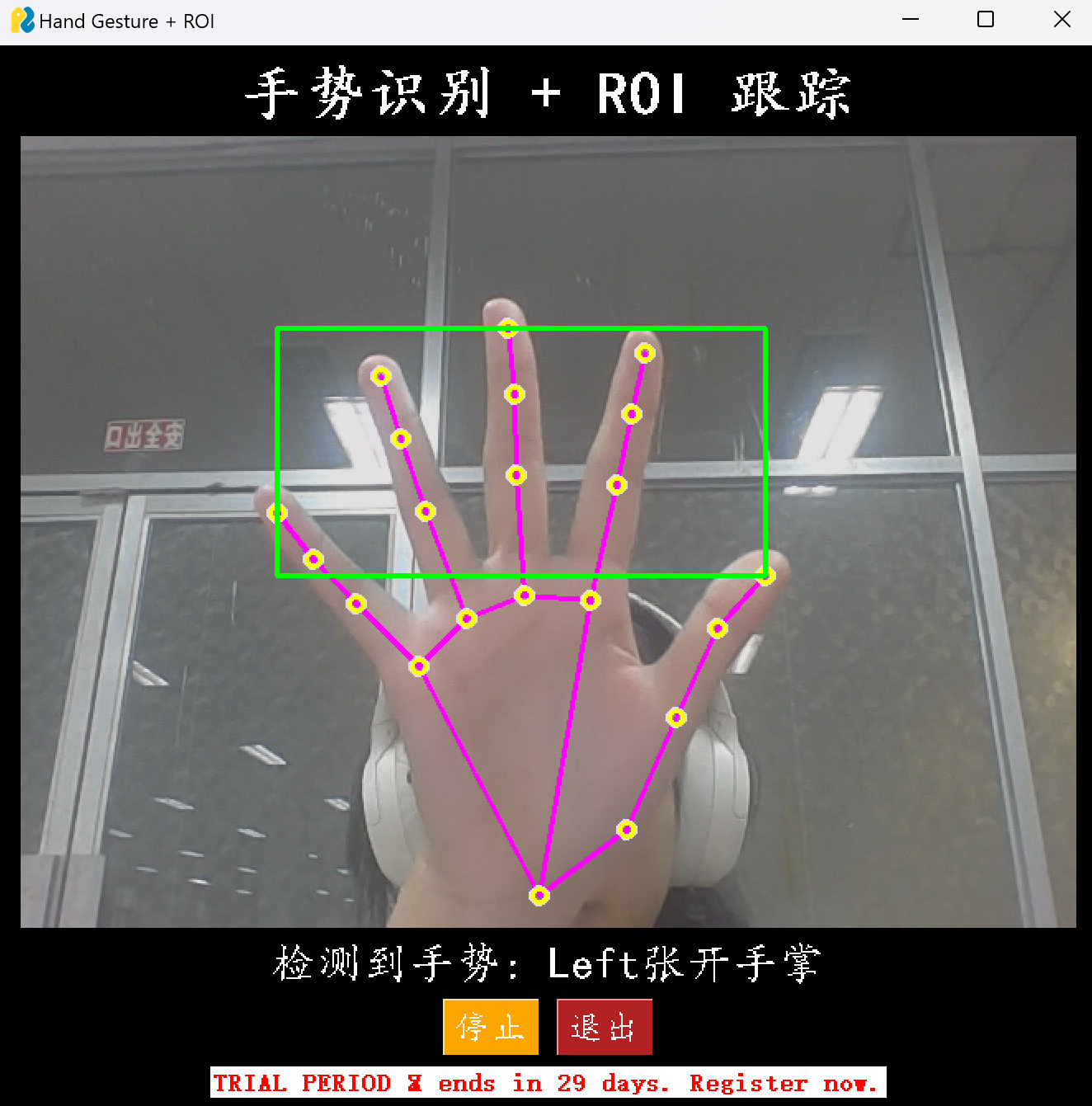
# 手势识别与 ROI 跟踪系统小报告

## 1. 项目背景

本实验实现了一个实时手势识别与 ROI 跟踪系统，通过摄像头捕捉手部动作，识别常用手势，并提取手部区域保存图像，用于后续分析。系统使用 Python、MediaPipe Hands、OpenCV 和 PySimpleGUI 实现跨平台运行。

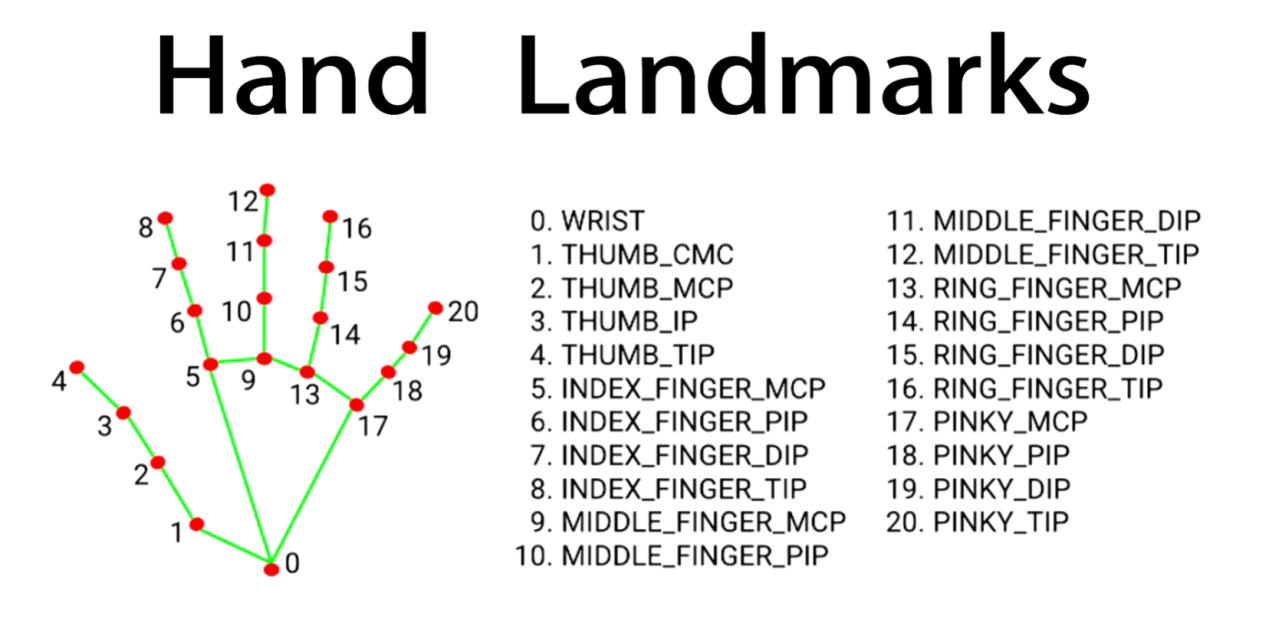
## **2. 系统方法与架构**

整体流程：摄像头视频流 -> MediaPipe Hands -> 手部关键点提取 -> 手势识别 -> ROI 提取 -> GUI显示



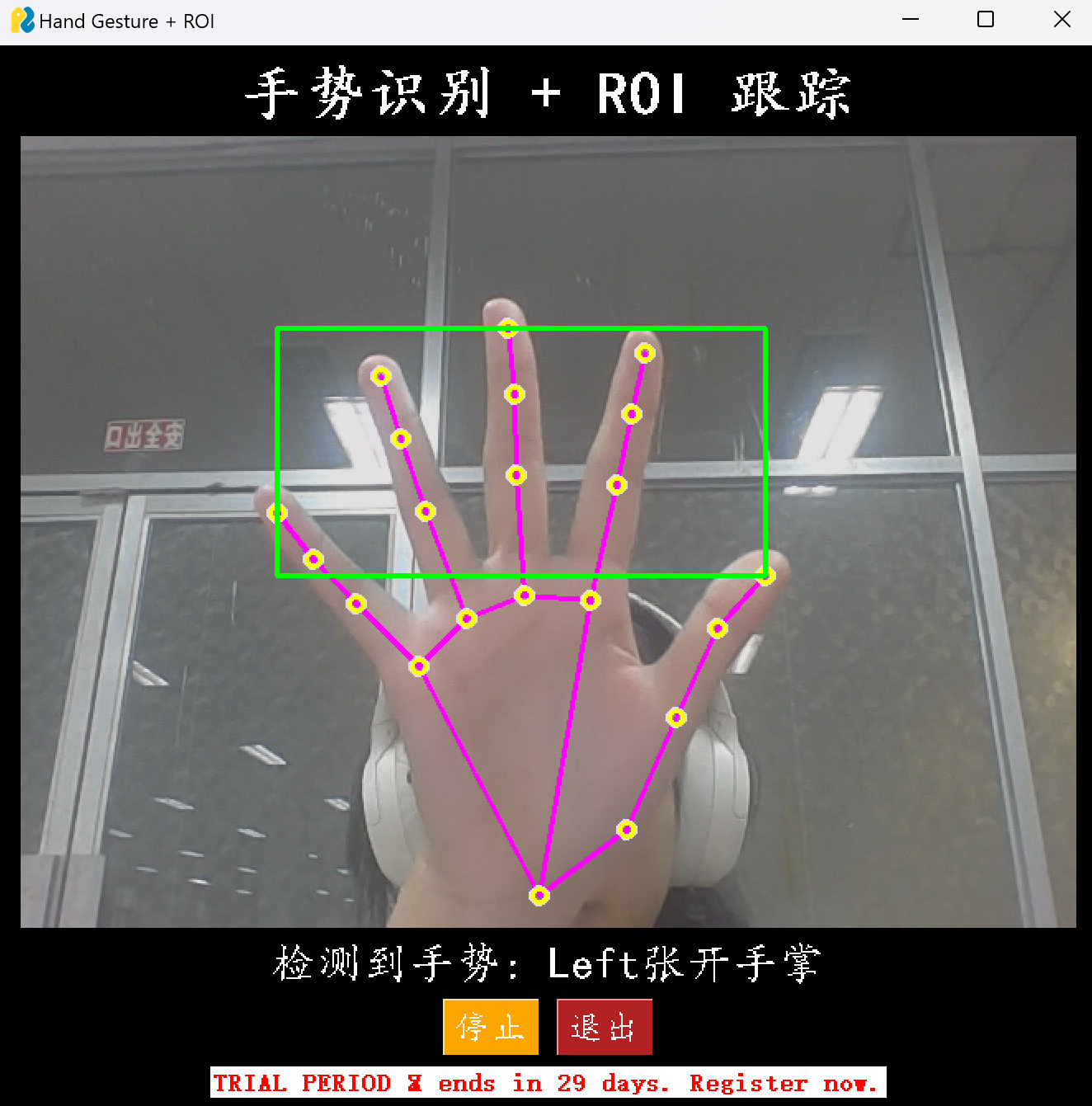
## **2.1 手势检测与关键点提取**

使用 MediaPipe Hands 预测手部 21 个关键点 (x, y, z)，x、y 为归一化坐标，z 为相对深度。Detector 阶段使用轻量卷积神经网络快速检测手部区域，输出 bounding box。Landmark 阶段在 ROI 内通过热力图回归预测关键点位置，z 坐标通过卷积回归获得。



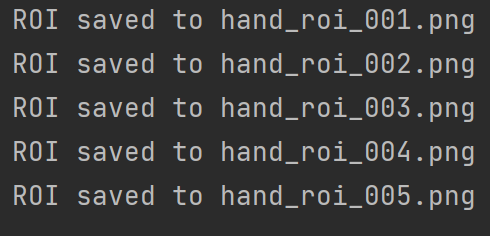
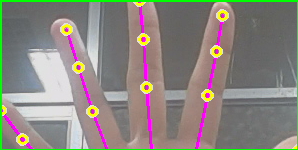
## **2.2 手势识别逻辑**

根据关键点判断手指伸直/弯曲状态。拇指根据左右手镜像逻辑判断，其他四指通过 TIP 和 PIP y 坐标判断。支持握拳、张开手掌、食指指向、胜利手势、竖大拇指等。



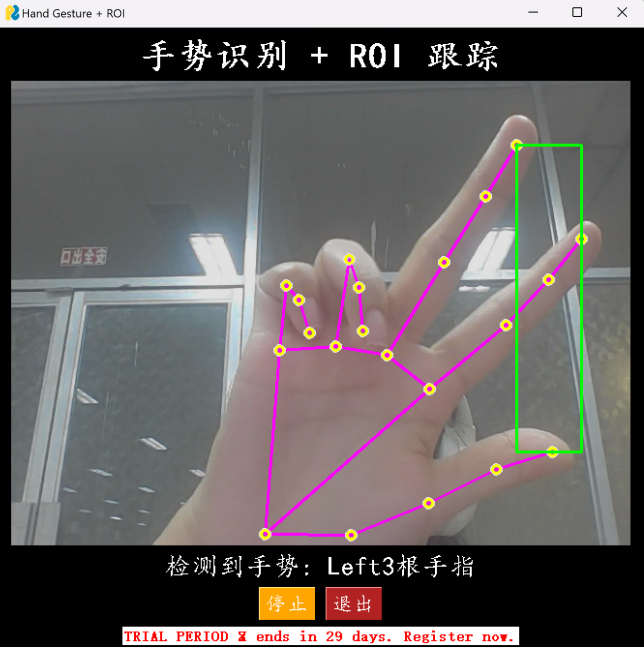
## **2.3 ROI 跟踪与图像保存**

提取伸直手指尖坐标构成 ROI，使用 IOU 跟踪器连续跟踪 ROI，多帧稳定后保存图像，避免抖动或快速移动造成的重复保存。



## **2.4 GUI 与实时显示**

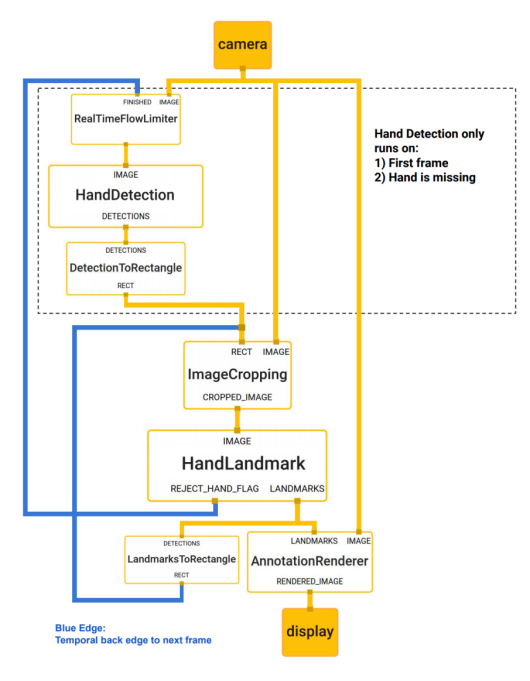
使用 PySimpleGUI 构建界面，显示视频流和实时手势识别结果，提供“开始/停止”和“退出”控制。



## **3. MediaPipe Hands 算法原理**

手部检测使用轻量卷积神经网络提取图像特征，快速回归手部 bounding box，并支持多手检测。关键点预测在 ROI 内生成 21 个关键点热力图，通过 soft-argmax 或最大值定位坐标，z 坐标通过卷积回归获得。局部预测和特征增强保证实时性和精度，多手预测支持左右手分类。

MediaPipe Hands 手部检测阶段详细原理



1、卷积神经网络特征提取（CNN Feature Extraction）

**手部检测阶段（Detector）**采用轻量级卷积神经网络（CNN）对输入的整张 RGB 图像进行特征提取。网络的前几层卷积（Conv2D）主要负责提取低级特征，如边缘、纹理、颜色变化等，并通过 Batch Normalization + ReLU 激活 或 Swish 激活 进行归一化和非线性映射，以稳定训练并加速收敛。后续卷积层和下采样层（如 Stride Conv、Max Pooling 或 Depthwise Separable Convolution）逐步提取更高层次的手部形状和结构信息，同时保留多尺度特征，以便网络对大手和小手区域都能敏感。CNN 的核心作用是将原始图像的空间信息编码为高维特征图，为后续候选手部位置预测提供丰富语义信息。

2、候选手部 ROI 生成（Region Proposal / Bounding Box Regression）

在 CNN 特征图上，Detector 会生成多个候选手部区域（ROI / bounding box）。可采用 **Anchor-based 或 Anchor-free 方法**：

Anchor-based：在特征图上**预定义**不同尺寸和长宽比的 anchor 框，每个 anchor 框预测手部存在概率和相对于 anchor 的偏移量。

Anchor-free：直接回归每个像素点的手掌中心坐标及宽高。

网络最终输出**每个候选框的中心坐标、宽度、高度**，以及手部存在的置信度。训练时通常采用 Smooth L1 Loss、IoU Loss 或 Focal Loss，以增强在小手或复杂背景下的检测鲁棒性。

3、非极大值抑制（Non-Maximum Suppression, NMS）

为了去除重复候选框，网络会对所有候选框按**置信度**进行排序，依次选择置信度最高的框，并剔除与其 **IoU 超过阈值**的其他框。经过 NMS 后，每只手只保留一个 ROI，对应其置信度最高的检测结果。输出包括每只手的 ROI 坐标及手部置信度 score。

4、ROI 的作用（Region of Interest）

ROI 用于缩小图像处理范围，保证 Landmark Regression 网络 的输入仅包含手部区域，减少背景干扰，提高关键点预测的精度和鲁棒性，同时提升计算效率。

5、**热力图回归与 ROI（Keypoint Heatmap Regression）**

尽管 ROI 是由 Detector 回归的 bounding box 得到的，但在关键点预测阶段，**Landmark CNN 会在 ROI 内使用热力图回归（Heatmap Regression）**。具体流程如下：

ROI 图像裁剪并归一化到固定尺寸（如 224×224）

CNN 提取手部关节的局部细节特征

每个关键点生成二维热力图，每个像素值表示该位置为关键点的概率

关键点坐标通过取热力图**最大值**或 **soft-argmax** 获取亚像素精度

深度方向（z 坐标）通过**全局池化卷积层或小型全连接网络回归**

6. Handedness 分类与 Tracking

Handedness：通过特征向量分类每只手是左手还是右手

Tracking：使用上一帧的关键点预测作为先验，通过**平滑**或 Kalman Filter 跟踪手部关键点位置，提高连续帧的稳定性和精度。

7. 总体流程

整体流程如下：

整图输入 → CNN 特征提取 → 候选框生成 → NMS → ROI 输出 → Landmark Heatmap 回归 → 关键点预测 → Handedness 分类与 Tracking

8. 核心特点

轻量化网络，适合 CPU 或移动端实时运行（30~60 FPS）

多尺度特征感知，可同时检测大手和小手

高鲁棒性，在复杂背景和不同手型下仍能准确定位

ROI 提供局部化输入，为关键点预测提供高精度、低干扰信息

连续帧跟踪，保证手势识别的平滑性和稳定性

## **4. 总结**

本系统实现了基于 MediaPipe Hands 的实时手势识别与 ROI 跟踪，通过卷积神经网络提取手部关键点，判断手指状态，实现多种手势识别，并稳定提取手部区域保存图像，界面简洁直观，具有良好实时性和可扩展性。