

排爆车智能视觉目标检测系统

白紫婵, 卫曦, 葛景仲

(北华航天工业学院电子与控制工程学院, 河北 065000)

摘要: 对复杂环境而言, 排爆车所作的作业要同时完成两项任务: 查清爆炸物、确定机械臂可用的位置。文中将实时检测同 6D 位姿估计结合起来, 设计为车载视觉方案。检测部分是经过改良的 YOLOv5s——在 SPPF 之前利用坐标注意力、把边界框回归损失由 CIOU 改作 EIOU, 以类别及边界框形式给出结果; 判断为高威胁的物体按框截取区域后即用 FoundationPose 处理, 结合已知三维模型的特征做 6 自由度位姿解算。所涉两个模块用优先级异步调度器连接。检测端针对未爆弹数据所获的指标是精确率 85.3%、召回率 67.7%、mAP77.0%, 比原版 YOLOv5s 各项数值分别多 0.7、1.4、1.1 个点, 其结果满足机械臂自动定位的需要, 验证了该方法的可行性与效率。

关键词: 排爆车; 目标检测; YOLOv5s; 6D 姿态估计; FoundationPose

基金项目: 北华航天工业学院创新资助项目“排爆车智能视觉目标检测系统”(YKY-2026-023)

DOI: <https://doi.org/10.68249/jstnei.2026.0014>

1 引言

爆炸物处置一直是城市安全和反恐处突里难以解决的一项重要问题。传统做法是人工进行近距离识别与处置人, 风险全压在操作人员身上。用智能化的无人平台替人进场, 是目前比较现实的解法^[1]。平台上的视觉系统有两项任务: 发现并锁定疑似爆炸物, 再把它的空间位置报给机械臂^[2]。

目标检测方面的技术路线已经比较成熟。YOLO^[3]、SSD^[4]一类单阶段网络速度和精度平衡较好, 是实时场景的常规选择; 对于移动端部署已经有多种轻量化改进得方法, 深度可分离卷积、通道重排、通道剪枝都是常用的手段^[5]。排爆场景用轻量化 YOLO 识别手榴弹一类制式弹药^[6]已经有先前研究, 也有针对 76mm 舰炮未爆弹改进 YOLOv5s 的工作^[7]。但直接进行应用效果并不理想: 弹种杂、外观差别大, 现场背景乱且常有遮挡, 漏检误判概率比较高。更根本的问题是, 只靠平面检测结果远不能满足机械臂作业需求, 机械臂需要准确知道目标在三维空间中的位置和朝向, 即 6 自由度 (6-DoF) 姿态。

6D 姿态估计有两种方法。PoseCNN^[8]、PVNet^[9]从单张 RGB 图回归或投票, 精度可以, 但换一批物体就得重训一次; FoundationPose^[10]大规模数据预训练之后, 只要给出三维模型或少量参考图, 没见过的物体也能估位姿并跟踪, 爆炸物形态杂、样本少, 后者更适用。

本文以排爆车为应用平台, 将二者结合并进行了验证: 检测端在 YOLOv5s 上嵌入坐标注意力、改用 EIOU 回归损失; 位姿端接 FoundationPose; 两级由异步调度协调算力。

2 系统架构设计

2.1 系统总体结构

系统分三层: 数据采集、算法推理、结果输出。采集层从车载相机取帧, 做预处理, 把标定参数和时间戳一并绑上。推理层放检测和位姿两个模块, 由调度器异步驱动。输出层把 2D 框和 3D 位姿合到一起, 经手眼标定换到车体坐标系, 交给决策模块。目标位姿用齐次变换矩阵描述, 如式 (1); 相机系到车体系的换算如式 (2),

作者简介: 白紫婵(2004—), 女, 硕士研究生, 研究方向为机器视觉;

卫曦(2002—), 男, 硕士研究生, 研究方向为机器视觉;

葛景仲(2002—), 男, 硕士研究生, 研究方向为机器人学。

通讯作者: 白紫婵, 邮箱: 782786425@qq.com

系统总体架构如图 1 所示。

$$T = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \in SE(3), R \in SO(3), t \in \mathbb{R}^3 \quad (1)$$

$$T_b^o = T_b^c T_c^o \quad (2)$$

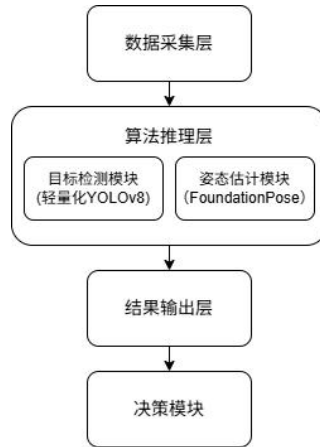


图 1 系统总体架构

2.2 级联工作流程

系统启动后先进行初始化：将轻量化 YOLOv5s 检测模型^[11]和 FoundationPose^[10]载入 GPU 显存，各类别的三维模型特征编码读进内存；初始化完成后相机开始采集，系统进入主循环；每轮取一帧 RGB 图，预处理后送入检测模块，得到类别、置信度和边界框等信息，调度器据此分流；若没有有效目标就取下一帧；有高威胁目标就按框裁区域，从特征库取对应类别的编码，一并送进位姿模块。得到的位姿信息换到车体坐标系，连同检测结果送往决策模块。

2.3 任务调度策略

由于车载平台算力紧张，系统调度采用了优先级异步机制。目标检测优先级最高，按固定频率持续运行，视野始终可以得到实时的监控；位姿模块按事件触发，出现新目标时创建新线程，初始化完成后转跟踪，更新频率相对较低；目标离开视野后，调度器及时回收对应的线程，保证单块 GPU 上检测的实时性和位姿精度可以同时得到兼顾。

3 核心算法实现

3.1 YOLOv5s 检测模型优化

YOLOv5 的骨干是 CSPDarknet，由 C3 模块堆叠、以 SPPF 收尾；颈部为 PANet 的双向特征融合；检测头是基于锚框的耦合头，分类与回归共用一条支路^[11]。s 版本参数量小、推理快，适合嵌入式部署。本文以 YOLOv5s 起步，先在排爆数据集上微调，再做两项改进，改进后的整体结构如图 2 所示。

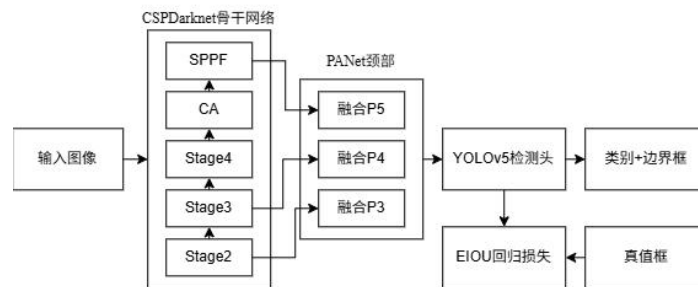


图 2 改进的 YOLOv5s 网络结构

一是在 SPPF 之前嵌入坐标注意力^[12]。它把水平、垂直两个方向分开池化，位置信息不会在通道压缩中丢掉。两个方向的池化如式（3），加权输出如式（4）。炮弹多为细长回转体，朝向判断依赖位置信息，这一点对本场景有用。

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i), \quad z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (3)$$

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (4)$$

二是把边界框回归损失从 CIOU 换成 EIOU^[13]，如式 (5)。CIOU 用一个长宽比因子间接约束形状，小目标上交集小、IOU 分母也小，数值容易抖；EIOU 把长宽比拆开，对宽和高分别惩罚，梯度更稳。排爆现场的目标常常又远又小，还带角度变化，EIOU 相对来说更适用。

$$L_{EIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \quad (5)$$

排爆场景样本少、背景杂，因此采用多样化的数据增强策略来扩充训练样本的多样性。Mosaic 把四张图随机缩放、裁剪再拼接，叠加随机翻转、旋转与色彩空间变换；输入统一缩放到 640×640 并归一化，标注框转为相对坐标。训练采用迁移学习方式，以 COCO 预训练权重初始化^[14]，优化器为 SGD，初始学习率 0.01、动量 0.937、权重衰减 0.0005，超参数按 hyp.scratch-low 配置，batch-size 取 8，共训练 50 个 epoch。

3.2 数据集构建

数据集构建采用仿真与采集相结合，仿真方面，利用 Blender 三维引擎搭建虚拟的排爆环境，获取不同种类炮弹的三维模型，控制相机在不同视角采样；随机调整光源参数来模拟不同的光照条件，在目标前方放置遮挡物来模拟实际场景的遮挡状况，每个类别生成约 1000 张渲染图像，自动记录 6D 姿态真值与边界框。使用 LabelImg 工具对收集到的图像标注边界框，最终按 8:1:1 随机将数据集划分为训练集、验证集和测试集^[7]。仿真数据集构成如图 3 所示，部分炮弹模型如图 4 所示。



图 3 数据集部分构成

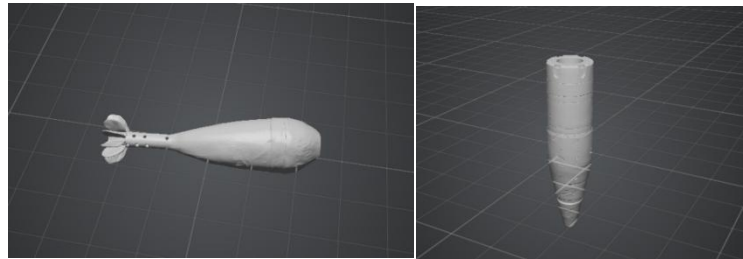


图 4 部分炮弹 3D 模型

3.3 FoundationPose 姿态估计

FoundationPose 的姿态估计能力依赖于对目标物体外观和几何特性的充分表示，离线阶段要为每类目标建好特征。CAD 模型先统一缩放、平移质心、重采样网格，再由编码分支提几何特征。炮弹是回转体，视觉上高度对称，位姿容易多义，还要标出对称轴、建等价集合。特征编码连同对称性元数据序列化后按类别入库，开机一次性载入内存。

在线推理阶段，系统需在缺少初始位姿信息的条件下完成初始化估计。FoundationPose 走由粗到精两步：姿态空间分层采样，用可微分渲染器^[15]把候选姿态渲染出来与观测区域比相似度，筛掉大部分；再以光度一致性为损失做梯度下降细调。这一路不依赖局部纹理，对纹理稀疏的金属炮弹外壳同样有效^[16]。粗搜索的筛选准则见式 (7)，细调的损失见式 (8)。对称目标从多个收敛解里选最优结果进行输出。

$$\hat{T} = \arg\max_{T \in \mathcal{T}} S(I_{\text{obs}}, \mathcal{R}(\mathcal{M}, T)) \quad (7)$$

$$L(\mathbf{T}) = \sum_u \|I_{\text{obs}}(u) - \mathcal{R}(\mathcal{M}, \mathbf{T})(u)\|^2, \mathbf{T}^* = \underset{\mathbf{T}}{\operatorname{argmin}} L(\mathbf{T}) \quad (8)$$

之后转为跟踪模式。以上一帧位姿 $\mathbf{T}_{k-1}$ 为初值，把模型渲染出来与当前帧作差，取轮廓及边界像素参与计算，高斯牛顿迭代求帧间增量，按式（9）复合得当前位姿。渲染一致性掉到阈值以下就重新初始化，在保证精度的同时实现高频连续的位姿更新。

$$\mathbf{T}_k = \Delta \mathbf{T}_k \mathbf{T}_{k-1} \quad (9)$$

4 实验测试与分析

4.1 实验环境配置与评估方法

实验硬件平台采用搭载高性能独立显卡的工控计算机，以模拟车载嵌入式计算环境，软硬件环境配置如表 1 所示。配置见表 1。

表 1 实验软硬件环境配置

项目	配置
计算平台	搭载高性能独立显卡的工控计算机（模拟车载嵌入式计算环境）
操作系统	Ubuntu20.04 LTS
检测模块	PyTorch+CUDA 训练，TensorRT（INT8 量化）部署
姿态估计模块	LibTorch（C++）推理，Nvdiffrastr 可微分渲染
数据与模型工具	Blender、Trimesh、Open3D、LabelImg

系统采用精确率、召回率和平均精度均值来评价检测性能；姿态估计精度由 ADD 平均距离和 n° -m cm 逐帧合格率两项指标来进行衡量——ADD 是预测位姿与真值位姿分别作用于模型采样顶点后两组点云的平均欧氏距离，ADD 定义见式（10）。指标定义和统计方式按照 BOP^[17]标准执行。端到端性能按各环节耗时进行评价。

$$\text{ADD} = \frac{1}{m} \sum_{x \in \mathcal{M}} \|(\mathbf{R}\mathbf{x} + \mathbf{t}) - (\tilde{\mathbf{R}}\mathbf{x} + \tilde{\mathbf{t}})\| \quad (10)$$

4.2 检测模块性能分析

为了验证 SE、CBAM 和 CA 三种注意力机制在 YOLOv5s 中的应用效果，并比较它们在不同位置的表现，设计实验来分析它们对目标检测性能的影响，各模型在 3.2 节所述的同一套训练集、验证集与测试集上训练评测。

以原版 YOLOv5s 为基线，把 SE、CBAM、CA 分别插入位置一（骨干末端、SPPF 之前）与位置二（颈部深层、第一个 concat 之后），结果见表 2。

表 2 注意力机制种类与插入位置对比

模型	P/%	R/%	mAP/%
YOLOv5s（基线）	84.6	66.3	75.9
+SE（位置一）	85.0	66.7	76.2
+SE（位置二）	84.8	66.7	76.0
+CBAM（位置一）	85.1	67.0	76.3
+CBAM（位置二）	84.8	66.8	76.1
+CA（位置一）	85.1	67.5	76.8
+CA（位置二）	85.2	66.9	76.3

三种模块都有增益，但是幅度不同。CA 插在位置一时最好：85.1%、67.5%、76.8%，比基线分别高 0.5、1.2 和 0.9 个百分点，召回率提升尤其明显；同位置的 SE、CBAM 的 mAP 为 76.2%、76.3%，均低于 CA。位置一整体优于位置二，骨干末端的特征还保留着较完整的空间分布，嵌入坐标注意力模块能够对位置信息进行重标定；颈部深层特征经过多次下采样与融合后空间细节有所损失，坐标注意力的增益相对有限。据此选定 CA 并插在 SPPF 之前。

回归损失与整体消融实验结果见表 3。

表 3 消融实验结果

CA	EIOU	P/%	R/%	mAP/%
—	—	84.6	66.3	75.9
✓	—	85.1	67.5	76.8
—	✓	84.9	66.8	76.7
✓	✓	85.3	67.7	77.0

把 CA 单独引入之后，所用模型的精确率及召回率、mAP 依次是 85.1%、67.5%与 76.8%，即比原来多出 0.5、

1.2 和 0.9 个百分点,由此可判断 CA 是特征提取方面的一个优良补充;若只使用 EIOU 来改进模型,那么精确率、召回率以及 mAP 为 84.9%、66.8%和 76.7%,所增幅度是 0.3、0.5、0.8 个百分点,改进边界框回归损失对定位及总体检测是有利的。注意力模块同 EIOU 的损失配合起来能提高检测效率。所称的协同效果正好适用于复杂情况中目标的识别和定位。检测效果如图 6 所示。



图 6 YOLOv8 检测模块

4.3 姿态估计模块性能分析

实验正式开始前,先构建目标的模型表示。使用一枚迷你炮弹做目标,从多个角度和方位录制视频,利用这些视频画面组成参考视图集,送进 FoundationPose 的编码分支建特征描述符,供初始化和跟踪调用。模型构建完成以后,对测试图像进行位姿估计实验。不遮挡时,制式弹和简易装置的匹配效果较好,渲染图和实拍比较吻合,ADD 误差不大。遮挡过半时 ADD 有所上涨,初始化还要在更大的参数空间里搜解,收敛变慢、响应时间拉长。整体来看,FoundationPose 依靠全局几何形状匹配机制,在多数场景下能够给出合理估计。转入跟踪后单帧耗时下降,满足动态跟踪的需要。效果如图 7 所示。

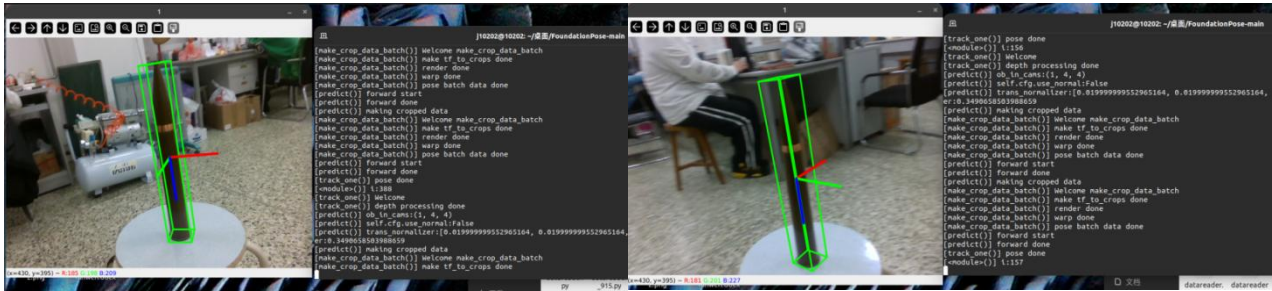


图 7 FoundationPose 姿态估计模块

4.4 系统整体运行评估

两个模块集成联调,输入用离线视频序列模拟实时流。画面无威胁目标时只运行检测模块,帧率和检测模型单独推理时一样。出现单个目标时,首帧初始化,之后转为跟踪模式。遮挡过半时初始化时间变长,帧率比较稳定。多个目标同时在画面时,多线程并行。检测与姿态估计两个模块能够配合运行,说明级联结构合理且稳定,系统整体性能可以满足大多数排爆作业场景的要求。

5 结论

本文对于排爆车研究了一套检测与 6D 位姿估计级联的视觉系统。检测端在 YOLOv5s 的 SPPF 前嵌入坐标注意力、将 CIOU 换成 EIOU 回归损失,在未爆弹数据集上精确率 85.3%、召回率 67.7%、mAP77.0%,较原版分别提升 0.7、1.4 和 1.1 个百分点;位姿端使用 FoundationPose,二者结合可以实现目标检测功能。

所得到的结果可用来证明轻量级检测器及一般性位姿基础模型是能配合使用的,所建仿真数据集亦可作同类工作的参考。目前模型试验仍有若干缺陷:有关检测的输入数据种类不够多,位姿相关部分基本未用实际资料,与现实状况相比有差距,考察的目标个数亦有限,所以接下来要以实物实验、对多目标同时处理为着眼点做分析,按阶段地把视觉识别同机械臂动作结合起来,形成完整控制逻辑。

参考文献:

- [1]曾俊,卢瑞涛,杨小冈,等.六旋翼无人机空基智能排爆系统设计与实现[J].电光与控制,2023,30(5):61-65.
- [2]于涵,李一染,毕书博,等.一种基于随动视觉的排爆机器人自主抓取系统[J].计算机工程,2021,47(1):298-304,311.
- [3]REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:779-788.
- [4]LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision,2016:21-37.
- [5]MA N, ZHANG X, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]//Proceedings of the European Conference on Computer Vision,2018:116-131.
- [6]邱锦,刘健,沈芸亦,等.基于 YOLOv5 手榴弹检测算法的轻量化研究[J].兵器装备工程学报,2023,44(6):35-41.
- [7]范凤欢.基于 YOLOv5s 的未爆弹识别算法研究[D].廊坊:北华航天工业学院,2025.
- [8]XIANG Y, SCHMIDT T, NARAYANAN V, et al. PoseCNN: a convolutional neural network for 6D object pose estimation in cluttered scenes[C]//Robotics: Science and Systems XIV,2018.
- [9]PENG S, LIU Y, HUANG Q, et al. PVNet: pixel-wise voting network for 6DoF pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019:4561-4570.
- [10]WEN B, YANG W, KAUTZ J, et al. FoundationPose: unified 6D pose estimation and tracking of novel objects[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2024:17868-17879.
- [11]JOCHER G. Ultralytics YOLOv5[EB/OL]. (2020-06-26) [2026-08-10]. <https://github.com/ultralytics/yolov5>.
- [12]HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021:13713-13722.
- [13]ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J].Neurocomputing,2022,506:146-157.
- [14]LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//Proceedings of the European Conference on Computer Vision,2014:740-755.
- [15]LAINE S, HELLSTEN J, KARRAS T, et al. Modular primitives for high-performance differentiable rendering[J].ACM Transactions on Graphics,2020,39(6):194.
- [16]HINTERSTOISSER S, LEPETIT V, ILIC S, et al. Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes[C]//Proceedings of the Asian Conference on Computer Vision,2012:548-562.
- [17]HODAN T, MICHEL F, BRACHMANN E, et al. BOP: benchmark for 6D object pose estimation[C]//Proceedings of the European Conference on Computer Vision,2018:19-34.

Intelligent Vision-based Object Detection System for Explosive Ordnance Disposal Vehicles

Zichan Bai, Xi Wei, Jingzhong Ge

School of Electronic and Control Engineering, North China Institute of Aerospace Engineering, Langfang 065000, China

Abstract: For complex environments, explosive ordnance disposal (EOD) vehicles must simultaneously accomplish two tasks: detecting explosives and identifying suitable positions for the robotic arm. This paper integrates real-time detection with 6D pose estimation into a vehicle-mounted vision system. The detection module is an improved version of YOLOv5s — enhanced by incorporating coordinate attention before the SPPF layer and replacing the bounding box regression loss from CIoU to EIou, outputting results in both class and bounding box formats. Objects identified as high-threat are cropped according to their bounding boxes and then processed by FoundationPose, which performs 6-degree-of-freedom pose estimation using features from known 3D models. The two modules are connected via a priority-based asynchronous scheduler. On unexploded munition datasets, the detection module achieves precision of 85.3%, recall of 67.7%, and mAP of 77.0%, surpassing the original YOLOv5s by 0.7, 1.4, and 1.1 percentage points respectively. These results meet the requirements for automatic robotic arm positioning, validating the feasibility and efficiency of the proposed method.

Keywords: Explosive Ordnance Disposal Vehicle; Object Detection; YOLOv8; 6D Pose Estimation; FoundationPose