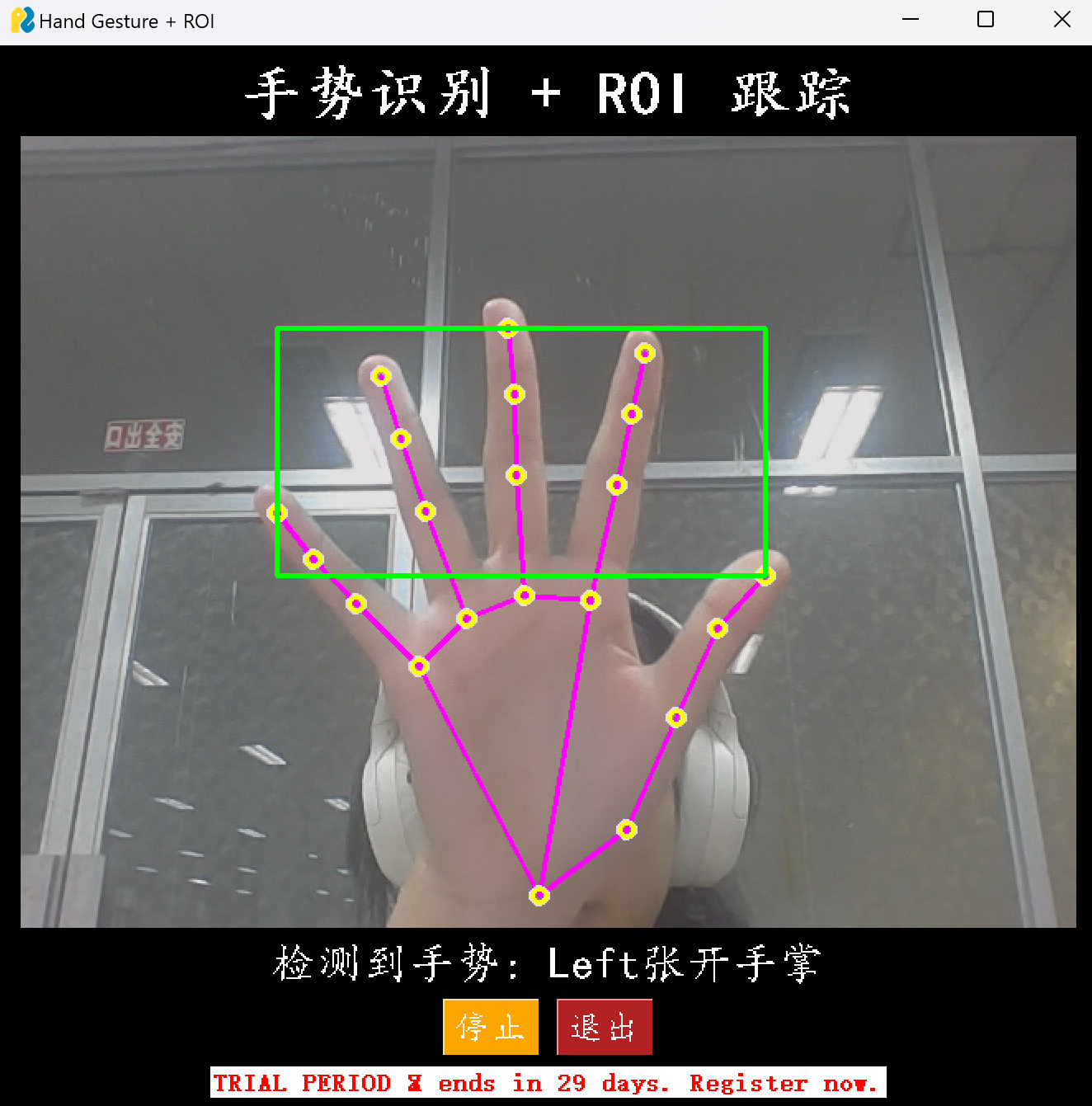
# 手势识别与 ROI 跟踪系统小报告

## 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。

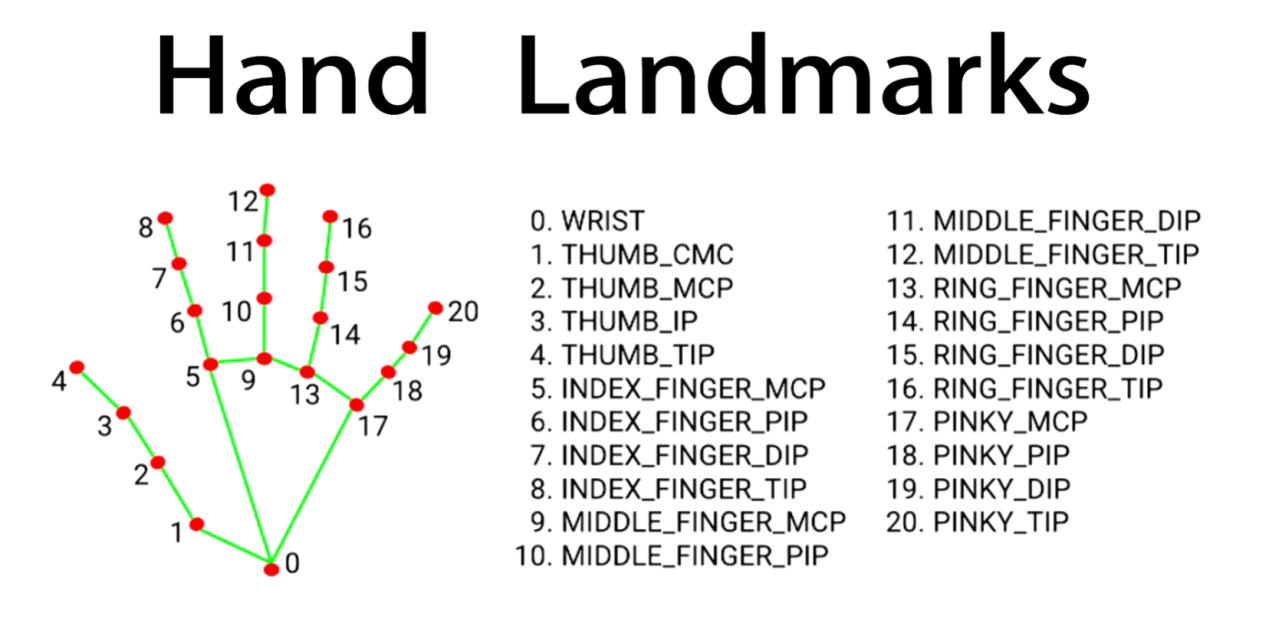
## **2. 系统方法与架构**

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



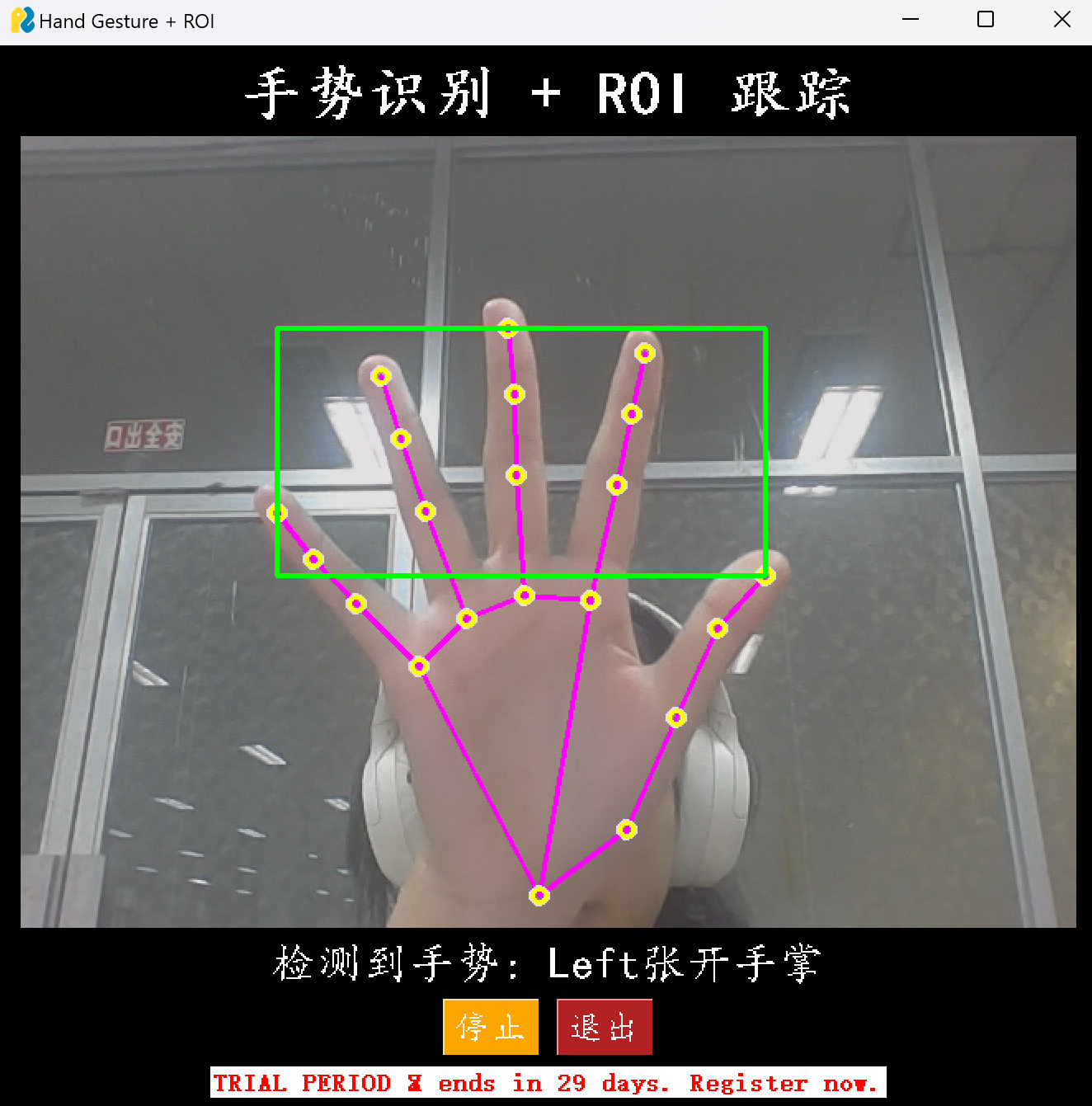
## **2.1 手势检测与关键点提取**

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。



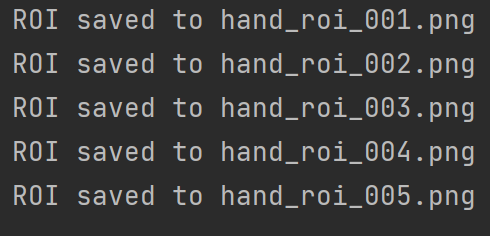
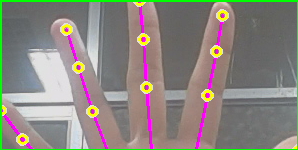
## **2.2 手势识别逻辑**

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。



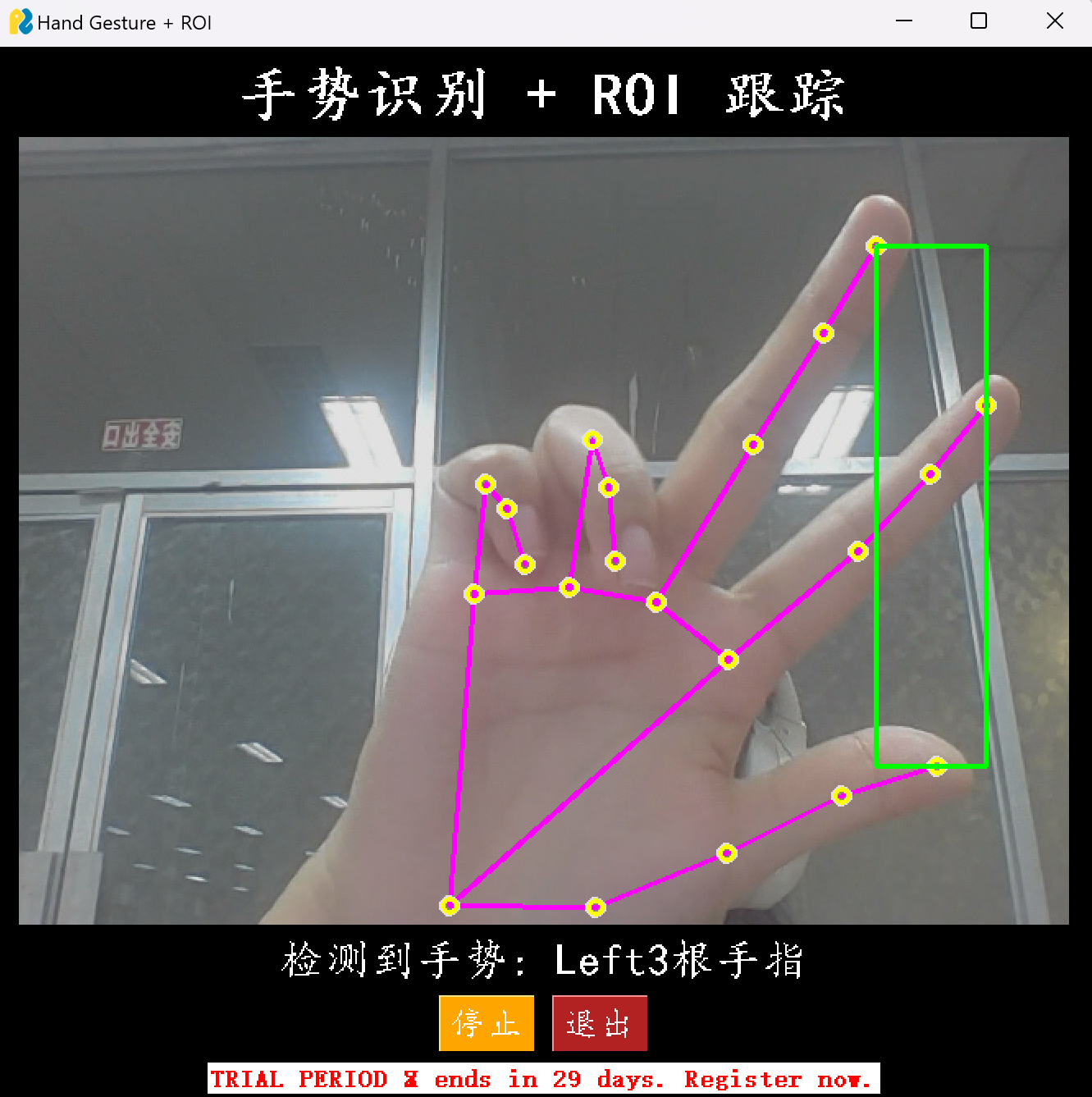
## **2.3 ROI 跟踪与图像保存**

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**

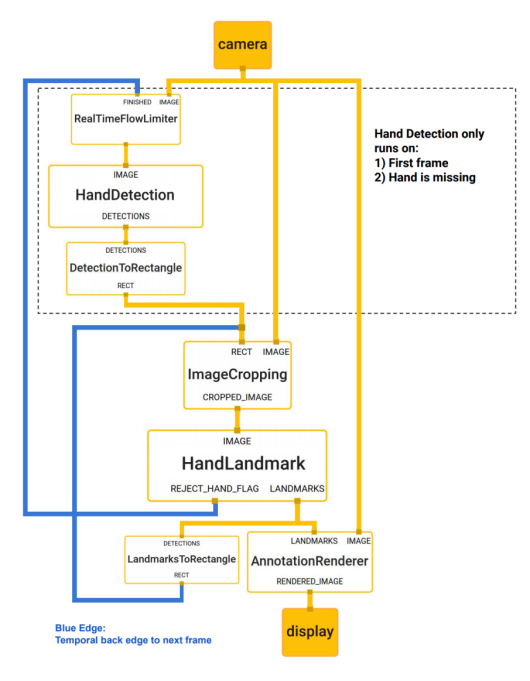
使用 PySimpleGUI 构建界面，显示视频流和实时手势识别结果，提供“开始/停止”和“退出”控制。



## **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

手部检测阶段（Detector）使用轻量级卷积神经网络（CNN）对输入图像进行特征提取。输入为整张 RGB 图像（通常为 256x256 或 320x320 分辨率）。卷积层初步提取低级特征（边缘、纹理、颜色等），后续卷积和池化层提取更抽象的手部形状和结构特征。多尺度特征融合通过不同分辨率的特征图使网络同时感知大手和小手区域。网络使用激活函数（如 ReLU 或 swish）和归一化（BatchNorm 或 LayerNorm）稳定训练和加速收敛。卷积神经网络的核心作用是将图像的空间信息编码为特征图，为候选手部位置预测提供基础。

2、候选手部 ROI 生成

在 CNN 特征图上，Detector 会生成多个候选手部区域（bounding box）。可以使用 anchor-based 或 anchor-free 方法。Anchor-based 方法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偏移量。Anchor-free 方法直接回归特征图每个像素点的中心坐标和宽高。网络输出每个候选框的中心坐标、宽度和高度，同时预测手部存在概率。训练时使用 smooth L1 loss、IoU loss 或 focal loss 提高小手和复杂背景下的检测鲁棒性。

3、非极大值抑制 NMS

为了去掉重复候选框，对所有候选框按置信度排序，依次选择置信度最高的框，并去掉与其 IoU 超过阈值的其他框，最终得到不重叠的 hand ROI，每个 ROI 对应一只手。输出包括每只手的 ROI 坐标和手部置信度 score。

4、ROI 的作用

ROI 缩小图像范围，提高关键点预测效率，保证 Landmark Model 输入的图像只包含手部，减少背景干扰，提高关键点预测精度和鲁棒性。

5、热力图回归与 ROI

虽然 ROI 是由 Detector 回归的 bounding box 得到的，但关键点预测阶段的卷积网络会在 ROI 内使用热力图回归。输入为裁剪并归一化到固定尺寸的 ROI 图像（如 224x224）。CNN 提取手部关节细节特征，每个关键点生成二维热力图，每个像素值表示该像素是关键点的概率。关键点坐标通过取热力图最大值或 soft-argmax 获得亚像素精度。z 坐标通过小型全连接网络或全局池化卷积层回归。

6、总结 Detector 阶段原理

输入整图 → CNN 特征提取 → 候选框生成 → NMS → ROI 输出。ROI 提供局部化输入，为热力图回归阶段的关键点预测提供高精度、低干扰的输入。网络特点包括轻量级、实时性高、多尺度感知、鲁棒性强，可在 CPU 或移动端实现 30~60 FPS。

## **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。