

参赛学生姓名：孙嘉一

中学：大连市第二十四中学

省份：辽宁省

国家/地区：中国/北方赛区

指导老师姓名：于喆

指导老师单位：大连市第二十四中学

论文题目：基于多级视觉复核和深度尺度约束的屏幕拍摄风险检测与显示干预系统

基于多级视觉复核和深度尺度约束的屏幕拍摄风险检测与显示干预系统

作者：孙嘉一

摘要

禁用截屏、静态水印和物理防窥可以减少一部分屏幕信息泄露，但如果有人直接使用手机等外部设备拍摄屏幕，这些方法仍然存在不足。针对这一问题，本研究设计并实现了一套基于 RK3588 的屏幕拍摄风险检测与显示干预原型。系统主要识别监测区域内能够被摄像头看到的手机等拍摄设备，不直接判断手机摄像头是否已经启动，也不能覆盖完全隐藏或严重遮挡的设备。

本项目最初从 Windows 笔记本上的 YOLO11n 手机检测原型开始。实际测试中，鼠标、遥控器、卡片等外观相似的物体经常被误认为手机；打印照片和屏幕画面中的手机图案，也可能被二维视觉模型识别为真实手机。针对这两类误报，系统逐步形成了多级复核流程：首先由轻量化 YOLO 模型生成人物和手机候选，当前原型采用 YOLO26n 验证整体流程；随后将手机候选区域输入 MobileNetV3，分为 phone、like_phone 和 negative 三类，用于过滤鼠标、遥控器等相似物；对于仍被判断为 phone 的候选，再利用 RGB-D 深度相机获取目标距离，并结合检测框和相机内参估算物理尺寸，从而排除尺寸明显异常的打印手机图案。

当前程序采用顺序门控方式进行判断。手机候选需要依次通过 MobileNetV3 分类、有效深度、距离和物理尺度检查；只有连续 3 帧均满足条件时，系统才触发手机告警，并保持黑屏 5 s。该方法避免了单帧误检立即造成显示切换，也便于根据程序日志检查候选在哪一级被过滤。

MobileNetV3 在排除训练集、验证集和测试集之间字节级完全重复图像后，对 1,411 张 ROI 图像进行测试，取得 98.51% 的总体准确率；phone 类的 Precision、Recall 和 F1-score 分别为 98.77%、98.36% 和 98.57%，非手机 ROI 被误判为 phone 的比例为 0.65%。典型测试中，YOLO26n 曾以 83.8% 的置信度将手持鼠标判断为手机候选，MobileNetV3 随后以 100.0% 的置信度将其判断为 like_phone，使该候选未进入后续手

机告警流程。对于打印照片中的手机图案，系统估算其物理尺寸约为 $5\text{ cm} \times 2\text{ cm}$ ，因尺寸明显不符合当前手机尺度条件而将其过滤。

早期基础链路曾在 0.5 m 、 1.0 m 、 2.0 m 和 2.5 m 四个距离点完成共 80 次正例事件测试，并记录约 $100\sim 200\text{ ms}$ 的单帧处理耗时。但这些数据形成于 MobileNetV3 接入主程序之前，因此只能说明早期候选检测、深度处理和显示控制链路能够运行，不能作为当前完整多级复核流程的端到端性能。

现有结果表明，YOLO 候选检测、MobileNetV3 语义复核、RGB-D 距离—尺度门控、连续帧确认和黑屏控制已经接入同一程序流程。系统在复杂光照、不同角度、严重遮挡、深度数据缺失和高仿真伪目标条件下的误报率、漏报率及端到端干预时延，仍需在 RK3588 上使用统一代码、模型和测试方法进一步验证。

关键词：屏幕拍摄风险检测；轻量化 YOLO；MobileNetV3；困难负样本；RGB-D；RK3588；边缘计算

目录

摘要	2
1 引言	8
1.1 本文研究与专利方案的对应关系	9
1.2 研究问题与适用边界	9
2 相关研究现状	10
2.1 现有屏幕防护技术	10
2.2 有代表性的研究与专利成果	12
2.3 研究空白与本文定位	13
3 方法	14
3.1 系统架构	14
3.1.1 不同部署架构分析与平台适配问题	16
3.1.2 算法—硬件—系统协同设计	17
3.2 轻量化 YOLO 候选模型选型	18
3.3 数据集构建	19
3.4 运行流程与线程结构	23
3.5 多源硬门控与黑屏判定	25
3.5.1 多源风险证据融合	26
3.5.2 指数移动平均抑制单帧波动	27
3.5.3 双阈值与持续时间判定	28
3.5.4 四态状态转移机制	28
3.5.5 状态机与显示防护联动	29
3.5.6 小结	29
3.6 深度感知与检测融合	29

3.6.1 候选目标获取与类别复核	30
3.6.2 候选区域深度获取	31
3.6.3 基于距离的实际尺寸估算	31
3.6.4 距离—尺寸联合约束	32
3.6.5 基于深度尺度约束的平面伪目标过滤示例	32
3.6.6 与状态机的融合方式	34
3.6.7 小结	34
4 实验与结果	35
4.1 实验平台与原型成本	35
4.2 实验设置	35
4.2.1 实验对象与设备	36
4.2.2 空间距离与测试位置	36
4.2.3 实验过程与记录方式	36
4.3 评价指标	37
4.3.1 目标检测有效性	37
4.3.2 检测触发率	37
4.3.3 误触发与伪目标过滤	37
4.3.4 深度测量与尺寸估算	37
4.3.5 系统响应时间	38
4.4 实验结果	38
4.4.1 正例事件的距离重复测试	38
4.4.2 近距离 RGB-D 联合检测结果	39
4.4.3 约 2.5 m 距离下的检测结果	39

4.4.4 多参与者真实场景检测结果	40
4.4.5 距离因素分析	41
4.4.6 平面伪目标与深度尺寸约束	41
4.4.7 环境感知实验小结	41
4.4.8 显示干预闭环验证	42
4.5 模块作用分析与案例对照	43
4.5.1 MobileNetV3 困难负样本复核与定量验证	43
4.5.2 典型场景下各模块的判定分工	44
4.5.3 平面伪目标的单案例模块对照	45
4.5.4 案例结果与实验边界	46
5 讨论	46
5.1 两阶段检测机制对误报控制的作用	47
5.2 RGB-D 深度信息的辅助价值	47
5.3 边缘端软硬件配合的意义	47
5.4 连续帧确认与黑屏保持的作用	48
5.5 系统局限性	48
5.6 应用价值与扩展方向	49
6 结论	50
7 参考文献	52
7.1 学术论文与预印本	52
7.2 专利	53
7.3 模型论文、专利补充、官方文档与工程资料	53
8 致谢	54

附录 A 复现实验环境与资源清单	57
附录 B 专利初步审查合格通知书	58
附录 C 软件著作权登记受理通知书	59
附录 D 多参与者真实场景测试样例	60

1 引言

外部摄像设备可以绕过禁用截屏、静态水印和物理防窥，直接获取屏幕内容。现有软件措施主要控制操作系统内部的数据复制或提供事后追溯，难以单独处理外部拍摄风险。本文据此设计一套基于 RK3588 的边缘端原型：显示处理通路通过 RGA、DMA-BUF 和硬件显示链路维持视频输出；环境感知通路执行人物与手机候选检测、ROI 复核和 RGB-D 物理尺度验证；统一状态机再根据多帧证据触发提示、水印、遮挡、画面模糊或黑屏等显示动作。

在本文实验平台和参数条件下，单路 1080p HDMI 显示链路可稳定保持约 60 FPS； 640×480 环境感知链路单帧处理耗时约为 100~200 ms。0.5 m、1.0 m、2.0 m 和 2.5 m 四个距离点共完成 80 次正例事件重复测试，均形成有效风险证据并触发状态机响应。MobileNetV3 在与训练目录分离的 1,420 张 ROI 测试图像上取得 98.73% 的准确率和 98.38% 的手机召回率，非手机 ROI 被误判为 phone 的比例为 0.65%。这些数据分别反映显示链路运行情况、受控正例触发重复性和 ROI 分类性能，不能替代开放环境下的系统级误报、漏报与端到端时延测试。

本文的主要贡献如下：

- (1) 多级视觉复核流程：以轻量化 YOLO 生成候选，使用 MobileNetV3 进行 ROI 语义复核，再以 RGB-D 距离与物理尺度合理性约束辅助排除明显异常目标。第一阶段模型可以替换，本文用 YOLO26n 验证整体流程；研究重点是不同证据的分工与组合。
- (2) 困难负样本驱动的语义误报抑制：针对鼠标、遥控器和卡片等 like_phone 样本，建立“误检发现—ROI 回收—分类标注—模型复训—分离测试”的迭代流程，并通过 phone、like_phone 和 negative 三分类降低相似物误报。
- (3) 多帧风险决策与显示干预闭环：采用多帧平滑、双阈值滞回和持续时间约束形成 S0~S3 风险状态，按等级驱动提示、动态水印、局部遮挡、画面模糊或黑屏。
- (4) RK3588 边缘端原型实现：构建 RGA、NPU 与 DMA-BUF 协同的数据链路，减少 CPU 数据搬运和显示路径开销，并记录原型所用的软件环境、模型与公开资源。

1.1 本文研究与专利方案的对应关系

本研究所依托的总体方案已申请发明专利《一种显示屏防拍摄方法及装置》。该申请围绕外部环境感知、风险判断和显示干预的总体流程及装置形态；本文记录轻量化候选检测、困难负样本驱动的 MobileNetV3 ROI 复核、RGB-D 距离与物理尺度合理性验证、统一状态机和 RK3588 边缘部署的工程实现与实验结果。专利申请情况用于披露研究时间线和知识产权关系，不作为论文原创性或实验有效性的证明。

1.2 研究问题与适用边界

本文围绕三个可检验问题展开：（1）在第一阶段保持候选召回的同时，MobileNetV3 能否拦截鼠标等高置信度相似物；（2）RGB-D 距离与尺度约束能否排除二维语义模型仍判为手机的明显异常平面图案；（3）多帧状态机能否把复核后的风险证据稳定转换为显示动作。

研究对象限定为监测区域内可见的手机等拍摄设备，实验范围为本报告记录的硬件、光照和 0.5~2.5 m 距离条件。本文不直接识别拍摄意图，不声称覆盖隐藏式、严重遮挡或远距离长焦设备，也不把正例触发次数等同于系统级防拍成功率。

2 相关研究现状

2.1 现有屏幕防护技术

(1) 可见水印：通过浮动水印震慑和溯源，但不能阻止拍摄；

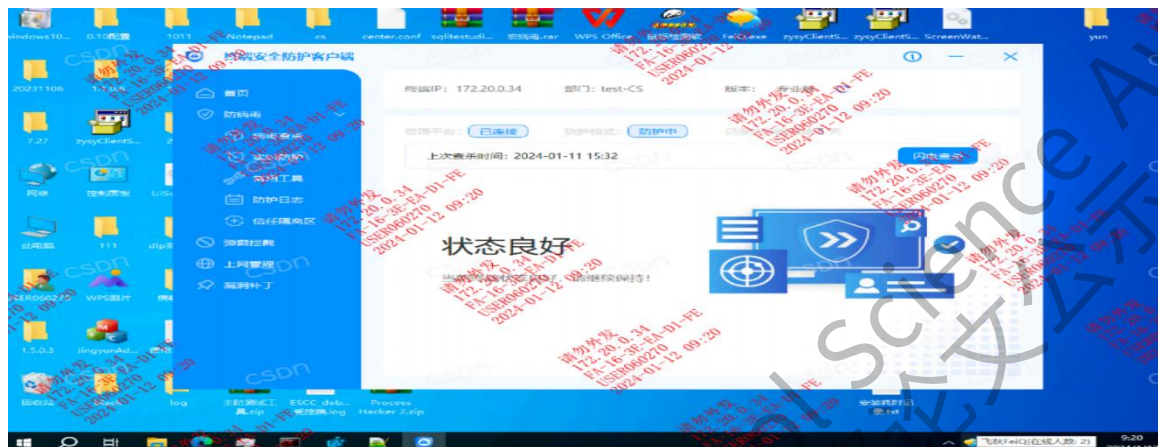


图 1 可见水印

(2) 隐形水印（暗水印）：对人眼透明，可跨媒介溯源，但仍是事后追责[8]；

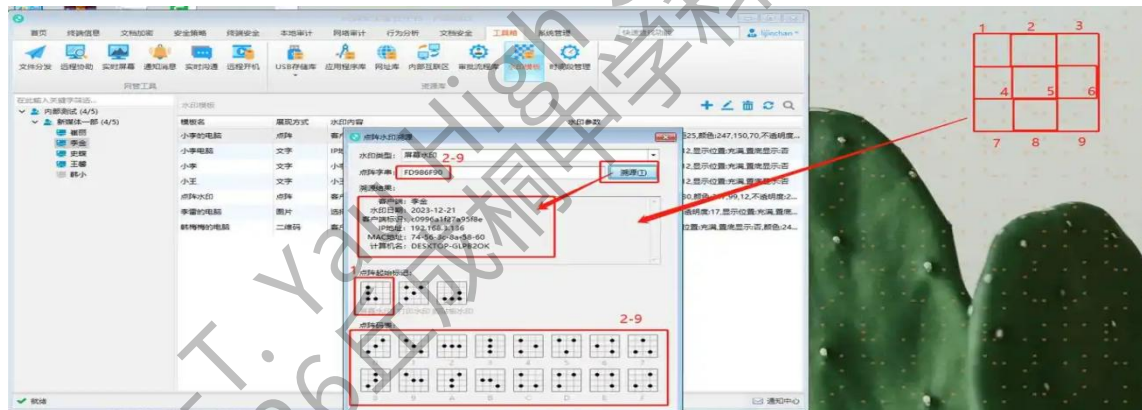


图 2 隐形水印（暗水印）

(3) AI 风险检测：利用摄像头检测人物、手机等目标，并结合多帧证据判断潜在拍摄风险，可在事中触发显示干预（本系统）；



图 3 AI 拍摄风险检测（本系统）

(4) 硬件防窥：通过限制有效可视角度降低旁观泄露风险，但对正面拍摄的防护能力有限。



图 4 硬件防窥

(5) 光学/信号干扰：通过调整显示刷新、亮度或时域/空域调制，使摄像机传感器采样与屏幕显示过程产生不匹配，从而降低拍摄图像质量。此类方法的效果与显示设备、相机快门和曝光参数有关，因此更适合作为主动检测与显示干预体系中的补充措施，而不宜作为唯一的安全机制。



图 5 光学/信号干扰

2.2 有代表性的研究与专利成果

表 1 代表性研究与专利成果

年份	类型	申请人/研究机构	内容概述	亮点/局限
2017	发明专利 CN106407827A[16]	合肥工业大学	将屏幕划分为多个区域，采用不同刷新频率，造成拍摄画面失真，但人眼视觉连贯	利用频率差扰动相机成像；适用效果受相机快门参数与设备差异影响
2019	学术研究[6]	中国航天系统科学与工程研究院	基于 Faster R-CNN 的屏幕防拍摄方法，构建动静结合的 VOC 数据集	引入双通道数据增强，提升检测精度；但速度不足，难以在嵌入式实时场景应用
2020	发明专利 CN111639592A[17]	苏州深信达网络科技有限公司	屏幕内置摄像头检测拍照行为后自动关闭显示并记录	采用拍照行为检测、清屏与记录流程；需要综合考虑误触发和使用体验
2022	发明专利 CN114863337A[18]	成都天奥集团有限公司	采用语义图、目标检测与行为分类识别拍摄行为并触发锁屏和证据留存	采用分割、检测与分类融合流程；公开专利未提供本文可直接复核的统一测试数据
2022	预印本论文[8]	<i>Screen-Shooting Resilient Watermarking</i>	在屏幕内容中嵌入水印，使拍摄照片中仍可识别用户身份	提供追溯能力，支持责任划分；但属于被动式方案，无法主动阻止泄密
2017	学位论文[7]	浙江大学	基于 VLF 辐射特征检测隐藏摄像头	可发现微型摄像设备，适合隐私保护场景；但仅检测摄像头存在，无法阻止屏幕拍摄
2024	发明专利 CN117786775A[12]	苏州涟漪信息科技有限公司	基于 YOLOv8 的实时检测与报警，自动关闭屏幕并触发告警	采用目标检测、报警与自动关屏流程；边缘端部署成本取决于模型与硬件配置
2024	实用新型 CN221615040U[14]	天津光电安辰	高安全防拍装置，含广角摄像头、识别模块和 HDMI 通道切换	采用摄像头、识别模块和 HDMI 切换构成硬件防护链路；需要额外硬件集成

年份	类型	申请人/研究机构	内容概述	亮点/局限
2024	发明专利 CN118194380A[25]	北京华中创世科技	主动防偷拍设备，基于 AI 图像识别与摄像头探针模组	采用 AI 图像识别与摄像头探针模组进行多源检测；系统组成较复杂
2023	发明专利 CN117037270A[15]	北京军安中科	基于图像处理与机密内容识别的屏幕防拍方法	结合图像处理、内容识别与环境监测；公开材料未充分说明动态场景适配性能
2025	学术论文	Ma et al., SID Symposium Digest	基于 CNN 构建面向保密会议场景的实时防拍系统，识别拍照风险后隐藏会议显示内容	验证了 AI 实时防拍的应用可行性；公开工作未重点讨论 RGB-D 物理尺度复核与 RK3588 软硬件协同链路

总体来看，现有研究与专利大体分为三类：

- (1) 基于图像扰动的防护机制（如不同刷新频率扰动）；
- (2) 基于传感器检测的防护机制（如 VLF 电磁辐射检测摄像头）；
- (3) 基于深度学习的检测与防护机制（如 Faster R-CNN、YOLOv8 等）[6]。

总体来看，前两类方法或依赖专用硬件，或适用范围受限；深度学习方法具有较强的场景适应性，但仍面临实时性、误报率和边缘端能效等问题。本文围绕这些约束，采用轻量化 YOLO 候选生成、MobileNetV3 ROI 复核、RGB-D 距离与物理尺度合理性验证和 RK3588 硬件加速相结合的方案，以提升系统的实时性、准确性与部署可行性。

2.3 研究空白与本文定位

公开论文和专利已经提出外部摄像头感知、AI 目标识别、告警或关屏、HDMI 通路切换等方案[5-8, 12, 14-18, 25]。因此，本文不把这些通用组成单独作为创新点。

现有公开材料较少同时报告困难负样本复核、距离一尺度约束、多帧决策和边缘端显示联动的实现细节与适用边界。本文的工作重点是把这些模块组合为可运行原型，并分别给出 ROI 分类指标、正例事件重复测试、典型伪目标案例和显示干预功能验证。

这些证据的粒度不同：MobileNetV3 报告独立 ROI 测试集指标；RGB-D 目前只有单案例模块对照；系统级测试主要覆盖正例触发，尚未形成统一负例、复杂环境和端到端时延统计。后续讨论和结论均按这一证据范围解释。

3 方法

3.1 系统架构

系统整体采用“双通路、单决策”架构：显示处理路径与环境感知路径并行运行，并由统一状态机进行协调，驱动后续防拍干预策略，如图 6 所示。

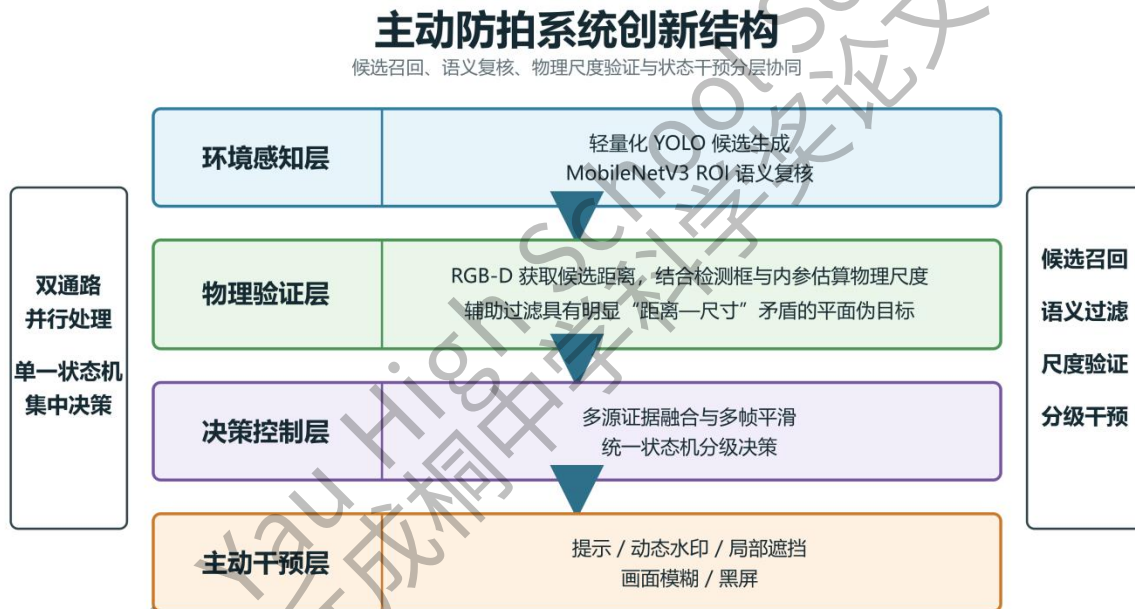


图 6 屏幕拍摄风险检测与显示干预系统结构

主显示信号采用物理串联拓扑：被保护主机 HDMI OUT → Orange Pi 5 Plus HDMI IN → RK3588（视频采集、RGA 图像处理及水印、遮挡或黑屏叠加）→ Orange Pi 5 Plus HDMI OUT → 显示器。开发板串联在视频信号源与显示终端之间，形成独立的显示

安全网关；RGB-D 摄像头作为旁路环境感知输入接入 RK3588，不替代主机输出的 HDMI 显示画面。

（1）显示处理路径

显示处理路径主要负责屏幕显示信号的低延迟采集、处理与输出。HDMI 输入信号经 v4l2src 完成视频采集，随后利用 RGA 硬件单元进行色彩空间转换、图像缩放等预处理，并通过 DMA-BUF 共享缓冲区在各处理模块之间传递图像数据，最终由 rkximagesink 完成硬件加速显示[19-20,23]。该链路尽量避免 CPU 参与逐帧内存复制，从而降低数据搬运开销和系统延迟，提高高分辨率视频场景下的实时处理能力。

显示处理路径不承担屏幕内容语义分类，其主要职责为显示信号处理与防护策略执行。当环境感知路径判定存在潜在拍摄风险后，统一状态机向显示处理路径下发控制指令，由显示处理路径根据当前风险等级执行水印叠加、局部遮挡、画面模糊、黑屏等防护操作。因此，该路径主要承担“显示内容承载与防护执行”功能。

（2）环境感知路径

环境感知路径负责对屏幕前方人物、手机及其他疑似目标进行实时检测与风险信息提取。监控摄像头采集的 RGB 或 RGB-D 图像首先输入轻量化 YOLO 候选检测器，获取人物、手机等候选目标的位置与置信度；本文实验部署选用 YOLO26n，其模型接口以 Ultralytics 官方文档为依据[21]。对于手机候选区域，进一步采用 MobileNetV3 对 ROI 进行 phone、like_phone 和 negative 三分类复核[3]；对于人物和手机候选，可结合 RGB-D 信息获取目标距离，并利用检测框与相机内参估算实际物理尺度[22]，用于排除明显异常的打印图像等伪目标。经复核的结果再结合多帧持续时间进入统一状态机进行风险判定。

（3）统一决策与状态控制

双通路在数据处理层面相互独立，在策略层面通过统一状态机协同。环境感知路径输出经复核的目标证据及多帧风险信息，状态机据此确定当前防护等级，并控制显示处理路径执行相应干预。系统由此形成“环境感知—目标复核—风险判定—状态决策—显示干预”的闭环流程，仅在风险证据达到设定条件时启动相应防护机制。

系统总体架构

双通路并行处理·统一状态机单点决策·闭环防拍干预

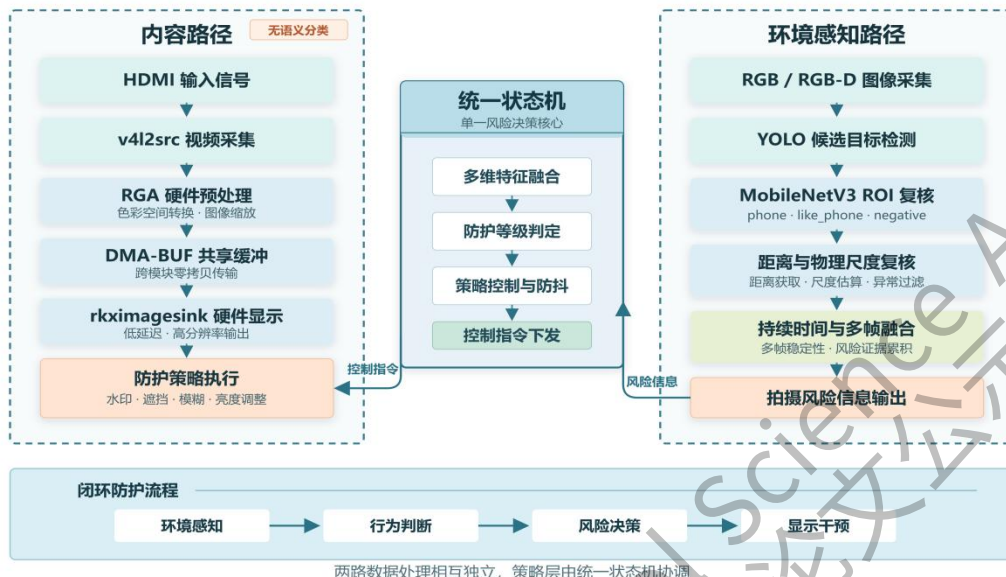


图 7 系统总体架构图

3.1.1 不同部署架构分析与平台适配问题

部署架构分析：

系统提供两种实现方式。（1）本地计算方案（早期验证）：在笔记本或台式机上使用摄像头监测，并在操作系统窗口中绘制水印，便于快速验证算法，但存在 CPU/GPU 占用较高、跨平台适配复杂和安全隔离不足等问题。（2）独立边缘端方案（本文）：由 RK3588 负责视频采集、NPU 推理、RGA 处理和硬件显示，感知结果通过状态机驱动显示处理路径干预。被保护主机只需通过标准 HDMI 接口输出视频，无需安装检测软件、显示驱动插件或调用宿主机操作系统内部接口；在 HDMI 格式与分辨率兼容的前提下，该物理串联方式可降低 Windows、Linux 和其他主机系统之间的软件适配差异。因此，该方案在平台兼容性、安全隔离和能效方面更适合长期部署。

表 2 跨平台实现的困难对比

平台	限制点	难度与问题
Windows	屏幕水印/遮挡需要调用 DirectX Overlay 或 GDI Hook	不同版本（Win7/10/11）图形栈差异大；权限要求高；存在被恶意软件绕过的风险
Linux	图形栈差异：X11 可用 Xlib/XComposite 抓取，但需要高权限；Wayland 默认禁止跨窗口绘制	桌面环境多样（GNOME/KDE/XFCE），适配成本高；Wayland 下几乎不可行
macOS	基于 Quartz/Core Graphics/Metal，系统默认不允许跨应用全局叠加绘制	无公开 API，仅能在应用内部实现；沙盒和安全策略严格，几乎无法系统级叠加水印

3.1.2 算法—硬件—系统协同设计

算法—硬件—系统协同设计。系统以端到端时延预算为约束，围绕“显示处理路径高帧率直出、感知路径低时延决策”展开三层协同：（1）RGA 在采集侧完成色彩转换、缩放和合成前处理；（2）YOLO 推理下沉至 NPU，CPU 负责轻量调度与消息分发；（3）DMA-BUF 在模块间共享缓冲区，减少用户态与内核态之间的数据复制；（4）rkximagesink 通过硬件路径直出 1080p 视频；（5）显示处理路径的关键链路被压缩为“RGA→可选水印/遮挡→硬件显示”，感知路径则形成“RGA/resize→NPU 推理→ROI 复核→状态机”，从而降低处理时延与 CPU 占用。相关硬件加速与缓冲区共享机制参见 Rockchip 和 GStreamer 工程文档[19-20, 23]。

如图 8 所示，硬件连接由一条串联式 HDMI 主视频通路和一条摄像头旁路感知通路组成。主视频通路负责连续承载被保护主机的显示信号，旁路感知通路负责产生风险证据；统一状态机将风险等级转换为控制指令，作用于 RK3588 的 RGA 与显示合成环节，从而在不中断宿主机视频输出逻辑的情况下执行水印、遮挡、模糊或黑屏等保护动作。

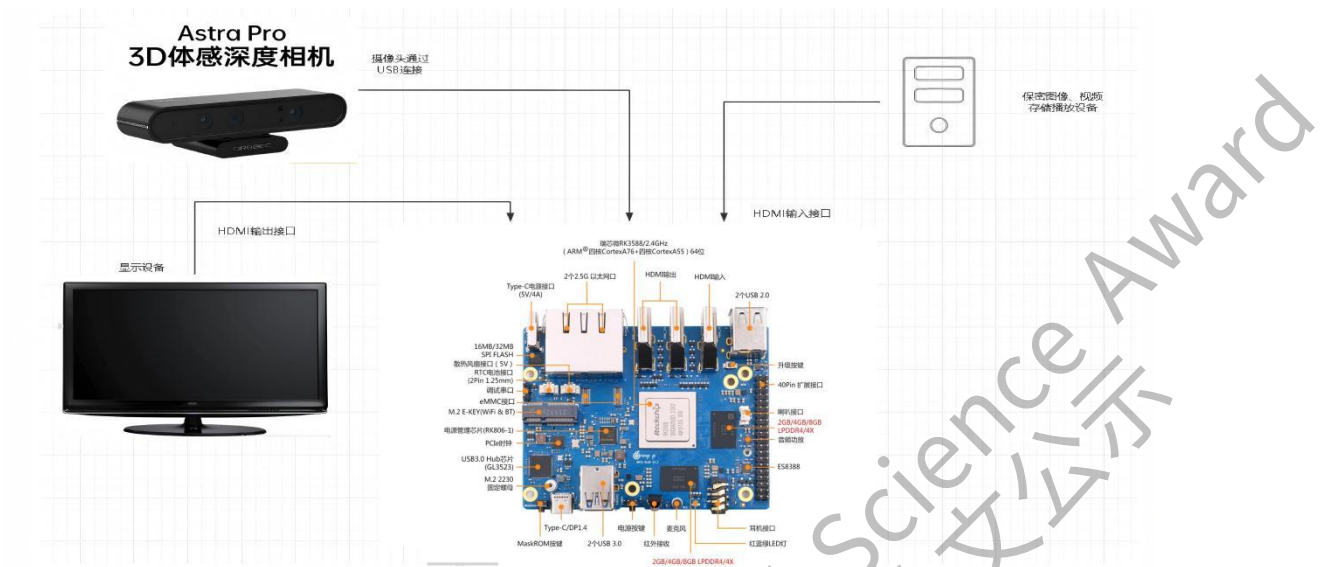


图 8 算法—硬件—系统协同与物理连接拓扑图

3.2 轻量化 YOLO 候选模型选型

本文将第一阶段目标检测器定义为可替换的候选生成模块，不把某一 YOLO 版本本身作为研究创新。该模块以较低阈值发现 person 和 cell phone 等候选，优先保证后续复核环节所需的候选召回能力；具体模型则根据权重可获得性、输入尺寸、RKNN 转换兼容性和端侧运行开销进行工程选型。

本文采用的 YOLO26n 为实验阶段使用的工程模型版本，主要用于验证整体系统流程，其模型版本和接口具有可替换性。由于本文创新重点不在目标检测网络结构改进，因此不对 YOLO26n 的模型性能进行单独排名比较。只要其他轻量化检测器保持相同的候选输出接口，即可接入后续 MobileNetV3、RGB-D 与状态机组成的多级验证框架。

为保证表述真实性，本文仅报告当前模型在本工程平台上的实际现象和运行数据，不以其他硬件平台的公开结果替代 RK3588 实测，也不以模型发布时间作为创新依据。若后续开展不同轻量化候选检测器的正式横向比较，应统一训练与测试数据、输入尺寸、INT8 量化方式和 RK3588 NPU 环境，并同时报告 Recall、mAP@0.5、纯 NPU 时延及 RKNN 模型大小。

在当前系统中，轻量化 YOLO 模型仅负责形成候选区域。第一阶段检测器输出标准目标框、类别候选及置信度信息，随后再经过 MobileNetV3 语义复核、RGB-D 物理尺度复核和状态机多帧判定。因此，系统效果来自分层验证与闭环决策，而非对单一检测模型能力的直接宣称。

YOLO26 的模型来源与结构说明见专门论文和官方文档[21, 24]。

3.3 数据集构建

本研究涉及两类数据资源。第一类用于说明第一阶段候选检测模型的类别基础及进行工程场景验证，包括 COCO 中的 person、cell phone 类别[10]及实际拍屏场景样例；本文未基于这些数据重新进行大规模 YOLO 训练。第二类用于 MobileNetV3 ROI 分类模型的迁移训练与测试[3]，包括 phone、like_phone 和 negative 三类。

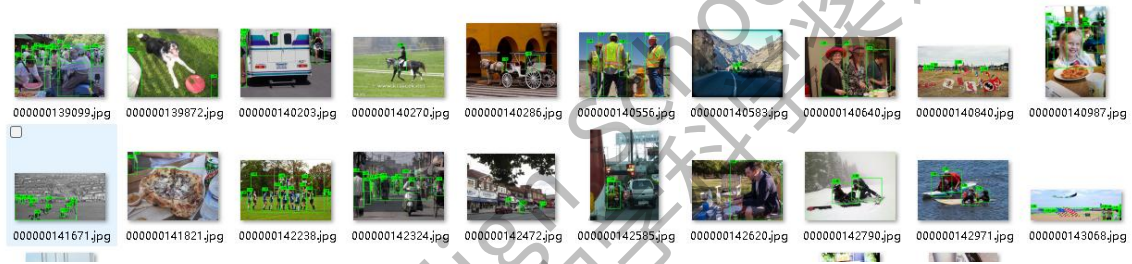


图 9 COCO 图像中的 person 分类



图 10 COCO 图像中的 cell phone 分类

COCO 原始类别包括 person (ID=0) 和 cell phone (ID=67) [10]。

COCO 中的 cell phone 类别覆盖通话、桌面放置、口袋携带等常见场景，但与“举起设备拍摄屏幕”的目标场景存在一定域差异。

person 类别提供较丰富的人员场景。上述 COCO 类别资源仅用于说明候选检测器的类别来源及工程验证基础；本文未基于其完整训练集重新训练 YOLO 检测器。

为减小域差异，系统在 COCO 数据基础上补充真实办公环境中的手机拍屏样本，并保留不同距离、角度、遮挡和光照条件下的候选区域。YOLO 阶段采用偏重召回的阈值策略，使易混淆目标也能进入后续复核；误检候选则反向进入 MobileNetV3 数据集，形成面向实际部署场景的数据闭环。

早期测试中，遥控器、计算器、工牌、卡片、鼠标、手机壳、电子设备外壳、金属牌和画面边缘等细长矩形物体容易被 YOLO 作为候选目标。为此，本研究不在 YOLO 检测头中强行压低阈值，而是在 MobileNetV3 中显式设置 phone 与 like_phone 类别：phone 收录真实手机 ROI；like_phone 专门收录外观近似手机、但不具备拍摄功能的困难负样本；negative 用于无明确目标或背景占主导的区域。该设计将“是否值得检查”与“是否确为手机”分开学习，有利于在保留候选进入复核机会的同时降低误报。



图 11 实际检测中的困难负样本

MobileNetV3 困难负样本优化流程如下：

(1) 候选样本生成。使用 YOLO26n 扫描训练图像，对检出的非人物目标进行置信度筛选，并根据端侧模型的实际输出形式完成候选框去重，保留可能与手机混淆的候选区域。

候选框按原图尺寸进行归一化记录，并在裁剪时向四周扩展约 8%，以保留握持手部和物体边缘等上下文信息。

(2) 类别组织。phone 类仅放入可确认的真实手机 ROI；like_phone 类放入遥控器、计算器、卡片、工牌、手机壳及其他相似矩形物体；negative 类放入背景或无明确目标区域。

三类数据按照训练集、验证集和测试集组织为 ImageFolder 目录，类别名称及索引随模型检查点一并保存，避免部署阶段出现标签顺序不一致。

(3) 模型训练。以 ImageNet 预训练的 MobileNetV3-Large 为骨干网络，将分类头替换为与实际类别数一致的输出层[3]。

训练采用交叉熵损失、AdamW 优化器和余弦退火学习率调度，并根据验证集准确率保存最佳模型。

(4) 在线复核。推理时同时计算整图与各候选 ROI 的分类结果，但事件判定以 ROI 结果为主；当任一候选被分类为 phone 时，系统才进入手机风险判断。

对于贴近画面边缘且 YOLO 原始标签不是手机的候选框，系统将告警等级由高危降为复核，降低局部截断目标引起的误触发。

(5) 迭代更新。将部署中高置信度误报的 ROI 经人工复核后加入 like_phone 类，将漏检的真实手机区域补入 phone 类，再重新训练并比较测试集准确率、手机召回率和非手机误报率。

这一闭环属于数据层面的困难负样本挖掘：困难样本通过明确类别参与标准监督训练，模型仍使用标准交叉熵损失，不额外引入 OHEM 或样本权重。通过持续补充

like_phone 样本, MobileNetV3 可重点学习手机镜头、屏幕边框、按键布局 and 整体结构等区分特征, 从而提高对相似物体的判别能力。该方法的效果应通过与训练目录分离的测试集和真实视频回放进行验证, 并分别报告 phone 召回率、总体准确率和非手机误报率。与单阶段直接告警相比, 两阶段架构将 YOLO 偏重召回的候选生成策略与 MobileNetV3 的细粒度分类能力结合, 便于在保留候选复核机会的同时逐步压低误报。训练和部署过程中均保留类别列表、模型结构和版本信息, 确保同一检查点能够被桌面验证程序与后续端侧转换流程一致解析。

为保证结论可复核, 本文不预设整套系统的固定性能增益。第 4.5 节使用工程模型注册记录中与训练目录分离的 1,420 张 ROI 测试图像, 对 MobileNetV3 给出准确率、手机召回率和非手机误报率; YOLO 与 RGB-D 模块尚未完成统一重复次数的整套系统定量对比, 因此继续采用典型案例进行定性分析, 并明确结论边界。

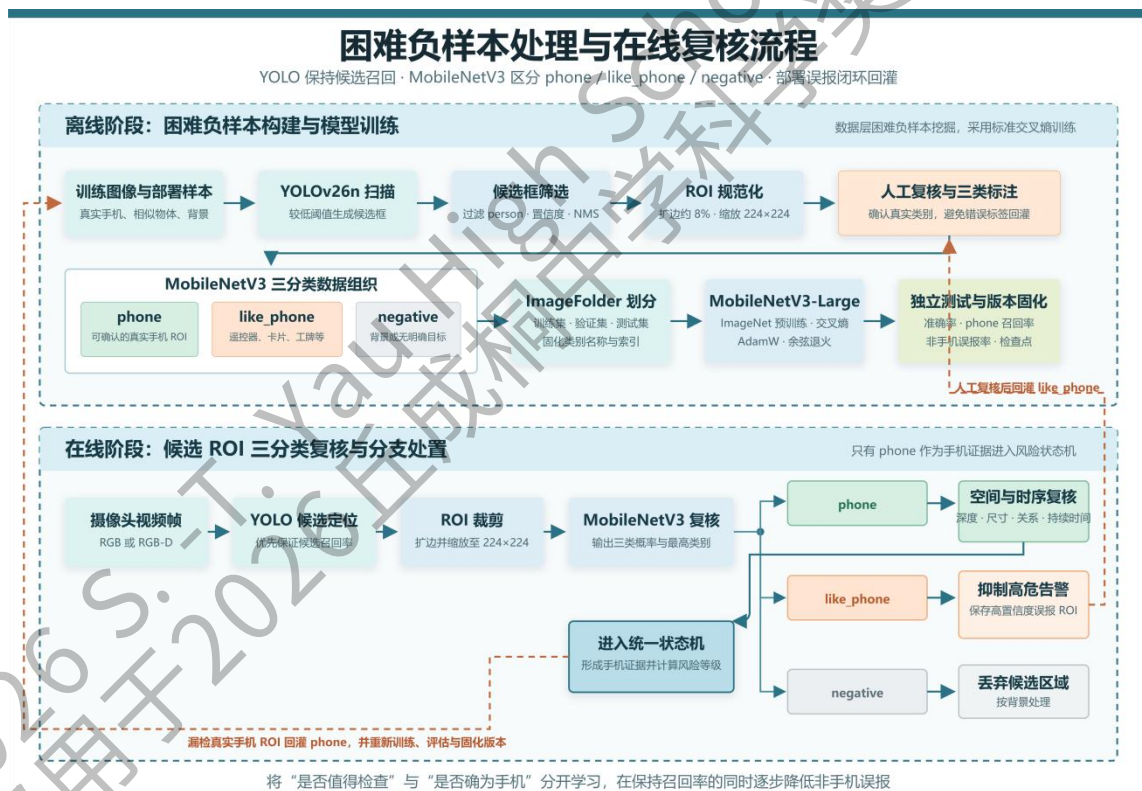


图 12 负样本优化流程图

相关数据目录结构、训练参数与复现实验步骤见附录 A。

3.4 运行流程与线程结构

为保证视频显示、目标检测与防护控制能够并行执行，系统采用多线程运行机制，将 HDMI 显示链路、环境感知链路和界面显示链路相互解耦。各线程通过共享帧缓存和状态变量交换数据，避免目标检测等耗时操作阻塞 HDMI 视频显示，从而提高系统整体的实时性与运行稳定性。

系统主要包括 HDMI 捕获线程、环境感知与目标检测线程以及合成显示线程，其功能划分如表 3 所示。

表 3 系统主要线程及功能

线程模块	主要输入	核心功能	主要输出
HDMI 捕获线程	1920×1080 HDMI 视频	视频采集、显示防护处理、水印或遮挡叠加	hdmi_frame
环境感知与目标检测线程	RGB/RGB-D 摄像头视频	YOLO 目标检测、ROI 复核、深度验证、风险信息提取	检测结果及风险状态
合成显示线程	hdmi_frame、检测结果、系统状态	HDMI 主画面显示、摄像头预览、检测框及运行信息可视化	最终显示画面

(1) HDMI 捕获线程

HDMI 捕获线程负责持续获取外部主机输出的屏幕画面，当前输入分辨率为 1920×1080。线程读取最新 HDMI 视频帧后，根据统一状态机输出的风险等级执行相应的显示防护操作，包括动态水印、告警提示、局部遮挡、画面模糊或黑屏等。

处理后的图像更新至共享帧缓存 hdmi_frame，供合成显示线程读取。该线程仅负责显示内容采集和防护策略执行，不承担人物、手机等环境目标的检测任务。

(2) 环境感知与目标检测线程

环境感知线程负责获取屏幕前方区域的 RGB 或 RGB-D 视频，并对潜在拍摄行为进行实时分析。摄像头图像完成预处理后输入部署于 RK3588 NPU 的 YOLO/RKNN 推理模块，对人物、手机等目标进行检测。

对于检测到的候选目标，系统进一步提取目标位置、检测置信度、人员数量及手机候选信息。当启用 MobileNetV3 二次复核时，对 YOLO 输出的手机候选 ROI 进一步进行 phone、like_phone 和 negative 分类；当启用 RGB-D 深度验证时，则结合目标距离和估算实际尺寸等信息对二维检测结果进行一致性验证。

上述信息经过多帧融合后送入统一状态机，由状态机综合判断当前拍摄风险等级。环境感知线程本身不直接修改 HDMI 显示内容，而是负责生成 检测结果和风险判定依据。

（3）合成显示线程

合成显示线程负责生成系统最终显示界面。正常运行时，以最新的 hdmi_frame 作为主显示画面；在实验或调试模式下，可同时叠加环境摄像头缩略图、目标检测框、人员数量、推理帧率、风险状态以及运行日志等信息。

为兼顾系统演示和实际部署需求，显示模块支持正常全屏模式、调试信息显示模式以及环境摄像头预览隐藏等工作方式。该线程仅负责 界面合成与运行状态可视化，与目标检测和风险决策逻辑保持相互独立。

（4）执行单元间的数据流与告警控制

系统在功能上分为显示处理和环境感知两条路径，但两条路径最终在同一主处理循环中完成汇合。主循环每轮读取环境摄像头图像、HDMI 输入画面以及深度接收线程更新的最新对齐深度帧，随后依次执行候选检测、语义复核、距离与尺度检查、连续帧确认和显示控制。

显示数据路径负责承载被保护主机的画面。HDMI 输入帧进入主循环后，在无有效告警时按正常方式输出；当手机告警或其他安全告警成立时，主循环切换为黑屏画面，并按照设定的保持时间维持显示保护。黑屏保持结束且没有新的告警时，系统恢复正常 HDMI 画面。水印叠加作为独立的全局显示功能，不随手机风险等级自动变化。

环境感知路径以 RGB 或 RGB-D 摄像头画面为输入。YOLO 首先生成人物和手机候选，其中只有标注为 cell phone 的候选区域进入 MobileNetV3 复核。候选 ROI 经扩边、裁剪

和缩放后，由 MobileNetV3 判断为 phone、like_phone 或 negative。只有分类结果为 phone 且置信度达到设定阈值的候选，才继续读取对应区域的深度数据，并根据目标距离、检测框像素尺寸和相机内参估算物理尺度。

若候选目标通过语义分类、有效深度、距离和尺寸等全部条件，则该帧形成一次有效手机证据；任一条件不满足，该候选在当前帧即被过滤。程序不依据单帧结果立即黑屏，而是要求有效手机证据连续出现 3 帧。连续条件成立后，phone_alarm 被置为真，并刷新 5 s 黑屏保持时间；若中间任一帧未形成有效手机证据，连续计数重新开始。

深度接收线程持续更新最新的对齐深度图，供主循环在需要时读取；Web 服务线程负责参数、日志和预览接口。这两个辅助线程不直接决定是否触发黑屏，最终告警仍由主循环按照统一规则产生。

因此，当前系统的实际数据流可以概括为：

环境 RGB 图像

- YOLO 手机候选
- MobileNetV3 语义复核
- RGB-D 距离与物理尺度门控
- 连续 3 帧确认
- 手机告警
- HDMI 黑屏保持

与此同时，HDMI 输入画面始终由主循环读取并输出。环境感知结果只通过告警变量改变最终显示内容，不直接修改检测或采集模块，从而使候选检测、风险判断和显示控制之间保持清晰的职责边界。

3.5 多源硬门控与黑屏判定

为减少单帧误检对显示结果的影响，当前系统没有采用“检测到手机即黑屏”的直接规则，也没有计算加权综合风险分数。手机候选需要按照固定顺序通过 YOLO 类别筛选、

MobileNetV3 语义复核、有效深度检查、距离限制和物理尺度限制。只有同一帧中的候选满足全部条件，才形成手机风险证据。

在时间维度上，系统进一步采用连续帧确认机制。只有连续 3 帧均出现通过全部门控的手机候选，程序才触发手机告警；任一帧失败，连续计数立即清零。告警成立后，系统将黑屏保持时间更新为 5 s，新告警可继续延长保持时间。该设计将空间与语义条件同时确认结合起来，在不引入复杂概率模型的情况下，降低偶发误检和检测框抖动直接引起显示切换的可能性。

与简单采用“检测到手机即黑屏”的规则不同，本系统将风险判定划分为安全、观察、预警和干预四个状态，使系统能够在响应速度与误触发率之间取得平衡。

表 4 系统四态风险状态机

状态码	状态名称	主要判定条件	系统响应
S0	安全状态	未检测到有效手机候选，或综合风险值低于安全阈值	正常显示
S1	观察状态	出现疑似手机候选，但类别复核、可用的距离一尺度证据或持续时间证据不足	保持正常显示，继续累积多帧证据
S2	预警状态	手机候选通过部分复核，综合风险持续超过预警阈值	显示告警提示或增强动态水印
S3	干预状态	手机候选通过类别复核，综合风险持续达到高风险条件；有效 RGB-D 信息可用时纳入距离一尺度证据	局部遮挡、画面模糊、黑屏及安全告警

3.5.1 多源风险证据融合

环境感知线程首先由 YOLO26n 检测人物和疑似手机目标。对于手机候选区域，系统进一步使用 MobileNetV3 进行 phone、like_phone 和 negative 三分类复核，并利用 RGB-D 深度信息对目标距离与物理尺寸合理性进行辅助验证。为避免将任一单一检测结果直接作为防护触发条件，系统构造综合风险评分，如式（1）所示：

$$R_t = w_1 C_{det} + w_2 C_{cls} + w_3 C_{depth} , w_1 + w_2 + w_3 = 1, w_i \geq 0 \quad (1)$$

- R_t : 第 t 帧的综合风险评分;
- C_{det} : YOLO26n 对手机候选目标的检测置信度;
- C_{cls} : MobileNetV3 将候选区域判断为 phone 的分类置信度;
- C_{depth} : RGB-D 距离和估算物理尺寸是否符合目标类别合理范围的可信度;
- w_1 、 w_2 、 w_3 : 各证据项对应的权重。

各项证据经归一化后参与风险计算。当深度模块因无效深度点等原因暂时不可用时, 可将对应权重设置为 0, 并对其余权重重新归一化; 多帧持续时间不直接作为单帧风险评分项, 而在后续 EMA 平滑和状态持续时间判定中体现。

3.5.2 指数移动平均抑制单帧波动

目标检测模型在连续视频中可能受到光照变化、目标遮挡、运动以及检测框抖动等因素影响, 使原始风险评分在相邻帧之间产生波动。为提高状态判定的稳定性, 系统采用指数移动平均 (Exponential Moving Average, EMA) 对综合风险评分进行平滑, 如式 (2) 所示:

$$\bar{R}_t = \beta \bar{R}_{t-1} + (1 - \beta) R_t \quad (2)$$

- R_t : 第 t 帧经 EMA 平滑后的风险评分;
- $\bar{R}_{t-1}$: 上一帧经 EMA 平滑后的风险评分;
- R_t : 由式 (1) 计算得到的第 t 帧原始综合风险评分;
- β : 平滑因子, 满足 $0 < \beta < 1$ 。

例如, 当 $\beta = 0.8$ 时, 当前帧原始风险评分 R_t 占 20% 的权重, 上一时刻的平滑风险评分 $\bar{R}_{t-1}$ 占 80% 的权重。因此, 单帧检测结果发生短暂波动时, 不会立即导致平滑风险评分产生同等幅度的变化, 有助于降低瞬时误检、短时遮挡和检测置信度抖动对状态机判定的影响, 减少系统在不同风险状态之间频繁切换。

在状态机初始化时, 可令 $\bar{R}_0 = R_0$, 即以首帧原始风险评分作为初始平滑值。

3.5.3 双阈值与持续时间判定

仅采用单一风险阈值容易在临界值附近产生状态反复切换。为此，系统采用进入阈值和退出阈值相互分离的滞回控制机制。设高风险进入阈值为 T_{high} ，风险解除阈值为 T_{low} ，且满足式（3）：

$$T_{high} > T_{low} \quad (3)$$

当平滑风险值满足以下条件，并持续达到设定进入时长 T_{enter} 时，系统才进入高级风险状态，具体条件如式（4）所示：

$$\bar{R}_t \geq T_{high} \wedge \Delta t_{high} \geq T_{enter} \Rightarrow S_k \rightarrow S_{\min(k+1, 3)} \quad (4)$$

当系统已经处于预警或干预状态时，只有当平滑风险值低于退出阈值，并持续达到 T_{exit} 后，系统才降低风险等级或恢复正常显示，具体条件如式（5）所示：

$$\bar{R}_t \leq T_{low} \wedge \Delta t_{low} \geq T_{exit} \Rightarrow S_k \rightarrow S_{\max(k-1, 0)} \quad (5)$$

式中， Δt_{high} 和 Δt_{low} 分别表示平滑风险值连续高于进入阈值、连续低于退出阈值的累计时长； S_k 表示当前风险状态， $k \in \{0, 1, 2, 3\}$ ，分别对应安全、观察、预警和干预状态。 $\min$ 与 $\max$ 约束保证状态编号始终处于 0~3 范围内。

式（4）和式（5）用于描述相邻风险等级之间的通用滞回控制形式。实际状态转换除综合风险值外，还需结合候选目标是否存在、MobileNetV3 类别复核结果以及可用的 RGB-D 物理尺度证据，各状态的具体进入条件由状态机规则共同决定。

这种双阈值设计形成了状态切换的“滞回区间”，可避免风险值在阈值附近轻微变化时导致系统状态反复跳变。实际阈值和持续时间参数应通过验证集与真实视频回放实验确定，而不宜仅凭经验固定。

3.5.4 四态状态转移机制

系统状态按照风险证据逐步升级，而不是由单帧结果直接从安全状态跳转至最高风险状态。未发现有效手机候选时，系统保持 S0 安全状态；当 YOLO26n 检测到疑似手机，但 MobileNetV3 分类结果、可用的 RGB-D 距离与尺度证据或持续时间尚不足以确认风险时，进入 S1 观察状态，并继续累积后续帧证据。

当候选目标连续多帧被判断为真实手机，且综合风险评分超过预警阈值后，系统由 S1 进入 S2。此时可执行提示或动态水印等较温和的保护动作。当风险继续升高，手机候

选通过类别复核、综合风险持续达到高风险条件，并满足高风险持续时间要求后，系统进入 S3；当有效 RGB-D 信息可用时，同时纳入距离与物理尺度合理性证据。

当风险目标消失后，系统不会立即恢复正常状态，而是在连续多帧满足低风险条件后逐级降低风险等级，以避免目标短暂遮挡或瞬时漏检造成画面快速恢复。

风险升高：S0 安全 → S1 观察 → S2 预警 → S3 干预

风险降低：S3 干预 → S2 预警 → S1 观察 → S0 安全

3.5.5 状态机与显示防护联动

状态机仅负责风险判定，不直接完成视频处理。其输出以控制指令形式发送至显示处理路径，由 HDMI 显示线程执行相应的防护操作。不同状态对应的显示策略如下：

- S0 安全状态：保持原始 HDMI 画面正常显示；
- S1 观察状态：保持正常显示，仅在后台继续累积检测证据；
- S2 预警状态：显示安全提示、增加动态水印或其他低影响防护措施；
- S3 干预状态：根据系统配置执行局部遮挡、画面模糊、黑屏及告警记录等高强度防护措施。

通过将环境感知、风险判定与显示执行相互解耦，系统能够避免目标检测线程直接控制显示链路，从而提高软件结构的可维护性和后续扩展能力。

3.5.6 小结

本系统的状态机并非独立于检测模型存在，而是对多源检测结果进行时间融合和风险映射。YOLO26n 采用偏重召回的阈值策略生成候选目标，MobileNetV3 负责抑制鼠标、遥控器、卡片等 like phone 困难负样本，RGB-D 信息进一步提供目标距离与物理尺度合理性证据；状态机则通过 EMA 平滑、双阈值滞回和持续时间约束控制保护动作。该设计在机制上降低单帧误检造成的画面频繁切换风险，并保留对持续高风险目标的响应能力。

3.6 深度感知与检测融合

为降低仅依赖二维图像检测所产生的误报，系统在 YOLO26n 候选检测基础上引入 RGB-D 深度信息作为辅助约束。其中，手机候选先由 MobileNetV3 进行 ROI 分类复核；人物候选不经过 MobileNetV3，而可直接进入距离与尺寸验证。深度信息不用于直接识别

“举起手机拍摄”等复杂行为，也不承担独立目标分类任务，其主要作用是获取候选目标距离，并结合二维检测框估算实际物理尺寸，从而排除尺度明显不合理的平面伪目标。设备接口与深度数据获取方式参照 Orbbec SDK 文档[22]。

这一设计针对实际测试中出现的一类典型误检：例如在纸张上打印一个尺寸很小的手机图像，并将纸张放置在摄像头较近位置。此时，二维 RGB 图像中的手机图案可能具有较大的像素尺寸，YOLO26n 仍可能将其识别为手机候选；但通过深度相机获得目标距离后，可以结合候选框的像素宽高估算其对应的实际物理尺寸。若估算尺寸明显小于真实手机的合理范围，则将该候选目标降级或过滤，从而降低打印图像、卡片图案等目标造成的误触发。

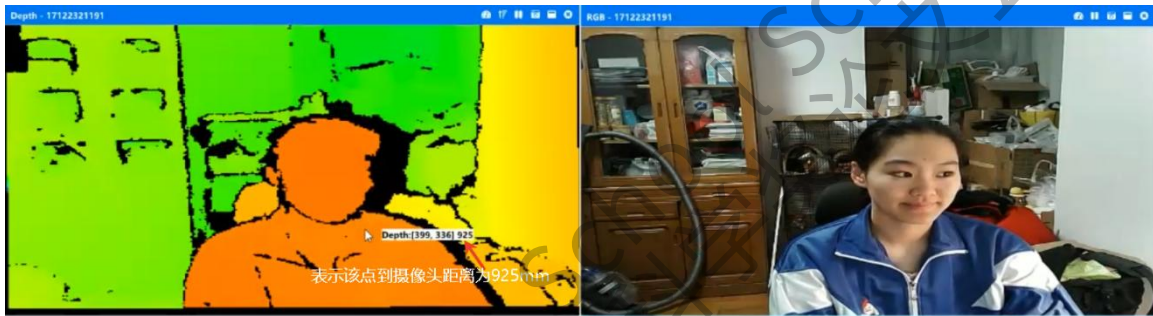


图 13 RGB 图像与对齐深度图像示意

3.6.1 候选目标获取与类别复核

系统首先使用 YOLO26n 对环境摄像头图像进行目标检测，以较低检测阈值获取人物和疑似手机候选区域。对于手机候选 ROI，经过适当扩边后缩放至 224×224 ，再由 MobileNetV3 输出 phone、like_phone 或 negative 三类结果；只有分类为 phone 的手机候选继续作为有效手机证据。对于 person 候选，则直接保留 YOLO26n 的检测结果，并可进入后续 RGB-D 距离与尺寸验证。因此，MobileNetV3 仅承担手机候选的类别复核，深度尺寸约束则可同时用于人物和手机候选。

深度模块不替代 YOLO26n 与 MobileNetV3 的视觉判断，而是在二维模型产生候选结果后增加基于距离与物理尺度的合理性检查。

3.6.2 候选区域深度获取

对于需要进行距离与尺度复核的人物或手机候选框，系统在与 RGB 图像对齐的深度图中提取对应区域的有效深度值。由于深度图可能存在无效点、边缘噪声或少量异常值，实际处理时不直接使用单个像素作为目标距离，而是选取候选框中心区域或有效深度像素集合的中位数作为目标距离估计值。

$$Z_{roi} = median(D(x, y))$$

- Z_{roi} : 候选目标的估计距离;
- $D(x, y)$: 深度图中像素 (x, y) 对应的深度值;
- $median(\cdot)$: 对候选区域内有效深度值取中位数。

采用中位数而不是简单平均值，可以降低少量无效深度点和边缘测距噪声对距离估计的影响。若候选区域有效深度像素数量过少，则该候选不直接依据深度结果作强制过滤，而是保留二维视觉结果并降低深度证据权重。

3.6.3 基于距离的实际尺寸估算

获得候选目标距离后，系统利用深度相机标定得到的相机内参，将二维检测框的像素尺寸近似换算为目标在该距离下对应的实际物理尺寸。设候选框的像素宽度和高度分别为 w_{px} 和 h_{px} ，相机在 x 、 y 方向的焦距参数分别为 f_x 和 f_y ，目标距离为 Z_{roi} ，则候选目标的实际宽度 W 和高度 H 可近似表示为：

$$W \approx w_{px} \cdot Z_{roi} / f_x$$

$$H \approx h_{px} \cdot Z_{roi} / f_y$$

- W 、 H : 候选目标估算得到的实际宽度和高度;
- w_{px} 、 h_{px} : YOLO26n 候选框在 RGB 图像中的像素宽度和高度;
- Z_{roi} : 深度图得到的候选目标距离;
- f_x 、 f_y : 相机标定得到的水平和垂直焦距参数。

上述计算并不要求得到毫米级精确尺寸，其主要用途是判断候选目标的物理尺度是否合理。例如，一张纸上打印的手机图像即使在近距离拍摄时占据较大的像素区域，由于其

实际打印尺寸可能只有数厘米，结合短距离 Z_{roi} 计算得到的 W 、 H 仍会明显小于正常手机尺寸。因此，该候选可被判定为尺寸异常，从而避免仅凭二维外观触发防护。

3.6.4 距离—尺寸联合约束

为提高尺度复核的可靠性，系统联合使用候选目标距离与二维检测框尺寸。对每个候选目标，首先由 RGB-D 数据获取距离 Z_{roi} ，再结合检测框像素宽高和相机内参估算其近似物理宽度 W 和高度 H 。若估算尺度位于对应类别的经验合理范围内，则保留该物理尺度证据；若估算尺度明显偏离常见手机或人体目标的合理范围，则过滤该候选或降低其风险权重。当前实现主要用于发现明显的“距离—尺寸”矛盾，不对候选内部三维结构或深度连续性进行显式建模。

$$W_{min} \leq W \leq W_{max}$$

$$H_{min} \leq H \leq H_{max}$$

对于手机候选，系统重点判断其估算宽高是否明显偏离常见手机的合理范围；对人物候选，仅将距离与估算尺度作为辅助证据，不据此进行精确身高或体型分类。

表 5 深度信息辅助过滤流程

处理阶段	判定内容	处理结果
YOLO26n 候选检测	检测 person 与疑似手机目标	生成人物/手机候选 ROI
MobileNetV3 复核	仅对手机候选进行 phone / like_phone / negative 分类	phone 保留；like_phone / negative 过滤
深度距离获取	对需复核的人物或手机候选提取有效深度	得到 Z_{roi}
尺寸合理性判断	利用 Z_{roi} 与检测框像素尺寸估算 W 、 H	与对应类别尺度一致则保留；明显异常则过滤或降权

3.6.5 基于深度尺度约束的平面伪目标过滤示例

在实际测试过程中，二维目标检测模型不仅可能受到手机相似物体的干扰，对于打印照片、海报以及屏幕中显示的人物图像，也可能产生错误检测。图 14 给出了一个典型测试场景：在 A4 纸张上打印人物图像，并将纸张置于摄像头前方。由于打印图像保留了人体轮廓、头部、躯干和四肢等明显视觉特征，YOLO26n 在仅使用 RGB 图像进行检测时，可能将纸张中的人物图像识别为一个独立的 person 目标。

然而，该目标在真实空间中的物理尺度与正常人体明显不符。系统利用 RGB-D 相机获取候选区域对应的深度距离 Z_{roi} ，并结合检测框在 RGB 图像中的像素宽度和高度，以及相机标定得到的内参参数，估算目标对应的实际物理宽度和高度。

对于图 14 所示的 A4 纸打印人像，二维检测器可能将纸张中的人物图案识别为 person 候选。系统依据候选距离和检测框像素尺寸估算其实际宽高；当估算尺度与正常人物目标的合理范围存在明显矛盾时，将其作为低可信伪目标过滤或降权。该判断只针对明显的“距离—尺寸”矛盾，不用于精确估计人体尺寸。

当前实现不对纸张平面或候选内部深度连续性进行显式建模；对齐深度图仅用于获取候选区域的距离，主要判定依据是由距离、检测框和相机内参得到的估算物理尺度。

这一实验说明，仅依赖二维视觉模型容易受到打印照片等平面图像干扰；通过将视觉语义与距离、物理尺寸结合，可辅助降低明显尺度异常的平面伪目标误报。对于二维外观和估算尺度均接近真实目标的仿真对象，当前规则不能保证排除。

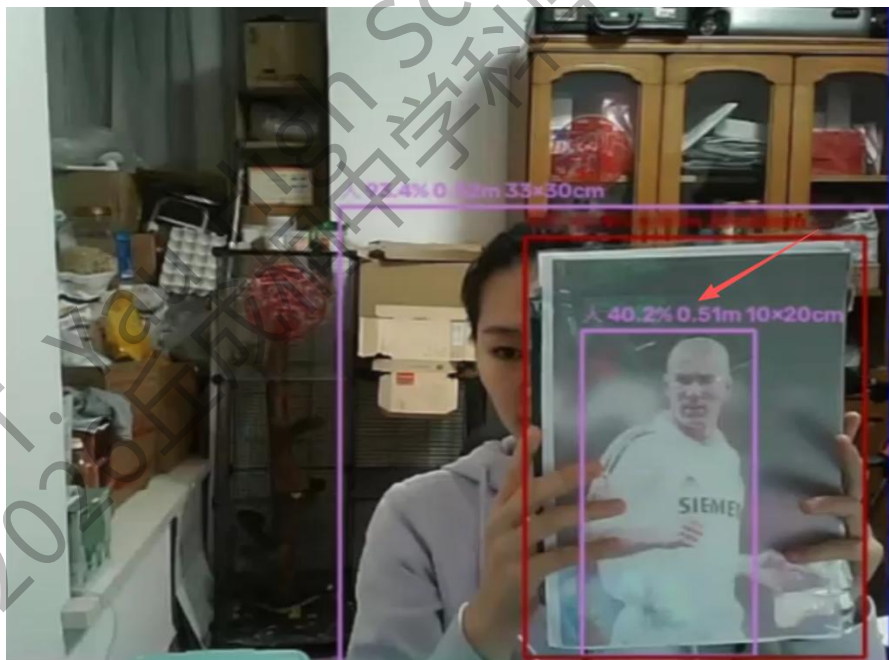


图 14 RGB-D 距离与物理尺度复核结果

3.6.6 与状态机的融合方式

深度验证模块不直接决定系统状态，而是为二维视觉结果提供距离与物理尺度合理性证据。YOLO26n 首先产生人物和手机候选，MobileNetV3 对手机候选进行语义复核，系统再结合 RGB-D 数据获取距离并估算物理尺度；人物信息只用于辅助检查明显的尺度矛盾，不承担精确人体尺寸分类。

当候选目标通过距离与尺寸一致性验证时，其检测结果被保留，并继续参与后续的多帧风险判定；当估算得到的实际尺寸明显异常时，例如 A4 纸张中的打印人物、尺寸过小的打印手机图像或其他平面伪目标，系统将该候选判定为低可信目标，并对其进行过滤或降低其风险权重。由此可避免二维模型对打印图像等伪目标的误检结果直接影响状态机决策。

经过 MobileNetV3 类别复核，并在有效 RGB-D 信息可用时补充距离与物理尺度复核后，系统将候选结果交由统一状态机进行多帧决策。手机候选主要用于风险判定；人物候选仅辅助检查明显的距离—尺度矛盾。

3.6.7 小结

综上，本研究将 RGB-D 深度信息作为二维视觉检测后的辅助排除条件。手机候选依次经过“YOLO 候选生成—MobileNetV3 类别复核—深度距离获取—粗略尺寸估算—尺度合理性过滤—状态机判定”；人物候选则由 YOLO 检测后进行距离获取与明显尺度矛盾检查。该方法用于排除明显异常的 A4 打印人像、打印手机图案等二维伪目标，不将人物尺度估计表述为精确人体分类能力。

4 实验与结果

4.1 实验平台与原型成本

实验平台由 Orange Pi 5 Plus (RK3588)、Astra Pro Plus RGB-D 摄像头、HDMI 采集与显示链路以及必要的存储、电源和连接配件组成。按项目采购时的实际记录，原型硬件成本约为 800 元。

(1) 原型成本记录：系统未额外配置独立显卡，主要硬件由 RK3588 开发板、RGB-D 摄像头及连接配件构成，项目采购时记录的原型硬件成本约为 800 元，具体成本会随采购时间和配置变化。

(2) 硬件加速能力：RK3588 集成 NPU 与 RGA，可将部分神经网络推理和图像处理任务从 CPU 侧卸载，适合持续运行。

(3) 复现条件明确：核心软硬件均可通过公开渠道获得，附录 A 给出了系统环境、依赖组件及主要复现资源。

(4) 便于独立部署：系统作为独立边缘设备运行，不依赖宿主机操作系统内部接口，适合会议、金融、医疗及研发等需要屏幕外部拍摄防护的场景。

上述平台能够支持本文从视频采集、AI 推理、深度复核到显示干预的完整原型链路，为验证主动式屏幕防拍方法的工程可行性提供了基础。

4.2 实验设置

为验证系统在实际屏幕防拍场景中的检测能力与工程可用性，本文在室内环境中搭建实验平台，重点测试人物与手机目标在不同距离条件下的检测稳定性、RGB-D 深度测量有效性以及状态机触发能力。在当前实验平台和摄像头安装条件下，0.5~2.5 m 范围内的典型测试均能够获得人物与手机的有效检测结果。

4.2.1 实验对象与设备

(1) 主要检测对象：以智能手机为主要防拍检测对象，同时记录场景中的人物目标，用于后续人员存在判断和手机风险状态分析。

(2) 扩展测试对象：平板电脑、数码相机或微单可作为其他成像设备的泛化测试对象，但不作为当前 phone 分类模型的主要训练类别。

(3) 环境感知设备：采用 RGB-D 深度摄像头同步获取 RGB 图像和深度图像，用于目标检测、目标距离测量及实际尺寸辅助估算。

(4) 边缘计算平台：采用基于 RK3588 的 Orange Pi 开发板运行 YOLO26n、MobileNetV3、深度信息处理及状态机模块。

(5) 显示设备：采用实际受保护显示器、笔记本屏幕或一体机屏幕输出待保护内容。

4.2.2 空间距离与测试位置

测试位置包括近距离 0.5 m、1.0 m、中距离约 1.5 m、2.0 m 和远距离约 2.5 m。其中，1.5 m 用于典型场景观察，不纳入本节 80 次标准化重复实验统计。

测试过程中，被测人员手持手机处于环境摄像头有效视野范围内，系统同步记录 YOLO26n 检测框、目标置信度、RGB-D 深度距离以及根据检测框与深度信息估算的目标实际尺寸。实际测试表明，在约 2.5 m 的较远测试位置，系统仍能够同时获得人物与手机的有效检测结果。

本阶段实验主要采用摄像头正前方位置，尚未开展统一重复次数的水平与垂直角度覆盖统计。

4.2.3 实验过程与记录方式

每次实验均从无目标或低风险状态开始，使被测人员携带手机进入指定测试位置，待系统完成目标检测和深度信息获取后记录结果。为减少连续视频帧被重复计数的问题，每次完整的“目标进入—检测—状态确认—目标退出”过程记为一次独立测试。

当前实验用于验证系统在不同距离下能否稳定发现人物和手机目标，以及深度信息能否提供合理的距离与尺寸估算。本文在 0.5 m、1.0 m、2.0 m 和 2.5 m 四个距离点分别

完成 20 次独立重复实验，共 80 次，并记录成功次数与检测触发率；平均响应时间尚未完成统一统计。

4.3 评价指标

结合当前系统已完成的实验内容，本文主要从目标检测有效性、深度测量结果、风险触发情况和系统响应性能等方面评价系统。尚未实际完成的视觉质量主观评价指标不纳入本阶段正式实验结果。

4.3.1 目标检测有效性

目标检测有效性用于判断人物和手机目标是否能够在当前测试条件下被正确检测。对于系统级实验，重点记录不同测试距离下是否能够稳定生成与目标位置相符的检测框，并进一步获得有效的深度信息。

4.3.2 检测触发率

本文采用检测触发率评价真实手机进入监测区域后，系统成功形成有效风险证据并进入预警或干预状态的比例：

$$\text{检测触发率} = \text{成功触发次数} / \text{总测试次数} \times 100\%$$

需要指出的是，检测触发率仅表示系统能否成功形成风险证据并完成状态切换，并不等同于最终拍摄图像不可辨识的“防护成功率”或“防拍成功率”。

4.3.3 误触发与伪目标过滤

误触发主要关注遥控器、卡片、打印照片以及其他外观相似目标是否会错误进入高风险状态。其中，MobileNetV3 用于降低手机相似物体造成的类别误报，RGB-D 深度模块用于利用距离和物理尺寸约束排除 A4 打印人像、打印手机图案等物理尺度明显不合理的平面伪目标。

4.3.4 深度测量与尺寸估算

对于获得有效深度信息的目标，记录目标距离以及根据检测框像素尺寸和相机内参估算得到的实际宽高。该指标的主要用途不是获得高精度测量值，而是验证目标物理尺度是否与实际人物或手机的合理范围一致，从而作为二维检测结果的辅助真实性约束。

4.3.5 系统响应时间

本文区分环境感知链路的单帧处理耗时与系统端到端干预时延。在 640×480 输入条件下，当前环境感知链路的单帧处理耗时约为 $100 \sim 200$ ms，主要包括图像预处理、候选检测及相关复核过程。需要特别说明的是，本文所述 $100 \sim 200$ ms 不是“风险出现到显示端完成保护”的端到端干预时延；后者还包括多帧证据累积、状态机持续时间约束以及显示动作执行，目前尚未进行统一统计。

4.4 实验结果

4.4.1 正例事件的距离重复测试

根据当前实际实验结果，在本实验平台、环境摄像头安装位置和模型参数条件下，系统在 $0.5 \sim 2.5$ m 的测试范围内均能够获得人物和手机的有效检测结果。因此，本文将 2.5 m 作为当前原型系统的主要实验有效距离上限。该结论仅针对本文实际测试条件，不表示在所有光照、遮挡和设备条件下均具有完全相同的检测性能。

为观察正例事件在不同距离下能否重复触发，实验固定摄像机位置、光照条件、模型版本、阈值和状态机参数，每次完整执行“目标进入—状态确认—目标退出”流程。 0.5 m、 1.0 m、 2.0 m 和 2.5 m 四个距离点各进行 20 次，共 80 次，结果见表 6。该实验未加入无手机负例，因此不用于计算系统级误报率。

表 6 不同距离下正例事件的重复触发结果

测试距离	独立测试次数	成功判定标准	记录指标	正例触发结果
0.5 m	20 次	检测到有效手机目标，并根据状态机规则进入预警或防护状态	成功次数、检测触发率	20/20，均完成有效风险触发。
1.0 m	20 次	检测到有效手机目标，并根据状态机规则进入预警或防护状态	成功次数、检测触发率	20/20，均完成有效风险触发。
2.0 m	20 次	检测到有效手机目标，并根据状态机规则进入预警或防护状态	成功次数、检测触发率	20/20，均完成有效风险触发。
2.5 m	20 次	检测到有效手机目标，并根据状态机规则进入预警或防护状态	成功次数、检测触发率	20/20，均完成有效风险触发。

表 6 显示, 在本文实验平台、光照条件、模型版本和参数设置下, 四个距离点的 80 次正例事件均形成有效风险证据并触发状态机响应。该结果只支持受控条件下的正例触发重复性, 不能据此推断没有手机时的误触发率, 也不代表其他光照、遮挡、拍摄姿态和硬件条件下具有相同性能。平均端到端响应时间仍需通过统一计时方法统计。

4.4.2 近距离 RGB-D 联合检测结果

图 15 给出了近距离多人环境下的典型检测结果。系统能够同时检测场景中的多个人物, 并识别手持手机目标。在该测试画面中, 手机目标的深度距离约为 1.02 m, 系统同时给出了约 20 cm $\times$ 11 cm 的尺寸估计; 人物目标也能够获得相应的距离和尺寸信息。

该结果表明, 二维目标检测结果能够与 RGB-D 深度数据形成有效对应。

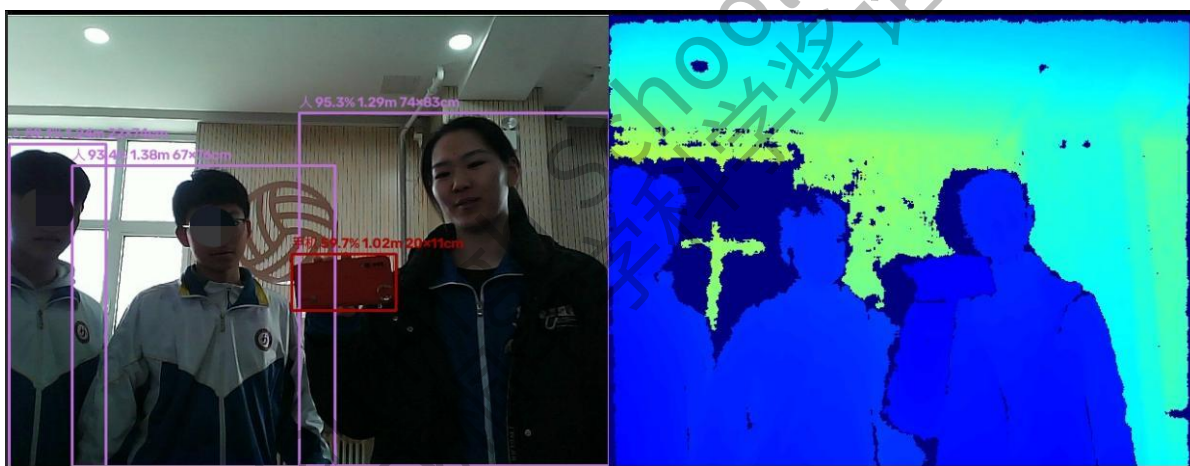


图 15 近距离多人场景下人物与手机的 RGB-D 联合检测结果

4.4.3 约 2.5 m 距离下的检测结果

图 16 给出了较远距离下的典型测试结果。此时人物目标距离约为 2.45 m, 手持手机目标距离约为 2.36 m。虽然手机在 RGB 图像中的像素面积明显小于近距离场景, 但系统仍能够形成有效手机检测框, 并从深度图中获得对应距离信息。图中手机估算尺寸约为 8 cm $\times$ 16 cm 量级, 与常见手机尺寸范围一致, 说明距离增加至约 2.5 m 时, 当前系统仍可完成手机检测和基本的物理尺度复核。



图 16 约 2.5 m 距离下人物与手机目标的检测及深度测量结果

4.4.4 多参与者真实场景检测结果

为进一步验证系统在不同人员、不同持机姿态和不同空间位置条件下的适应能力，本文邀请多名参与者在实际室内环境中进行手机持握与模拟拍摄测试。不同参与者在身高、服装、站立位置、手机型号及持机方式等方面存在差异，能够在一定程度上模拟真实应用环境中人员和拍摄方式的多样性。



图 17 多参与者、多姿态真实场景下的人物与手机检测结果

如图 17 所示, 系统能够在多人同时出现、人物尺度变化以及手机横持、竖持、单手持握和双手持握等不同情况下检测人物与手机目标。对于进入有效检测范围的目标, 系统同时输出二维检测框、检测置信度, 并结合 RGB-D 数据获得相应的距离和物理尺寸估计。

从实际测试结果可以看出, 系统并非仅针对单一测试人员或固定拍摄姿态进行优化。在不同人员、不同持机方式以及不同目标距离条件下, YOLO26n 均能够形成有效候选检测, 并为后续 MobileNetV3 复核、深度尺寸约束及状态机风险判断提供基础信息。

该组图像说明, 当前环境感知链路能够处理若干人员外观和常见持机姿态变化。由于该部分未按参与者数量、姿态类别和重复次数进行分层统计, 本文不据此计算跨参与者的系统性能指标。

4.4.5 距离因素分析

从当前实验现象看, 在 2.5 m 以内, 人物与手机目标均能够保持有效检测。随着距离增加, 手机在 RGB 图像中的像素尺寸仍会逐渐减小, 因此远距离小目标检测仍是影响系统性能上限的重要因素。但在本文已完成的约 2.36 m 手机测试中, YOLO26n 仍能够正确定位目标, RGB-D 模块也能够提供有效深度距离及合理的尺寸估算。

4.4.6 平面伪目标与深度尺寸约束

系统随后利用候选区域的 RGB-D 距离、检测框像素尺寸及相机内参估算其实际物理宽高。若估算尺度与正常人物或手机目标的合理范围存在明显矛盾, 则在进入状态机前对该候选进行过滤或降权。当前实现不对纸张平面或候选内部深度连续性进行显式建模。

4.4.7 环境感知实验小结

在当前平台和模型参数条件下, 0.5~2.5 m 范围内均获得人物与手机有效候选, 并可利用 RGB-D 信息获取距离和估算物理尺度。四个距离点的 80 次正例事件均进入预警或防护状态; 第 4.4.8 节进一步验证状态机能够把复核结果转换为显示动作。后续仍需

扩大光照、遮挡和持机姿态范围，并加入统一负例测试，统计误报率、漏报率及端到端响应时间。

4.4.8 显示干预闭环验证

为验证系统在发现风险后能够真正驱动显示保护，本文在原型联调中构造由无有效手机、出现单帧疑似候选、持续有效手机证据到高风险确认的输入序列，记录统一状态机的状态变化及显示处理路径接收的动作。该实验属于功能性闭环验证，重点检查“感知—复核—决策—干预”链路是否贯通，不将尚未进行重复统计的响应次数描述为拦截率。

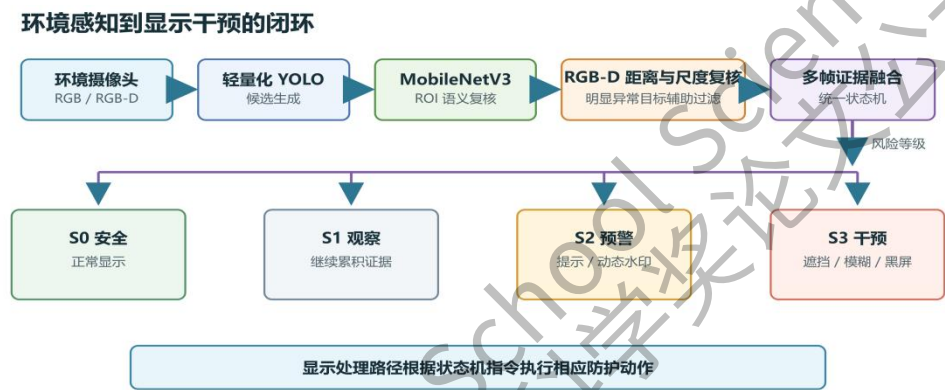


图 18 统一状态机驱动显示干预的闭环验证流程

如图 18 所示，环境感知路径仅提供经过语义与距离—尺度复核的风险证据，显示处理路径不直接响应单帧 YOLO 输出。当证据不足时，系统保持正常显示并继续观察；当风险持续累积至预警或干预阈值时，状态机才分别下发动态水印、局部遮挡、画面模糊或黑屏等动作。工程联调中，各状态对应的控制指令均能够按既定逻辑传递至显示处理路径，验证了系统从检测到干预的闭环完整性。

表 7 显示干预闭环功能验证

输入与复核条件	状态	显示处理动作	联调观察
无有效手机证据	S0 安全	正常显示	显示链路保持原画面
单帧疑似候选，证据不足	S1 观察	正常显示并累积后续帧	未因单帧候选立即干预
有效手机证据持续超过预警阈值	S2 预警	提示或增强动态水印	状态机下发预警级动作
类别复核通过，综合风险持续达到高风险条件；有效 RGB-D 信息可用时纳入距离—尺度证据	S3 干预	局部遮挡、模糊或黑屏并告警	状态机下发干预级动作

该验证说明，原型能够通过统一状态机把复核后的风险证据转换为显示动作。该项实验验证的是控制链路功能，不等同于拍摄行为识别或防拍成功率；端到端响应时间、动作成功率和用户体验影响仍需在固定试验次数下统计。

4.5 模块作用分析与案例对照

为理清 YOLO26n、MobileNetV3 与 RGB-D 在三级检测链路中的作用，本节采用三类证据：与训练目录分离的 ROI 定量结果、系统级典型案例和平面伪目标的单案例模块对照。MobileNetV3 报告独立 ROI 测试集指标；YOLO 与 RGB-D 的系统级作用通过案例说明，不把单案例外推为整套系统的固定增益。

4.5.1 MobileNetV3 困难负样本复核与定量验证

本文采用基于 ImageNet-1K 预训练权重的 MobileNetV3-Large 作为手机候选区域的 ROI 精细分类网络。具体实现基于 timm 库中的 mobilenetv3_large_100 模型，宽度倍率为 1.0，输入尺寸为 224×224 。模型初始化时加载 timm/mobilenetv3_large_100.ra_in1k 预训练权重，并根据本文数据集类别自动替换分类头，最终构建 phone、like_phone 和 negative 三分类模型。训练环境采用 timm 1.0.27、PyTorch 2.8.0 和 torchvision 0.23.0。该模型通过迁移学习方式，使网络在保留 ImageNet-1K 通用视觉表征能力的基础上，进一步学习真实手机与鼠标、遥控器、卡片等困难负样本之间的细粒度差异[3]。

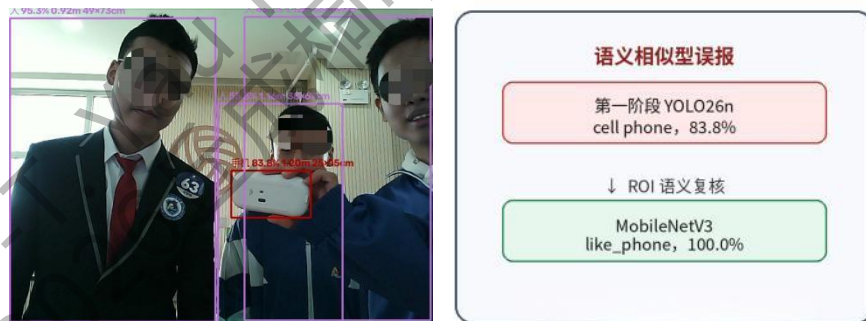


图 19 实验部署的 YOLO26n 将手持鼠标误检为手机候选的典型案例

如图 19 所示，实验部署的 YOLO26n 在多人场景中能够正确形成人物候选，但同时将被试者手持的白色鼠标误识别为手机候选。第一阶段检测结果中，该目标被判定为 cell phone，候选置信度为 83.8%。需要说明的是，该数值仅表示第一阶段检测器对该候选属于 cell phone 类别的预测置信度，并不代表系统已经完成最终手机判定。

该候选 ROI 随后输入 MobileNetV3 进行二次语义复核。根据程序运行日志，MobileNetV3 将该目标判定为 like_phone，分类置信度为 100.0%。因此，该候选在第二阶段被归入非手机类别，不会作为有效 phone 证据进入统一状态机，也不会直接参与后续高风险防护状态的触发。

程序日志对应结果为：YOL026n 判定 cell phone，置信度 83.8%；MobileNetV3 判定 like_phone，置信度 100.0%。

该案例说明，两阶段检测结构能够将第一阶段候选检测器的目标发现能力与第二阶段细粒度分类能力结合起来。即使鼠标等外观相似物体在第一阶段获得较高的手机候选置信度，经过针对困难负样本训练的 MobileNetV3 复核后，仍可被识别为非手机目标，从而降低此类相似物体直接进入高风险状态机所造成的误触发风险。

模型训练完成后，在与训练目录分离的 test 目录上对 1,420 张 ROI 图像进行测试。测试集由人工复核标注，类别标签经过二次检查，包含 phone、like_phone 和 negative 三类。

表 8 MobileNetV3 与训练目录分离的 ROI 测试集混淆矩阵

真实类别	预测 phone	预测 like_phone	预测 negative
phone	487	5	3
like_phone	2	422	1
negative	4	3	493

根据表 8 的混淆矩阵，模型在 1,420 张 ROI 图像上的总体准确率为 98.73%；phone 类 Precision 为 98.78%，Recall 为 98.38%，F1-score 为 98.58%；非手机 ROI 误报为 phone 的比例为 0.65%。这些结果说明 MobileNetV3 能够有效过滤大部分 like_phone 和 negative 候选，但该结果仍属于与训练目录分离的 ROI 测试，不直接等同于完整视频系统的端到端检测性能。

4.5.2 典型场景下各模块的判定分工

表 9 汇总了三类代表性场景中各模块的输出与系统最终状态。YOL026n 负责以偏重召回的阈值策略产生候选，MobileNetV3 在语义层排除鼠标等外观近似物，RGB-D 则利用目标距离与估算物理尺寸提供尺度合理性约束。各模块的输出最终由统一状态机融合，而不是由任一单帧检测结果直接触发显示干预。

表 9 典型场景下各模块判定结果对比

场景	YOL026n	MobileNetV3	RGB-D 距离与物理尺度复核	最终状态
真实手机拍摄	phone（高置信度）	phone	目标距离与估算尺寸处于合理范围	预警/干预

场景	YOLO26n	MobileNetV3	RGB-D 距离与物理尺度复核	最终状态
手持鼠标等近似物	phone	like_phone	语义层已排除，无需进入尺度复核	安全
手持照片中手机图案	phone（高置信度）	phone	估算尺寸约 $5\text{ cm} \times 2\text{ cm}$ ，明显偏离真实手机合理尺寸范围	安全

表 9 表明，三个模块处理的是不同类型的不确定性：第一阶段解决“是否值得继续检查”，第二阶段解决“二维外观上是否为手机”，第三阶段解决“该候选的距离—尺寸关系是否符合真实手机的合理物理尺度”。这种分工使系统既能保留真实手机候选，又能分别抑制语义相似物和物理尺度明显异常的平面伪目标。

4.5.3 平面伪目标的单案例模块对照

为观察 RGB-D 距离与物理尺度约束对平面伪目标的作用，本文构造一个案例：测试者手持一幅包含人物与手机图案的平面照片，并使其位于摄像头有效感知范围内。图 20 对比 RGB 通路的候选检测结果与对齐深度图。



图 20 手持平面照片中“手机”伪目标的模块对照与物理尺度复核结果

(a) YOLO26n 将照片中的“手机”图案检测为手机候选，MobileNetV3 复核后仍判定为 phone；(b) 对齐深度图中，照片主体区域的深度值整体较为接近。系统进一步获取“手机”候选区域的距离，并结合检测框像素尺寸和相机内参估算其实际物理尺寸约为 $5\text{ cm} \times 2\text{ cm}$ ，明显偏离常见真实手机的合理尺寸范围。

该案例记录依次加入 MobileNetV3 和 RGB-D 物理尺度复核后的判定变化，具体如下。

(1) 配置 A，仅使用 YOLO26n。YOLO26n 在 RGB 图像中以较高置信度检测到照片中的“手机”图案并输出边界框。由于图案轮廓完整、二维外观与真实手机相近，候选检测器无法仅依据二维图像判断其是否为真实拍摄设备，系统因而存在误报风险。

(2) 配置 B, 加入 MobileNetV3 语义复核。MobileNetV3 对候选 ROI 进行细粒度分类后仍输出 phone。照片中的手机图案在纹理、颜色和长宽比等二维特征上与真实手机接近, 说明仅依赖二维分类不足以稳定排除印刷照片或屏幕显示图像。

(3) 配置 C, 加入 RGB-D 物理尺度复核。系统从“手机”候选区域获取距离 Z_{roi} , 并结合候选框像素宽高和相机内参估算实际物理尺寸。本案例的估算值约为 $5\text{ cm} \times 2\text{ cm}$, 明显小于常见真实手机的合理尺寸范围, 形成“距离—尺寸”矛盾。RGB-D 模块据此对该候选降权或过滤, 统一状态机维持安全状态。对齐深度图中照片主体区域的深度值整体接近, 可作为平面属性的辅助观察; 当前实现未对候选内部三维结构或深度连续性进行显式建模。

需要说明的是, 图 20 中的“ $5\text{ cm} \times 2\text{ cm}$ ”为当前测试条件下依据 RGB-D 距离、检测框和相机内参得到的近似物理尺寸。该数值主要用于发现明显尺度矛盾, 而非进行高精度几何测量; 若复制目标的二维外观和估算物理尺寸均接近真实手机, 仅依赖当前距离—尺寸规则不能保证排除。

4.5.4 案例结果与实验边界

综合表 8、图 19、表 9 与图 20 可以看出, 三级检测链路具有明确且互补的职责: YOLO26n 提供候选入口; MobileNetV3 在语义层排除鼠标、遥控器等困难负样本; RGB-D 在物理层利用目标距离与估算尺寸识别明显的尺度异常。删除 MobileNetV3 会增加相似物误报及后续复核负担; 删除 RGB-D 尺度约束时, 二维语义上接近真实手机的打印图案更容易形成错误风险证据。

本节的 MobileNetV3 指标来自与训练目录分离的 1,420 张 ROI 测试图像; RGB-D 的作用说明来自平面伪目标的单案例模块对照。受样本规模与实验周期限制, 本文尚未在统一端到端测试集上完成配置 A、B、C 的多轮重复统计, 因此不报告未经验证的系统级误报率或漏报率差值。后续可固定设备、距离、光照和重复次数, 统一量化各配置的触发率、误报率、漏报率与端到端响应时间。

5 讨论

本文围绕屏幕外部拍摄这一传统信息安全措施难以完全覆盖的泄露途径, 构建了一套基于边缘人工智能的主动式屏幕防拍系统。当前实验表明, 该组合能够在 RK3588 原型平台上形成完整闭环, 并在典型误报案例中提供多级过滤能力; 其在复杂开放场景中的稳定性仍需通过更大规模标准化测试进一步验证。

5.1 两阶段检测机制对误报控制的作用

单独依赖目标检测模型时，遥控器、卡片、手机壳、计算器以及其他细长矩形物体可能因外观与手机相似而被误识别。为此，本文将目标检测与细粒度分类分离：YOLO26n 采用偏重召回的阈值策略产生候选区域，MobileNetV3 则进一步将候选目标划分为 phone、like_phone 和 negative。

这种两阶段结构在保持第一阶段候选召回倾向的同时，对实际部署中的困难负样本进行进一步过滤，避免单纯提高检测阈值带来的漏检风险。第 4.5 节与训练目录分离的 ROI 测试中，MobileNetV3 在 1,420 张图像上取得 98.73% 的准确率和 98.38% 的手机召回率，非手机 ROI 误报率为 0.65%；图 19 的鼠标案例则直观展示了二次复核对高置信度误检的拦截作用。RGB-D 的系统级增益仍需通过统一测试集上的定量对照实验进一步测定。

5.2 RGB-D 深度信息的辅助价值

只看二维图像时，模型有时难以区分真实手机和照片中的手机。因此，本研究加入 RGB-D 深度信息，在判断目标外观的基础上，再检查它的距离和实际尺寸是否合理。修正后的程序利用检测框确定搜索范围，再根据深度前景估计目标的长短边，不直接把检测框的宽高当作手机尺寸。

深度信息也有局限。尺寸明显异常的手机图案，可以通过尺度检查进一步排除；但如果复制图案与真实手机大小接近，仅靠深度仍可能无法区分。另外，测到一个大小合理的手机，并不能说明它正在拍摄屏幕。因此，深度模块主要提供补充判断依据，不能代替拍摄行为识别。其实际效果还需要在完成相机标定和图像对齐检查后，通过实机测试确认。

5.3 边缘端软硬件配合的意义

系统运行速度不仅取决于检测模型。视频采集、图像缩放、MobileNetV3 复核和显示输出都需要时间，即使 YOLO 推理较快，其他环节也可能拖慢整个系统。

本项目以 RK3588 为边缘平台，保留了 YOLO 的 RKNN/NPU 部署路径。当前 MobileNetV3 在板端仍通过 CPU 运行，因此，不能用 YOLO 的推理耗时代表整套系统的响应时间。后续需要分别测量各环节耗时，找出主要影响速度的部分，再判断是否需要进一步优化。

这也说明，选择模型时，除了准确率，还要考虑模型大小、设备是否支持相关运算，以及部署后的实际速度。模型更新，并不一定意味着它更适合当前设备。

5.4 连续帧确认与黑屏保持的作用

如果单帧检测到手机就立即黑屏，一次短暂误检也可能打断正常使用。为此，当前程序对手机风险设置了连续帧确认：在默认启用 MobileNetV3 的情况下，连续三帧都有候选通过视觉复核和深度尺度检查，才产生手机告警。中间某一帧不满足条件，计数就会清零。这里确认的是每帧是否存在风险目标，并没有跟踪同一台手机。

告警触发后，程序保持黑屏；出现新的告警时重新计时，没有新告警时则在最后一次告警五秒后恢复显示。这一设置是为了减少画面反复切换，但也会延长显示中断的时间。连续帧确认能够在规则上过滤孤立的单帧结果，不过它究竟能减少多少误触发、会增加多少响应延迟，还需要通过统一测试来评价。

5.5 系统局限性

虽然当前系统已经能够实现较完整的主动防拍闭环，但仍存在一定局限性。本研究主要用于验证主动式屏幕防拍系统的工程可行性，并不建立能够覆盖所有拍摄威胁和使用环境的绝对安全模型。

首先，远距离条件下手机在环境摄像头图像中的像素尺寸会持续减小，小目标检测能力仍是影响系统有效工作距离的重要因素。受当前实验场地空间限制，本文标准化重复实验的最远测试距离为约 2.5 m，尚未对更远距离条件下的系统性能进行系统验证，因此不能据此确定系统的最大有效检测距离。同时，在严重遮挡、极端光照或拍摄设备仅露出局部区域等情况下，仍可能出现候选目标漏检或识别置信度下降。

其次，当前实验主要针对摄像头正前方及若干典型位置进行了验证，尚未按照统一测试次数对不同水平角度和垂直俯仰角条件下的检测性能进行统计。因此，本文不对系统的最大角度覆盖范围作定量结论，后续仍需通过标准化角度实验进一步评估不同观察方向对手机候选检测和风险触发能力的影响。

再次，RGB-D 深度测量可能受到测量距离、目标表面反射特性、红外干扰、遮挡以及无效深度点等因素影响。因此，本文仅将 RGB-D 获取的距离及由此估算的物理尺度作为辅助复核证据，而不将其作为唯一的风险判断依据。当前“距离—尺寸”规则主要用于发现物理尺度明显异常的打印图像等伪目标；若复制目标的二维外观和估算物理尺寸均接近

真实手机，仅依赖现有尺度约束不能保证将其排除，这类高仿真伪目标仍需结合更多传感信息、目标姿态或行为证据进一步判断。

此外，当前系统主要针对环境中可见的智能手机等拍摄设备进行检测。对于远距离长焦设备、微型隐藏摄像头、被其他物体严重遮挡的摄像设备，以及当前分类模型未充分覆盖的其他成像设备，系统仍无法保证有效发现。因此，本文所研究的方法更适用于可见外部摄像设备威胁下的主动屏幕防护，而不能替代针对隐藏摄像头或其他特殊窃密设备的专用检测手段。

环境感知摄像头在检测手机的同时，也可能拍到周围人员和其他无关信息。因此，实际使用时应尽量在设备本地完成图像分析，避免上传和长期保存原始视频。对于需要保留的告警记录，也应控制内容和数量，并明确谁可以查看、保存多久、何时删除。具体管理方式还需要根据使用场景确定。

已有原型展示了从环境图像采集、目标检测到显示干预的基本流程，但这并不代表系统在各种条件下都能可靠工作。当前多级复核版本仍需进一步实测，包括不同距离和角度下的检测效果、复杂环境中的误报与漏报，以及对接近真实手机尺寸的复制图案的判断能力。这些测试需要采用统一的记录和统计方法，才能比较不同方案的实际效果。

5.6 应用价值与扩展方向

本方案通过独立的边缘设备处理图像和显示信号，不必把全部防护功能安装在被保护的主机上。这有助于减少对 Windows、Linux 等操作系统的依赖，但不同主机和显示设备之间的兼容性仍需测试。

这类方案可以面向会议室、实验室、研发场所等环境开展进一步测试。金融、医疗等场景也有保护屏幕信息的需求，不过，这些场景对可靠性和隐私保护要求较高，不能仅凭目前的原型结果就判断系统已经适合实际使用。

后续改进应先解决当前方法的不足。例如，增加手机朝向和持机姿态的判断，帮助区分正常使用手机与疑似拍屏；在不同水平、垂直角度下重复测试，确定摄像头能够有效观察的范围。完成这些验证后，再考虑结合其他传感器或加入屏摄鲁棒水印，探索拍摄后追查信息来源的可能性。

其中，屏摄鲁棒身份水印和复杂三维行为识别目前只是后续研究方向，本文不将它们列为已经实现或验证的功能。

6 结论

本文针对禁用截屏、静态水印和物理防窥难以覆盖外部设备直接拍摄屏幕的问题，设计并实现一套基于 RK3588 的屏幕拍摄风险检测与显示干预原型。

系统采用“双通路、单决策”架构：显示处理通路与环境感知通路相互独立，统一状态机集中决策。环境感知部分以可替换的轻量化 YOLO 模型生成候选，实验阶段采用 YOLO26n 验证流程；随后由 MobileNetV3 对手机 ROI 进行 phone、like_phone 和 negative 三分类复核，并利用 RGB-D 获取候选目标距离，结合检测框与相机内参估算物理尺度，完成尺度合理性复核。统一状态机结合多帧证据驱动不同等级的显示保护动作。

在系统实现方面，本文围绕 RK3588 的硬件特点构建了由 NPU、RGA 和 DMA-BUF 等模块组成的边缘端处理链路。其中，NPU 负责神经网络推理，RGA 承担图像预处理与缩放，DMA-BUF 用于减少不同模块之间的数据复制，从而使环境感知与 HDMI 显示链路能够并行运行。统一状态机综合多帧检测结果及辅助验证信息，将风险划分为安全、观察、预警和干预等不同状态，并根据风险等级驱动水印提示、局部遮挡、画面模糊或黑屏等显示保护措施。

原型已在 RK3588 边缘设备上完成闭环联调。640×480 环境感知链路单帧处理耗时约为 100~200 ms；该数值不包括多帧证据累积、状态机持续时间约束和显示动作执行。0.5 m、1.0 m、2.0 m 和 2.5 m 四个距离点的 80 次正例事件均形成有效风险证据并触发状态机响应。MobileNetV3 在与训练目录分离的 1,420 张 ROI 测试图像上取得 98.73% 的准确率、98.38% 的手机召回率和 0.65% 的非手机 ROI 误判为 phone 的比例；鼠标案例经二次复核后未进入 phone 分支。平面伪目标单案例和显示干预联调分别说明了距离—尺度复核与状态驱动控制的作用。

本研究完成了从候选检测、ROI 复核、距离—尺度验证和多帧决策到边缘端显示控制的原型实现，并分别记录了模块指标、案例结果和系统联调现象。现有证据支持该技术链

路在受控条件下可运行，但尚不足以给出开放环境中的系统级误报率、漏报率和防拍成功率。

当前系统仍存在远距离小目标检测能力有限、深度数据受环境条件影响以及对隐藏式拍摄设备检测能力不足等问题。后续研究可进一步提高小目标检测性能，扩充不同设备、人员和场景的数据集，并研究手机朝向、人体姿态和多模态行为信息的融合方法。同时，可进一步引入具有屏摄鲁棒性的身份水印技术，使系统由当前的主动风险检测与显示干预，进一步扩展至泄露内容的事后追踪。

综上，该原型在受控条件下实现了“环境感知—目标复核—风险决策—显示干预”闭环，为后续开展统一负例、复杂环境、角度覆盖和端到端时延测试提供了可复核的工程基础。

7 参考文献

7.1 学术论文与预印本

- [1] 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.
- [2] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 7263-7271.
- [3] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 1314-1324.
- [4] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. 2020. arXiv:2004.10934.
- [5] MA Y, YU C, LIU G, et al. Real-time anti-photography system based on convolutional neural network model[J]. SID Symposium Digest of Technical Papers, 2025, 56(S1): 349-352. DOI:10.1002/sdtp.18804.
- [6] 王晓媛, 张文涛, 韩磊. 基于 Faster R-CNN 的屏幕防手机拍摄算法研究[J]. 航天控制, 2019, 37(2): 66-72.
- [7] 汪嘉恒. 面向防拍摄的摄像检测技术[D]. 杭州: 浙江大学. 2017
- [8] GE S, XIA Z, TONG Y, et al. A screen-shooting resilient document image watermarking scheme using deep neural network[EB/OL]. 2022. arXiv:2203.05198.
- [9] SHOTTON J, FITZGIBBON A, COOK M, et al. Real-time human pose recognition in parts from single depth images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2011: 1297-1304.
- [10] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common Objects in Context[C]//European Conference on Computer Vision. 2014: 740-755.

7.2 专利

- [11] 一种水印添加方法、装置及屏幕图像泄漏追踪方法、装置[P]. CN113496451B, 2024.
- [12] 吴天鹏, 刘晓稳, 肖峰, 等. 一种显示器屏幕防拍方法、装置、设备及存储介质[P]. CN117786775A, 2024.
- [13] 张志祥. 一种防拍摄装置[P]. CN114598848A, 2022-06-07.
- [14] 贾玉凤, 胡双喜, 杨才山, 等. 一种高安全屏幕防拍摄装置[P]. CN221615040U, 2024-08-27.
- [15] 胡凡, 张建, 纪朋朋. 基于图像处理的屏幕防拍摄系统及方法[P]. CN117037270A, 2023-11-10.
- [16] 金施群, 邢金玉, 王行芳, 等. 一种基于频率差的屏幕防偷拍设备[P]. CN106407827A, 2017-02-15.
- [17] 李云飞. 一种电子显示屏幕防拍照系统及方法[P]. CN111639592A, 2020-09-08.
- [18] 陶冠宏, 范振军. 一种新型屏幕防拍照识别方法[P]. CN114863337A, 2022-08-05.

7.3 模型论文、专利补充、官方文档与工程资料

- [19] ROCKCHIP. RGA hardware manual (RGA2) [R]. Rockchip official technical documentation.
- [20] ROCKCHIP. GStreamer user guide: gstreamer-rockchip and rkximagesink[R]. Rockchip official technical documentation.
- [21] ULTRALYTICS. Models supported by Ultralytics[EB/OL]. <https://docs.ultralytics.com/models/>.
- [22] ORBBEC. Orbbec SDK documentation[EB/OL]. <https://orbbec.github.io/OrbbecSDK/>.
- [23] GSTREAMER PROJECT. DMA buffer sharing (DMA-BUF) design documentation[EB/OL]. <https://gstreamer.freedesktop.org/documentation/>.
- [24] JOCHER G, QIU J, LIU M, et al. Ultralytics YOLO26: Unified Real-Time End-to-End Vision Models[EB/OL]. 2026. arXiv:2606.03748. DOI:10.48550/arXiv.2606.03748.
- [25] 一种主动防偷拍的检测及显示设备[P]. CN118194380A, 2024.

8 致谢

作为一名已经高三的女生，能够把一个从好奇心出发的想法，逐步做成可以运行的软硬件原型，对我来说是一段很特别的经历。

我最早关注屏幕防拍摄问题，是因为中考后在参观贵州平塘 FAST 时多次看到“禁止拍摄”的提示。当时我开始思考：电脑可以禁止截屏，但如果有人直接用手机拍摄屏幕，现有措施还能起到多大作用？带着这个问题，我从 2024 年开始尝试进行研究。项目早期，我在 Windows 笔记本和 GPU 环境下使用 YOLO11n 搭建手机检测原型。最初看到程序能够识别手机时，我非常兴奋，但实际测试很快暴露出许多问题：鼠标、遥控器、卡片等物体会被误认为手机，单纯增加训练样本或调整阈值，也很难同时兼顾真实手机的检出和相似物的过滤。

为了处理这类误报，我收集实际运行中出现的候选区域，将样本划分为 phone、like_phone 和 negative 三类，并训练 MobileNetV3 对 YOLO 输出的 ROI 进行二次复核。后来我又发现，打印照片或屏幕画面中的手机图案，有时仍会同时通过 YOLO 和 MobileNetV3 的二维判断。针对这一问题，我进一步加入 RGB-D 深度相机，利用候选目标距离、检测框像素尺寸和相机内参估算物理尺度，对尺寸明显异常的平面复制目标进行过滤。之后，我又将检测、复核、深度门控、连续帧判断和黑屏控制逐步移植到 RK3588 边缘计算设备上。整个项目经历了从 YOLO11n 单阶段检测，到 YOLO26n 与 MobileNetV3 语义复核，再到加入 RGB-D 距离—尺度约束和边缘端显示干预的多次迭代。

在这个过程中，我越来越清楚地认识到，屏幕拍摄风险检测并不是更换一个模型或提高一个指标就能够解决的问题。数据采集、样本标注、目标检测、分类模型、深度测量、硬件接口、显示链路和程序控制之间都会相互影响。项目中出现过误检、模型加载失败、检测框偏大、深度无效、程序速度下降以及边缘设备兼容等问题。面对这些问题，我需要重新检查原始图像、程序日志、模型输出和硬件运行状态，逐项排查原因。虽然这一过程有时很反复，但也让我逐渐学会了如何记录实验、核对证据、区分推测与事实，并对自己写入论文的每一个结论负责。

本项目为个人参赛项目。作为参赛学生，我负责选题提出、资料检索、研究方案设计、样本采集与标注、困难负样本整理、模型训练、程序编写与调试、RGB-D 距离和尺度计算、RK3588 硬件搭建与部署、实验设计、测试记录整理、数据计算、图表制作以及论文撰写和修改。我也负责核对不同阶段的代码、模型和实验记录，区分早期基础链路结果与后期多级复核链路结果，避免把不同版本的数据混合为未经验证的整机结论。

于喆老师是本校指导教师，与我是正常的师生指导关系，对本项目的指导为无偿指导。本项目未接受培训机构、公司或外包团队的有偿研究服务。于喆老师主要在研究问题梳理、技术路线检查、实验设计审阅、科研规范、论文结构、展示视频制作和模拟答辩等方面提出意见，帮助我发现表述不清、实验依据不足、数据口径不一致以及部分结论超出实际实现范围的问题。指导教师没有代替我编写核心程序、训练模型、采集或处理实验数据，也没有代替我形成研究结论。

项目研究中遇到的主要困难，也推动了技术方案的形成。早期单阶段检测器会持续误报鼠标、遥控器和卡片，我没有只依靠提高检测阈值，而是回收误检 ROI，建立 like_phone 类并重新训练分类模型；打印照片中的手机图案仍可能通过二维语义判断，因此我加入距离和物理尺度门控；在 RK3588 上部署时，模型格式、NPU 兼容、深度相机、HDMI 输入输出和运行速度又形成了新的问题，我通过查阅技术文档、拆分程序模块、保存运行日志和逐项测试进行排查。项目持续时间较长，不同阶段的代码和实验条件也发生过变化，因此我在论文修改时重新核对原始记录，对尚未在当前完整程序上复测的结果进行单独说明，不将其写成已经完成的整机性能。

感谢参与系统测试的老师、同学和其他人员。不同参与者在身高、服装、手机型号、站立位置和持握方式等方面存在差异，为原型系统提供了更加多样的测试条件，也帮助我发现多人场景、不同距离和不同持机姿态下的问题。

在项目研究和论文撰写过程中，我也阶段性使用了大语言模型等人工智能工具。自 2024 年起，我曾使用 ChatGPT（早期界面显示为 GPT-4）、豆包和 DeepSeek，主要用于提供文献检索线索、解释技术概念、分析程序报错、讨论代码修改思路和辅助润色部分文字；在后期论文和代码集中修改阶段，还使用过 WPS AI 助手和 OpenAI Codex，辅助核对参赛规则、整理章节结构、检查语言与版式、分析代码与论文表述是否一致、草拟测试用例和图表说明。上述工具为阶段性、非连续使用，没有固定的每日或每周使用频率；对于没有保存或无法确认的更细版本信息，我不会进行推测或补造。

需要特别说明的是，本项目是一项具有明显设备依赖性的软硬件系统研究。人工智能工具不能直接连接和操作本项目使用的 RK3588 开发板、NPU、RGA、HDMI 输入输出链路和 RGB-D 深度相机，也不能替代模型转换、硬件接线、驱动适配、深度对齐、实机计时、持续运行观察和真实场景重复实验。因此，人工智能工具给出的代码、参数、性能判断和硬件配置建议，只能作为问题排查和思路讨论的参考，不能直接作为实验结果。凡涉及模型能否运行、目标距离、物理尺度、推理速度、连续帧触发、黑屏控制和系统稳定性等内容，均以实际设备运行、程序日志、原始图像和实验记录为依据；当人工智能建议与实机结果不一致时，以实际测试结果为准。

人工智能工具没有用于生成、修改或替代实验数据，也没有代替真实模型训练和硬件实验。项目的选题动机、研究目标、技术路线、样本和实验事实、关键方案选择以及最终结论，均由我依据实际研究过程确定。对于人工智能工具辅助生成或修改的文字、公式、代码、测试用例和图表，我需要先理解其含义，再与源代码、技术文档、原始数据和设备运行结果逐项核对；没有亲自核验、无法复现或不能解释的内容，不应进入最终提交材料。最终论文由我确认，我对其中研究内容的真实性、准确性和学术规范性负责。

感谢学校为本项目提供学习、实验和展示环境，感谢于喆老师在研究过程中给予的耐心指导，也感谢所有参与测试并给予帮助的老师、同学和其他人员。这个项目陪伴了我高中阶段很长一段时间，也让我从最初只关注“模型能不能检测到手机”，逐步学会关注问题是怎样产生的、证据是否充分、程序是否真正实现，以及结论是否能够经得起核查。

谨向所有在本项目研究和成长过程中给予我指导、帮助和支持的人，表示诚挚的感谢。

附录 A 复现实验环境与资源清单

本附录记录报告中已明确的软件、硬件和公开资源。复现时还需逐项核对候选阈值、状态机阈值、持续时间、相机内参、权重文件校验值和随机种子；本报告不对原始记录中未明确的参数作推断。

(1) 边缘平台：Orange Pi 5 Plus (RK3588)。

http://www.orangepi.cn/orangepiwiki/index.php/Orange_Pi_5_Plus

(2) 操作系统：Ubuntu 24.04，所用镜像版本页面如下。

<https://github.com/Joshua-Riek/ubuntu-rockchip/releases/tag/v2.4.0>

(3) 训练与评估环境：timm 1.0.27、PyTorch 2.8.0、torchvision 0.23.0。

(4) OpenCV 5.x 源码与编译资源：<https://github.com/opencv/opencv>

(5) GStreamer RGA 转换插件：<https://github.com/BlueBrightWind/gstreamer-rgaconvert>

(6) MobileNetV3 预训练权重标识：[timm/mobilenetv3_large_100.ra_in1k](https://github.com/timm/mobilenetv3_large_100.ra_in1k)。

(7) COCO 数据集：<https://cocodataset.org/#download>

(8) 补充手机图像数据集来源：<https://universe.roboflow.com/boxes-dataset-minmax/mobilephones/browse>

(9) YOLO26 官方说明：<https://docs.ultralytics.com/models/yolo26/>

(10) Ultralytics 官方实现：<https://github.com/ultralytics/ultralytics>

附录 B 专利初步审查合格通知书



国家知识产权局

100080 北京市海淀区苏州街1号9层9132号 北京鲁班天下专利代理有限公司 张露(16710390762)	发文日: 2024年12月24日
 	
申请号或专利号: 202411677870.2	发文序号: 2024121900456360
申请人或专利权人: 孙嘉一	
发明创造名称: 一种显示屏防拍摄方法及装置	

发明专利申请初步审查合格通知书

上述专利申请,经初步审查,符合专利法实施细则第50条的规定。
申请人于2024年11月22日提出提前公布声明,经审查,符合专利法实施细则第52条的规定,专利申请进入公布准备程序。

初步审查合格的上述发明专利申请是以:

2024年11月22日提交的说明书摘要
2024年11月22日提交的权利要求书
2024年11月22日提交的说明书
2024年11月22日提交的说明书附图
为基础的。

提示:

- 1.发明专利申请人可以自申请日起3年内提交实质审查请求书、缴纳实质审查费,申请人期满未提交实质审查请求书或者期满未缴纳或未缴足实质审查费的,该申请被视为撤回。
- 2.专利费用可以通过网上缴费、银行/邮局汇款、直接向代办处或国家知识产权局专利局缴纳。缴费时应当写明正确的申请号/专利号、费用名称及分项金额,未提供上述信息的视为未办理缴费手续。了解缴费更多详细信息及办理缴费业务,请登录国家知识产权局官方网站。

审查员:朱雪冬
联系电话:010-53960259

审查部门:专利审查协作北京中心



210304
2023.03

纸件申请,回函请寄:100088 北京市海淀区蓟门桥西土城路6号 国家知识产权局专利局受理处收
电子申请,应当通过专利业务办理系统以电子文件形式提交相关文件。除另有规定外,以纸件等其他形式提交的文件视为未提交。

附录 C 软件著作权登记受理通知书



软件登记受理通知书

流水号: 2025R11L0803413

软件名称: 显示屏幕防拍摄系统V1.0

登记类型: 计算机软件著作权登记

申请人: 孙嘉一

根据《计算机软件著作权登记办法》第十九条的规定, 对申请人提出的上述计算机软件著作权登记申请, 中国版权保护中心予以受理。

受理号: 2025R11S1068500

经核实确认中国版权保护中心收到如下申请文件:

打印签字或盖章的登记申请表	4
程序鉴别材料 - 一般交存	80
文档鉴别材料 - 一般交存	20
著作权人 - 孙嘉一的居民身份证复印件	1
其他相关证明文件	5

中国版权保护中心软件著作权部
2025年05月22日

附录 D 多参与者真实场景测试样例



图 D-1 多参与者、多姿态、多距离条件下的人物与手机检测结果汇总