

基于YOLOv5的手术器械末端识别研究

林元玺^{1*}, 颜 敏²

¹上海理工大学健康科学与工程学院, 上海

²成都中医药大学眼科学院, 四川 成都

收稿日期: 2025年5月11日; 录用日期: 2025年6月3日; 发布日期: 2025年6月11日

摘 要

目的: 在手术过程中, 医生需要实时掌握手术器械的位置。传统机器学习方法与深度学习模型受遮挡等因素影响, 其场景适应性和漏检率仍有优化的空间。本研究提出一种基于YOLOv5的手术器械末端识别方法。方法: 通过光学相机采集不同背景、光照条件下的手术器械图像, 利用Roboflow平台标注器械末端, 构建符合条件的数据集进行训练。其次, 将数据集按7:2:1比例划分训练集、验证集和测试集, 采用精确度、召回率、mAP等指标对比识别器械整体与末端的效果。结果: 在器械被遮挡的场景下, 末端识别置信度较高于整体识别, 目标检测的mAP@0.5为92.7%。结论: 在手持器械导致的器械被遮挡情况下, 通过识别器械末端辅助医生找到手术器械的位置, 为解决手术环境中的遮挡问题提供了初步探索 and 思路。

关键词

手术器械末端, YOLOv5, 动态遮挡, 目标识别, 深度学习

Research on the Identification of Surgical Instrument Ends Based on YOLOv5

Yuanxi Lin^{1*}, Min Yan²

¹School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai

²School of Ophthalmology, Chengdu University of Traditional Chinese Medicine, Chengdu Sichuan

Received: May 11th, 2025; accepted: Jun. 3rd, 2025; published: Jun. 11th, 2025

Abstract

Objective: During surgery, doctors need to know the position of surgical instruments in real time. Traditional machine learning methods and deep learning models are affected by factors such as occlusion, and their scene adaptability and missed detection rate still need to be optimized. This study proposes

*通讯作者。

文章引用: 林元玺, 颜敏. 基于 YOLOv5 的手术器械末端识别研究[J]. 建模与仿真, 2025, 14(6): 67-74.

DOI: 10.12677/mos.2025.146476

a method for identifying the end of surgical instruments based on YOLOv5. Methods: An optical camera was used to collect images of surgical instruments under different background and illumination conditions. The Roboflow platform was used to mark the end of the surgical instruments to construct a qualified dataset for training. Secondly, the data set was divided into training set, validation set and test set according to the ratio of 7:2:1, and the accuracy, recall rate, mAP and other indicators were used to compare the effect of identifying the whole and the end of the device. **Results:** In the scene where the device was occluded, the confidence of end recognition was higher than that of overall recognition, and the mAP@0.5 of object detection was 92.7%. **Conclusions:** In the case of instrument occlusion caused by hand-held instruments, it assists the doctor to find the position of the surgical instrument by identifying the end of the instrument, which provides a preliminary exploration and idea for solving the occlusion problem in the surgical environment.

Keywords

Surgical Instrument End, YOLOv5, Dynamic Occlusion, Object Recognition, Deep Learning

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来, 计算机视觉作为智能医疗的关键技术, 在手术器械精检测领域受到越来越多的关注。该技术通过对手术录像的实时分析, 实现手术器械的高精度识别和定位, 在术后器械清点、手术导航、多器械协同控制等环节具有非常重要的意义。

目前, 根据数据分析方式, 计算机辅助识别手术器械方法主要分为两种类型: 一种是基于传统机器学习的检测方法, 其依赖于人工设计的特征提取算法, 如阈值分割、边缘处理、形态学操作等; 另一种是基于深度学习的检测方法, 以卷积神经网络(Convolutional Neural Networks, CNN)为核心, 通过端到端训练机制自动提取图像特征, 在提升检测精度的同时大幅降低对人工特征工程的依赖。典型代表包括YOLO系列算法(如YOLOv7x、YOLOv8)[1], 通过改进网络结构提升检测精度。这两种方法在识别手术器械的过程中, 都具有以下缺陷: 动态场景适应性差, 即, 当术中器械处于复杂运动或遮挡的情况下, 容易受到干扰, 影响识别结果。

为了解决上述问题, 有学者提出一种基于运动矢量追踪的手术器械定位方法, 该方案利用手术器械上标记物形成的不同几何形状, 通过计算标记物间的边长比进行注册识别; 再根据前后帧标记物的运动矢量, 实现对手术器械的实时跟踪[2], 与传统机器学习的检测方法相似, 其缺陷仍然是手术器械被遮挡, 便无法识别, 并且当手术器械发生较大角度旋转或快速复杂运动时, 标记物形成的几何形状会发生变化, 可能超出预设的边长比范围, 导致识别失效。汪睿等人提出的机器视觉器械清点分类方法, 通过形态学骨架提取与像素滑动检索技术, 确定器械数量与位置, 并预先建立搜索框作为先验信息, 结合模板匹配实现分类[3]。该方法仍存有局限性: 当器械种类或外观改变时, 需重新调整特征提取算法并更新模板库, 导致模型泛化性差。此外, 孟晓亮等人提出基于改进YOLOv5s的手术器械检测方法, 通过Gamma校正算法校正图像亮度和对比度, 解决手术器械反光和阴影遮挡问题, 然后在YOLOv5s基础上, 添加CBAM和动态卷积模块, 增加重要特征权重, 优化空间金字塔池化模块以扩大感受野, 提高目标检测准确度、减少漏检率[4]。由于术中器械经常被手部遮挡, 与手术场景中的光影干扰叠加, 加剧模型对目标特征的提取难度, 导致器械跟踪连续性下降。

以上方案主要聚焦于手术器械整体特征的检测,而近年来陆续有很多学者通过识别器械局部特征,解决术中器械被遮挡的难题。宋霜等人提出一种基于单目内窥镜影像的手术器械追踪方法,通过 HSV 颜色空间变换、阈值识别和器械分割等技术,确定器械末端位置点,并结合标尺变换和角点检测技术,实现器械的实时追踪[5]。然而,HSV 颜色分割易受到光照强度的影响,影响末端位置点的计算,无法实现连续跟踪。Ullah 等人研发一种导丝尖端追踪系统,该系统采用双阶段网络架构,通过网络定位检测目标区域,并在连续帧的局部图像块中,运用分割网络提取时空特征[6]。该方法虽实现了 0.36 mm 的定位精度,但对分割网络依赖性强,当器械发生非刚性形变时,需频繁调用检测网络进行重新定位,难以满足实时性要求。

不同于以上的识别、跟踪原理,本研究提出了通过 YOLOv5 算法识别手术器械末端的方法,构建包含复杂遮挡场景的手术器械数据集,增强模型对部分遮挡特征的鲁棒性。实验表明,该方法在器械遮挡的场景下,末端识别置信度比识别手术器械整体较高,初步解决了传统方法和现有深度学习模型在遮挡环境下的漏检难题,为手术中动态器械的精准识别提供了思路和技术支持。

2. 方法

2.1. 基于 YOLOv5 的手术器械末端识别算法流程

本文提出的手术器械末端识别算法是在 YOLOv5 架构的基础上优化实现的,具体流程包含数据采集、特征增强、模型训练和实时推理四个阶段。首先,通过光学相机采集不同背景和光照条件下的手术器械图片,利用在线平台 Roboflow 对图像中的器械末端和手术器械整体进行标记,建立数据集并进行训练。在训练过程中,损失值收敛到一定程度时,结束训练。最后,用连接计算机的相机实时拍摄台面上的器械,通过训练好的网络模型对其进行识别,观察是否能达到理想的识别效果。

2.2. YOLOv5 算法介绍

YOLOv5 算法由 Ultralytics 公司于 2020 年提出[7],作为 YOLO (You Only Look Once)系列的最新成员,其在 YOLOv4 的基础上进行了多项优化,包括引入 Mosaic 数据增强、自适应锚框计算、Focus 模块和 CSP (Cross Stage Partial)结构等。该算法以单目标检测框架为核心,通过网格划分和锚框机制实现端到端的实时检测,其损失函数采用 GIoU (Generalized Intersection over Union)优化定位精度,并利用非极大值抑制(Non-Maximum Suppression, NMS)筛选预测框[8]。发展至今,YOLOv5 已迭代至 6.1 版本,衍生出 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 等多个子版本,通过调整网络深度和宽度(depth_multiple 与 width_multiple 参数)平衡速度与精度,其中 YOLOv5s 以轻量化著称,适合移动端部署,而 YOLOv5x 则在高精度场景中表现优异。

近年来,已有很多研究者将 YOLOv5 算法成功应用于手术器械检测与分割当中。例如,朱俊玲等人 YOLOv5 模型的基础上,采用 Soft-NMS 算法改进模型,提升手术器械检测精度[9];Jiang 等人 YOLOv5 模型的基础上,提出通过添加损失函数和注意力模块,使 mAP 提升至 88.7%,验证了其在手术器械自动化清点中的可行性[10]。此外,王炎等人提出一种基于适用于手术器械检测的深度学习算法,通过采用 BiFPN 结以及最新的 α -IoU Loss 函数,识别抓取钳、分离钳的平均精度分别达 98.7%和 99.5%,进一步佐证了该技术在医疗场景的应用潜力[11]。

本文 YOLOv5 的算法结构分为四部分:输入层(Input)、主干网络(Backbone)、颈部网络(Neck)和输出端(Head)。Input 模块采用 Mosaic 数据增强方式和自适应调节锚框提升训练整体效率;Backbone 模块基于 CSPDarknet 结合 Focus 模块实现快速提取特征信息;Neck 模块融合特征金字塔(feature pyramid network, FPN)和路径聚合网络(path aggregation network, PAN)结构,增强特征融合能力;Head 部分的功能为输出

定位和分类结果[12]。其原理核心在于将图片划分为不同的网格，每个网格预测边界框和类别概率，通过 CIOU 损失函数提高真实框和预测框的匹配程度。

3. 实验结果与分析

3.1. 实验环境

本研究的显卡为 NVIDIA GEFORCE GTX 4070，处理器为 12th Gen Intel Core i7-12800HX 2.00 GHz，运行内存 32 GB，操作系统为 Windows 10，64 位。整个实验在 Pytorch 2.2.2，Python 3.10，win11 环境下进行。

3.2. 制作数据集

在本次实验中，由于当前网上缺乏公开手术器械末端识别专用数据集，我们便自主拍摄构建了多模态手术器械图像数据库。我们使用光学相机垂直器械平面 90°拍摄手术器械视频，利用编程软件 Visual Studio Code 编写关键帧提取算法，按照每 5 帧为一个单位对拍摄视频截取关键帧，生成原始图片 263 张，并将其中手术器械较为模糊的图像剔除。本文采用在线平台 Roboflow 对图像中的器械末端和手术器械整体进行标注，通过平台自带的特征增强模块，增加对比度和图像去噪，制作满足训练条件的样本图片。通过在不同背景和光照条件下拍摄图片，来增强算法的泛化能力，使分割结果更加精确，并且为了防止样本过于单一，本实验采取了翻转、旋转等方法对数据集进行了扩充，最后数据集达 450 张。

3.3. 评测指标

在目标检测领域，通常使用以下核心指标量化算法性能，包括精确度(precision, P)、召回率(recall, R)、F1 得分值(F1 score)、平均精度(average precision, AP)和平均精度均值(mean average precision, mAP)。

精度又称查准率，计算公式如下：

$$P = \frac{TP}{TP + FP} \times 100 \quad (1)$$

式中，TP 为真正例数目，即检测框与真实目标的交并比(intersection over union, IoU)大于某个阈值的例数；FP 为假正例数目，即检测框的置信度高于阈值但与任何真实目标的交并比都低于阈值的例数。

召回率又称查全率，计算公式如下：

$$R = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

式中，FN 为假负例数目，即未被任何检测框覆盖的真实目标数目。

理想状态下，精确度和召回率均达到最高值，分割效果最好，但是提高精确度往往导致召回率下降，反之亦然。所以为了找到二者的最佳组合，我们引入 F1 值作为综合评估指标，其本质是精确度与召回率的调和平均数，计算方法如下：

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (3)$$

AP 是精度 - 召回率曲线下的面积，所有类别的 AP 取平均值即为 mAP，mAP 被用来评估模型在不同 IoU 阈值下检测多个类别目标时的检测精度。通常情况下该 IoU 阈值设定为 0.5，对应的指标为 AP50 和 mAP50。

3.4. 制作数据集

本研究按照 7:2:1 的比例将数据集随机划分为训练集(315 张)，验证集(135 张)和测试集(45 张)。初始

学习率设置为 0.01, 动态系数设置为 0.937, 权重衰减为 0.0005, Batch Size 设置为 16, 训练过程持续 300 轮, 对比基于 YOLOv5 识别手术器械整体和末端的识别效果。

3.5. 结果与分析

使用 YOLOv5 模型在数据集上进行实验。实验得到的 P-R (precision-recall) 曲线如图 1 所示, 其横轴是召回率, 纵轴是精确率, 各曲线在置信度较低时 F1 值较高, 随着置信度增加, F1 值逐渐下降, 表明模型在中等置信度下能较好平衡精确率与召回率。基于 YOLOv5 的识别手术器械整体和末端的 F1 曲线如图 2 所示, 识别器械末端(instrument end)的 mAP@0.5 为 0.917, 识别手术器械(surgical instrument)为 0.936, 显示模型对两类目标检测具有较高精确率与召回率。基于 YOLOv5 的各类别手术器械的混淆矩阵图如图 3 所示, 识别手术器械整体的准确率会比识别器械末端的准确率高。

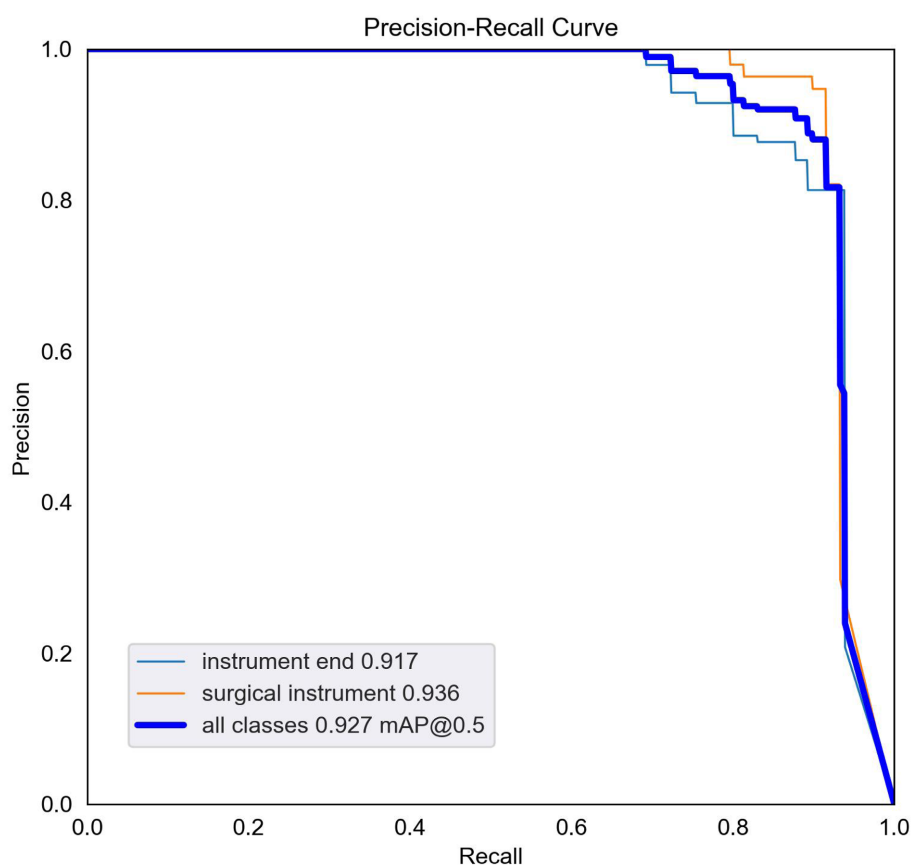


Figure 1. YOLOv5-based P-R curve for identifying the whole surgical instrument and the end of the instrument
图 1. 基于 YOLOv5 的识别手术器械整体和器械末端的 P-R 曲线图

为了比较术中识别手术器械整体和器械末端的效果, 随机拍摄 4 张图片在模型上进行测试, 其中两张为器械平放在桌面上的照片, 另外两张为手持器械的照片, 对比结果如图 4 所示。从图 4 可以看出, 在器械没有遮挡时, 由于器械末端较于手术器械整体结构复杂、尺寸偏小, 导致识别器械整体会比器械末端较容易一些; 在器械存在部分遮挡时, 手术器械整体特征受到干扰, 导致置信度下降, 而手术器械末端特征, 受光照、角度等因素的影响较小, 器械整体受到部分遮挡, 只要末端特征可见, 模型仍能有效识别从而提高了检测的准确性。

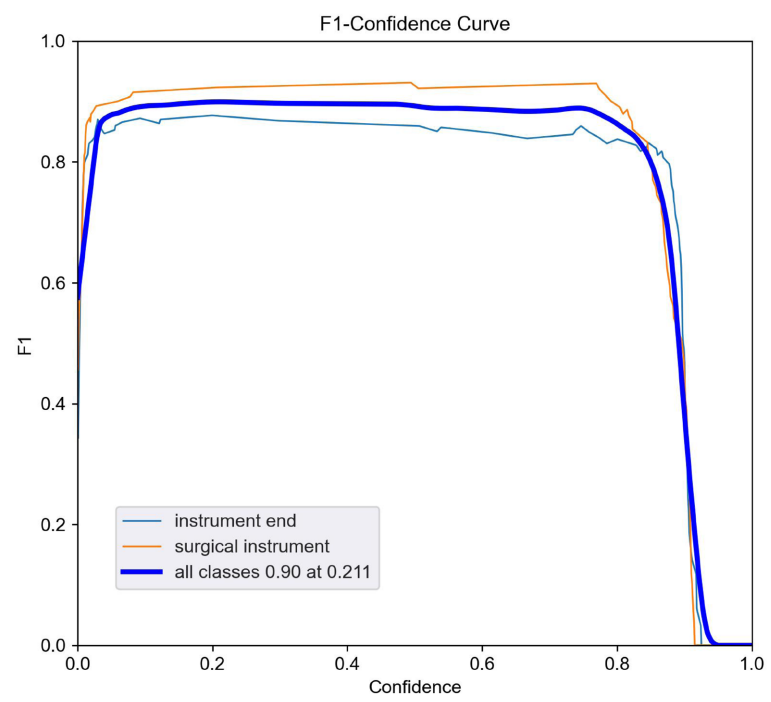


Figure 2. F1 curve of YOLOv5-based identification of the whole surgical instrument and the end of the instrument
图 2. 基于 YOLOv5 的识别手术器械整体和器械末端的 F1 曲线图

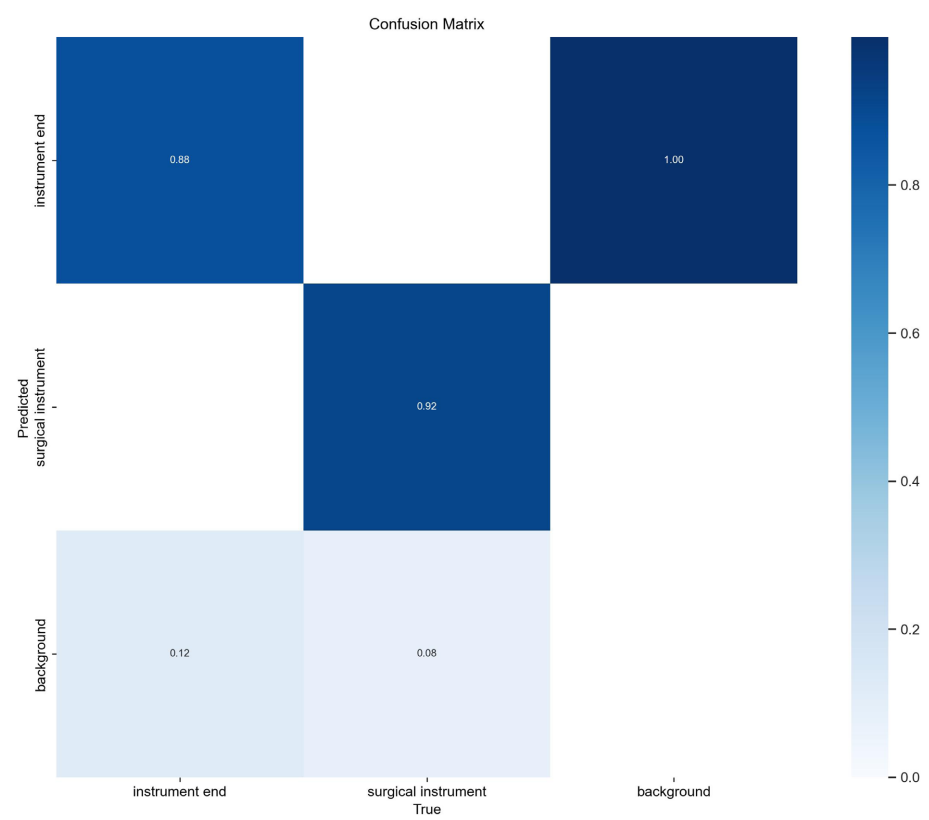


Figure 3. Confusion matrix diagram based on YOLOv5 for identifying surgical instruments as a whole and at the end of the instrument
图 3. 基于 YOLOv5 的识别手术器械整体和器械末端的混淆矩阵图



Figure 4. Comparison of the recognition results of flat (left) and handheld (right) surgical instruments
图 4. 平放手术器械(左)和手持手术器械(右)的识别结果对比图

本文 YOLOv5 算法模型通过以下两个模块来提升识别性能，图像去噪、增加对比度，为了验证在线平台 Roboflow 各个数据增强模块的有效性，在数据集上进行消融实验，结果如表 1 所示，加粗数据表示最优结果。

Table 1. Ablation experiments in different modules on the dataset
表 1. 数据集上不同模块的消融实验

模块		精确度	召回率	F1 值	mAP@0.5
图像去噪	增加对比度				
√		0.953	0.921	0.937	0.9
		0.957	0.928	0.942	0.910
	√	0.954	0.928	0.941	0.905
√	√	0.961	0.935	0.948	0.927

注：加粗数据表示最优结果。

上述结果显示，模型不启用数据增强模块，其精准率、召回率、F1 值和 mAP@0.5 分别为 95.3%、92.1%、93.7%、90%。而将两个模块均加入模型则精准率、召回率、F1 值和 mAP@0.5 相较于原模型分别提升 0.8%、1.5%、1.2%和 3%。表明模型中加入图像去噪可以减少噪声干扰、增强特征识别能力，加入增加对比度可以帮助模型更好地检测边缘，两者结合可最大化提升目标检测性能。该模型在手术器械末端识别任务中展现出良好的性能，可检测实时性与可靠性要求较高的器械，验证了其高效的学习能力。

4. 结论

4.1. 研究总结

本文提出了一种基于 YOLOv5 的手术器械末端识别方法。通过构建包含复杂遮挡场景的多模态数据集,并结合 Roboflow 平台的数据增强策略,模型在遮挡环境下展现出显著优势。实验结果表明,器械末端识别的 mAP@0.5 达到 92.7%, 精准率与召回率分别达 96.1%和 93.5%, 尤其在手持器械的动态遮挡场景中,末端识别置信度较整体识别提升约 10%, 验证了局部关键特征学习对干扰的鲁棒性,为传统方法与现有深度学习模型在遮挡环境中的漏检难题,提供了初步探索和思路,也为智能医疗领域提供了技术支持。

4.2. 未来方向

在后续的研究中,可在算法优化方面结合 Transformer 架构或 RepLKNet 大卷积核,增强模型对长距离依赖关系的建模能力,进一步提升复杂遮挡下的检测精度;扩充和增强数据集以提升泛化能力;通过与医疗机构合作将其集成到实际设备和系统进行应用;还可考虑融合多模态信息,推动该领域的持续发展与完善。

参考文献

- [1] Frey, S., Facente, F., Wei, W., Ekmekci, E.S., Séjor, E., Baqué, P., *et al.* (2025) Optimizing Intraoperative AI: Evaluation of Yolov8 for Real-Time Recognition of Robotic and Laparoscopic Instruments. *Journal of Robotic Surgery*, **19**, Article No. 131. <https://doi.org/10.1007/s11701-025-02284-7>
- [2] 张梦诗, 贾博奇, 梁楠, 等. 基于运动矢量追踪的多手术器械光学定位算法[J]. 北京生物医学工程, 2018, 37(4): 345-350.
- [3] 汪睿, 苗玉彬. 基于改进模板匹配的外科手术器械清点方法[J]. 机电一体化, 2022, 28(Z2): 51-57.
- [4] 孟晓亮, 赵吉康, 王晓雨, 等. 基于改进 YOLOv5s 的手术器械检测与分割方法[J]. 液晶与显示, 2023, 38(12): 1698-1706.
- [5] 哈尔滨工业大学(深圳). 胃部消化道下手术器械追踪和实时预警的方法和系统[P]. 中国, 202110988923.2. 2023-08-01.
- [6] Ullah, I., Chikontwe, P. and Park, S.H. (2019) Real-Time Tracking of Guidewire Robot Tips Using Deep Convolutional Neural Networks on Successive Localized Frames. *IEEE Access*, **7**, 159743-159753. <https://doi.org/10.1109/access.2019.2950263>
- [7] 孙歆, 王晓燕, 刘静, 等. 经典 YOLO 系列目标检测算法及其在乳腺癌检测中的应用[J]. 计算机系统应用, 2023, 32(12): 52-62.
- [8] Zhou, Y. (2023) IYOLO-NL: An Improved You Only Look Once and None Left Object Detector for Real-Time Face Mask Detection. *Heliyon*, **9**, e19064. <https://doi.org/10.1016/j.heliyon.2023.e19064>
- [9] 朱俊玲, 瞿芳, 钱贝, 等. 基于改进 YOLOv5 的外科手术器械识别系统的开发[J]. 护理研究, 2024, 38(21): 3923-3928.
- [10] Jiang, K., Pan, S., Yang, L., Yu, J., Lin, Y. and Wang, H. (2023) Surgical Instrument Recognition Based on Improved YOLOv5. *Applied Sciences*, **13**, Article No. 11709. <https://doi.org/10.3390/app132111709>
- [11] 王炎. 基于深度学习的手术器械视觉检测与跟踪技术研究[D]: [硕士学位论文]. 天津: 天津理工大学, 2022.
- [12] 曲英伟, 刘锐. 基于 YOLOv5-MobileNetV3 算法的目标检测[J]. 计算机系统应用, 2024, 33(7): 213-221.