

基于 Web 前端技术的 YOLO 实例分割标注系统设计及效率评估

李延胜¹ 张旭^{1*} 柏洪雪¹ 战晓林¹

LI Yansheng ZHANG Xu BAI Hongxue ZHAN Xiaolin

摘要

YOLO 系列模型的普及,使得高质量图像标注数据集的需求大幅增长。在利用人工智能算法开展水面漂浮物监测、地质影像相关研究的过程中,传统标注工具存在部署复杂、操作繁琐等问题,严重影响了小型团队的项目推进效率。基于此,文章设计实现 YOLO Vision Pro 智能视觉标注系统,该系统基于 HTML5 Canvas 与 JavaScript 开发,支持浏览器端免安装运行,通过极简交互逻辑优化标注操作。经 1 500 张图像并发测试验证,系统大幅降低学习门槛,且原生支持 YOLO 格式归一化 TXT 文件导出,为计算机视觉数据准备提供了轻量化、高效率的工程化解决方案。

关键词

图像标注工具;人机交互;标注效率;数据工程;Canvas 渲染

doi: 10.3969/j.issn.1672-9528.2026.07.029

0 引言

在深度学习技术驱动的计算机视觉研究与应用中,高质量的标注数据是保障模型性能的核心基础,当前市面上的实例分割标注工具主要分为两大类型,且均存在明显的应用短板。第一类是以 LabelMe^[1]为代表的本地端标注软件,这类工具功能较为全面,但需要本地配置 Python 运行环境,对用户的计算机操作基础有一定要求,且跨平台兼容性较差,难以在不同操作系统间无缝使用;第二类是以 CVAT 为代表的云端标注平台,虽支持多用户协同标注,但后端部署依赖 Docker 等容器技术,环境配置繁琐,且其交互逻辑设计偏向全功能化,工具栏与菜单层级复杂,对于追求标注速度的纯手动标注场景而言,过多的操作步骤大幅降低了标注效率。

在利用人工智能算法开展水面漂浮物监测^[2]、地质影像相关研究的过程中,发现传统标注工具存在部署复杂、操作繁琐等问题,严重拖累了小型团队的项目推进效率。针对上述痛点,团队研发了一款轻量化的 YOLO Vision Pro 智能视觉标注系统(以下简称“本系统”)^[3],核心设计目标是实现“零安装、低学习、极速标注”的应用效果。本文通过重构传统标注系统的人机交互流程,简化冗余的点击确认操作,

设计了基于鼠标与键盘的“盲操”标注模式,让用户无需频繁关注界面按钮,仅通过键鼠配合即可完成全流程标注;同时,系统的标注数据输出格式直接对接 YOLO 实例分割的 TXT 标注标准,从根源上省去了后续的格式转换工作,适配 YOLO 系列模型的直接训练需求。

1 系统架构与设计逻辑

1.1 技术栈选择

为实现“零安装、跨平台”的核心需求,本系统摒弃了传统的服务器端渲染架构,采用纯前端技术栈进行开发,各技术层的设计与功能如下:

(1) 渲染层:基于 HTML5 Canvas API 构建图像与标注层的分层渲染机制,通过画布分层隔离图像展示与标注操作,确保在万级像素分辨率的高清图像标注场景中,系统仍能保持 60 帧/s 的流畅操作帧率,无卡顿延迟。

(2) 逻辑层:以 JavaScript 来处理标注的核心逻辑,采用内存对象的方式实时保存所有图像的坐标数据。因为避免了频繁读写磁盘,卡顿问题从根本上得到了缓解。

(3) 存储层:接入 JSZip 前端压缩库,在浏览器里就能把标注文件批量打包下载。既符合浏览器的安全限制,无需依赖后端服务介入,也提升了多文件导出的效率,适合大批量图像的标注场景。

1.2 极简交互逻辑设计

本系统的交互设计核心原则是“减少用户视觉焦点的偏移”,让熟练的标注人员能够实现无视觉依赖的“盲操”标注,

1. 保山学院工程技术学院 云南保山 678000

[基金项目] 云南省教育厅 2026 年度科学研究基金项目“腾冲地块古生代构造亲缘性及对特提斯洋演化的约束研究”(2026J1058)

整体采用“右手鼠标打点、左手键盘控制”的键鼠协同操作设计思路，具体交互规则如下：

(1) 左右手配合操作：用户通过右手操控鼠标，完成标注范围的确定，包括多边形顶点点击、矩形框定位等核心操作；左手敲键盘完成各种指令，比如空格键保存当前标注并翻到下一张图，鼠标右键撤销上一步，数字键快速切换标注类别。这样不用来回移动鼠标点击按钮，操作很连贯。

(2) 两点成矩的快捷方式：如果目标物体形状比较规整，用户只需要点两个对角顶点，然后按一下空格键，系统就会自动补全成一个顺时针排列的四个顶点的矩形框。相比手动点四个点，大幅简化操作步骤。

(3) 坐标归一化：在画布上标注时，显示的是像素坐标，方便用户看图定位。等用户保存的时候，系统会自动把这些像素坐标换算成 0~1 之间的相对坐标，满足 YOLOv8-seg 这类分割模型的数据格式要求。

2 系统实现细节

2.1 坐标转换算法

为解决不同屏幕分辨率、浏览器窗口缩放比例对标注精度的影响，确保标注坐标与图像原始尺寸的一致性，本系统设计并实现了数据的动态比例映射。

设图像的原始像素尺寸为 (W_i, H_i) ，Canvas 画布中图像的显示尺寸为 (W_c, H_c) ，用户在画布上的鼠标点击位置坐标为 $(x_{\text{click}}, y_{\text{click}})$ ，则该点击位置对应的图像原始像素坐标 $(x_{\text{img}}, y_{\text{img}})$ 的计算公式为：

$$x_{\text{img}} = x_{\text{click}} \cdot \frac{W_i}{W_c} \quad (1)$$

$$y_{\text{img}} = y_{\text{click}} \cdot \frac{H_i}{H_c} \quad (2)$$

为适配 YOLO 系列模型的训练要求，最终保存的标注坐标为归一化相对坐标，计算公式为：

$$x_{\text{norm}} = \frac{x_{\text{img}}}{W_i} \quad (3)$$

$$y_{\text{norm}} = \frac{y_{\text{img}}}{H_i} \quad (4)$$

2.2 内存管理与文件读取

为适配大规模图像的标注需求，本系统支持基于文件夹批量上传模式，用户可一次性将数千张待标注图像导入系统，无需单张逐个上传。在图像加载与渲染方面，采用懒加载技术，仅将当前视图中展示的图像数据渲染至 Canvas 画布，未展示的图像暂不加载，降低浏览器的内存占用。^[4] 在标注数据存储方面，构建全局 JSON Map 数据结构，以图像文件名为键名、标注坐标数据为键值，将所有标注数据统一维护在内存中，实现标注数据的轻量化存储与快速调取^[5]。

3 实验验证与效率分析

3.1 实验设置

为定量评估本系统的实际标注效能，本文设计了对比实验，实验选取 1 500 张包含塑料漂浮物、生物漂浮物等目标的实例分割待标注图像，所有图像均处在不同程度的复杂背景下，贴合实际应用场景。

(1) 实验参与人员为 5 名具备计算机操作知识、但无专业数据标注经验的测试者（化名：段学、李师、李小、李学、熊学），确保实验对象和本系统的后续使用者相似。

(2) 实验流程：所有测试者于北京时间 20:10 同时开始标注，每人分配 300 张待标注图像，将其平均分到 5 个子文件夹中（每个文件夹 60 张图像），测试者按文件夹顺序依次完成标注，工具自动记录各文件夹的完成时间。

3.2 耗时数据分析

通过记录每位测试者完成各文件夹标注的耗时，得到详细的标注效率统计数据，具体结果如表 1 所示。

表 1 测试人员分批次耗时详表

测试人员	单位: min					个人均值
	文件夹 1	文件夹 2	文件夹 3	文件夹 4	文件夹 5	
段学	10	7	7	5	5	6.8
李师	9	8	6	6	5	6.8
李小	10	7	7	6	5	7.0
李学	10	6	6	6	6	6.8
熊学	9	7	6	6	5	6.6
整体均值	9.6	7.0	6.4	5.8	5.2	6.8

3.3 效率趋势与学习曲线

从表 1 的实验数据中，可清晰地观察到测试者使用本系统的标注效率呈现明显的提升趋势，存在显著的学习效应，具体可分为三个阶段：

(1) 初期探索阶段：完成第 1 批次 60 张图像标注的平均耗时为 9.6 min，折算下来单张图像标注耗时约 9.6 s。此阶段测试者主要处于探索工具快捷键、鼠标打点手感的磨合期，对交互逻辑尚未形成操作记忆，标注效率相对较低。

(2) 适应操作阶段：完成第 3 批次 60 张图像标注的平均耗时降至 6.4 min，相较于初期阶段，标注速度提升 33.3%。此阶段测试者已基本掌握工具的键鼠协同操作方式。

(3) 熟练操作阶段：完成第 5 批次 60 张图像标注的平均耗时仅为 5.2 min，折合单张图像标注耗时约 5.2 s，所有测试者均能在 5~6 min 内完成 60 张图像的标注。此阶段测试者无需频繁关注界面，可实现简单的“盲操”标注，操作流程的衔接性大幅提升。

实验结论：本系统的极简交互逻辑设计符合人体操作习惯，操作过程极易被用户内化为肌肉记忆，随着操作熟练度的提升，标注效率持续提高，熟练操作阶段的标注效率较初期探索阶段提升约 85.1%，具体结果如图 1 所示。

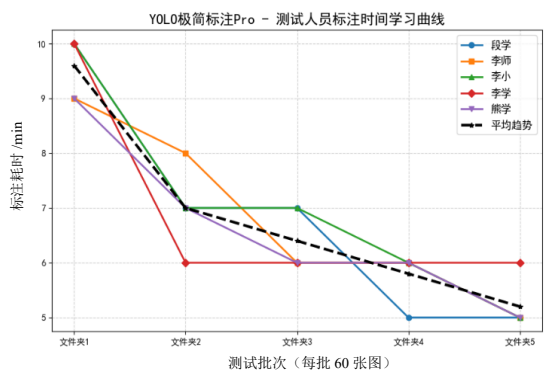


图 1 本系统测试人员效率趋势与学习曲线分析

3.4 稳定性与产出物验证

实验结束后，本系统成功将 1 500 张图像的标注结果导出为包含 1 500 个 TXT 标注文件的 ZIP 压缩包，文件命名与原始图像文件严格对应，方便后续的数据集管理。所有 TXT 标注文件的内部数据格式均符合 YOLO 实例分割的标准要求，以“class_{id}、x₁、y₁、x₂、y₂...”的格式存储标注类别与归一化坐标，无需任何后处理操作，可直接导入 YOLO task=segment 训练命令中，用于 YOLO 系列分割模型的训练，验证了工具的输出稳定性与格式适配性。

4 综合对比分析

为进一步验证本系统的应用优势，将 YOLO 标注系统与目前主流的 LabelMe、CVAT 标注工具从部署成本、格式适配、交互逻辑、标注效率、适用场景五个维度进行定性对比，具体结果如表 2 所示。

表 2 主流标注工具与本系统的定性对比表

对比维度	LabelMe	CVAT	本系统
部署成本	需配置 Python 及各类依赖库，部署繁琐	需 Docker 容器或云端服务器部署，依赖较重	纯网页端运行，零部署成本
格式适配	默认输出 JSON 格式，需编写脚本转换为 YOLO 格式	支持多种格式输出，但 YOLO 格式需复杂配置	原生支持 YOLO TXT 格式导出，无需格式转换
交互逻辑	操作过程中频繁弹出确认窗口，点击步骤冗余	工具栏功能繁杂，菜单层级深，操作路径长	快捷键 + 鼠标右键驱动，盲操模式，无冗余操作
适用场景	适用于对标注精度要求极高的精细学术研究	适用于多团队协同的大规模视频帧标注场景	适用于快速迭代的个人 / 小团队计算机视觉项目

5 结论

本文针对 YOLO 实例分割标注场景中，传统工具部署繁琐、交互复杂、效率低下的问题，设计并实现了一款标注系统。该系统基于 HTML5 Canvas 与 JavaScript 技术构建纯前端架构，实现了零安装、跨平台的运行效果；通过重构人机交互逻辑，设计了键鼠协同的“盲操”标注模式，简化了标注操作步骤；同时实现了标注坐标的归一化，原生支持 YOLO TXT 格式导出，适配模型直接训练需求。本系统的极简交互逻辑设计符合人体操作习惯，极易被用户内化为肌肉记忆，随着操作熟练度的提升，标注效率持续优化，后期极致阶段的标注效率较初期适应阶段提升约 85.1%。

本文所设计的 YOLO 标注系统，为小团队及快速迭代型计算机视觉项目的 YOLO 模型训练，提供了一套轻量化、高效率、高适配性的数据标注解决方案，也为计算机视觉领域轻量化标注工具的设计提供了参考思路。

参考文献：

[1] 吉江燕,方挺.基于Labelme的参考图像的手工分割[J].微型机与应用,2015,34(17): 49-51.

[2] 贺晓倩.水面漂浮物目标检测算法的研究与设计[D].西安:西安工业大学,2022.

[3] 保山学院.YOLO Vision Pro 智能视觉标注系统: V1.0 [CP/DK].北京:中华人民共和国国家版权局,2026-05-20.软著登字第 17825290 号.

[4] 迪力夏提·多力昆,张太红,冯向萍.LabelMe 标注核对系统的设计与实现[J].计算机技术与发展,2022,32(3): 214-220.

[5] 项阳阳.网站图片加载技术的设计与实现[J].冶金与材料,2019,39(5): 77-78.

【作者简介】

李延胜(1988—)，男，云南大理人，硕士，实习研究员，研究方向：智慧水利。

张旭(1991—)，通信作者(email:bsxy11027@bsc.edu.cn)，男，贵州遵义人，博士，讲师，研究方向：工程地质。

柏洪雪(1988—)，女，云南保山人，硕士，讲师，研究方向：智慧水利。

战晓林(1977—)，男，黑龙江齐齐哈尔人，硕士，讲师，研究方向：智慧水利。

(收稿日期：2026-06-01 修回日期：2026-07-14)