Joy Seeker

JoySeeker知美而行

基于深度学习的自动化人像航拍解决方案

深度学习期末汇报

刘锦松、廖彦铭、罗则宁、肖淞元、杨蒻、赵宸昊、张郡杰、邹贤哲

2025 年 12 月 22 日

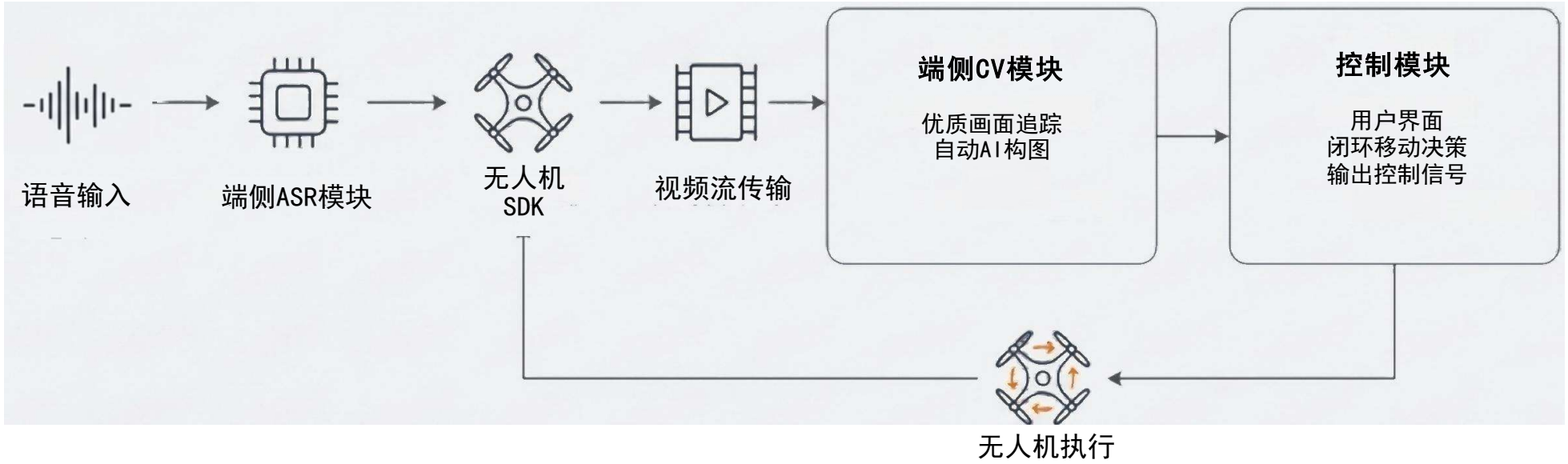
传统活动摄影业务

行业背景

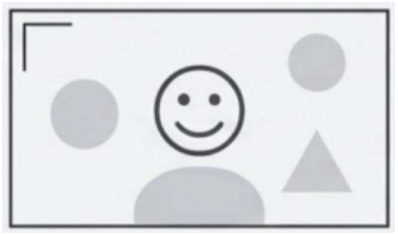

MINI 5 PRO


IF
极客公园
创新大会
2026
Founder Park 特别环节
AI 产品快闪
Theme:
拍照「邪修」
AI 构图
Finch
Doka 相机
创始人
GEEKPARK
INNO FEST
2026

技术方案


语音输入
端侧ASR模块
无人机
SDK
视频流传输
端侧CV模块
优质画面追踪
自动AI构图
控制模块
用户界面
闭环移动决策
输出控制信号
无人机执行

深度学习策略：对象识别



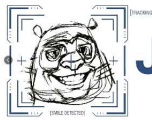
阶段一：目标定位

- 技术：OpenCV Haar Cascade
- 作用：从完整画面中检测并裁剪出人脸区域

阶段二：表情分类

- 技术：MobileNetV3(Large,Pre-trained)
- 作用：对裁剪出的人脸图像进行精细化的7分类表情识别。

表情分类：数据集选择（Kaggle）



“开心”样本：





“中性”样本：





“悲伤”样本：





表情分类：模型选型（MobileNet V3）

Joy Seeker

Input	Operator	exp size	#out	SE	NL	s
$224^2 \times 3$	conv2d	-	16	-	HS	2
$112^2 \times 16$	bneck, 3x3	16	16	-	RE	1
$112^2 \times 16$	bneck, 3x3	64	24	-	RE	2
$56^2 \times 24$	bneck, 3x3	72	24	-	RE	1
$56^2 \times 24$	bneck, 5x5	72	40	✓	RE	2
$28^2 \times 40$	bneck, 5x5	120	40	✓	RE	1
$28^2 \times 40$	bneck, 5x5	120	40	✓	RE	1
$28^2 \times 40$	bneck, 3x3	240	80	-	HS	2
$14^2 \times 80$	bneck, 3x3	200	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	480	112	✓	HS	1
$14^2 \times 112$	bneck, 3x3	672	112	✓	HS	1
$14^2 \times 112$	bneck, 5x5	672	160	✓	HS	2
$7^2 \times 160$	bneck, 5x5	960	160	✓	HS	1
$7^2 \times 160$	bneck, 5x5	960	160	✓	HS	1
$7^2 \times 160$	conv2d, 1x1	-	960	-	HS	1
$7^2 \times 960$	pool, 7x7	-	-	-	-	1
$1^2 \times 960$	conv2d 1x1, NBN	-	1280	-	HS	1
$1^2 \times 1280$	conv2d 1x1, NBN	-	k	-	-	1

Table 1. Specification for MobileNetV3-Large. SE denotes whether there is a Squeeze-And-Excite in that block. NL denotes the type of nonlinearity used. Here, HS denotes h-swish and RE denotes ReLU. NBN denotes no batch normalization. s denotes stride.

模型架构

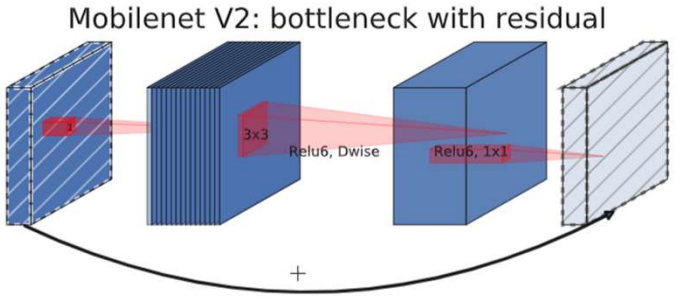


Figure 3. MobileNetV2 [39] layer (Inverted Residual and Linear Bottleneck). Each block consists of narrow input and output (bottleneck), which don't have nonlinearity, followed by expansion to a much higher-dimensional space and projection to the output. The residual connects bottleneck (rather than expansion).

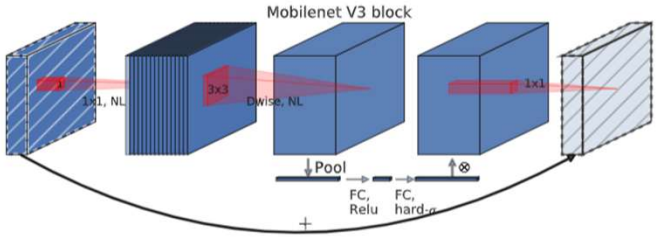


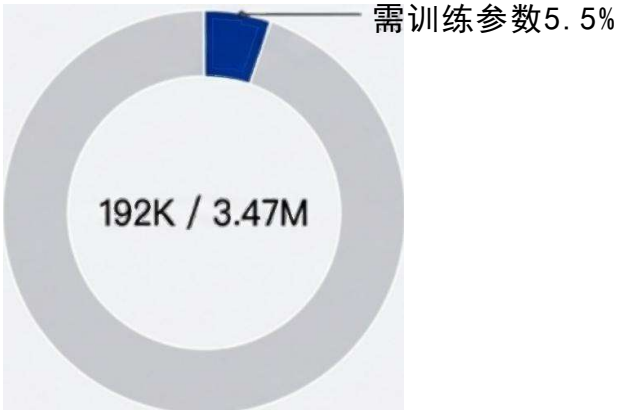
Figure 4. MobileNetV2 + Squeeze-and-Excite [20]. In contrast with [20] we apply the squeeze and excite in the residual layer. We use different nonlinearity depending on the layer, see section 5.2 for details.

Bottleneck块

表情分类：迁移学习训练

Joy Seeker

- **迁移学习：**使用ImageNet 预训练权重，**冻结模型前70%的层**，仅微调与任务强相关的顶层参数。
- **参数效率：**总参数3.47M，实际需训练参**数仅192K（占比5.5%）**，大幅缩短训练时间并有效防止过拟合。

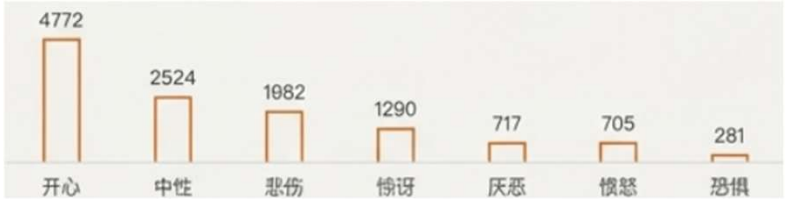


数据增强（DA）

实施了随机水平翻转、随机旋转、色彩抖动（亮度、对比度）、随机擦除等多种变换。

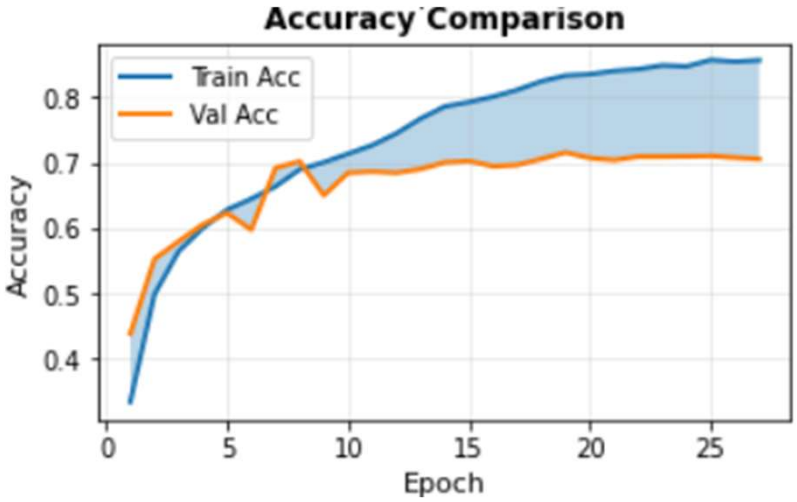
损失函数优化

Focal Loss，加大对“开心”类别的惩罚权重



表情分类：训练效果

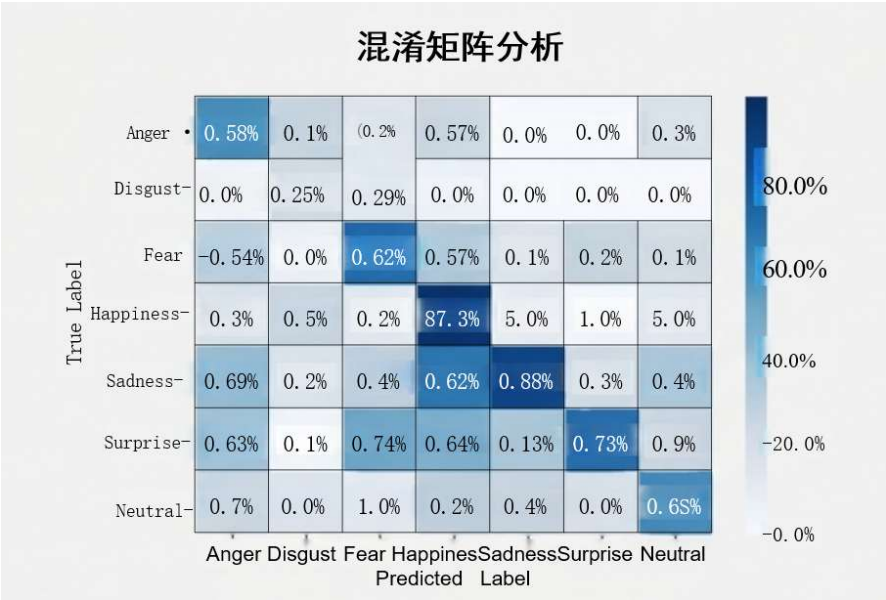
Joy Seeker



表情分类：模型验证

Joy Seeker

类别	精确率	召回率	F1分数
愤怒	0.5870	0.5000	0.5400
厌恶	0.2582	0.2938	0.2749
开心	**0.8387**	**0.8734**	**0.8557**
恐惧	0.5493	0.5270	0.5379
悲伤	0.6990	0.6025	0.6472
惊喜	0.6331	0.7447	0.6844
中性	0.6739	0.6382	0.6556
加权平均	**0.7078**	**0.7070**	**0.7059**



工程实现：稳定滤波器

原始AI输出的问题：不稳定

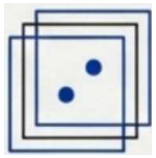
时间抖动

开心

非开心

时间抖动： 用户的微表情、瞬间的光影变化或模型偶尔的误判，会导致识别结果在“开心”与“非开心”之间高频切换。

空间抖动



空间抖动： OpenCV 人脸检测框的位置在每一帧都有微小像素的跳变。

若直接用于无人机控制→导致飞行指令的剧烈震荡和目标频繁丢失，用户体验极差。

我们的解决方案：稳定滤波器

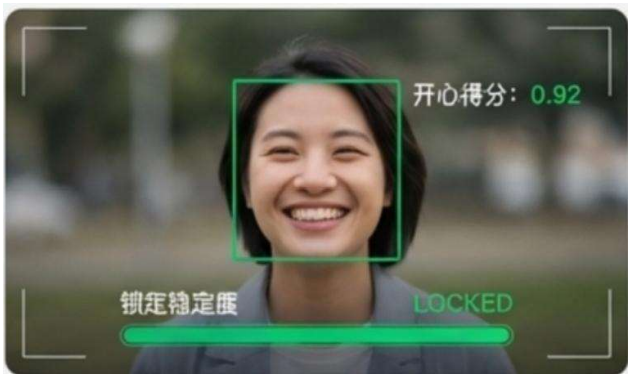
时间防抖 (StabilityFilter)

- **机制：引入“能量条”累积机制。**
- **逻辑：**只有当模型连续在多个帧中检测 到“开心”时，能量条才会充满，系统才确认“锁定”目标。单帧的 **抖动或丢失不会立即取消锁定。**

空间防抖 (SpatialStabilizer)

- **机制：**位置平滑滤波。
- **逻辑：**结合当前帧检测到的目标中心点和历史帧的位置，通过加权平均输出一个平滑、稳定的目标坐标，滤除高频位置跳变。

产品实现



应用场景与展望

个人/家庭用户：打造“AI摄影师”功能。自动跟拍Vlog、旅行记录、家庭聚会、宠物互动，解放用户双手。

活动/赛事主办方：提供低成本的自动化拍摄方案。在婚礼、派对、小型体育比赛中，多机位自动捕捉精彩瞬间。

更丰富的构图理解：从识别表情扩展到理解全身姿态、多人互动关系，实现更复杂的电影级运镜（如过肩镜头、拉伸镜头）。

模型持续轻量化：探索LoRA、模型量量化**等前沿技术，进一步降低**功耗，在同等电池下实现更长续航**或搭载更复杂的模型。**

跨领域应用拓展：将核心技术迁移至**智能安防**（识别异常行为）、机器人交互（感知用户情绪）、**智慧零售**（分析顾客满意度）等**颁域。**

Joy Seeker

Q&A

谢谢大家！