面向智能家居的动态手势识别轻量化系统优化

刘姝怡[1] 徐梦溪[1*] 刘姝悦[1] 刘梓莹[1]
LIU Shuyi XU Mengxi LIU Shuyue LIU Ziying

摘 要 动态手势识别在向复杂家居环境迁移时面临实时性、低功耗与鲁棒性挑战。文章针对端到端模型资源受限设备部署难的问题，提出一种基于“检测 - 跟踪 - 特征提取 - 分类”分阶段流水线策略的动态手势识别方法（DTFC-DGRM），构建全流程协同技术框架，设计轻量化识别系统。系统以单目 RGB 视频为输入，通过 YOLOv5 精准定位多手区域（640×640 分辨率下 mAP@0.5 达 83.6%），利用 MediaPipe 稳定跟踪 21 个手部关键点（轨迹完整率 93.2%），融合 HOG 手工特征、关键点时序特征与 CNN 语义特征后由 SVM 分类优化。实验显示，系统在 Light-HaGRID 及自定义数据集上识别准确率达 94.1%，在树莓派 4B 平台实现 32 帧 /s 处理速度，且功耗 <5 W，验证了方法有效性，为智能家居轻量化交互提供可行方案。

关键词 动态手势识别；智能家居系统；轻量化系统；深度学习；SVM

doi：10.3969/j.issn.1672-9528.2026.08.006

0 引言

物联网及人工智能正促成智能家居的发展，而无接触动态手势交互是优化用户体验的关键手段。要识别动态手势必须利用有关手部的时序信息[1-3]，目前已有方法有其局限：传统手工设计的特征不适用于复杂情形的鲁棒分析；3D-CNN 或 Transformer-LSTM 等端到端方法所占参数太多，低算力设备未必能实时应用[4-6]；现有轻量化方案只处理单一模块，不能满足家居 200~800 lx 光照及杂乱背景的要求。

为此本文提出 DTFC-DGRM 分阶段流水线式的识别方案，按序利用改进 YOLOv5、MediaPipe、简型 VGG16 和 SVM 对手部做检测、对关键点做跟踪、将多特征结合以判别手势。文中给出全流程配合优化的思路，把流水线并行及动态资源分配纳入算力管理，设计家居适用的手势数据库，又借助树莓派 4B 对各方面加以试验分析，实践证明该轻量级方法在识别质量、执行帧数及能耗上都较合理，具有作为家居嵌入式装置应用的实际价值。

1 相关工作

单阶段端到端模型（如 3D-CNN、Transformer-LSTM）通过单一网络直接完成检测、特征提取与分类，流程简洁且精度较高，展现了深度时序建模的潜力[7]。但其参数量大、推理延迟高，难以在树莓派 4B 等边缘设备实时运行，且对 200~800 lx 光照波动和遮挡场景适应性差，难以平衡精度与轻量化需求[8]。

双阶段方案将检测与分类解耦，可灵活组合模型，降低背景干扰，工程应用广泛。但现有双阶段方法大多将检测与分类视为串行孤立模块，未充分考虑后续跟踪与特征提取的全链路协同，串行计算带来额外开销，且缺乏对家居光照和复杂背景的针对性设计，整体效率有较大提升空间。

面向智能家居嵌入式部署需求，轻量化动态手势识别方法旨在降低模型参数量与计算开销，适配树莓派 4B 等资源受限设备[9-10]。现有轻量化手段围绕网络设计、模型压缩及硬件优化展开，取得了一定进展。然而，多数方法仅聚焦单一模块优化，缺乏全流程协同设计；部分模型精度损失明显，且未充分结合 200~800 lx 家居光照特点进行鲁棒性优化；多数方案缺少嵌入式环境下的功耗与帧率实测，工程实用性不足。在传感层面，单目 RGB 成本低但易受光照和背景干扰，深度与多模态方案精度高却增加系统复杂度，不利于轻量化部署。尽管家居手势识别研究已呈现场景化适配趋势，但现有工作在多模块协同、特征融合与动态资源调度方面仍缺乏系统性验证，未能很好解决轻量化、实时性与鲁棒性的平衡问题，这正是本文研究的核心出发点。

2 DTFC-DGRM 方法与技术框架

针对智能家居所涉及的某些具体应用中工程限制及边缘设备资源不足的情况，本文提出一个以“检测 - 跟踪 - 特征提取 - 分类”为分阶段流水线思路的动态手势识别方案（即 DTFC-DGRM）来处理任务，把各部分作模块化安排，在时间、

1. 南京工程学院 江苏南京 211167
[基金项目] 江苏省大学生创新创业训练计划项目（202511276125Y）

空间、资源上同时做整体协调，用此办法克服传统流水线不连贯、资源空转、对干扰不敏感等缺陷。

2.1 DTFC-DGRM 整体架构设计

DTFC-DGRM 的整体架构如图 1 所示，所采用的识别方法是按手势检测、手部关键点跟踪、特征分析及 SVM 分类这四个环节来串行处理的，各模块接口标准化，可分别加以改进，不依赖于端到端大模型（其延迟、能耗都较高）。同时搭建“时间 - 空间 - 资源”三者协同优化的理论模型：

（1）时间维度：用流水线方式安排并行调度，多帧数据分时并行处理，消除模块空闲等待，降低整体时延；（2）空间维度：融合手工几何特征、关键点时序特征、CNN 语义特征，依靠特征互补提升手势区分度；（3）资源维度：负载感知动态算力调度，根据设备实时负载分配 CPU、内存资源。

框架目标函数为 min Cost(R, T) s.t. ACC ≥ 94%，其中 Cost(R, T) 为功耗、延迟综合代价，在满足精度阈值前提下最小化系统开销。

RGB 视频流依次输入 YOLOv5 检测手部边界框，裁剪手部区域送入 MediaPipe 提取 21 个关键点时序坐标，融合手工几何特征送入轻量化 CNN 编码为 512 维特征向量，最终由 SVM 输出手势类别。系统层引入流水线并行、动态资源调度，实现多帧多模块同步处理，最终在树莓派 4B 实现 32 帧 /s、功耗＜ 5 W 的运行指标。

2.2 分阶段流水线

手势检测阶段。选用改进 YOLOv5-Conv 作为手部检测模型，一阶检测架构兼顾精度与推理速度，适配嵌入式算力。输入固定图像尺寸 640×640，在自建家居数据集微调后 mAP@0.5 达到 83.6%，可精准定位画面内多只手区域，裁剪手部 ROI 消除全图背景冗余计算，相比 NanoDet 等轻量检测器更适配家居光照波动场景。该模块单帧推理耗时 10.4 ms，整体占系统总耗时 35%。

手部跟踪与关键点定位阶段。采用 MediaPipe Hands 框架，在 YOLO 提供的裁剪手区基础上作实时分析，能给出 21 个三维手关键点随时间变化的坐标（轨迹完整率 93.2%），每帧处理延迟少于 10 ms，同轻量型 OpenPose 相比，这里只对局部图像做输入处理，计算负担大为减轻，又可准确地追踪手的连续动作，得到时序信息。模块单帧耗时 7.4 ms，占总耗时 25%。

特征提取阶段。采用手工几何特征 + 轻量化裁剪 VGG16 融合方案：先基于关键点时序坐标计算指关节角度、手指间距、运动速度等手工特征；再将手工特征构造特征图输入轻量化 VGG16 网络。原 VGG16 参数量 138×10^6，本文删减冗余卷积、全连接层，仅保留三层卷积、两层全连接层，参数量压缩至 8.2×10^6，最终输出 512 维语义特征向量。

手工特征可解释性强、计算开销低，但易受环境干扰；关键点时序特征抗背景干扰，但特征维度低；CNN 深层语义特征抽象判别能力强。三类特征融合后特征区分度显著提升，方差分析 F 值由单一特征 ≤ 3.2 提升至 8.6，验证了融合的有效性。该模块单帧耗时 5.9 ms，占总耗时 20%。

手势分类阶段。采用 SVM 作为分类器，适配家居有限手势类别的小样本场景，泛化能力强；相比 RNN、CNN 深度分类器推理开销极低，单帧分类仅 3.2 ms。多项式核 SVM 无需迭代时序状态，内存占用小，不会大幅增加系统功耗，模块耗时占总流程 20%。

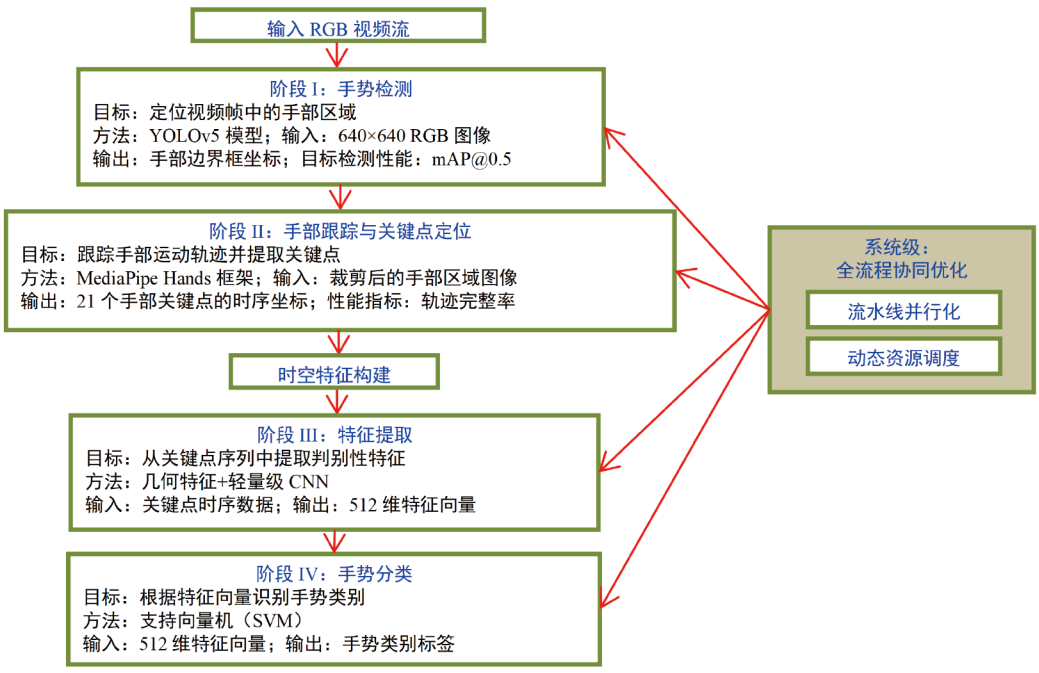


图 1 DTFC-DGRM 整体架构

2.3 全流程协同优化

2.3.1 流水线并行化

将四大模块构建多级流水线，不同帧数据分配至不同模块同步处理，依靠时间戳对齐保障时序一致性，时间戳偏差超过阈值时采用线性插值补全关键点数据，模块间最大时序误差控制在 5 ms 内。

定义串行总耗时为：

$$T_{\text{parallel}} = \max(t_1, t_2, t_3, t_4) + \Delta t \quad (1)$$

式中：Δt 为模块间的同步时间开销，本研究中 $\Delta t \leq 5$ ms。

系统效率提升率 η 的计算公式为：

$$\eta = \frac{T_{\text{serial}} - T_{\text{parallel}}}{T_{\text{serial}}} \times 100\% \quad (2)$$

代入本研究的实际实验数据，t_1=10.4 ms，t_2=7.4 ms，t_3=5.9 ms，t_4=5.9 ms，Δt=5 ms，计算可得 T_{serial}=29.6 ms，T_{parallel} = 15.4 ms，效率提升率 η 达 48%，验证了流水线并行化的优化效果，为系统实现 32 帧 /s 的高处理速度提供了核心支撑。

2.3.2 动态资源调度机制

实时采集 CPU 利用率 U、单帧推理耗时 T、视频帧率 F 三项负载指标，采样周期 50 ms，取三次均值判定负载等级：（1）低负载时 $U < 50\%$、$T < 25$ ms、$F \geq 30$ 帧 /s，系统空闲，CPU 降频至 1.0 GHz、间隔跳帧降低功耗；（2）中负载时 $50\% \leq U < 80\%$、$25\text{ ms} \leq T < 35$ ms、$20\text{ 帧 /s} \leq F < 30$ 帧 /s，常规手势交互，为检测、跟踪模块分配独立 CPU 核心；（3）高负载时 $U \geq 80\%$、$T \geq 35$ ms、$F < 20$ 帧 /s，连续快速手势，分辨率降至 416×416、实时回收空闲内存。

调度遵循“检测、跟踪优先级最高”原则，形成采集 - 判定 - 调度 - 反馈闭环，与主识别流程并行运行，无额外延迟，平衡系统实时性与整机功耗。

3 系统实现与实验设置

3.1 软硬件部署环境

软件开发基于 Python 实现，核心工具链：PyTorch 训练 YOLOv5 与轻量化 VGG16；OpenCV 处理视频流与图像预处理；MediaPipe 提取手部关键点；Scikit-learn 搭建 SVM 分类器；TensorRT 完成模型 16 位量化加速，量化后 YOLOv5 推理耗时由 45 ms 降至 28 ms，轻量化 VGG16 由 32 ms 降至 19 ms。

硬件平台选用树莓派 4B（四核 CortexA72、4 GB LPDDR4 内存），模拟智能家居中控终端，支持多线程并行、动态 CPU 调频，满足多模块协同运行的算力与功耗测试需求。

系统集成流程：首先 YOLOv5 输出手部 ROI 坐标，其次将裁剪图像送入 MediaPipe 提取 21 个关键点并编码为 128 维时序特征，同时提取 3 780 维 HOG 手工几何特征，并将手部 ROI 输入轻量化 VGG16 提取 512 维 CNN 语义特征；将上述四类特征拼接为 4 420 维联合特征向量，经 PCA 降维（保留 95% 方差信息）与 ***Z***-Score 标准化后，输入 SVM 输出手势类别；多模块分配独立线程并行流水线执行，时间戳对齐保证时序一致性。

3.2 实验数据集

数据集由公开数据集与自建家居手势库组合而成，统一按照 7:1.5:1.5 划分训练、验证、测试集，每类手势样本量差值控制在 15% 以内，配套随机旋转、亮度调整、高斯噪声、局部遮挡多层数据增强。

公开数据集：LightHaGRID 静态手势数据集 [11]、Jester 动态手势数据集 [12]，用于模型预训练，提升通用特征提取能力；

自建家居手势库：分为静态、动态两类，覆盖开灯、调亮、中档亮度、调暗、关灯五类数字手势 1~5，采集视角包含正面、俯视、45° 侧视，背景纹理系数 0.3~0.8；静态手势单类采集 10 帧图像，动态手势单类录制 30 帧 /s、3 s 视频（90 帧图像）；总样本 2 500 张，静态 250 张、动态 2 250 张，Labelme 标注后转换为 VOC 标准格式。自定义智能家居动态手势库的典型图像如图 2 所示。

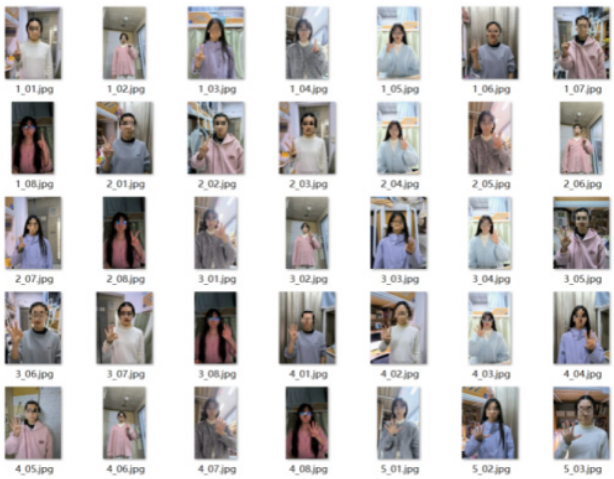


图 2 自定义智能家居动态手势库的典型图像

3.3 评估指标与对比基线

性能评估指标为识别准确率 ACC、处理帧率 FPS、整机功耗 W、手部检测 mAP@0.5、关键点轨迹完整率、端到端响应延迟、误检率。对比基线模型包括：传统单算法 KNN/SVM 模型 [13-15]、双层轻量化 CNN[16]、轻量化端到端模型 - 改进 YOLOv5-Conv[17-18] 一体化手势识别网络。

4 实验结果与分析

本研究构建的轻量化动态手势识别系统基于 DT-FC-DGRM 完成精度测试，统一采用树莓派 4B、200~800 lx 光照、五类控制手势开展实验。系统在 Light-HaGRID 公开数据集与自定义智能家居手势库的综合测试中，识别准确率达 94.1%。

4.1 识别精度分析

Light-HaGRID 数据集识别准确率 95.3%，手指区分度高的 1、2、3 类手势准确率超 97%，4、5 类因手指间距小、姿态易偏移，准确率分别为 93.2%、92.8%；自建家居数据集

常规场景准确率 96.7%，复杂纹理背景（0.8 系数）下仍保持 93.5%。测试集总样本 1 090 份，正确识别 1 051 份，综合准确率 94.1%，混淆矩阵（如图 3 所示）可见各类手势误判样本极少，仅 2、3，4、5 存在少量混淆，对角线置信度接近 1，分类边界清晰。

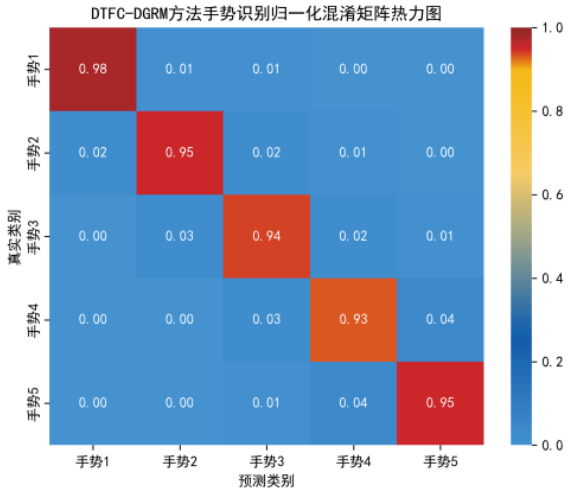


图 3 DTFC-DGRM 方法手势识别归一化混淆矩阵热力图

不同模型在核心测试任务中的性能对比如表 1 所示。DTFC-DGRM 识别准确率相较 KNN/SVM 提升 21.1%，相较浅层 CNN 提升 11.1%，相较端到端 YOLOv5-Conv 提升 1.8%，精确率、召回率与 F_1 得分均衡稳定。KNN/SVM 依托 HOG 手工特征，无法提取时序运动信息，相似手势混淆问题突出，光照波动会造成 12.5% 的精度浮动；浅层 CNN 不具备时序建模能力，手部快速运动易造成特征模糊，多视角样本泛化效果较差；改进 YOLOv5-Conv 依靠像素特征完成识别，难以分辨细微手势差异，低光场景精度衰减明显，且端到端串行推理存在算力冗余问题。

表 1 不同模型在核心测试任务中的性能对比

单位：%

模型	准确率	精确率	召回率	F_1- score
KNN/SVM	73.0	71.0	67.0	69.0
CNN	83.0	85.0	83.0	83.0
改进 YOLOv5-Conv	92.3	91.8	90.5	91.1
DTFC-DGRM	94.1	93.8	93.5	93.6

4.2 系统实时性与功耗分析

4.2.1 帧率与延迟

处理速度与延迟对比如表 2 所示，DTFC-DGRM 帧率 32 FPS，较端到端模型提升 28%，端到端延迟仅 31.2 ms，远低于 100 ms 家居交互舒适阈值。核心提升来自流水线并行多线程调度与 TensorRT 模型量化，硬件资源利用率提升 42%。

表 2 不同模型处理速度与延迟对比

模 型	平均处理速度 /(帧·s^{-1})	端到端延迟 /ms	单帧推理耗时 /ms
KNN/SVM	18	55.6	48.2
CNN	22	45.5	42.8
改进 YOLOv5-Conv	25	40	38.6
DTFC-DGRM	32	31.2	29.7

4.2.2 功耗测试

功耗与能耗对比如表 3 所示。DTFC-DGRM 整机功耗 4.7 W，满足＜ 5 W 轻量化指标，功耗性能比 0.15 为四组最优，同等算力下可处理更多视频帧；满载温度 46.8℃，避免高温降频，5 000 mAh/5V 电池可连续运行 5.3 h，续航优于改进 YOLOv5-Conv 方案。

表 3 不同模型功耗与能耗对比

模 型	平均功耗 /W	满负载发热温度 /℃	功耗 / 性能比 /(W·帧$^{-1}$·s)
KNN/SVM	3.8	42.5	0.21
CNN	4.2	45.3	0.19
改进 YOLOv5-Conv	5.8	51.2	0.23
DTFC-DGRM	4.7	46.8	0.15

4.3 消融研究

为验证 DTFC-DGRM 核心组件的必要性，本研究设计消融实验，以完整模型为基准，在统一测试环境下从准确率、速度与功耗三个维度量化各组件贡献。

分类器选择方面：将 SVM 替换为含 2 层 GRU 单元的 RNN 后，识别准确率从 94.1% 略降至 93.7%，处理速度由 32 帧 /s 降至 24 帧 /s，功耗由 4.7 W 升至 5.3 W，单帧推理耗时由 3.2 ms 增至 8.5 ms。RNN 对 four/five 相似手势的误判率为 5.1%，高于 SVM 的 4.8%。SVM 在精度损失可忽略的前提下效率与功耗更优。

特征类型方面，将 640 维关键点特征替换为 4096 维原始像素特征后，识别准确率由 94.1% 降至 88.2%，复杂背景与低光场景准确率分别下降 11.2 和 10.8 个百分点，CNN 特征提取耗时增加 47.4%，处理速度降至 26 帧 /s。关键点特征过滤背景干扰能力更强，是系统兼顾精度与效率的核心设计。

若将流水线并行及动态资源调度都予以关闭，处理速度就从 32 帧 /s 减少为 20 帧 /s，端到端延迟从 31.2 ms 增至 50 ms，CPU 利用率也由 89% 降到 52%。单靠串行方式功耗只减少 0.2 W，而连续手势的响应时间变为 1.8 s（原为 1.1 s）。流水线并行能防止各模块无用等待，动态调度又可让 CPU 核不发生冲突，二者的结合正是实时交互所必需的基础。有关性能比较如表 4 所示。

表 4 协同优化策略性能对比

执行模式	处理速度 /(帧 ·s^{-1})	端到端延迟 /ms	连续 1 000 帧处理耗时 /s	平均功耗 /W	CPU 资源利用率 /%	动态手势响应时间 /s
并行调度	32	31.2	31.2	4.7	89	1.1
串行执行	20	50	50	4.5	52	1.8

三组消融实验已就各基本组件做必要性的分析：SVM 能在几乎不损失精度时达到最小的推理开销及功耗；有关关键点的特征比原像素多获 5.9% 识别正确率，而所占维度仅是 6.4 倍的量级；协同优化把速度提高 37.5%、CPU 使用比例由 52% 变为 89%。三者协同构成 DTFC-DGRM 的技术优势。

4.4 家居环境鲁棒性分析

光照适应性（200~800 lx）：DTFC-DGRM 的准确率在 200 lx 微光时 90.6%，在 800 lx 强光时 92.8%，全光照变化只导致最大衰减 3.5%；对 YOLOv5-Conv 类方法来说低光衰减有 10.5%，CNN 及 KNN/SVM 分别衰减 9.3%、6.2%。按关键点几何思路处理亮度问题，此方案在光照上比基线具有明显鲁棒性。

复杂背景抗干扰：有关背景纹理的系数是 0.8，在高复杂度场合中 DTFC-DGRM 的准确率仍达 93.5%，只比简单背景低 2.4%；手工方法（如 KNN 或 SVM）遇到纹理即失效，所获准确率约 57.9%，用端到端 YOLOv5-Conv 时误检率上升 6.3%。若对检测所得手部 ROI 作裁剪、按关键点剔除背景多余成分，复杂家居识别会更稳定。

静态手势的准确率是 94.1%，慢速动态 93.8%，快速动态 92.3%，对连续多步动态作识别 91.2%，最大性能损失 4.1%；改进 YOLOv5-Conv 连续动态召回率衰减 10.8%，而 CNN 或传统法不考虑时序，连续识别效率大为降低。MediaPipe 按时序追踪关键点，对快、连续动作有良好辨识

5 结论

本文针对智能家居边缘设备算力受限、复杂环境识别鲁棒性不足的问题，提出 DTFC-DGRM 分阶段轻量化手势识别流水线，将任务拆分为检测、跟踪、特征提取、分类四大模块，搭配流水线并行、动态资源调度实现全局协同优化。树莓派 4B 实测结果显示，系统识别准确率 94.1%，运行帧率 32 帧 /s，整机功耗低于 5 W；相较于 KNN/SVM、浅层 CNN、端到端 YOLOv5-Conv，在识别精度、实时性与能效上优势显著，适用于智能家居无接触手势交互落地。

本文创新分为三点：搭建时间、空间、资源三维协同优化模型，统筹分配各模块算力；设计负载感知闭环调度机制，适配嵌入式设备动态负载；依托特征互补性融合几何、关键点时序、CNN 语义特征，增强光照与复杂背景下识别稳定性，为边缘轻量化视觉识别提供工程参考。

研究尚存不足：手部重度遮挡、极速运动时易出现关键点跟踪丢失；系统仅适配预设手势，拓展新手势需重新训练分类器；不同用户手部尺寸会造成识别泛化波动。后续将改进时序跟踪算法，提升极端场景稳定性；引入增量在线学习，无需重训即可适配新手势；采用硬件感知神经网络搜索自动压缩模型，进一步均衡识别精度与推理开销。

参考文献：

[1] Pisharady P K, Saerbeck M. Recent methods and databases in vision-based hand gesture recognition: a review[J]. Computer Vision and Image Understanding, 2015, 141: 152-165.

[2] Rahman M M, Uzzaman A, Khatun F, et al. A comparative study of advanced technologies and methods in hand gesture analysis and recognition systems[J/OL]. Expert Systems with Applications, 2025[2026-07-29].https://www.sciencedirect.com/science/article/pii/S0957417424027969.

[3] 付智凯 , 李文新 , 罗新奎 . 基于视觉的动态手势识别技术综述 [J]. 计算机测量与控制 ,2025,33(1):9-19.

[4] Hashi A O, Hashim S Z M, Asamah A B. A systematic review of hand gesture recognition: an update from 2018 to 2024[J]. IEEE Access, 2024, 12: 143599-143626.

[5] Qi Jing, Ma Li, Cui Zhenchao, et al. Computer vision-based hand gesture recognition for human-robot interaction: a review[J]. Complex & Intelligent Systems, 2024, 10: 1581-1606.

[6] Emporio M, Ghasemaghaei A, Laviola J J Jr, et al. Continuous hand gesture recognition: benchmarks and methods[J/OL]. Computer Vision and Image Understanding, 2025[2026-07-29].https://www.sciencedirect.com/science/article/pii/S1077314225001584.

[7] Zhang Wenjin, Wang Jiacun, Lan Fangping. Dynamic hand gesture recognition based on short-term sampling neural networks[J]. IEEE/CAA Journal of Automatica Sinica, 2020, 8(1): 110-120.

[8] Cui Cui, Sunar M S, Goh E S. Deep vision-based real-time hand gesture recognition: a review[J/OL]. Peer J Computer Science,2025[2026-07-29].https://www.semanticscholar.org/paper/Deep-vision-based-real-time-hand-gesture-a-review-Cui-Sunar/69af07de6dc6e678c30494324a5263eb4595d0ab.

[9] Qi Wen, Ovur S E, Li Zhijun, et al. Multi-sensor guided hand gesture recognition for a teleoperated robot using a recurrent neural network[J]. IEEE Robotics and Automation Letters, 2021, 6(3): 6039-6045.

[10] 薛阳 , 李清 , 卢秋红 , 等 . 基于改进 YOLOv5-Conv 手势识别研究 [J]. 信息技术与信息化 , 2024(7): 207-211.

[11] Kapitanov A, Makhliarchuk A, Kvanchiani K. HaGRID–HAnd gesture recognition image dataset[PP/OL]. (2022-06-14) [2026-07-29]. https://arxiv.org/pdf/2206.08219v2.

[12] Materzynska J, Berger G, Bax I,et al. The jester dataset: a large-scale video dataset of human gestures [C/OL]// 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Piscataway: IEEE, 2019[2026-07-29]. https://ieeexplore.ieee.org/document/9022297.DOI: 10.1109/ICCVW.2019.00349.

[13] 张硕 . 基于 KNN 算法的空间手势识别研究与应用 [D]. 长春 : 吉林大学 ,2017.

[14] 陈嘉伟 , 韩晶 , 郝瑞玲 , 等 . 基于改进 KNN 算法的动态手势识别研究 [J]. 中北大学学报 (自然科学版), 2020, 41 (3): 232-237.

[15] Zhou Kai, WanYi, Huang Hailong. Research on recognition and application of hand gesture based on skin color and SVM[J]. Journal of Computational Methods in Sciences and Engineering, 2019, 20(1): 269-278.

[16] 王庆鑫 . 基于轻量级 CNN 的手势识别算法研究 [D]. 哈尔滨 : 哈尔滨理工大学 ,2023.

[17] 王路遥 . 基于改进 YOLOv5s 的静态 / 动态手势识别算法探究 [J]. 信息系统工程 ,2024(12):138-141.

[18] 张常斌 , 孙连山 , 唐景琰 . 基于改进 YOLOv5 的轻量级手势识别算法 [J]. 智能计算机与应用 ,2025,15(6):184-189.

【作者简介】

刘姝怡（2004—），女，江苏苏州人，本科，研究方向：数字媒体技术，email:2133993942@qq.com。

徐梦溪（1983—），通信作者（email: xumx@njit.edu.cn），女，江苏南京人，博士，教授，研究方向：图像处理、智能信息处理、多媒体系统与应用软件。

刘姝悦（2004—），女，江苏苏州人，本科，研究方向：数字媒体技术，email:3137631284@qq.com。

刘梓莹（2005—），女，广东广州人，本科，研究方向：数字媒体技术，email:3375788023@qq.com。

（收稿日期：2026-06-24 修回日期：2026-08-03）

（上接第 27 页）

参考文献：

[1] 张欢 , 刘静 , 冯毅博 , 等 . U-Net 及其在肝脏和肝脏肿瘤分割中的应用综述 [J]. 计算机工程与应用 , 2022, 58(2): 41-52.

[2]Khan S U R, Lee H. Deep learning-based liver tumor segmentation: a scoping review [J]. Convergence Hepatology, 2026, 2 (1): 29-55.

[3]Qi Yingyi, Xiong Wei, Leow K, et al. Semi-automatic segmentation of liver tumors from CT scans using bayesian rulebased 3D region growing[J/OL].The MIDAS Journal, 2008[2026-02-04].https://doi.org/10.54294/rofibw.

[4]Seo H, Huang C, Bassenne M, et al. Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images [J]. IEEE Transactions on Medical Imaging, 2020, 39 (5): 1316-1325.

[5] 钟思华 , 郭兴明 , 郑伊能 . 改进 U-Net 网络的肺结节分割方法 [J]. 计算机工程与应用， 2020, 56 (17): 203-209.

[6]Li Wenkun, Jia Maoning, Yang Chen, et al. SPA-UNet: A liver tumor segmentation network based on fused multi-scale features [J]. Open Life Sciences, 2023,18 (1): 1245-1256.

[7]Ruan Jiacheng, Xie Mingye, Gao Jingsheng, et al. EGE-UNet: an efficient group enhanced UNet for skin lesion segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2023. Cham: Springer, 2023: 481-490.

[8]Wang Fang, Cheng Xueli, Luo Ningbin, et al.Attention-guided context asymmetric fusion networks for the liver tumor segmentation of computed tomography images[J]. Quantitative Imaging in Medicine and Surgery, 2024, 14(7): 4825-4841.

【作者简介】

肖露（1997—），女，河南南阳人，硕士，助教，研究方向：计算机视觉、图像分割。

王亚茹（1994—），女，河南信阳人，硕士，助教，研究方向：机器学习、大数据分析。

刘锦博（1999—），男，河南郑州人，硕士，助教，研究方向：计算机视觉、人工智能。

邓锴（1997—），男，河南南阳人，硕士，助教，研究方向：算力网络、未来网络。

黄星帆（2005—），女，河南商丘人，本科在读，研究方向：计算机视觉。

（收稿日期：2026-07-14 修回日期：2026-08-12）