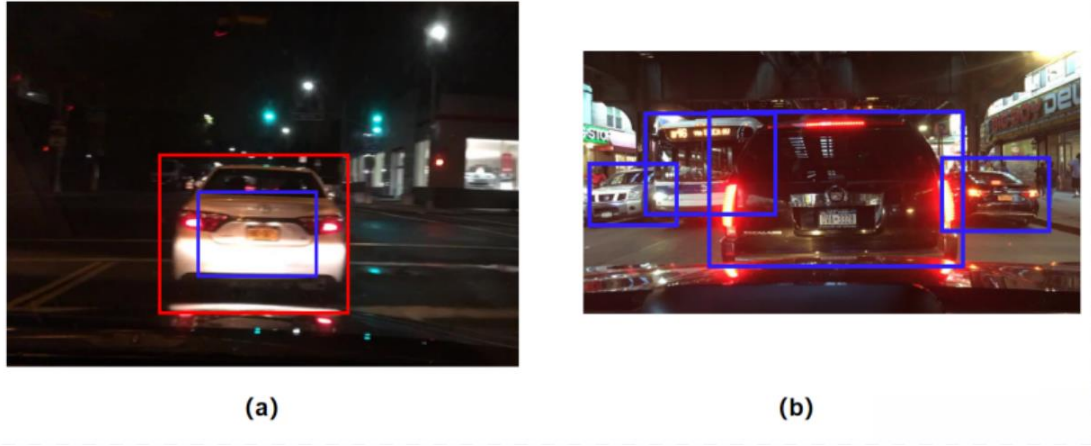


图 3-3 CIoU 失效的情况

为了解决上述问题，研究者深入挖掘了水平矩形的几何特性，提出了一种基于最小点距离的边界框相似度量方法——MPDIoU^[9]。该方法融合了现有各类损失函数所考虑的关键因素，例如重叠区域、中心点间距以及边界框的宽高差异，同时在计算上更加简洁高效。

MPDIoU 通过优化预测框与真实框左上角与右下角坐标点之间的距离，实现更精确的边界框拟合。在训练阶段，它通过公式迫使模型预测的每个边界框（公式 3.2）接近其实际边界框（公式 3.3）：

$$L = \min_{\Theta} \sum_{B_{gt} \in \mathcal{B}_{gt}} L(B_{gt}, B_{prd} | \Theta) \quad (3.1)$$

$$B^{prd} = [x^{prd}, y^{prd}, w^{prd}, h^{prd}]^T \quad (3.2)$$

$$B_{gt} = [x_{gt}, y_{gt}, w_{gt}, h_{gt}]^T \quad (3.3)$$

其中, B_{gt} 是实际边界框的集合, Θ 是用于回归的深度模型的参数。 L 的典型形式是 $L_n - norm$ 。

MPDIoU 的损失函数公式如下:

$$d_1^2 = (x_1^B - x_1^A)^2 + (y_1^B - y_1^A)^2 \quad (3.4)$$

$$d_2^2 = (x_2^B - x_2^A)^2 + (y_2^B - y_2^A)^2 \quad (3.5)$$

$$MPDIoU = \frac{A \cap B}{A \cup B} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (3.6)$$

$$L_{MPDIoU} = 1 - MPDIoU \quad (3.7)$$

其中, $(x_1^A, y_1^A), (x_2^A, y_2^A)$ 表示边界框 A 的左上角和右下角的坐标, $(x_1^B, y_1^B), (x_2^B, y_2^B)$ 表示边界框 B 的左上角和右下角的坐标。

实验结果表明, 在将 MPDIoU 损失函数引入 YOLACT (实例分割) 和 YOLOv7 (目标检测) 等先进模型后, 在 PASCAL VOC、MS COCO 和 IIT5k 等数据集上的性能表现均优于传统损失函数。

综上, MPDIoU 不仅简化了边界框之间相似度的计算过程, 同时在处理重叠与非重叠框的回归任务中也展现出更强的适用性。因此, 在二维或三维计算机视觉任务中, MPDIoU 有望作为交并比的有效替代方案, 用于衡量整体性能。

下面是 MPDIoU 在验证集上的表现, 训练设置与 CIoU 保持一致, 统一在同一验证集上进行性能评估, 得到如下结果与 PR 曲线:

表 3-3 两种模型在验证集上的表现

损失函数	类别	Precision	Recall	mAP@0.5	mAP@0.5:0.95
CIOU	benign	0.839	0.901	0.928	0.531
	tumor	0.901	0.898	0.952	0.625
	total	0.870	0.900	0.940	0.578
MPDIoU	benign	0.859	0.904	0.934	0.573
	tumor	0.900	0.888	0.939	0.641
	total	0.880	0.896	0.937	0.607

从上述结果可以看出：

- 在 mAP@0.5:0.95 这一综合评价指标上，MPDIoU 相较 CIOU 实现了约 5.0%的提升（0.578 → 0.607），说明 MPDIoU 在目标框回归的精细度方面具有更强的表现；
- 虽然在部分指标上（如 benign 类的 Recall）两者接近，但总体精度（Precision）从 0.870 提升至 0.880，说明 MPDIoU 有助于减少误检；
- tumor 类别下的 mAP@0.5:0.95 更是从 0.625 提升至 0.641，进一步验证了其在边界框拟合精度方面的优势。

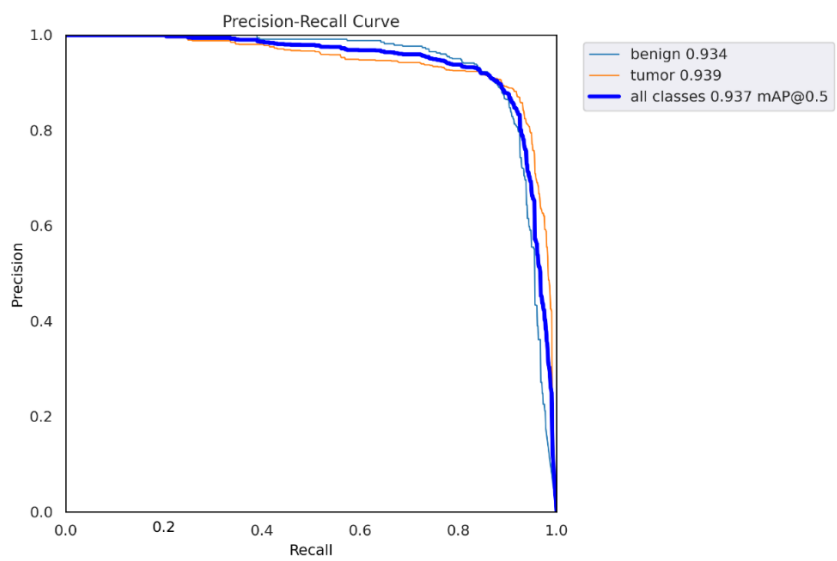


图 3-4 MPDIoU 训练下的 PR 曲线

因此,MPDIoU 作为一种更具惩罚性与区分度的 IoU 变体,显著提升了 YOLOv5 在医学图像目标检测中的回归性能,在特定场景中具备替代 CIoU 的潜力。

3.4.2 更换损失函数之 NWD

在观察数据集的过程中,我发现了一个现象(图 3-5):与恶性肿瘤相比,许多良性肿瘤的体积较小,在临床中表现为小结节,这给自动化诊断方法的特征提取带来巨大困难。

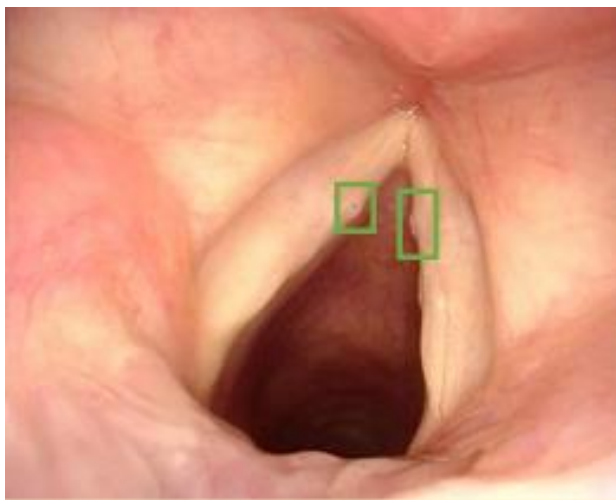


图 3-5 良性肿瘤目标太小

在小目标场景下, IoU 会经常为 0 或非常不稳定,原因如下:

- 预测框稍微偏一点,交集就为 0。
- 框很小, IoU 对宽高或位置误差特别敏感。
- IoU 本身不是一个“连续光滑”的函数,梯度不稳定,训练难收敛。

所以用 IoU 类损失(比如 DIoU/CIoU)训练小目标时,模型容易学不到有用信息。

为了更好地提升模型的性能,我查找到了 NWD 这一个损失函数。NWD (Normalized Wasserstein Distance) 是一种全新的边界框相似度度量方法。它通过将边界框建模为高斯分布,并采用 Wasserstein 距离来计算分布间的相似度,从而替代传统的 IoU 指标。相比于 IoU, NWD 即使在两个框几乎没有重叠

甚至完全不重叠的情况下，也能有效评估相似性。此外，该方法对目标的尺度变化不敏感，特别适用于小目标检测任务，表现更为稳定^[10]。

NWD 更适于小目标检测的主要原因有以下几点：

① 它不依赖重叠面积

NWD 把框看成高斯分布，用中心点 + 尺寸表示，然后计算 Wasserstein 距离：

$$NWD(N_a, N_b) = \exp\left(-\frac{\sqrt{w_2^2(N_a, N_b)}}{C}\right) \quad (3.8)$$

这意味着即使两个框完全没有重叠（IoU = 0），NWD 也能提供一个“连续、可微”的距离反馈，告诉模型“你还差多远”。其中，（3.8）公式中的 $w_2^2(N_a, N_b)$ 表示的是在假设边界框 $R = (cx, cy, w, h)$ 的情况下，对于两个边界框来说的 2 阶 Wasserstein 距离：

$$W_2^2(N_a, N_b) = \left\| \left[cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \right]^T, \left[cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \right]^T \right\|_2^2 \quad (3.9)$$

② 它对小尺度的变化不敏感

通常小目标的框尺寸非常小，稍有偏差的话 IoU 就会失效，但 NWD 使用的是“距离 + 方差”结构。在这种情况下，平移误差和尺寸误差被建模成两个高斯分布的差异，即使预测框和真实框大小不一样，NWD 也能平滑地学习。

③ 它是连续、光滑、全可导的

Wasserstein 距离是数学上良定义的度量，它对位置和形状变化都很敏感，但又不会“跳跃”，所以模型更容易收敛。

④ 它进行了归一化处理，对尺度不敏感

它引入了归一化常数 C ，通常设为图像对角线长度，使得所有目标（不论大小）都落在类似的尺度区间，提高训练稳定性。

综上，相比于 l_{IoU} ，NWD 有以下好处：

- 尺度不变性。
- 对于位置的差别变换平缓。
- 具有度量不相交的框的相似度的能力

表 3-4 五种模型在验证集上的表现

损失函数	类别	Precision	Recall	mAP@0.5	mAP@0.5:0.95
CIOUS	benign	0.839	0.901	0.928	0.531
	tumor	0.901	0.898	0.952	0.625
	total	0.870	0.900	0.940	0.578
MPDIoU	benign	0.859	0.904	0.934	0.573
	tumor	0.900	0.888	0.939	0.641
	total	0.880	0.896	0.937	0.607
NWD	benign	0.851	0.909	0.942	0.570
	tumor	0.913	0.921	0.949	0.640
	total	0.882	0.915	0.945	0.605
0.5NWD + 0.5CIOUS	benign	0.868	0.906	0.937	0.573
	tumor	0.903	0.892	0.944	0.646
	total	0.885	0.899	0.940	0.609
0.3NWD + 0.7CIOUS	benign	0.867	0.905	0.937	0.572
	tumor	0.909	0.897	0.951	0.653
	total	0.888	0.901	0.944	0.613

本节选取 NWD 作为替代损失函数进行实验，并尝试与 CIOUS 进行融合来提升模型整体性能。为全面评估 NWD 在本任务中的表现，设置了以下三组实验：

- 纯 NWD：完全以 NWD 作为回归损失函数；
- NWD 与 CIOUS 等权重融合（0.5 NWD + 0.5 CIOUS）；

- NWD 与 CIoU 加权融合 ($0.3 \text{ NWD} + 0.7 \text{ CIoU}$), 以偏重 CIoU 提供的几何约束能力, 因为小目标的图像数量相对来说较少, 所以我希望 CIoU 的权重更大一些。

各方案在验证集上的性能如表 3-4 所示, 并与 CIoU 和 MPDIoU 的表现进行对比。

从实验结果可以看出:

- 采用纯 NWD 损失函数的模型在 benign 类上的 recall 值最好, 确实发挥了其在小目标检测上的优势。通过观察发现有些指标提升不多, 可能是数据集中的小目标 benign 类数量不多的原因导致提升不太明显。但是, 这并不能否认其在小目标检测中的优势
- 融合策略 ($0.5 \text{ NWD} + 0.5 \text{ CIoU}$) 能在兼顾两者优点的同时获得较稳定的检测性能, 特别是在各指标上接近 MPDIoU 的水平;
- 加权融合 ($0.3 \text{ NWD} + 0.7 \text{ CIoU}$) 在两个类别的检测效果上表现都不错, 其 Recall (0.897) 和 $\text{mAP}@0.5:0.95$ (0.653) 均为当前最优, 表明适当偏重 CIoU 能提升目标定位的精度, 而 NWD 提供了较好的鲁棒性补充。

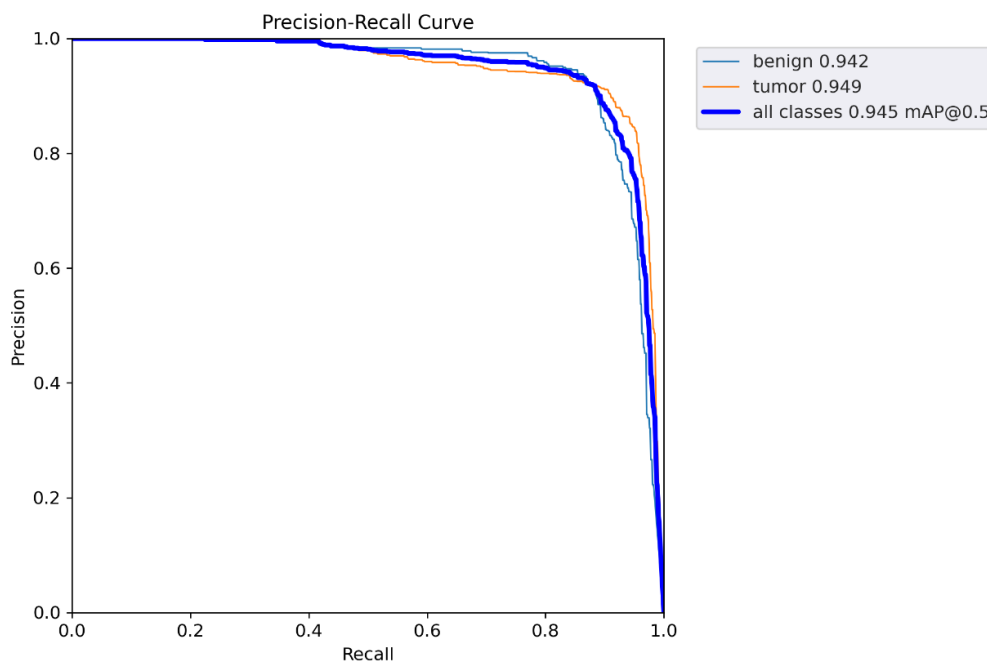


图 3-6 NWD 训练下的 PR 曲线

综上可见，NWD 与 CIOU 的组合相较于单一损失函数更具优势，尤其适用于本任务中目标尺度变化大、边界复杂等实际问题。该结论为后续模型进一步优化提供了有效的方向。

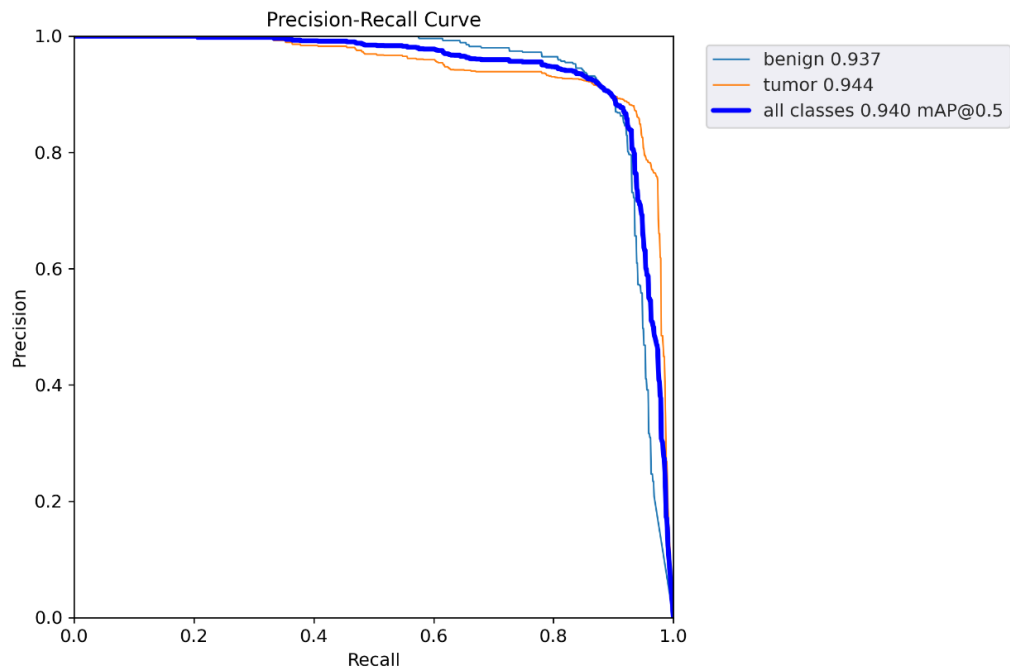


图 3-7 0.5NWD + 0.5CIOU 训练下的 PR 曲线

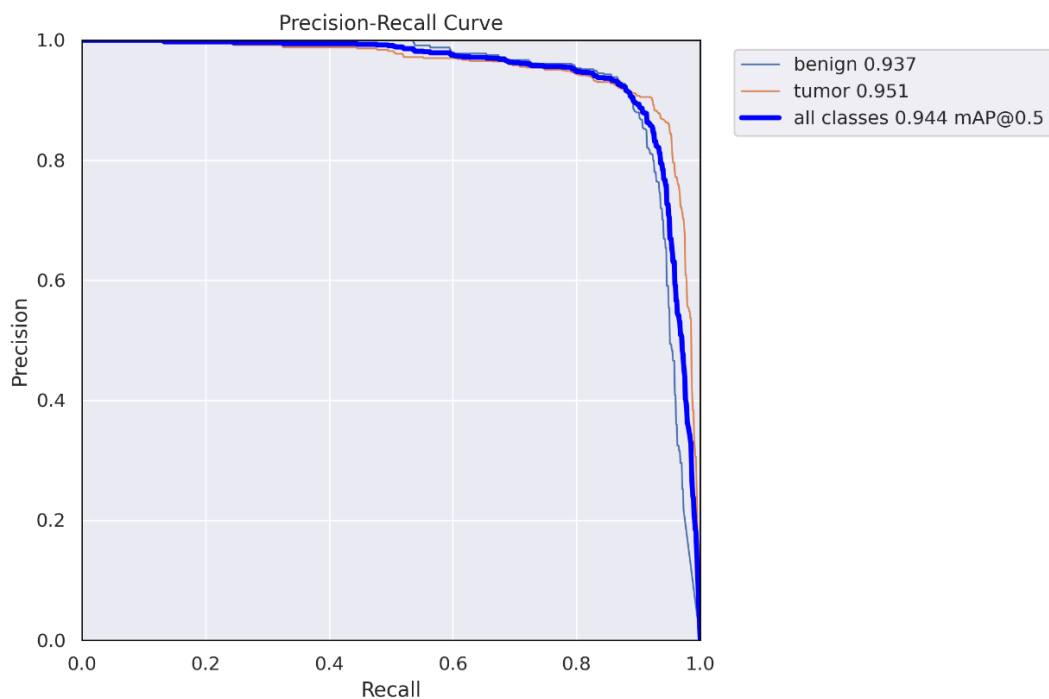


图 3-8 0.3NWD + 0.7CIOU 训练下的 PR 曲线

3.4.3 增加小目标检测层

首先，引入“小目标”的概念。以通用目标检测数据集 COCO 的标准为例，小目标通常被定义为尺寸小于 32×32 像素的目标；中等目标为尺寸在 32×32 至 96×96 之间的目标，而尺寸大于 96×96 的则视为大目标。在实际应用中，研究者更倾向于使用相对尺寸的方式来定义小目标：若目标标注框的长宽乘积开根号后，占原图对角线长度的比例小于 3%，则可视为小目标。

在小目标检测任务中，常见的实际场景主要包括以下两类：

(1) 大小目标混合的场合

在此类场景中，一张图像中包含少量大目标和大量小目标。因此，我们往往会遇到能够准确地检测大目标但检测不到小目标和小目标的 recall 率很低等问题。

(2) 仅包含小目标的场景

在此类场景中，所有目标均为小尺寸目标。常见挑战与上述类似，包括：recall 较低、小目标密集导致特征干扰、模型在训练过程中难以有效学习小目标特征等。

对于以上的情况，我们可以通过增加模型的小目标检测层来解决。图 3-9 是 YOLOv5-6.0 的网络结构，由输入端+Backbone+Neck+Head 组成：

图中我们可以看到原始的 YOLOv5 有 3 个检测头：分别是 20×20 （大目标）、 40×40 （中目标）和 80×80 （小目标），我们要增加小目标检测层，就可以在 80×80 的上一步，也就是 160×160 尺寸增加（图 3-10）。

在增加一层检测层以后，网络结构就变成了图 3-11。其中有两个部分要注意：

① 红框部分表示删除区域。由于原始网络的 Neck 模块并不包含 160×160 的特征图，因此我们对 80×80 的特征图进行一次上采样操作，从而生成一个新的 160×160 特征图。在未增加检测层之前，该部分是网络的最末端，检测

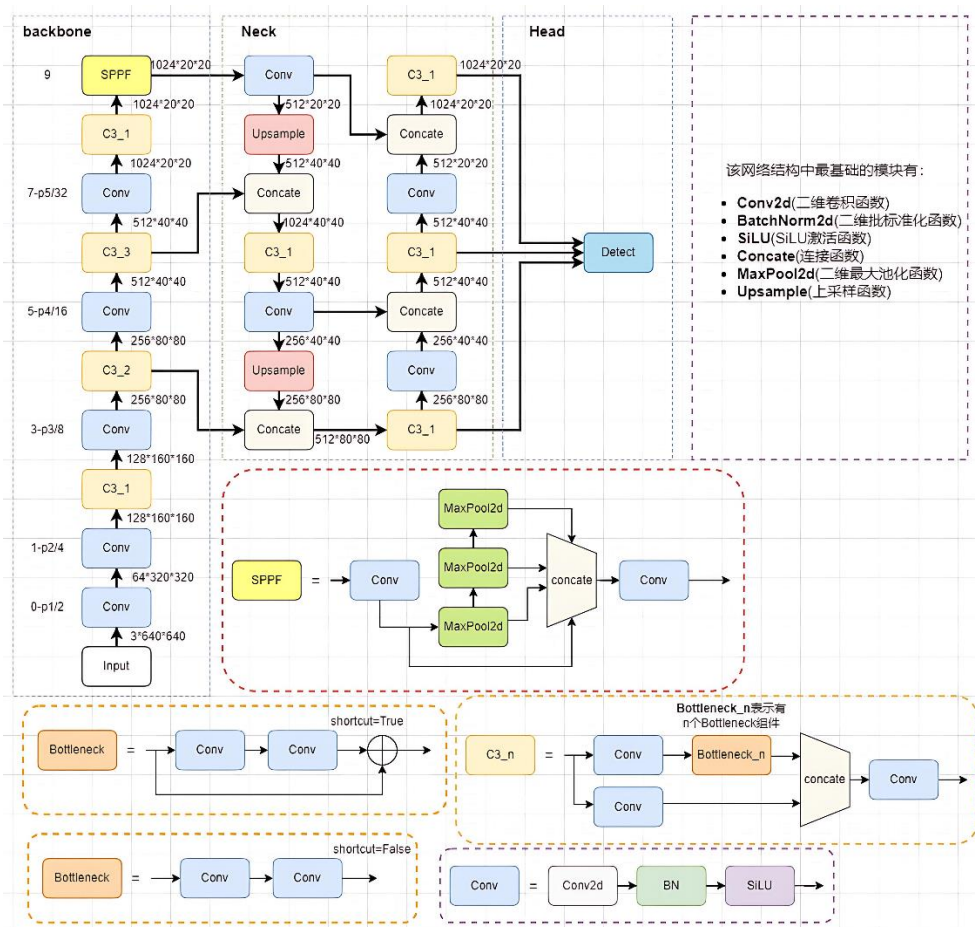


图 3-9 YOLOv5-6.0 的网络结构

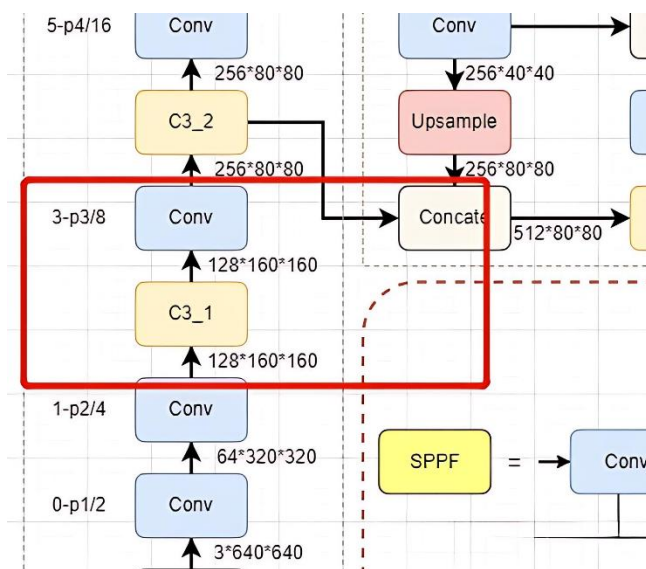


图 3-10 增加的部位

头可直接获取该特征图；而现在引入新的检测层后，该特征图不再处于末端，需从其下层获取。

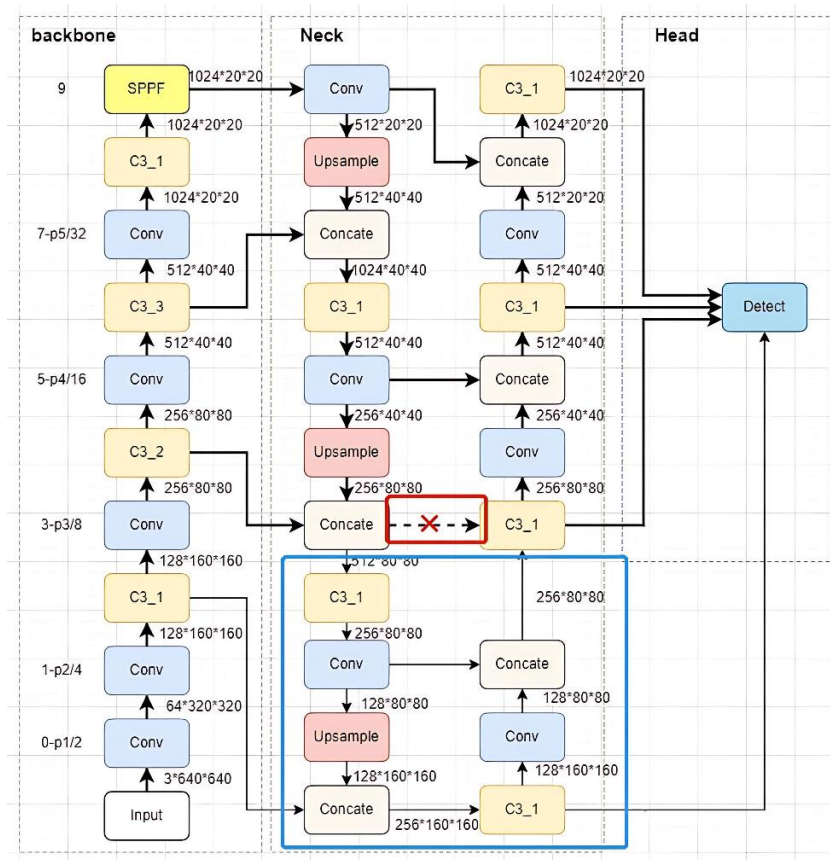


图 3-11 增加检测层后的网络结构

② 蓝框部分是新添加的检测层。自上而下地看，先对 Neck 中的 80×80 特征图进行上采样，得到 160×160 尺寸的特征图。该特征图再与 Backbone 第二层的 160×160 特征图进行 concat 融合，形成新的 160×160 特征图。接着向右传递，通过一次下采样重新生成 80×80 的特征图（FPN + TPN），并经由 C3 模块传输至检测头。

实验结果如表 3-5 所示：

从实验数据可以看出，增加小目标检测层的模型在 Precision (0.891)、mAP@0.5 (0.949)、mAP@0.5:0.95 (0.619) 等关键指标上均取得了当前所有实验方案中的最高值，在 Recall (0.911) 方面也排名第二，仅次于纯 NWD。这表明，针对内窥镜图像中存在大量尺寸较小、边界模糊的喉部肿瘤目标，通过引入额外

的小目标检测层,模型能够更有效地捕捉微小特征区域,显著提升整体检测性能,为后续的临床辅助诊断提供了更为可靠的技术支撑。

表 3-5 六种模型在验证集上的表现

改进方法	类别	Precision	Recall	mAP@0.5	mAP@0.5:0.95
CIOU	benign	0.839	0.901	0.928	0.531
	tumor	0.901	0.898	0.952	0.625
	total	0.870	0.900	0.940	0.578
MPDIoU	benign	0.859	0.904	0.934	0.573
	tumor	0.900	0.888	0.939	0.641
	total	0.880	0.896	0.937	0.607
NWD	benign	0.851	0.909	0.942	0.570
	tumor	0.913	0.921	0.949	0.640
	total	0.882	0.915	0.945	0.605
0.5NWD +	benign	0.868	0.906	0.937	0.573
0.5CIOU	tumor	0.903	0.892	0.944	0.646
	total	0.885	0.899	0.940	0.609
0.3NWD +	benign	0.867	0.905	0.937	0.572
0.7CIOU	tumor	0.909	0.897	0.951	0.653
	total	0.888	0.901	0.944	0.613
小目标检测 层	benign	0.872	0.896	0.940	0.585
	tumor	0.909	0.926	0.957	0.654
	total	0.891	0.911	0.949	0.619

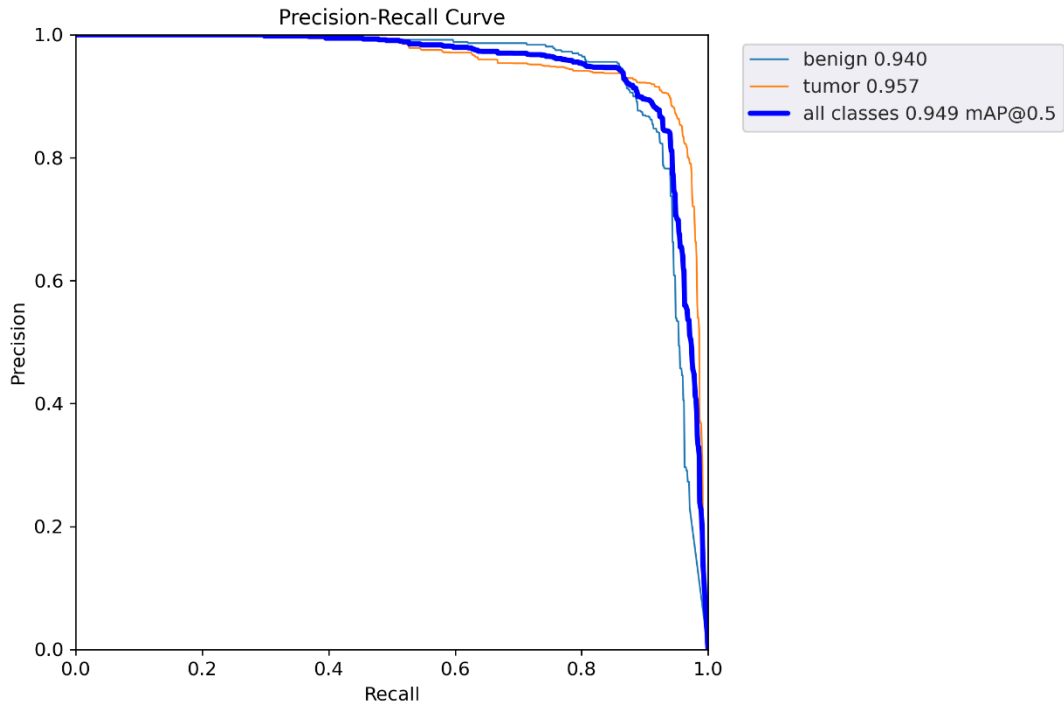


图 3-12 增加小目标检测层训练下的 PR 曲线

3.5 小结

本章围绕基于内窥镜图像的 YOLO 建模与优化方法展开了深入探讨，主要分为模型构建与三类优化策略两个部分。

在建模阶段，首先对数据集进行了系统整理与预处理，确保模型训练的有效性与鲁棒性。随后选择了 YOLOv5 模型作为检测框架，结合其轻量化结构、高效推理能力以及对小目标良好的检测性能，为本研究提供了可靠的技术基础。在模型训练与验证部分，构建了适配内窥镜图像的训练流程，并在验证集上评估了模型初始性能。

在优化研究部分，本文提出了三种改进方案：其一，采用 MPDIoU 替代原有的 CIoU 损失函数，在一定程度上提升了定位精度；其二，引入 NWD 损失函数并设置与 CIoU 的组合权重，充分融合两者的优点，提高了模型的综合表现；其三，增加小目标检测层，从结构上提升模型对喉部微小肿瘤的检测能力。通过多组对比实验验证，这些优化策略在不同维度上有效增强了模型性能，表现出良好的实用价值和扩展潜力。

综上，本章完成了从 YOLO 建模到性能优化的系统工作，不仅实现了内窥镜图像下喉部肿瘤的初步检测目标，也为后续的性能评估与部署应用奠定了坚实基础。

4. 实验结果分析与讨论

4.1 实验环境与设置

为验证所提出的 YOLOv5 模型在喉癌内窥镜图像检测任务中的有效性，本研究在软硬件环境均可控的条件下进行了系统的实验设置与训练测试。实验采用 PyTorch 深度学习框架，在具备 GPU 加速的计算服务器上完成模型训练、测试与评估。

4.1.1 硬件环境

本实验运行于一台配置较高的服务器，其主要硬件配置如下：

- 操作系统：Ubuntu 20.04.6 LTS
- 处理器（CPU）：AMD EPYC 7282 16-Core Processor
- 显卡（GPU）：NVIDIA RTX A6000
- CUDA 版本：11.8

GPU 的引入大幅提升了模型的训练效率，保证了大规模图像数据集在深度网络中的高效处理。

4.1.2 软件环境

本研究基于 Python 语言构建模型训练与分析流程，主要的软件配置如下：

- 编程语言：Python3.10.4
- 深度学习框架：Pytorch 2.2.2
- 图像处理与可视化：OpenCV、Matplotlib、Seaborn
- 其他依赖库：NumPy、Pandas、tqdm 等

YOLOv5 模型采用官方开源版本，并在其基础上进行部分定制优化，包括修改配置文件、训练超参数调整、检测层改进等。

4.1.3 模型训练设置

在 YOLOv5 模型训练过程中，设置的核心训练参数如下：

- Batch Size: 16（表示每个训练 batch 里包含 16 张图像）
- 初始学习率: 0.01，采用默认的线性衰减（Linear Decay）策略动态调整学习率
- 优化器: SGD（动量为 0.937，权重衰减 0.0005）
- 损失函数: 采用原有的 CIOU 和后面引进的 MPDIoU
- 训练轮数（Epoch）: 200
- Anchor 调整: 启用自动聚类锚框机制（AutoAnchor）

4.2 实验结果与讨论

在本节中，我们基于测试集对不同模型改进策略进行全面评估与比较，以验证各方法的泛化能力和实际应用效果。通过记录 Precision、Recall、mAP@0.5 与 mAP@0.5:0.95 等核心指标，我们旨在从多个维度分析各方案的性能优劣，并为后续系统部署提供数据支持。

表 4-1 各模型在测试集上的表现

方法	Precision	Recall	mAP@0.5	mAP@0.5:0.95
CIOU（基线）	0.861	0.891	0.923	0.562
MPDIoU	0.869	0.888	0.926	0.580
NWD	0.871	0.900	0.930	0.578
0.5NWD + 0.5CIOU	0.873	0.894	0.928	0.581
0.3NWD + 0.7CIOU	0.876	0.907	0.934	0.588
增加小目标检测层	0.880	0.902	0.939	0.602

从上述表格可以看出，不同方法在测试集上均表现出对基线模型（CIOU）的明显改进效果：

- CIOU 作为原始回归策略，虽在 Recall 和 $mAP@0.5:0.95$ 上略逊一筹，但依然维持了较好的基础检测水平， $mAP@0.5$ 为 0.923， $mAP@0.5:0.95$ 达到 0.562，为后续改进提供了坚实的比较基线。
- MPDIOU 的引入使得模型对目标定位的回归精度有所提升，尤其在 $mAP@0.5:0.95$ 上升至 0.580，优于基线模型。这说明 MPDIOU 能够在边界框回归过程中提供更加平滑且合理的梯度反馈，提升模型对肿瘤边界的刻画能力。
- NWD 损失函数作为一种衡量目标中心点距离的策略，能够缓解小目标偏移带来的误差。其 Precision 与 Recall 水平均有小幅提升，表现出对检测鲁棒性的增强，但在综合指标上略逊于权重融合方法。
- $0.5 \text{ NWD} + 0.5 \text{ CIOU}$ 和 $0.3 \text{ NWD} + 0.7 \text{ CIOU}$ 两种融合策略中，后者的表现更为优异：其 Recall 达到了 0.907，为六组方法中最高，说明其在不漏检的前提下仍具备较强的检测能力。这也体现了 CIOU 在保持大体定位能力的同时，辅以 NWD 提供的局部目标中心约束，可以有效提升召回率。
- 增加小目标检测层 的方法表现最为突出，其 Precision 为 0.880， $mAP@0.5$ 达 0.939， $mAP@0.5:0.95$ 达到 0.602，为所有策略中最高。这说明通过结构性改进提升模型对不同尺度目标（尤其是小目标）的感知能力，是提升模型检测性能的重要手段。尤其在面对分辨率低、肿瘤目标微小的内窥镜图像中，增强对小目标的捕捉能力对最终诊断结果有着积极作用。

4.3 模型性能评估

为全面评估所提出模型在喉部肿瘤检测任务中的性能表现，本文从多个角度对模型进行性能分析与对比，涵盖检测精度、召回能力、目标定位能力与整体鲁棒性等关键维度。实验中选择了不同的损失函数组合策略（如 MPDIOU、NWD、NWD 与 CIOU 融合）以及结构改进手段（如增加小目标检测层），并基于统一的验证集与测试集，确保评估结果的客观性与可比性。

首先，从平均精度均值（ $mAP@0.5$ ）来看，所有改进方法均较基线模型（使用 CIOU）有不同程度的提升，表明无论是引入更合理的回归损失，还是增强小

目标检测能力，都对整体检测性能具有积极影响。尤其是引入小目标检测层的模型，在 $mAP@0.5$ 上达到了 0.939，较 CIOU 提升约 1.6%，在 $mAP@0.5:0.95$ 上提升也较为显著，说明其对不同尺度目标的检测均具有良好适应性。

其次，从召回率 (Recall) 角度观察，0.3 NWD + 0.7 CIOU 权重融合模型表现最为突出，达到了 0.907，反映该模型在最大程度降低漏检率方面表现最为优异，对于医学图像辅助诊断场景具有更高实用价值。而 Precision (精度) 方面，增加小目标检测层模型以 0.880 的最高值位居第一，表明该结构对于减少误检具有明显优势。

此外，在评估过程中还引入了 $mAP@0.5:0.95$ 指标以衡量模型对边界框精度的更严格要求。相比 $mAP@0.5$ ，该指标更加考验模型在不同 IoU 阈值下的稳健性。其中，结构改进带来的提升最为显著，进一步验证了增强特征层次感知能力对于复杂内窥镜图像处理的重要性。

综合上述指标可得，不同方法在多个维度上各有优势，具体表现如下：

- 增加小目标检测层：综合性能最优，尤其在检测精度与定位能力上具备强大优势，适合部署在目标尺度分布广泛的实际应用中；
- 0.3 NWD + 0.7 CIOU：召回能力最强，有效缓解漏检问题，适用于医疗影像中高敏感性要求的场景；
- MPDIoU / NWD / 0.5 NWD + 0.5 CIOU：性能稳定，整体略优于基线，适合作为中间方案；
- CIOU (基线)：尽管性能略低，但作为对比参考具有良好的代表性，为改进方法的有效性提供了评估依据。

综上所述，本节通过系统的指标对比，验证了不同模型优化策略的有效性与适应性，为后续模型部署与系统落地提供了坚实的理论与实验支持。

4.4 小结

本章围绕本文提出的基于 YOLO 的喉部肿瘤检测方法，进行了系统的实验设计与结果分析。首先，明确了模型训练与测试所依赖的软硬件环境和训练参数设置，保证了实验的可复现性和公平性。随后，通过在验证集和测试集上对不

同模型优化方案（如更换 IoU 计算方式、调整损失函数权重、增加小目标检测层）进行对比实验，评估了各模型在精确度、召回率及 mAP@0.5 等多个指标上的性能表现。

实验结果表明，通过引入小目标检测层和合理设计的混合损失函数（如 $0.3 \text{ NWD} + 0.7 \text{ CIoU}$ ）可以显著提升模型在喉部肿瘤检测任务中的整体识别效果，尤其在保持高精度的同时进一步提升了对小目标的召回能力。此外，相较于传统的 CIoU 方案，MPDIoU 与 NWD 等改进策略在部分指标上也展现出良好的适应性和泛化能力。

通过本章的实验与分析，为后续模型在临床辅助诊断中的应用提供了有效的依据，也为模型的进一步优化和扩展打下了坚实基础。

5. 喉部肿瘤辅助诊断系统设计与实现

5.1 功能模块实现

为了实现对喉部内窥镜图像中肿瘤的自动检测与辅助诊断，本文基于优化后的 YOLO 模型设计并开发了一个完整的软件系统。该系统通过集成图像与视频的上传、肿瘤检测、结果可视化等模块，旨在为临床医生提供便捷高效的辅助诊断工具。

系统采用 Flask 框架进行后端开发，前端采用静态页面技术 (HTML/CSS/JS) 进行可视化界面设计并使用 Bootstrap 框架进行样式美化，整体界面简洁直观，操作便捷。主要功能模块将在后面的几小节中详细描述。

5.1.1 图像上传模块

图像上传模块是系统的第一步入口，用户可通过系统界面上传待检测的内窥镜图像或视频文件。

功能特点：

- 支持多种格式 (.jpg, .png, .mp4) 上传；
- 上传后自动保存至服务器指定目录；
- 显示预览，便于用户确认图像内容是否符合要求。

5.1.2 肿瘤检测模块

该模块是系统的核心部分，基于优化后的 YOLOv5 模型对上传的图像进行目标检测。模型已在喉部内窥镜数据集上进行训练与优化，能够对图像中的肿瘤区域进行定位与分类。

功能特点：

- 一键调用模型进行推理；
- 自动识别图像中的肿瘤类型 (良性/恶性)；
- 输出检测框位置、类别及置信度等信息；

- 支持图像与视频两种输入形式。

5.1.3 结果可视化模块

检测结果通过可视化方式展示在前端页面，便于用户直观理解模型判断。系统会将检测框叠加在原始图像上，并以不同颜色标注肿瘤类型。同时显示每个检测目标的置信度得分。

功能特点：

- 检测框颜色区分肿瘤类型（如良性为蓝色，恶性为青色）；
- 显示类别与概率信息；
- 支持检测结果保存与下载；

5.2 页面展示

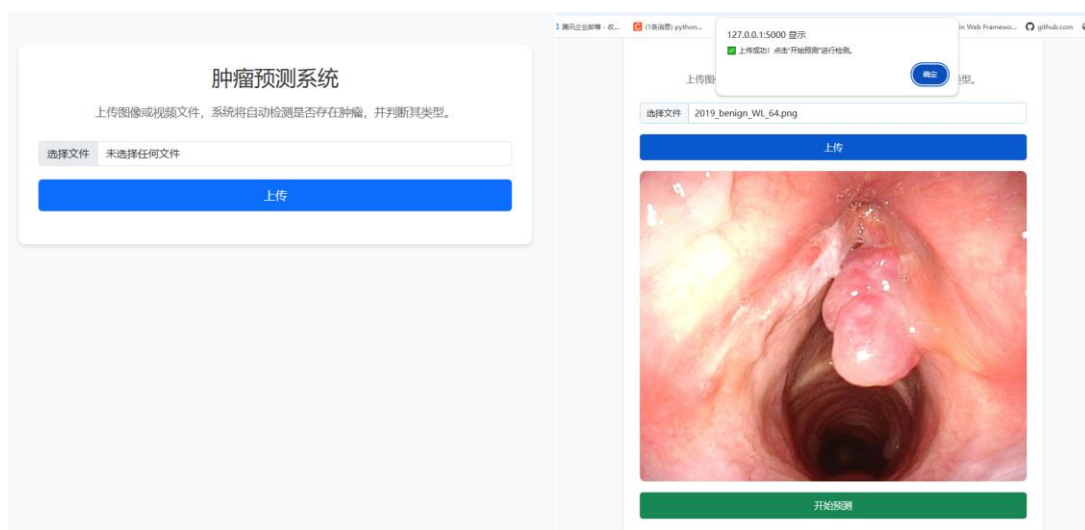


图 5-1 预测系统首页

图 5-2 用户上传内窥镜图像

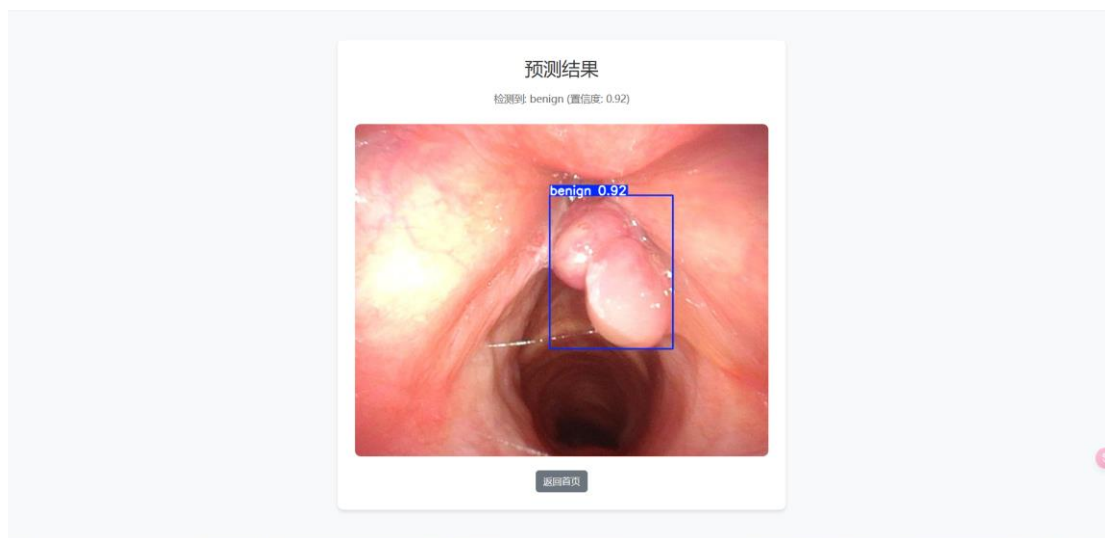


图 5-3 预测结果展示

整个系统的运行流程如下：

1. 用户上传内窥镜图像；
2. 模型进行检测并返回预测结果；
3. 可视化模块将检测结果叠加在图像上并展示；
4. 用户可选择保存或重新上传图像；
5. 可选择返回预测界面继续预测。

界面设计方面，系统采用简洁的交互风格，操作按钮明确，检测结果区域突出，便于医生在最短时间内获取有效信息。

5.3 小结

本章介绍了喉部肿瘤辅助诊断系统的整体架构与主要模块实现。该系统在前端易用性与后端模型准确性之间取得良好平衡，初步实现了从图像输入到检测结果展示的一体化流程。后续将继续完善系统功能，拓展多模态输入与用户交互性，提升其在临床应用中的实用价值。

6. 总结与展望

6.1 研究总结

本文围绕“基于内窥镜图像的 YOLO 建模与优化研究”这一主题，系统地开展了从基础模型构建到实际应用系统开发的一系列研究与实验，旨在提升喉部肿瘤的自动检测效率与准确率，为临床提供辅助诊断工具，缓解医生工作负担。

首先，本文梳理了内窥镜图像处理的技术背景与目标检测算法的发展现状，深入分析了 YOLOv5 模型在医学图像目标检测任务中的应用潜力。随后，选取 YOLOv5s 作为基础架构，对其进行了建模训练，并以 CIOU 损失函数作为初始基线，通过在验证集上的表现验证其具备较高的准确性和稳定性。

其次，为进一步优化检测性能，本文从两个方向进行了改进实验：一是替换与融合不同的边界框回归损失函数，如 MPDIoU、NWD 以及 NWD 与 CIOU 的加权组合；二是引入小目标检测层结构改进，增强模型对尺寸较小病灶区域的感知能力。通过一系列对比实验，验证了改进方案在 mAP@0.5、Recall、mAP@0.5:0.95 等多个指标上的提升效果，尤其是在召回率与小目标检测精度方面取得了显著进步。

此外，本文还基于优化后的模型，开发了一个具备实用性的喉部肿瘤辅助诊断系统，实现了从图像上传、肿瘤检测到结果可视化的完整流程，提升了模型的工程落地能力，展示了算法研究与系统实现的有效融合。

总体而言，本文完成了以下主要研究工作：

- 1、构建了基于 YOLOv5s 的喉部肿瘤检测模型，并实现了针对内窥镜图像的适配与训练；
- 2、分别从损失函数设计和模型结构优化两个方面提升模型性能，系统对比了各改进方案的优劣；
- 3、设计并实现了具备良好交互性与可视化能力的辅助诊断系统，验证了模型在实际场景中的应用价值。

本研究在提升模型检测性能的同时，也为医学图像目标检测领域提供了具有

参考价值的思路与实验路径，对后续深入优化 YOLO 模型在医疗场景下的适应性具有一定借鉴意义。

6.2 研究不足与未来展望

尽管本研究在基于内窥镜图像的喉部肿瘤检测建模与优化方面取得了一定成果，但仍存在一些不足与待改进之处：

（1）模型改进相对局部

本文主要围绕损失函数与检测结构（小目标检测层）进行优化，但主干网络部分仍使用 YOLOv5s 默认架构，尚未尝试引入更深层次的结构改造，如轻量化网络、注意力机制等。

（2）系统功能仍较基础

当前构建的辅助诊断系统实现了基本功能，但在用户体验、智能交互方面仍有提升空间。例如结果追踪、病例管理、多图批量处理、结果报告导出等模块尚未完善。

（3）模型推理效率未深入评估

尽管改进方案在精度上取得一定提升，但对于模型在不同硬件平台上的推理速度与资源消耗等工程性指标未进行系统测试，影响其在终端设备或实际医院部署中的可用性分析。

基于以上几点的不足，我认为未来仍有广阔的探索空间，可以从以下几点出发：

（1）引入更先进的网络结构

可考虑将当前 YOLOv5 模型替换或融合更具表现力的结构，如 YOLOv8、EfficientDet、PP-YOLOE 等，甚至结合 Transformer、Swin Transformer 等视觉编码架构，提升对复杂特征的提取能力。

（2）多损失函数融合与动态权重机制

在损失函数方面，可以引入动态加权机制，自动根据训练状态调整 NWD、CIoU

等函数的权重比例，进一步提高训练稳定性和检测效果。

（3）实现多模态诊断系统

基于图像之外，未来可引入患者基本信息、声带视频、临床问诊文本等多模态数据，构建融合式辅助诊断系统，提升系统综合判断能力。

（4）开展临床验证与模型部署

推动模型在实际临床环境中的部署与试用，收集真实反馈，调整系统性能，逐步实现从研究型原型向临床级工具的转变。

参考文献

- [1] CATTARUZZA M,MAISONNEUVE P,BOYLE P. Epidemiology of laryngeal cancer[J].European Journal of Cancer Part B: Oral Oncology,1996, 32(5): 293-305.
- [2] 邓敏鑫. 窄带成像技术与专业图像增强技术在咽喉部恶性肿瘤的应用价值[D]. 南方医科大学, 2019.
- [3] LI Y,GU W,YUE H,et al. Real-time detection of laryngopharyngeal cancer using an artificial intelligence-assisted system with multimodal data[J].Journal of Translational Medicine, 2023,21(1): 698.
- [4] 倪晓光, 贺舜, 高黎, 等. 窄带成像内镜在喉咽癌早期诊断中的应用[J]. 中国耳鼻咽喉头颈外科, 2009, 16(10): 550-554.
- [5] 谭鑫平, 高志辉, 韩航迪, 等. 基于改进 YOLOv5 的荧光图像细胞智能检测研究[J]. 半导体光电, 2023, 44(5): 709-716. DOI:10.16818/j.issn1001-5868.2023052801.
- [6] Mohamed N , Almutairi R L , Abdelrahim S ,et al.Automated Laryngeal Cancer Detection and Classification Using Dwarf Mongoose Optimization Algorithm with Deep Learning[J].Cancers, 2024, 16(1):18.DOI:10.3390/cancers16010181.
- [7] Ali S , Dmitrieva M , Ghatwary N ,et al.Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy[J].Medical Image Analysis, 2021, 70(1):102002.DOI:10.1016/j.media.2021.102002.
- [8] Redmon J , Divvala S , Girshick R ,et al.You Only Look Once: Unified, Real-Time Object Detection[C]//Computer Vision & Pattern Recognition.IEEE, 2016.DOI:10.1109/CVPR.2016.91.
- [9] Ma S , Xu Y .MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression[J]. 2023.DOI:10.48550/arXiv.2307.07662.
- [10] Wang J , Xu C , Yang W ,et al.A Normalized Gaussian Wasserstein Distance for Tiny Object Detection[J]. 2021.DOI:10.48550/arXiv.2110.13389.
- [11] Zheng Z , Wang P , Liu W ,et al.Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[J].arXiv, 2019.DOI:10.1609/aaai.v34i07.6999.
- [12] Rezatofighi H , Tsoi N , Gwak J Y ,et al.Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression[C]//2019 IEEE/CVF Conference on Computer Vision

- and Pattern Recognition (CVPR).IEEE, 2019.DOI:10.1109/CVPR.2019.00075.
- [13] Lin T Y , Goyal P , Girshick R ,et al.Focal Loss for Dense Object Detection[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.DOI:10.1109/TPAMI.2018.2858826.
- [14] Yu J , Jiang Y , Wang Z ,et al.UnitBox: An Advanced Object Detection Network[J].ACM, 2016.DOI:10.1145/2964284.2967274.
- [15] 刘秋蕊, 赵宇. 人工智能技术在喉癌诊疗领域中的应用[J]. 中国耳鼻咽喉颅底外科杂志, 2021, 27(5): 530-533
- [16] 岳泉. 内窥镜影像系统设计及图像处理的研究[D]. 天津工业大学, 2007. DOI:CNKI:CDMD:2. 2007. 072288.
- [17] 李勇. 内窥镜图像技术在早期喉癌及其癌前病变诊断中的应用研究[C]//2005 年浙江省耳鼻咽喉科学学术会议. 0[2025-05-03]. DOI:ConferenceArticle/5aa3ca9dc095d72220bc4cf6.
- [18] 刘居林, 李德豪, 张振豪, 等. 基于改进 YOLOv5s 的花生仁检测系统[J]. 信息技术, 2024, 48(9):168-175.
- [19] 陈光, 乔梁, 黄晓明, 等. 基于 YOLOV5 的安全帽佩戴识别检测系统的研究和设计[C]//江西省电机工程学会 2023 年年会. 国能九江发电有限公司; 国能丰城发电有限公司, 2024.
- [20] 吴桐. 基于 Flask 框架的物资管理系统的设计与实现[J]. 南京大学, 2016.
- [21] 王译庆. Flask 框架下成品油销售系统设计与实现[D]. 西安电子科技大学, 2015. DOI:10. 7666/d. D01069068.
- [22] 舒后, 熊一帆, 葛雪娇. 基于 Bootstrap 框架的响应式网页设计与实现[J]. 北京印刷学院学报, 2016, 24(2):6. DOI:10. 3969/j. issn. 1004-8626. 2016. 02. 013.
- [23] 周萍, 赵娜, 李慕. Bootstrap 框架在响应式 Web 设计中的应用[J]. 软件导刊, 2017, 16(6):3. DOI:10. 11907/rjdk. 162838.
- [24] 马继昌, 李建兵, 刘玉冬. 基于 Python 语言及 Flask 框架快速开发网站的研究[J]. 数码世界, 2018, 000(012):50. DOI:10. 3969/j. issn. 1671-8313. 2018. 12. 045.
- [25] 吴小凤, 王天淼, 朱文秀. 基于 Flask 框架的监控平台可视化设计研究[J]. 工业控制计算机, 2024, 37(3):90-91.
- [26] 王宁. 基于 Bootstrap 框架的高校学生就业信息管理系统设计[J]. 自动化技术与应用,

2021. DOI:10.3969/j.issn.1003-7241.2021.04.040.

- [27] 不公告发明人. 一种基于 Bootstrap 框架设计的商城系统:CN201611157578.3[P]. CN108230069A[2025-05-03].
- [28] 高鑫,杨玉婷,袁韵洁. 基于 Flask 框架的 ad-hoc 网络管理系统实现[J]. 信息产业报道, 2023(2):0001-0003.
- [29] 王凌艳. Bootstrap 前端框架设计与实现[J]. 2020.
- [30] 叶青. HTML+CSS+JavaScript 实用详解[M]. 电子工业出版社, 2008.
- [31] 王艺燕. 基于 HTML+CSS+JS 的网页版 2048 的实现[J]. 电子世界, 2017(9):2. DOI:10.3969/j.issn.1003-0522.2017.09.020.
- [32] Thakur A , Dhiman K .Chat Room Using HTML, PHP, CSS, JS, AJAX[J]. 2021.DOI:10.6084/M9.FIGSHARE.14869167.[1] 塔库尔 A , 迪曼 K .使用 HTML、PHP、CSS、JS、AJAX 的聊天室[j].2021.DOI: 10.6084/M9.FIGSHARE.14869167.

致谢

欲买桂花同载酒，终不似，少年游。

时光荏苒，四年如歌。在这段奔赴梦想、书写青春的旅程中，我从懵懂走向成熟，从迷茫走向坚定。回首论文完成的日日夜夜，既有代码调试的艰辛，也有灵感乍现的欣喜。而这一切，离不开许多人给予我的帮助与陪伴。

首先，我要衷心感谢我的导师樊小毛老师。在整个课题研究过程中，您不仅引领我进入了科研道路，更在学术上给予我严谨的指导与悉心的教诲。每一次的细致批注、每一次的耐心答疑，都是我不断前行的动力源泉。您的敬业精神和治学态度，将是我今后学习与工作中不变的楷模。

其次，感谢我的家人，是你们始终在背后给予我最坚实的支持与最无条件的爱。每一次远方的牵挂，每一声温暖的问候，都是我静心沉淀、安心研究的力量所在。

最后，感谢自己，感谢那个在图像框里寻找意义、在函数曲线上追求真理的自己。愿今后的路上，仍能秉持热爱，勇敢前行，不负韶华，不负理想。

谨以此文，献给所有在这段旅途中给予我帮助与温暖的人。