1. 研究进展总述

自开题报告确立研究方向以来，本研究围绕"双光路采血管分层检测系统"的核心目标，在专业采血管双光路医疗数据集搭建与双光路分割算法创新方面取得了显著进展。项目聚焦于医院检验科在庞大医疗数据与高通量样本周转压力下暴露的痛点，即传统人工目检血清、血浆与凝胶样本既费时又易出错，难以满足现代检验流水线的效率与精度要求，而现有的基于单光路输入的计算机视觉分割方法，在该场景下难以实现对样本分层的精确定位，特别是在单光图白膜层区域出现反光、模糊、白膜层厚度过薄的特殊样本情况下。为此，本研究将双光路输入的神经网络模型融入采血管预处理环节，以可靠识别分层边界和样本类型，为后续自动移液、检测等环节提供结构化数据支撑。

在技术路径上，本研究重点突破了双光路模态特征融合的关键技术瓶颈。通过构建白光与蓝光双通道成像系统，充分利用白光图像在边界细节上的清晰度优势和蓝光图像在血液透光性方面的特性，设计了基于跨注意力机制的YOLO分割等架构。项目当前已完成从原始数据采集、标注处理、数据集构建到模型训练、评估验证以及模型流水线部署的完整技术链条，建立了包含502张高分辨率双模态图像的标准数据集，实现了多种特征融合策略的对比验证。在最优配置下，双模态跨模态注意力模型在血清/血浆层检测上达到97.50%的检测率，IoU指标达到0.79，显著超越单模态基线方法，为血液检测流水线的自动化改造奠定了坚实的算法基础。

2. 数据集构建与处理策略

数据集构建是本研究的重要基础工作，涉及多阶段的数据采集、预处理与增强流程。项目采用白光/蓝光双通道成像设备，获取采血管在不同光照条件下的高分辨率图像对，确保双模态数据在时间与空间上的严格对齐。

2.1 原始数据采集与标注处理

数据采集阶段建立了完整的双模态成像协议。通过`scripts/annotate\_and\_crop.py`脚本实现JSON标注文件的解析与可视化标注，支持蓝光图像（T5通道）与白光图像（T3通道）的自动配对。该脚本基于文件名模式匹配机制，实现了"2022-04-15\_084806\_41\_T5\_2438.json"格式标注文件对应图像的精确定位，并通过裁剪区域标准化（crop\_left=800, crop\_top=250, crop\_width=1216, crop\_height=1504）确保所有样本在相同的感兴趣区域内进行分析。标注处理流程采用多边形轮廓标注方案，针对血液分层的复杂几何特性，完成了对各个数据样本中对象的准确轮廓标注。

2.2 双模态数据集创建流程

双模态数据集的创建通过`dual\_dataset/d\_dataset\_creation.py`中的`DualModalDatasetCreator`类实现。该创建器负责将Roboflow预处理的JPG格式数据与原始BMP裁剪数据进行匹配转换。关键技术包括文件名提取算法，通过正则表达式解析格式，提取关键ID作为匹配前缀，确保蓝光与白光图像的准确对应；数据集构建过程实现了BMP到JPG的高质量转换，保持图像细节的完整性；标签拷贝与重命名机制确保YOLO格式标注文件与图像文件的一一对应关系。通过`d\_dataset\_config.py`配置管理器，支持多版本数据集的灵活切换，当前主要使用"Dual-Modal-1504-500-1"版本，包含训练、验证、测试三个标准分割。

Une image contenant bouteille, mur, intérieur, Bleu cobalt

Le contenu généré par l’IA peut être incorrect.

2.3 数据增强与质量控制

数据增强策略通过`dual\_dataset/d\_dataset\_augmentation.py`实现，设计了九种增强变体以提升模型的泛化能力。增强策略包括：原始图像（策略0），二元组合增强（策略1-4）结合5度/-5度旋转与模糊、曝光调整，三元组合增强（策略5-8）整合10度/-10度旋转、亮度调整与多重变换。每种策略确保双模态图像的同步变换，维持空间对应关系的一致性。

旋转增强采用OpenCV的`getRotationMatrix2D`实现像素级精确变换，标注坐标通过齐次坐标系统进行同步旋转，确保几何一致性。亮度与曝光调整采用HSV颜色空间操作与Gamma校正，模拟不同成像条件下的光照变化。模糊增强使用高斯滤波模拟成像系统的光学特性变化。

通过6通道数据预处理（`d\_dataset\_concat\_6ch.py`），将双模态3通道图像拼接为单一6通道张量，显著提升训练效率。最终构建的数据集包含基础版本502张图像，增强后版本达到4518张图像，为模型训练提供了充足的样本多样性。质量控制通过可视化验证（`d\_dataset\_visulize.py`）与后处理检查（`d\_dataset\_postprocess.py`）确保数据集的标注准确性与格式一致性。

几个不同颜色的手机

AI 生成的内容可能不正确。

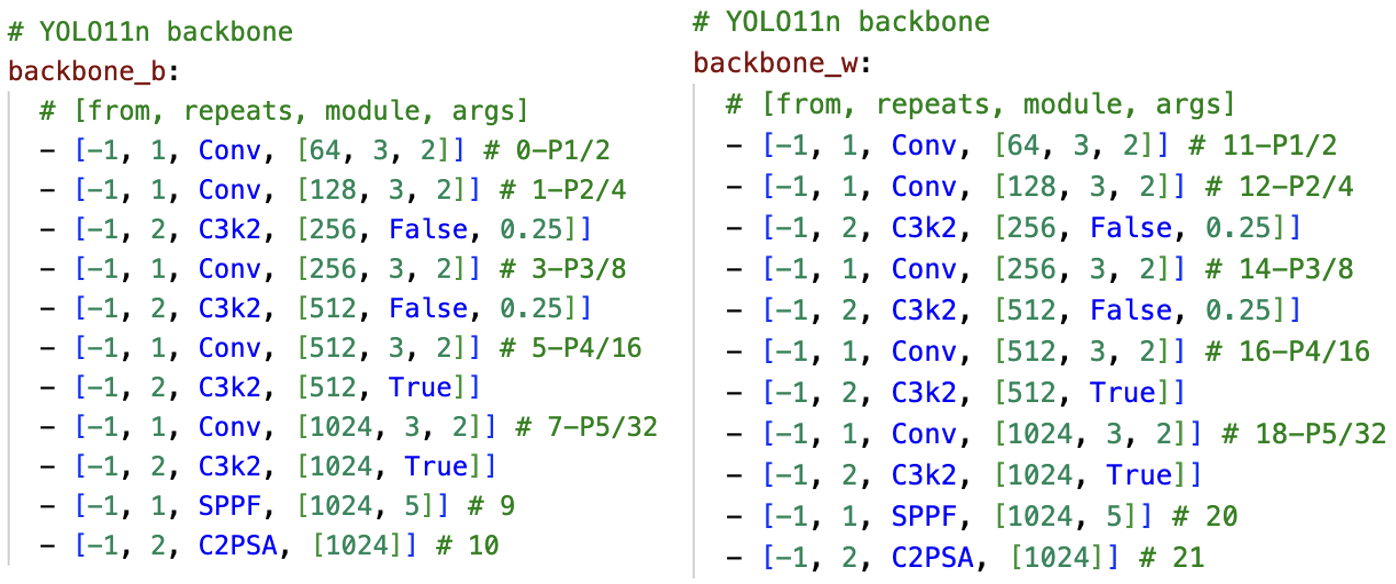
3. 双模态模型架构设计

双模态YOLO架构设计是本研究的核心技术创新，通过跨模态特征融合机制实现了白光与蓝光图像信息的深度整合。架构采用双骨干网络设计，分别处理蓝光与白光输入，并在多个尺度上实现特征融合。

3.1 双骨干网络结构

模型架构基于YOLO11-X基础框架进行双模态扩展。蓝光骨干网络（backbone\_b，层0-10）与白光骨干网络（backbone\_w，层11-21）采用完全对称的设计，每个骨干网络包含Conv下采样、C3k2特征提取、SPPF空间金字塔池化与C2PSA位置感知注意力模块。这种对称设计确保了两个模态在特征抽取能力上的均衡性，避免单一模态主导的问题。

根据`dual\_yolo/models/yolo11x-dseg-crossattn-precise.yaml`配置，模型支持6通道输入（ch\_b=3，ch\_w=3），在P3/8、P4/16、P5/32三个关键尺度进行特征融合。backbone\_b处理蓝光通道特征，经过逐步下采样获得不同分辨率的特征表示；backbone\_w处理白光通道特征，遵循相同的特征提取流程。双骨干网络的参数量达到YOLO11-X规模（depth=1.00，width=1.50），确保充足的模型容量捕捉复杂的血液分层特征。



3.2 跨模态特征融合机制

跨模态特征融合通过`ultralytics/nn/modules/fusion.py`中的多种融合算子实现。项目设计了四种核心融合策略：拼接压缩融合（ConcatCompress）、加权融合（WeightedFusion）、跨模态注意力融合（CrossModalAttention）以及多层跨模态注意力融合（MultiLayerCrossModalAttention）。



`CrossModalAttention`模块是本研究的关键实验所在，实现了基于Token的局部注意力机制。该模块将蓝光特征作为Query查询白光特征中的互补信息，通过"用注意力加权的白光信息增强蓝光特征"实现跨模态特征增强。具体实现中，特征图被分割为token\_size×token\_size的Token单元，每个Token在neighbor\_size×neighbor\_size的局部邻域内进行注意力计算，有效降低了计算复杂度（相比全局注意力减少91%）。

注意力计算采用三步流程：特征投影生成Query、Key、Value表示；局部邻域提取通过unfold操作实现高效的邻域特征聚合；加权求和通过einsum操作实现注意力权重与特征值的精确融合。残差连接设计确保原始蓝光特征与跨模态增强特征的有效组合，避免信息丢失。

3.3 模型配置与优化策略

针对血液分层检测的精度要求，项目设计了`crossattn-precise`精确配置方案。该配置在P3、P4、P5三个融合点采用不同的token\_size与neighbor\_size参数：P3融合点使用token\_size=2与neighbor\_size=3，实现更精细的局部特征对齐；P4融合点使用token\_size=3与neighbor\_size=5，平衡计算效率与感受野覆盖；P5融合点使用token\_size=4与neighbor\_size=7，获取更大的全局上下文信息。

权重初始化采用预训练策略，通过`d\_weights\_transfer.py`实现单模态权重向双模态模型的迁移。该迁移过程保持骨干网络的预训练特征提取能力，仅对融合模块进行随机初始化，加速模型收敛并提升最终性能。多层注意力扩展（MultiLayerCrossModalAttention）通过渐进式特征精化进一步提升融合效果，每层注意力都能获得来自白光的新信息，实现层次化的模态关系建模。

4. 训练流程与实验结果

模型训练采用系统化的配置管理与多策略对比验证方法，通过`dual\_yolo/d\_model\_train.py`实现了从零开始训练、预训练权重加载、骨干网络冻结等多种训练模式的支持。

4.1 训练配置与实验设置

训练配置采用高分辨率输入（imgsz=1504）以保证血液分层边界的精确检测，批次大小设置为4以适应GPU内存限制，训练轮数为30轮确保充分收敛。关键训练参数包括：关闭混合精度训练（amp=False）以保证跨模态注意力计算的数值稳定性；多GPU训练支持（device=[0,1,2,3]）提升训练效率；数据加载并行度设置为4个工作进程。

实验采用6通道预处理数据集（`datasets/Dual-Modal-1504-500-1-6ch`）实现六通道图片数据载入的训练过程。训练模式支持三种策略：从零开始训练（scratch）用于基线对比；预训练模式（pretrained）加载`dual\_yolo11x\_bw.pt`权重进行微调；冻结骨干网络模式（freeze\_backbone）仅训练融合模块与检测头，适用于快速验证融合策略效果。

融合策略对比实验涵盖五种核心方法：concat-compress拼接压缩融合、weighted-fusion加权融合、crossattn标准跨模态注意力、crossattn-precise精确跨模态注意力，以及单模态基线（id-blue，id-white）。每种策略采用相同的训练配置确保公平对比。

4.2 评估方法与指标体系

评估系统通过`dual\_yolo/d\_model\_evaluate\_v2.py`实现，采用基于YOLO训练标签的精度评估方法，分组计算血清/血浆层（Class 0，2）与白膜层（Class 1）的检测效果。评估方法采用"精确一次检测"医学标准，要求每个血液分层类别必须被检测到恰好一次，不允许漏检或多检现象。

关键评估指标包括：检测率（Detection Rate）衡量成功检测样本占总样本的比例；IoU指标评估分割精度，通过预测掩码与真实掩码的交并比计算；表面差异指标通过上下表面位置偏差（像素级）量化检测精度。评估过程支持多置信度阈值（0.5-0.8）的对比分析，确保结果的鲁棒性验证。

旋转增强样本的评估采用坐标系转换策略，将预测结果与真实标注都转换到原始坐标系进行表面差异计算，确保评估结果的准确性。可视化功能支持成功检测与失败检测样本的分别保存，便于错误案例分析与模型改进。

4.3 实验结果与性能分析

根据`dual\_yolo/evaluation\_results\_v2`中的实验结果，各融合策略在不同置信度阈值下表现如下：

在置信度0.65条件下，**Dual Yolo CrossAttn** 在血清/血浆层检测达到最佳性能：检测率100.00%，IoU均值0.98±0.01，上表面差异4.7±4.5像素，下表面差异3.6±2.8像素。该配置在白膜层检测达到97.50%检测率，IoU均值0.76±0.07。Yolo11-Blue单模态基线在血清/血浆层达到99.72%检测率，IoU均值0.98±0.01， Yolo11-White达到100.00%检测率但IoU略低（0.97±0.01）。

在置信度0.7条件下， Dual Yolo CrossAttn 保持血清/血浆层100.00%检测率优势，白膜层检测率为94.72%。Dual Yolo Weighted策略表现稳定，血清/血浆层检测率98.89%，表面差异控制在较低水平（上表面4.0±2.9像素，下表面3.5±2.6像素）。

Dual Yolo CrossAttn 精确配置在置信度0.65下血清/血浆层检测率91.67%，IoU达到0.97±0.01，但白膜层检测率较低（0.56%）。这表明精确配置在特定类别上具有优势，但需要进一步优化以平衡多类别检测性能。

对比分析显示，跨模态注意力机制（crossattn系列）相比简单融合策略（concat-compress，weighted-fusion）在检测准确性上具有显著优势，验证了注意力机制在跨模态特征融合中的有效性。30轮训练相比默认训练轮数显著提升了模型性能，说明跨模态注意力机制需要充分的训练时间以学习复杂的模态关系。

5. 推理部署与工具支撑

为支撑实际应用场景的部署需求，项目开发了完整的推理服务框架与远程验证工具链，实现了从模型加载、推理计算到结果返回的全流程自动化支持。

5.1 Flask推理服务架构

推理服务通过`scripts/yolo\_service.py`实现，基于Flask框架构建RESTful API接口。服务架构采用模型预加载策略，在服务启动时一次性加载YOLO模型权重，避免每次推理请求的重复加载开销。推理端点`/infer`接受JSON格式请求，包含输入图像路径与输出图像路径参数，支持灵活的文件路径配置。

推理配置针对血液检测任务进行优化：置信度阈值设置为0.6，在检测精度与召回率间取得平衡；图像尺寸配置为[1504, 1504]，适配实际采血管成像分辨率；GPU设备指定为"cuda:0"，确保推理速度；可视化选项支持检测结果的图像保存，便于结果验证与问题排查。

服务支持多种输出格式控制，包括边界框显示、置信度显示、标签显示等可选功能，满足不同应用场景的需求。错误处理机制提供详细的异常信息返回，支持输入文件存在性检查、推理过程异常捕获与结果保存状态验证。

5.2 部署配置与性能优化

推理服务配置支持多GPU环境的灵活部署，通过设备参数调整适配不同硬件环境。服务监听配置为"0.0.0.0:5000"，支持远程访问与负载均衡部署。模型加载采用异常处理机制，确保服务启动的稳定性。

性能优化策略包括：模型权重的一次性预加载减少推理延迟；批处理支持提升多图像推理的并行效率；GPU内存管理确保长时间运行的稳定性。推理结果的缓存机制支持重复请求的快速响应，适用于交互式应用场景。

6. 下一阶段计划

基于当前研究进展与实验结果分析，下一阶段研究将围绕数据扩容、模型优化、性能提升与论文撰写四个核心方向展开，确保项目在技术深度与应用广度上的进一步突破。

6.1 数据扩容与质量控制

数据集扩容计划将当前502张基础样本扩展至1000张以上，覆盖更多血液样本类型、成像条件与设备差异。扩容策略包括多设备数据采集，引入不同厂商采血管与成像设备的样本差异；多条件数据补充，涵盖不同存储时间、温度条件下的血液分层变化；边界案例收集，重点补充分层边界模糊、异常样本等挑战性案例。

质量控制体系将建立多级标注验证机制：一级标注由专业技术人员完成；二级交叉验证通过多人标注一致性检查确保标注质量。

数据增强策略优化将基于当前实验结果进行精细化调整，重点优化旋转角度分布、光照变化范围与模糊程度设置。新增增强策略包括弹性变形模拟成像设备差异、颜色域转换模拟光源色温变化、噪声注入模拟实际成像环境等。

6.2 跨模态消融实验与模型优化

跨模态注意力机制的深度消融实验将通过`dual\_yolo/d\_model\_train.py`扩展功能实现。计划开展的消融实验包括：token\_size参数对比（2×2、3×3、4×4、6×6），分析不同粒度Token化对特征融合效果的影响；neighbor\_size范围测试（3×3、5×5、7×7、9×9），优化局部注意力的感受野配置；多层注意力深度实验（1层、2层、3层、4层），验证渐进式特征精化的最优层数。

注意力机制优化将探索更高效的计算方式：稀疏注意力通过重要性采样减少计算复杂度；自适应注意力根据特征相似度动态调整邻域大小；分组注意力在通道维度进行分组计算提升并行度。融合策略创新将设计混合融合机制，在不同网络层采用不同融合策略的组合方案。

预训练参数载入消融实验将系统分析不同预训练策略的效果：单模态预训练权重的最优组合方式；冻结策略对比（冻结backbone、冻结部分层、全参数微调）；学习率调度策略优化（余弦退火、分段衰减、自适应调整）。

6.4 论文撰写与成果输出

当前阶段的论文撰写工作将集中于 **IEEE Journal of Biomedical and Health Informatics (JBHI)** 的期刊投稿准备。文章将重点阐述跨模态注意力机制的技术创新与血液检测应用的临床价值。写作过程中将突出方法设计的科学性、实验结果的可靠性以及在医学信息学领域的潜在推广价值，力求形成具有完整性与影响力的学术成果。

技术专利申请将围绕跨模态注意力融合方法、双光路成像系统、血液分层自动检测算法等核心技术展开。软件著作权申请将涵盖完整的软件系统，包括数据处理工具、模型训练框架、推理部署系统等。

开源代码发布将在GitHub平台建立项目主页，提供完整的代码实现、预训练模型、数据集样例与使用文档。技术文档撰写将包括算法原理说明、实现细节介绍、部署指南与常见问题解答等。用户手册编写将面向实际用户提供系统安装、配置、使用的详细指导。

项目验收材料准备将按照学位论文要求，完善实验数据、技术文档、应用案例等支撑材料，确保研究成果的完整性与规范性。