

# 组会汇报

2025 年 11 月 25 日

张鹏

厦门大学航空航天学院



## ① 课题背景与目标

## ② 数据与方法

## ③ 实验设计与结果

## ④ 问题与讨论

## ⑤ 下一步计划

## ⑥ 参考文献

## ① 课题背景与目标

项目背景

研究目标与指标

## ② 数据与方法

## ③ 实验设计与结果

## ④ 问题与讨论

## ⑤ 下一步计划

## ⑥ 参考文献

## 项目背景

## 项目背景

- 智慧安防、智能医疗等场景迫切需要准确的多模态行为理解能力。

## 项目背景

## 项目背景

- 智慧安防、智能医疗等场景迫切需要准确的多模态行为理解能力。
- 现有方案在跨场景泛化和轻量部署方面表现不足，缺乏可解释的融合策略。

## 项目背景

## 项目背景

- 智慧安防、智能医疗等场景迫切需要准确的多模态行为理解能力。
- 现有方案在跨场景泛化和轻量部署方面表现不足，缺乏可解释的融合策略。
- XMU 航空航天学院承担的纵向项目要求在 2024 年底前完成核心算法的工程化验证。

## 研究目标与指标

## 研究目标与指标

## 阶段目标

- 构建覆盖视觉、语音、惯导传感的多模态数据集与基准管线。
- 设计鲁棒的表征学习与跨模态对齐策略，支撑复杂行为识别。
- 打通端到端推理流程，实现 30 FPS 以上的实时识别能力。

## 评估指标

- 精度：Top-1 准确率  $\geq 88\%$ ，F1-score  $\geq 0.85$ 。
- 性能：推理延迟  $\leq 33ms$ ，模型大小  $\leq 120$  MB。
- 部署：支持 Jetson Orin 与 X86 GPU 的双平台运行。

## ① 课题背景与目标

## ② 数据与方法

数据来源与采集

系统方案概览

## ③ 实验设计与结果

## ④ 问题与讨论

## ⑤ 下一步计划

## ⑥ 参考文献

## 数据来源与采集

## 数据来源与采集

数据模态	采集频率	样本量	备注
RGB 视频	30 FPS	1 200 Clips	室内/户外混合
IMU 传感	200 Hz	3 600 序列	穿戴式采集
音频	16 kHz	1 800 片段	附带噪声标签

表 1: 本阶段已完成的数据采集与清洗情况

- 数据标注采用三人交叉审核机制，保障事件级别一致性。
- 通过自动脚本生成时间戳对齐文件，方便后续融合。

## 系统方案概览

## 系统方案概览

- 视觉分支：基于 Video Swin Transformer，支持剪枝与蒸馏。
- 语音分支：采用 Conformer 编码器，与视觉特征共享注意力。
- 传感器分支：轻量级 TCN，强调时间依赖建模。
- 融合策略：多尺度对齐 + 动态门控融合，配合多任务损失。

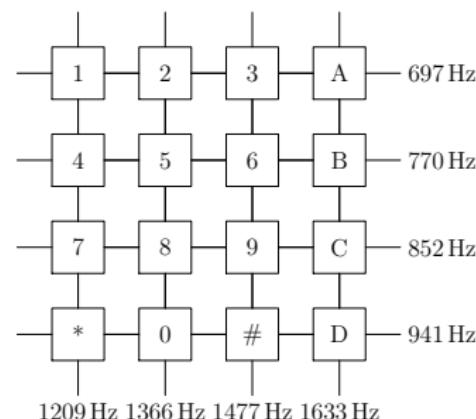


图 1：流程示意：数据同步  
→ 特征提取 → 跨模态融合  
→ 行为预测

## ① 课题背景与目标

## ② 数据与方法

## ③ 实验设计与结果

实验设置

性能对比

可视化与案例

## ④ 问题与讨论

## ⑤ 下一步计划

## ⑥ 参考文献

## 实验设置

## 实验设置

- 训练环境:  $8 \times$  A100, AdamW 优化器, Cosine 学习率, batch=64。
- 数据增强: 视觉使用 RandAugment+Mixup, 音频使用 SpecAugment, IMU 使用抖动/裁剪。
- 评价协议: 5 折交叉验证 + 跨场景留一验证, 确保泛化性。
- 发布两条对照实验: 单模态基线、全模态融合。

## 性能对比

## 性能对比

方法	Top-1(%)	F1-score	时延 (ms)
视觉单模态	80.6	0.74	21.3
语音 +IMU	82.1	0.77	18.9
全模态融合 (当前)	<b>87.9</b>	<b>0.84</b>	28.5
目标 (2024Q4)	88.0+	0.85+	33 以下

表 2: 关键结果对比，融合模型已逼近项目考核指标

- 视觉分支限制了总体表现，后续需进一步蒸馏与剪枝。
- 推理延迟仍受多模态对齐开销影响，需要算子优化。

# 可视化与案例

## 典型案例

- 多人协同：融合模型在“交接物品”场景准确率提升 12%.
- 异常行为：通过 IMU 通道捕获微动信息，降低误报。
- 噪声环境：语音分支结合噪声标签提升鲁棒性。



图 2: 示意：可根据实际需求替换为行为可视化截图

## ① 课题背景与目标

## ② 数据与方法

## ③ 实验设计与结果

## ④ 问题与讨论

挑战与风险

讨论要点

## ⑤ 下一步计划

## ⑥ 参考文献

# 挑战与风险

- 数据问题：户外场景样本不足，需补采夜间与极端光照数据。
- 算法问题：跨模态对齐仍然依赖手工超参，自动调度策略缺失。
- 工程问题：Jetson 端推理频率不稳定，TensorRT 插件仍需优化。
- 协同问题：生态合作方接口迟迟未定，影响整合测试。

## 讨论要点

## 讨论要点

## 需要决策

- 是否投入额外预算扩充夜间采集?
- 模型蒸馏优先支持视觉还是 IMU 分支?
- 平台优化优先 Jetson 还是 X86?

## 拟定支持

- 申请学院 GPU 集群的额外夜间使用时段。
- 与合作院校共享噪声标注工具链。
- 争取工程团队协助推理框架合并。

## ① 课题背景与目标

## ② 数据与方法

## ③ 实验设计与结果

## ④ 问题与讨论

## ⑤ 下一步计划

里程碑安排  
本周行动项

## ⑥ 参考文献

## 里程碑安排

## 里程碑安排

- 6月：完成夜间与极端场景数据补采，固化清洗规范。
- 7月：实现跨模态自适应对齐模块，完成蒸馏实验。
- 8月：交付 Jetson 优化版本，时延压缩到 25 ms 以内。
- 9月：联合生态合作方开展系统联调和用户测试。
- 10-11月：撰写论文与技术报告，准备验收材料。

## 本周行动项

## 本周行动项

- 数据组：推进夜间场景采集脚本开发。
- 算法组：完成多模态门控网络初版上线。
- 工程组：定位 TensorRT 插件的内存泄漏。

## 风险缓解

- 预留 2 天回归窗口验证性能波动。
- 维护实验记录与配置，避免重现实验困难。

## ① 课题背景与目标

## ② 数据与方法

## ③ 实验设计与结果

## ④ 问题与讨论

## ⑤ 下一步计划

## ⑥ 参考文献

参考文献

课题背景与目标

○○○

数据与方法

○○○

实验设计与结果

○○○○

问题与讨论

○○○

下一步计划

○○○

参考文献

○●

致谢

○

## 参考文献

*Thanks!*