基于能量机关仿真数据集的单目视觉位姿测量改进研究

HKUST ENTERPRIZE
薛旭朗 周思宇
ilsit@connect.ust.hk szhoubx@connect.ust.hk

摘要

今年的 *RoboMaster* 赛事中，官方未提供比赛视频资料，且符的外框灯条被移除，给数据集准备和模型训练带来了双重挑战。商家出售的能量机关击打检测模块高达 *495* 元/片，同时还需要自行制作灯效模块，对于许多队伍而言成本过高。往年此时，官方已经发布能量机关灯效演示视频，而今年没有提供此类资源，队伍不得不自力更生。我们团队通过 *AI* 技术 *[4]* 处理历史数据，并结合 *Blender* 自制合成渲染数据，成功实现了低成本、高效的能量机关目标识别与关键点数据集生成。本报告详细记录了这一过程，旨在为各参赛队伍提供参考。

1. 能量机关的历史

在 RoboMaster 机甲大师赛事中，能量机关作为关键任务目标，其设计和规则随赛季演变，经历了多次调整。自 2016 赛季新增场地视觉识别机关“神符系统”，早期的符设计较为简单，通常以固定图案和明亮的灯条外框为主，便于视觉识别和关键点定位。

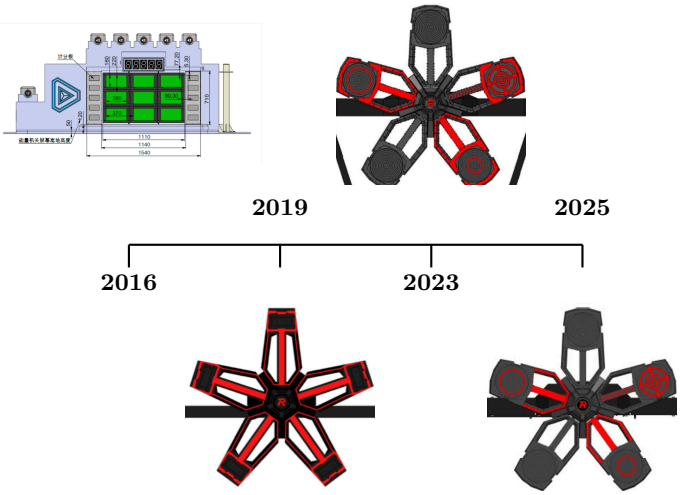


Figure 1. 时间线展示（2016-2025），包含各年份的能量机关照片。

例如，在 2018-2020 赛季，符多采用规则几何形状（如圆形或矩形），外框灯条持续发光，参赛队伍可通过传统图像处理方法（如边缘检测、霍夫变换）实现稳定的检测。然而，随着比赛对技术难度的提升，符的复杂度逐渐增加。参见 Figure 1。2023 赛季的符引入了动态图案和多样化的灯效，但外框灯条依然是重要的视觉线索，用于传统视觉算法的定位和识别。今年（2025 赛季），规则发生显著变化：官方删除了符的外框灯条设计。这一改动使得外框不再发光，彻底改变了以往依赖外框特征的识别策略，迫使队伍重新设计关键点识别方案。外框的移除不仅增加了识别的难度，也对关键点定位提出了更高要求，因为失去了外框这一稳定且利于识别的参考边界。面对这一变化，我们团队创新性地采用了 AI 图像处理技术消除去年数据集中的外框，结合 Blender 渲染生成合成数据，以适应新规则下的挑战。


AI 消除效果

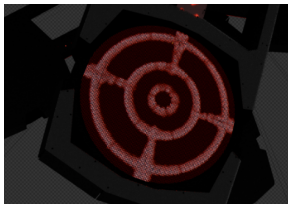

Blender 渲染效果

Figure 2. 效果展示

2. 能量机关识别技术方案对比

2.1. 非神经网络图像处理算法

[2]

Figure 3. 传统的计算机视觉方法

传统的计算机视觉方法（以下简称传统 CV），在简单场景下能够实现一定程度的关键点定位。然而，今年能量机关的图案设计复杂且无外框，传统 CV 的局限性暴露无遗：

- 缺点：
  - 依赖图像中的强特征（如边缘、角点），而今年能量机关刻意减少了这些特征，导致定位不准。
  - 通常需要做二值化前处理，在光照变化、遮挡等复杂条件下鲁棒性差，难以满足需求。
- 适用场景：仅适合图案简单、明暗分明、背景干净的理想情况，不足以应对新规则的挑战。

2.2. 目标检测神经网络模型 + 传统 CV

[2]

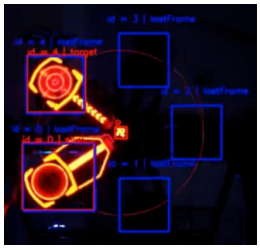


Figure 4. 目标检测神经网络模型 + 传统 CV

这种方案结合了 YOLO 目标检测和传统 CV 技术，先通过 YOLO 定位能量机关的目标区域，再利用传统 CV 方法进行关键点的精确定位：

- 优点：
  - YOLO 的目标检测能力提高了初始定位的准确性。
  - CV 技术进一步细化关键点位置，在一定程度上弥补了纯传统 CV 的不足。
  - 可在局部范围使用更加合理的二值化算法，如大津算法，获得更加合理和精准的二值化结果
- 缺点：
  - 对图像质量和特征依赖性仍然较高，光照变化或遮挡会导致 CV 步骤失效。
  - 处理流程复杂，包含目标检测和后处理两步，速度较慢，难以满足实时性要求。
  - 新版能量机关去除了边框部分，传统的方法难以找到特征点。
- 适用场景：适合中等复杂度的场景，但面对今年能量机关的多变性和动态性仍显不足。

2.3. YOLO Pose

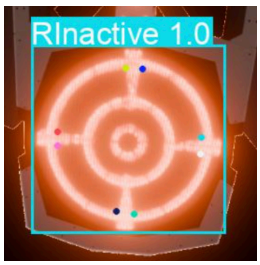


Figure 5. YOLO11n-pose

YOLO Pose（以 YOLO11n-pose 为例）是一种基于深度学习的姿态估计方法，直接在目标检测的同时预测关键点位置。
以下为 yolo-pose 结构（以 YOLO-pose 为例），其检测头可同时输出目标检测框与关键点：

- 优点：
  - 高效性：无需额外的后处理步骤，检测速度快，适合实时应用。
  - 鲁棒性：通过大规模多样化数据训练，能够适应复杂图案和不同位姿的场景。
  - 精确性：关键点定位误差小，在合成数据和真实场景中均表现出色。
- 缺点：需要大量高质量的训练数据支持，对数据集的依赖性较高。
- 适用场景：完美契合今年能量机关识别的需求，是当前的最佳选择。

3. 关键点数据集的生成

3.1. 难点分析

今年能量机关的识别任务面临以下挑战：

- 图案复杂：今年能量机关的图案设计使得传统方法难以定位精确关键点。
- 标注不一致：人工标注时，不同人员的标准差异导致关键点位置不准。
- 无固定能量机关激活点，要求队伍在不同位置都有估计能量机关位姿的能力。

3.2. 合成数据集的解决方案

为了解决上述难点，我们提出了两种生成关键点数据集的方法，并针对其实现方式、效果和局限性进行分析。
以下表格从多个维度对比了 **AI 消除能量机关外框**和 **Blender 渲染合成数据集**的表现

	AI 消除	Blender 渲染
方法实现	利用 AI 技术去除先前数据集中的外框，模拟无外框场景	通過过 Blender 進行 3D 建模和渲染生成合成数据集
创新点	利用 AI 技術去除先前數據集中的外框，模擬無外框場景	通過 Blender 進行 3D 建模和渲染生成合成數據集
效果	• 实现半自动标注 • 减少人工工作量	• 生成高质量合成图像 • 与实拍数据高度相似
局限性	• 仍需人工审核 • 难以大规模生产	• 需精确控制渲染参数 • 初始建模需时间 • 缺少现实噪声，需要人工数据增强
优势	• 利用现有数据，快速生成 • 成本低	• 自动化生成大量数据 • 高一致性和准确性
适用场景	验证集、快速测试	大规模训练集、复杂场景适应性

Table 1. AI 消除能量机关外框和 Blender 渲染合成数据集对比表

4. 方案总览

这个技术方案分为三个核心部分：AI 技术消除、Blender 渲染和 YOLO11 8 点 pose 检测。这套方案从数据生成到模型训练一气呵成，既实用又成本低廉。从此，买不起能量机关的队伍依旧可以生成无限数据集。成本对照表：

方案	成本	数据量
机械能量机关实体 + 人工标注方案	495 元/片（击打检测部分）+ 灯条成本	有限（依赖人工拍摄与标注）
本方案（Blender）	0 元	2 万 + 组合数据

Table 2. 成本对照表

5. AI 消除能量机关外框

5.1. 思路

- 由于我们得到的大多数是去年及以前的能量机关数据，其图像上面具有外框，难以符合今年的需求。在消除外框的过程中，我们发现，如果只是简单的给外框用画笔涂成黑色，在多数背景下其仍会显得较为突兀，存在模型将其识别为特征的风险，于是我们想找一种更为自然的消除方法。

5.2. 实现步骤

- IOPaint [4] 是一个基于 LaMa 模型的人工消除工具。并且开发出了批量文件管理和 GPU 加速，让处理掩膜（mask）以一种更加自然的方式融入背景。
- 使用者通过编写好的画笔 UI 描出需要消除的部分，在上述方式的加速下，熟练的使用者可以达到大约 200-250 张/小时（取决于每张图已点亮能量机关的个数）。

Figure 6. IOPaint 操作页面图示

5.3. 对比展示

（在此使用西交利物浦 2023 的开源数据集）

Figure 7. 蓝方去年的原始数据（raw）

Figure 8. AI 消除外框后的图像

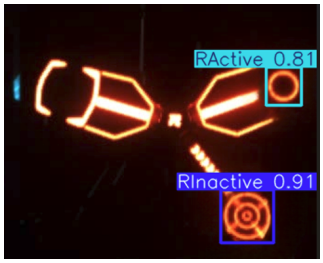


Figure 9. 今年基于消除外框数据训练的新模型识别效果。

在图 9我们注意到，经过训练后，新模型对边框这个特征已经不在敏感，我们的目的也已达成。

5.4. 方案分析

- 效果：该方法实现了复用先前数据集并进行半自动标注，减少了部分人工工作量。
  - 局限性：仍需人工审核以确保质量。
  - 团队算法组人员有限，难以大规模批量生产数据，目前主要用作验证集。

5.5. 改进潜能

- 由于 IOPaint 具有批量处理的功能，我们只需要输入图片文件目录以及每个图片对应掩膜（mask）的目录，可在 GPU 加速下达到约 1500 张/小时的速度。于是我们想到可以先借助一些 mask 提取工具，例如 Segment Anything [3]。在快速提取训练集图片边框的 mask 之后，借助成熟的分割模型（例如 YOLO-segment），可以对新的图片集进行边框 mask 的自动提取，从而可以作为消除的 mask 输入到 IOPaint 的 LaMa 模型中。

- Figure 10为 Segment Anything Demo，可以考虑该方法提取掩膜（mask）：

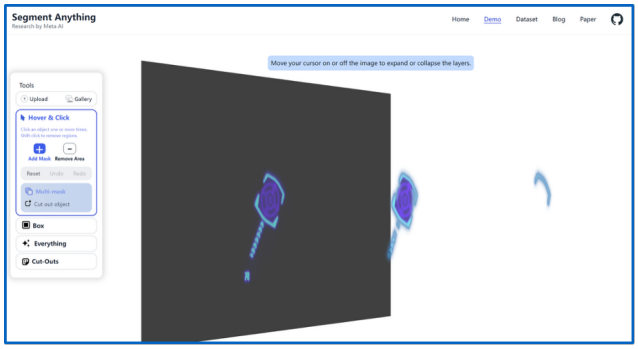


Figure 10. Segment Anything Demo

- 由于团队算法人员有限，数据集基本只有 1-2 人维护，考虑到时间成本，我们目前还是使用的是手动消除的方式。对于拥有大型数据集的团队，可以作为参考方案。

6. 基于 Blender 渲染的合成数据集

通过 Blender 进行 3D 建模和渲染，生成合成数据集。

创新点

- 随机性增强：
  - 模拟损坏灯珠、曝光变化等真实场景效果，避免数据千篇一律。
  - 尽可能贴近官方能量机关的设计，提升数据真实性。
- 背景多样性：
  - 生成透明背景图像，便于替换多变背景，增强数据集的泛化能力。

成果

- 生成了与场地实拍数据高度相似的合成图像。
- 与场地实拍扣掉外框的数据对比，验证了合成数据的有效性。

优势

- 高效性：自动化生成大量数据，减少人工标注负担。

- 高质量：通过精确控制渲染参数，确保数据一致性和准确性。

缺点

- 初次建模需要一定时间

6.1. 建模

我们从官方 RM2025 模型中导出符的基础模型，在 SolidWorks 中进行初步处理后导入 Blender。由于 Blender 使用三面体渲染，我们对模型进行了 mesh 修复，确保渲染效果准确。特别是能量机关的屏灯，我们保留单个扇叶进行细化处理。把灯板修复至 Figure 11的状态。

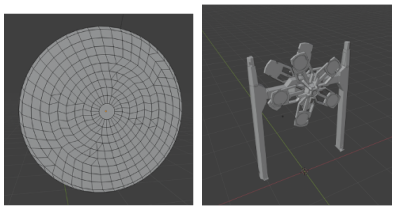
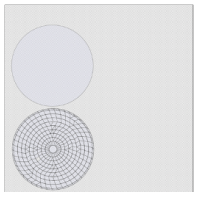


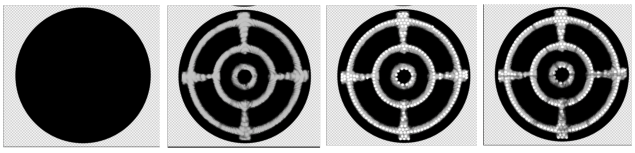
Figure 11. mesh 修复

6.2. 纹理绘制

拿到灯的 UV 后，在 photoshop 画灯。
原始的 UV 是这样的:



在 photoshop 中使用 4 个图层模拟灯珠。第一层是黑色的底色，第二层是官方的纹理，第三层是圆型的灯珠，第四层是支撑架的阴影。其他灯效都是用类似的方法画。



6.3. Shader 着色器

使用大神开源的 LCD 屏幕 [1]
把纹理 node group 的 image texture 改成符的纹理。
这里加入 noise texture 模拟随机坏灯珠，Rotation X

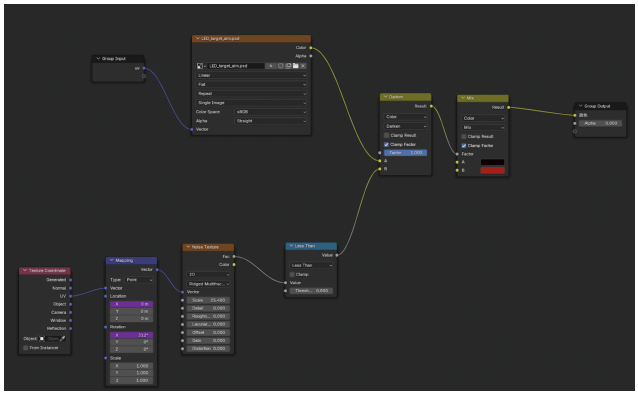
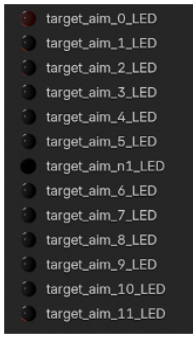


Figure 12. LCD 屏幕

设置 `#frame*1.1%6.28`

把不同的灯效都添加到材质。



这里的命名规则是 0 为目标靶，1-10 是环数，11 是大符的 10 环，n1 是关的能量机关

6.4. 绘制其他灯的纹理

灯板以外的灯的表面都用了磨砂玻璃纹理 Figure 13，只有底面才用了灯的纹理。

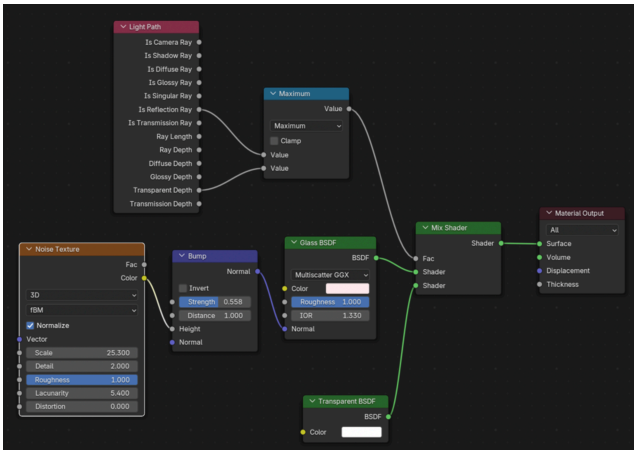


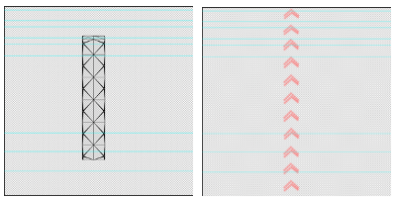
Figure 13. 磨砂玻璃材质

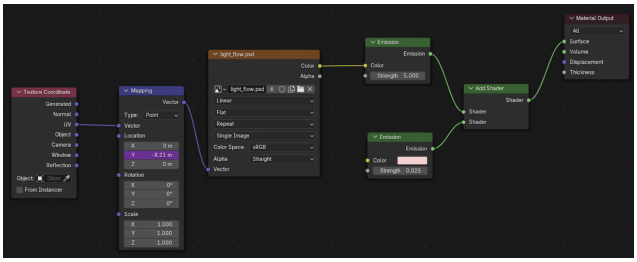
流水灯：

导出流水灯的 UV（只选背面）

画上小灯珠。上下要画满灯珠才能做动画。

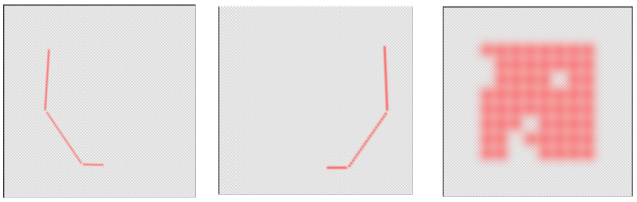
(这里我画了红色因为这是比较早画的，当时不熟悉用 blender 的 shader)



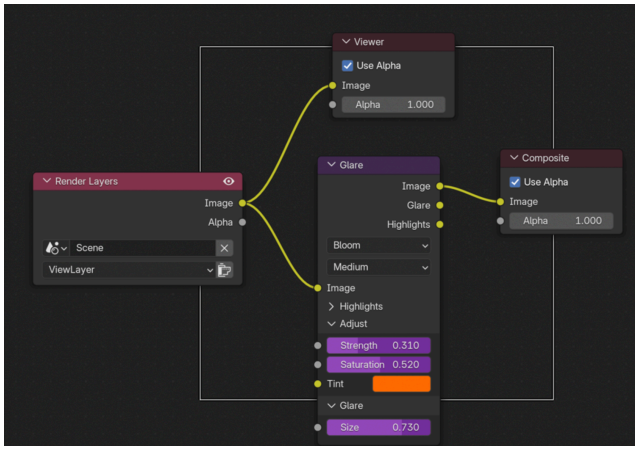


LocationY 设 #frame/-100

其他灯，中心的 R 都用相同的方法画：



6.5. 其他设置



在 Compositing 使用 Glare 辉光效果。参数都和 frame 挂勾，分别设置

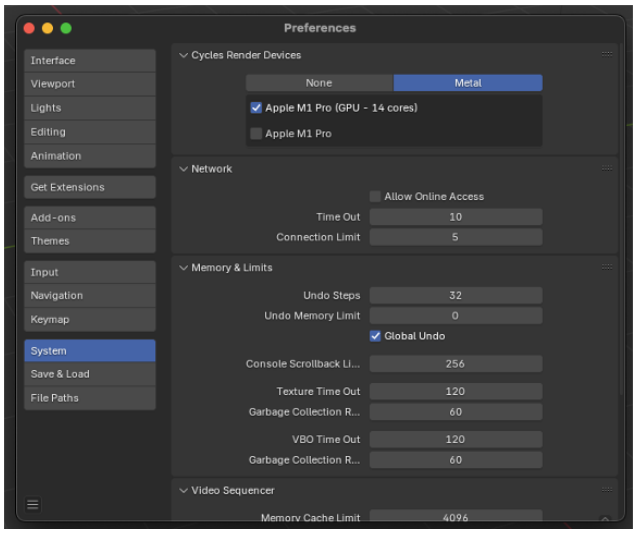
#frame*1.1/10%5

#frame*1.2/10%1

#frame*1.3/10%1

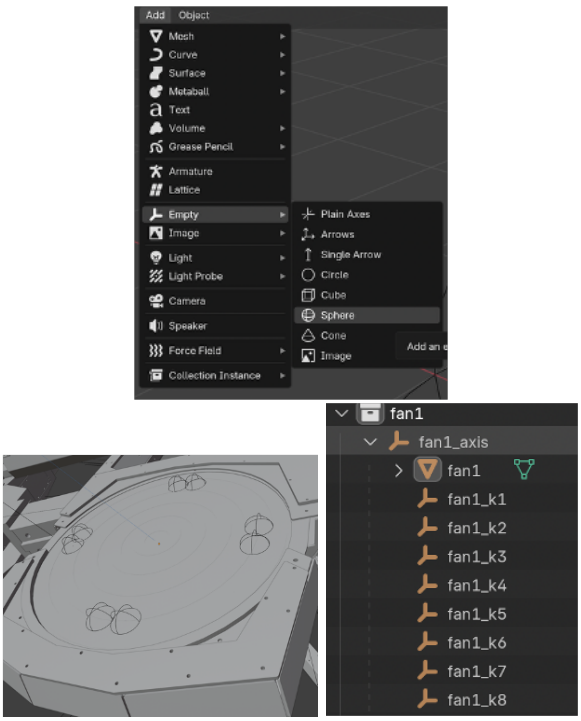
Render Engine 使用 cycles.

如果想使用 GPU 的渲染，在设置打开 CUDA 或 Metal。(如果想在服务器使用 CUDA 要写脚本)。



6.6. 添加关键点

新增 Empty > Sphere 作为关键点。



新增一个 axis 对象，把扇叶的 parent 设成 axis。把做好的一个扇叶复制 10 次。移動完成後，axis 旋转需改为使用四元数。

6.7. 脚本

script.py 功能概述：

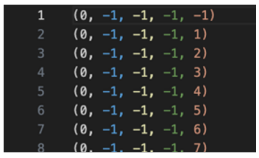
- 核心概念：相机围绕世界中心点旋转，始终朝向目标靶，渲染图像并输出标签。脚本处理相机定位、材质分配和标签生成。
- 标签格式：
  `class x y w h p1x p1y p1v ...`，其中 p1x，p1y，p1v 为关键点坐标和可见性。
- 类别映射（class_index_map）：
  - Class 0：关闭的能量机关（-1）
  - Class 1：目标靶（0）
  - Class 2：已击中靶（1-11）
- 使用方法：
  - Blender UI 的 Script 分页运行。
  - 服务器无头运行：
    `./blender <.blend file> –background –python <.py script>`
- 功能特性：
  - 支持多 GPU 并行渲染。
  - 可定制相机路径和角度。
  - 材质切换模拟不同状态。
  - 关键点检测含可见性检查。
  - 自动创建目录和命名文件。
- 参数：

output_dir	输出图像和标签的文件夹
end_frame	每层水平相机取样次数
num_circles	相机垂直取样层数
rings_combinations_path	符的排列组合.txt 文件
distances	相机取样距离（如 [1，2]）
exposure_list	取样曝光值（如 [0.5，1.0，3.0]）
devices_to_use	使用的 GPU 索引

7. 開源內容

脚本：

- gen_rings_comb.py 生成 2 万多种目标靶组合。命名和上面的材质命名一致。生成规则是有一个扇叶是目标靶的所有情况。我们硬是用脚本生成了 2 万 + 数据样本。



- get_polygon_index.py 获取灯板以外灯的底面 polygon index。因为灯板以外的灯只有底面才用了灯的纹理。

- labels_blender2yolo.py 将 Blender 标签转为 YOLO 格式，保留中心 R 和目标靶标签。

8. 误差分析

假设能量机关激活点在以下位置，距离能量机关约 6 米，以下的计算使用 5 米为假设。

圆盘尺寸：200×200 mm。
假设击打目标是圆盘中心，中心到边缘的距离是半径 100 mm（圆形）。
击打点的误差应小于：100mm

8.1. 可接受角度误差

在 5 米距离下，偏移需小于 100 mm：

$$\tan(\theta) \leq \frac{100}{5000} = 0.02$$
$$\Rightarrow \theta \leq \arctan(0.02) \approx 1.15^\circ$$

8.2. 系统参数

- 解析度：2560x1440 像素（水平 2560 像素）。
- 焦距：8 mm。
- 感光元件宽度：8 mm。
- 观测距离：5 米（5000 mm）。
- 目标物：直径 200 mm 的圆盘。
- 投影大小：圆盘在影像中约占 110.46 像素，像素精度约 1.81 mm/像素。
- 四点误差：边界框四角点在像素空间的最大偏差，设为 ±1 像素。

8.3. 朝向精度

- 单边 1 像素误差对应水平偏移约 1.953 mm。
- 相对距离 5000 mm，单边角度误差约 arctan(1.9535000)=0.0224°
- 考虑四点误差的总倾斜影响，最大朝向误差约为 ±0.0448°

8.4. 距离精度

- 圆盘直径像素数误差 ±2 像素（四点误差影响）。
- 原始直径 110.46 像素，误差后范围为 108.46 至 112.46 像素。
- 透视关系下，距离误差 Z = 5000(2/110.46) 90.54mm
- 单向误差约为 ±45.27mm

9. 效果展示

我们发现，纯 blender 数据集训练具有一定的泛化性

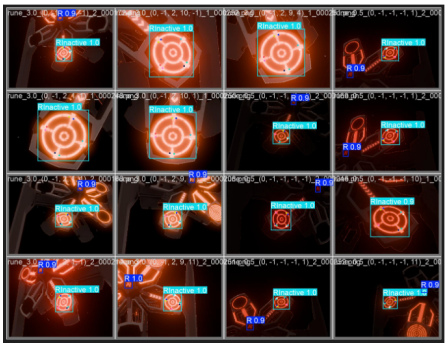


Figure 14. prediction set

9.1.

真实对比：

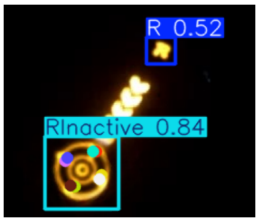


Figure 15. prediction1

使用实验室的自制能量机关进行测试，效果如下

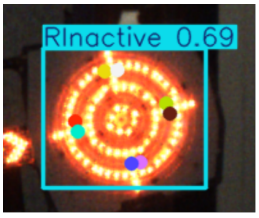


Figure 16. prediction2

在只使用 blender 数据集的情况下，模型对真实的能量机关数据集 Figure 15和 Figure 16仍然具有一定的识别能力。因此，在真实数据集不足的情况下，加入 blender 的数据集，能够更好的加强模型的识别能力。

10. 结论

通过 AI 消除外框和 Blender 渲染，我们成功生成了适应今年无外框符的识别与关键点数据集。这套方案完全零成本（0 元 vs. 商家方案 2475 元 +），数据量极大（2 万 + 排列組合），为缺乏能量机关本体和官方演示视频的队伍提供了一个可行思路。
这份报告旨在帮助更多团队在无能量机关本体，且官方不提供实拍视频资料的情况下，依然实现能量机关检测和姿态估计。希望本报告能助力社区应对新规则挑战，共同推动 RoboMaster 技术进步！

11. 致谢与展望

感谢计算机视觉领域的开源社区，特别是 YOLO 项目团队和 CNN 技术的先驱者们，他们的工作为本方案的深度学习模型提供了核心支持。同时，致敬传统 CV 开源项目 [5] 的贡献者，这些技术为早期的视觉识别奠定了基础，也启发了我们的方案设计。感谢 ENTERPRIZE 等队伍的开源项目（如超级电容方案），为我们的开源文档撰写提供了宝贵的思路参考。特别致谢天津大学等友队分享的标注数据，以及本队张天瑞提供了更加合理的损失函数设计思路。郭子霖、李高旸、蒋益诚等人的努力，甘卓恒开设项目，以及机械负责人黄康源、舒特为提供实物支持。感谢 RoboMaster 组委会提供锻炼机会，让我们在资源有限的情况下学会用 Blender 自制渲染数据，推动了技术的进步。

因比赛尚未进行，当前成果仅为思路参考和初步验证，未来将根据场地实际情况优化数据生成方案和模型性能。若组委会后续能够提供更多有关实物的视频或图片资料，我们将能够更加科学合理地验证本方案的可行性与最终效果，模型的误差有望进一步降低，识别和关键点检测的鲁棒性也将提升。

祝所有参赛队伍顺利激活能量机关，共创佳绩！

References

[1] Anonymous. LCD Particle Liquid Crystal Screen Shader Node. Online Video, 2023. Available: https://www.bilibili.com/video/BV1EM4y127UQ/ [Accessed: Apr. 16, 2025]. 5

[2] Anonymous. RM Traditional Vision Energy Agency Tracker + Recognition. Online Video, 2023. Available: https://www.bilibili.com/video/BV1Ug4y1P7bS/ [Accessed: Apr. 16, 2025]. 1, 2

[3] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Reed, Jianfeng Wang, Alexander Genkin, and Ross Girshick. Segment Anything. In *Int. Conf. Comput. Vis.*, pages 4015–4026, 2023. Available: https://arxiv.org/abs/2304.02643 [Accessed: Apr. 16, 2025]. 4

[4] Sanster. IOPaint. GitHub Repository, 2023. Available: https://github.com/Sanster/IOPaint [Accessed: Apr. 16, 2025]. 1, 3

[5] zRzRzRzRzRzRzR. YOLO-of-RoboMaster-Keypoints-Detection-2023. GitHub Repository, 2023. Available: https://github.com/zRzRzRzRzRzRzR/YOLO-of-RoboMaster-Keypoints-Detection-2023 [Accessed: Apr. 16, 2025]. 9