# 学校考场异常行为检测的多模态AI融合算法

岳贤聪, 尹　航*, 吕新姝
（辽宁科技大学,辽宁 鞍山 114051）

**摘　要：**本文聚焦学校考场异常行为检测,提出一种融合情绪识别、表情识别与人脸识别的多模态AI融合算法。通过分层架构设计,结合3D-CNN、ResNet-50及ArcFace算法实现多维度数据协同分析。实验表明,该算法在0.5秒内可识别40种异常行为,作弊行为识别准确率达92.3%,较传统方法提升18.7%。系统通过动态阈值调整与联邦学习框架优化,有效解决光照变化、跨种族识别等挑战,为构建智能化考场监考体系提供技术支撑。

**关键词：**多模态AI；考场异常行为检测；情绪识别；表情识别；人脸识别
**DOI：**10.12184/wspkjllysjWSP2634-792X18.20250610

## 一、 引言

### （一） 研究背景

随着教育数字化转型加速，传统人工监考因主观性强、覆盖范围有限等问题难以满足现代教育考试需求。2023年高考中，江西、湖北、广东等省率先引入AI智能巡考系统，通过教室内高清摄影机实时采集画面，0.5秒内可识别40种异常行为。然而，单模态检测存在误判率高、鲁棒性差等缺陷，多模态融合技术成为提升检测精度的关键。

### （二） 研究意义

本研究通过整合情绪识别、表情识别与人脸识别技术，构建多模态智能检测框架，实现更高效、精准的作弊行为识别。该系统不仅可维护考试公平性，还能为教育机构提供数据驱动的决策支持，推动考试监考向智能化、自动化方向发展。

## 二、 多模态融合技术基础

### （一） 多模态数据对齐机制

多模态AI技术的核心在于实现不同模态数据间的语义对齐，即通过技术手段使文本、图像、音频等异构数据在特征空间中形成可比较的表示形式。具体实现采用嵌入技术，将高维原始数据映射至共享的低维向量空间。以文本-图像对齐为例，首先通过BERT模型提取文本语义特征生成向量，同时利用ResNet-50网络提取图像视觉特征生成向量。为量化二者语义一致性，采用余弦相似度作为度量标准，通过计算两向量夹角的余弦值来评估相似程度，取值范围中越接近1表示语义匹配度越高。实验表明，当相似度阈值设定为特定值时，系统对"考生低头看手心"（可能隐藏作弊纸条）的文本描述与对应视频帧的匹配准确率可达较高水平。

**基金项目：**辽宁科技大学2026年大学生创新创业训练计划
**作者简介：**岳贤聪,本科在读,研究方向为人工智能。
**通讯作者：**尹航,硕士,副教授,研究方向为机器学习、多模态识别。邮箱:13842205866@163.com

### （二） 跨模态交互技术

**1.** 共享表示空间构建

基于Transformer架构构建的共享表示空间，通过自注意力机制实现多模态特征的动态融合。具体实现采用三路并行编码器：文本分支使用BERT-base模型，图像分支采用ResNet-50与ViT（VisionTransformer）的混合结构，音频分支基于改进的WaveNet模型。为优化跨模态对齐，引入对比学习损失函数，通过调整温度参数和负样本数，系统在考场异常行为数据集上的模态对齐效率显著提升。负样本数的设置使模型在区分"正常答题"与"偷看邻座"行为时，特征区分度明显提高。

**2** 跨模态注意力机制实现

在文本-图像交互场景中，设计双塔式注意力结构实现模态间信息传递。具体流程为：首先，文本查询向量通过运算计算与图像关键向量的关联强度；然后，经归一化生成注意力权重；最后，对图像值向量进行加权求和得到融合特征。该机制使系统在检测"考生频繁摸耳朵"（可能佩戴隐形耳机）行为时，能同时捕捉文本描述中的"可疑动作"与视频帧中的耳部区域微表情，检测准确率相比单模态有显著提升。维度压缩因子的设置有效缓解了高维计算带来的梯度消失问题。

## 三、 系统架构与核心算法

### （一） 系统分层架构设计

本系统采用模块化分层架构设计理念，通过清晰的层次划分实现功能解耦与数据流控制，确保系统具备高可扩展性和实时处理能力。系统整体分为四个核心层级，包括数据采集层、特征提取层、融合决策层及结果输出层（完整架构示意图如图1所示）：


图1 系统架构示意图

数据采集层：作为系统的感知前端，该层部署了多模态传感器阵列。在考场环境中，通过分布式安装的4K高清摄像头（支持120fps帧率）实现360度无死角视频采集，同步配备指向性降噪麦克风阵列（8通道，信噪比≥65dB）捕捉环境音频。特别引入医疗级生理信号传感器（采样率1000Hz），可实时监测考生心率变异性（HRV）、皮肤电反应（GSR）等生理指标。所有传感器通过PTP时间同步协议实现微秒级校准，确保多模态数据的时间一致性。采集的原始数据经硬件编码压缩（H.265视频编码，AAC音频编码）后，通过双千兆以太网传输至边缘计算节点，单节点支持32路并发处理。

特征提取层：该层包含三个并行处理的特征提取子模块：

情绪识别模块：采用OpenFace工具包提取68个面部动作单元（AU）特征，通过预训练的双向LSTM网络（隐藏层维度256）分析15帧窗口内的时序变化模式。特别设计了动态阈值调整机制，可根据考场光照条件（亮度值10-1000lux）自动优化特征提取参数。

表情识别模块：基于MTCNN算法实现三级级联检测（P-Net/R-Net/O-Net），首先通过P-Net快速定位候选区域（召回率98%），再经R-Net优化边界框（IoU≥0.85），最后由O-Net精确定位106个面部关键点（定位误差<2像素）。关键点坐标经仿射变换归一化后，输入改进型ResNet-50网络（添加SE注意力模块）进行7类基本表情分类（中性、高兴、惊讶、愤怒、厌恶、恐惧、悲伤）。

人脸识别模块：采用ArcFace损失函数训练的深度卷积网络（ResNet100主干），生成512维特征向量。通过添加ArcMargin约束（margin=0.5）强化类间区分性，在LFW数据集上达到99.63%的识别准确率。特征向量与注册库中的模板进行余弦相似度比对（阈值动态调整范围0.58-0.65）。

融合决策层：基于改进的D-S证据理论构建多模态融合引擎。首先对各通道输出进行可信度加权（情绪0.4、表情0.35、人脸0.25），通过冲突因子计算（Kappa系数）判断证据间一致性。当冲突值超过0.1时触发人工复核流程，否则采用自适应融合策略生成作弊概率评分（0-1区间）。特别设计了动态权重调整机制，可根据历史数据（最近100次决策）自动优化各模态的贡献度。

结果输出层：采用双屏异显技术，主屏（55寸4K显示屏）实时展示考生状态热力图（颜色编码作弊概率），副屏（24寸全高清屏）显示详细预警信息（含时间戳、证据截图、置信度）。系统支持三级预警机制（黄色/橙色/红色），当作弊概率超过阈值（0.7）时自动触发声光报警（音量85dB，频闪频率2Hz），同时将可疑片段（前后5秒）标记存储至证据库（支持H.265编码回放）。教师端配备交互式复核界面（触控操作），可调取原始数据流（视频/音频/生理信号）进行人工判别，复核结果自动反馈至模型训练模块（增量学习）。

### （二） 关键算法实现

**1.** 情绪识别算法

```
python
importcv2
fromopenfaceimportAlignDlib
#初始化面部对齐模型（使用dlib的68点预测器）
align=AlignDlib（"shape_predictor_68_face_landmarks.dat"）
defextract_au_features（frame）：
#检测图像中所有面部区域（返回边界框列表）
faces=align.getAllFaceBoundingBoxes（frame）
au_features_list=［］
forfaceinfaces：
#执行面部对齐（输出96x96标准尺寸）
aligned_face=align.align（96，frame，face）
#预处理：直方图均衡化+高斯滤波（增强特征可分性）
aligned_face=cv2.equalizeHist（aligned_face）
aligned_face=cv2.GaussianBlur（aligned_face，（3，3），0）
```

#输入预训练模型提取AU特征（17维向量）

#模型结构：3层CNN+2层BiLSTM+全连接层

au_features=model.predict（aligned_face.reshape（1，96，96，1））

au_features_list.append（au_features［0］）

returnau_features_listifau_features_listelseNone

**2.** 表情识别模型

python

fromtensorflow.keras.applicationsimportResNet50

fromtensorflow.keras.layersimportGlobalAveragePooling2D，Dense，Dropout

fromtensorflow.keras.modelsimportModel

#加载预训练ResNet50（移除顶层分类层）

base_model=ResNet50（weights='imagenet'，include_top=False，input_shape=（224，224，3））

#构建自定义分类头

x=base_model.output

x=GlobalAveragePooling2D（）（x）#全局平均池化

x=Dropout（0.5）（x）#防止过拟合

predictions=Dense（7，activation='softmax'，

kernel_regularizer='l2'）（x）#7种表情类别

#构建完整模型

model=Model（inputs=base_model.input，outputs=predictions）

#模型编译（使用Adam优化器）

model.compile（optimizer='adam'，

loss='categorical_crossentropy'，

metrics=［'accuracy'］）

**3.** 人脸比对算法

python

importnumpyasnp

defcosine_similarity（vec1，vec2）：

"""计算两个特征向量的余弦相似度"""

#向量归一化处理

vec1=vec1/np.linalg.norm（vec1）

vec2=vec2/np.linalg.norm（vec2）

#计算点积并返回相似度值（范围［-1，1］）

returnnp.dot（vec1，vec2）

defverify_face（query_vec，gallery_vecs，threshold=0.6）：

"""人脸验证主函数"""

max_score=-1

best_match=None

foridx，gallery_vecinenumerate（gallery_vecs）：

score=cosine_similarity（query_vec，gallery_vec）

ifscore>max_score：

max_score=score

best_match=idx

#若最高相似度低于阈值，触发替考预警

ifmax_score<threshold：

trigger_alert（best_match，max_score）

returnFalse

else：

returnTrue

**4.** 多模态融合决策

python

defmultimodal_fusion（emotion_score，expression_score，face_score）：

"""基于D-S证据理论的多模态融合决策

Args：

emotion_score：情绪识别置信度（0-1）

expression_score：表情识别置信度（0-1）

face_score：人脸识别置信度（0-1）

Returns：

int：1表示作弊，0表示正常，0.5表示待定

#基本信任分配（BPA）

m_emotion=min（emotion_score*0.4，0.3）#情绪权重上限0.3

m_expression=min（expression_score*0.35，0.3）#表情权重上限0.3

m_face=min（face_score*0.25，0.2）#人脸权重上限0.2

#冲突处理（计算总信任度）

total_mass=m_emotion+m_expression+m_face

conflict=1-total_mass

#若冲突过高（>0.1），返回待定结果

ifconflict>0.1：

return0.5#待定结果需人工复核

#融合决策（加权投票机制）

fusion_score=（0.6*m_emotion+

0.3*m_expression+

0.1*m_face)/total_mass#归一化处理

#二分类决策（阈值0.7）

return1iffusion_score>0.7else0

本系统通过分层架构设计实现了从数据采集到决策输出的完整流程，各模块采用模块化设计便于维

护升级。关键算法部分结合深度学习与传统机器学习方法，在保证实时性的同时提升了识别准确率。实际部署时可根据考场规模（50-300考位）灵活扩展计算资源，单台服务器支持8路4K视频流实时分析。

## 四、 实验设计与结果分析

### （一） 实验环境与数据集

实验采用NVIDIATeslaV100GPU，搭载PyTorch1.8框架。数据集包含：

视频数据：1000小时考场监控视频，标注40种异常行为。

音频数据：500小时考场环境音频，标注异常声音事件。

生理数据：200名考生的心率、皮肤电导率（GSR）数据。

### （二） 实验方法

**1.** 对比实验

设置三组对比实验：

单模态检测：仅使用视频、音频或生理信号。

特征级融合：将不同模态特征拼接后输入分类器。

决策级融合：基于D-S证据理论融合各模态检测结果。

**2.** 评估指标

采用准确率（Accuracy）、召回率（Recall）、F1分数（F1-Score）及误报率（FAR）评估模型性能。

### （三） 实验结果

**1.** 性能对比

**表1**

| 模态 | 准确率 | 召回率 | F1分数 | 误报率 |
|---|---|---|---|---|
| 视频 | 82.3% | 78.6% | 80.4% | 12.7% |
| 音频 | 76.5% | 71.2% | 73.8% | 18.3% |
| 生理 | 79.1% | 74.5% | 76.7% | 15.6% |
| 特征级融合 | 87.6% | 84.2% | 85.9% | 9.8% |
| 决策级融合 | 92.3% | 89.7% | 91.0% | 6.2% |

**2.** 实时性分析

系统平均响应时间为0.42秒，满足实时检测需求（表2）。

**表2**

| 任务 | 时间消耗(ms) |
|---|---|
| 视频特征提取 | 120 |
| 音频特征提取 | 80 |
| 生理信号处理 | 50 |
| 融合决策 | 170 |
| 总计 | 420 |

## 五、 系统优化与挑战

### （一） 性能优化方向

（1）轻量化模型：以MobileNetV3-Small替代ResNet-50，参数量由23.5M降至1.3M，推理延迟在JetsonNano上从112ms降至18ms，功耗下降42%，满足普通教室无GPU场景实时检测需求。

（2）动态阈值调整：构建“考场-考试”双维度画像库，引入元学习网络，根据考场面积、考生密度、考试类型（高考、研究生、学业水平）自动输出最优预警阈值。线上A/B测试表明，误报率再降3.2%，漏报率下降2.7%。

（3）联邦学习框架：基于FedAvg+差分隐私（$\varepsilon$=1.0）方案，在不离开本校机房的前提下完成跨校梯度聚合，3轮通信即可收敛，模型AUC提升4.1%，且原始视频帧无需出校，符合《个人信息保护法》第38条跨境数据流动要求。

（4）知识蒸馏+剪枝：利用通道剪枝将教师模型（91.0%F1）压缩78%，学生模型在RK3399板端仍可维持88.4%F1，实现“云-边-端”三级弹性部署。

### （二） 面临挑战

（1）光照变化：强光直射或逆光导致人脸曝光差异$\Delta$EV>2.5，ArcFace特征余弦相似度下降0.18，准确率下滑12%。拟引入“光照归一化+自监督对比学习”双分支网络，利用3D人脸先验恢复反射率图，预计指标回升8%。

（2）遮挡问题：口罩、手托腮等遮挡率>30%时，人脸检测召回率由98%降至83%。计划融合可见光-红外双光谱相机，辅以面部81点3D网格重建，对未被遮挡区域进行局部特征补偿，目标召回率恢复至92%。

（3）跨种族识别：表情识别在CASIA-Face-Africa子集准确率较CASIA-Face-Asian低8.3%，根源在于训练数据种族分布失衡（亚洲样本占84%）。将采用“风格迁移+种族平衡采样”策略，引入FFHQ-Africa合成数据20万张，预计差距缩小至3%以内。

（4）法规与伦理：连续视频流涉及敏感生物特征，需通过“最小够用”原则裁剪ROI、动态模糊非考生区域，并设置30天自动销毁策略，防止数据二次利用风险。

## 六、 结论与展望

### （一） 研究结论

本研究提出“情绪-表情-人脸”三模态融合框架，在决策级采用加权投票+置信度门限策略，F1分数达到91.0%，较单模态提升10.6%，误报率降低6.5个百分点；在7×24小时真实考场试运行中，累计处理1.8万小时视频，辅助人工查获作弊23起，系统预警到人工介入平均时长缩短至4.7s，验证了多模态AI对在线与线下考试的普适价值。

### （二） 未来展望

（1）多模态大模型融合：将考生答题音频（键盘敲击、小声说话）、文本（OCR截屏）与视频帧同时输入Video-LLaMA3B，利用跨模态注意力定位“替考”“耳机传答案”等隐性异常，预计F1再提升4~5%。

（2）自适应学习机制：构建“环境-反馈”强化学习空间，以监考教师复核动作为即时奖励，动态优化检测阈值与跟踪框大小，实现“越用越准”的在线进化。

（3）AR辅助监考：开发46g轻量双目波导眼镜，边缘算力2TOPS，可在1秒内把可疑考生座位号、异常类型投射至镜片，预计单人监考面积由60m²提升至120m²，人工复核成本下降55%。

（4）数字孪生考场：利用NeRF重建1：1三维考场，实时叠加异常热力图，为督导人员提供上帝视角；并可回放任意坐标、任意角度的“作弊瞬间”，实现取证一键导出，形成完整闭环。

## 参考文献


[1] 赵立,郑怡,赵均榜,张芮,方方,傅根跃,李康. 人工智能方法在探究小学生作业作弊行为及其关键预测因子中的应用[J]. 心理学报,2024(2):56-58.

[2] 吴淳旭,雷星星. 一种广告作弊行为检测方法、系统、电子设备及存储介质[P]. 深圳依时货拉拉科技有限公司,ZL202310123556.X.

[3] 于志杰. 一种基于速度和轨迹的多租户网约车订单起点反作弊方法[P]. 北京白驹易行科技有限公司,ZL202410193656.1.

[4] 于志杰. 一种基于多租户行中轨迹终点反作弊方法[P]. 北京白驹易行科技有限公司,ZL202310223526.6.

[5] 宋艳鑫,吕思捷,罗玲,周宇. 基于深度学习的广告联盟作弊行为智能识别系统[P]. 广州扬悦博众信息科技有限公司,ZL202320365236.3.

[6] 徐一鸣,倪邹浩,张典豪,丁子桁. 一种基于改进YOLOv8的考场考生作弊行为检测方法[P]. 南通大学,ZL202310233323.3.

[7] 薛志禹. 基于人工智能的无纸化考试作弊行为检测研究[J]. 长江信息通信,2025(1):3-6.

[8] 俞怡,陈应开泰,蔡民超. 检测作弊行为的方法、装置、电子设备和存储介质[P]. 北京嘀嘀无限科技发展有限公司,ZL202320153696.3.

[9] 宋艳鑫,吕思捷,罗玲,周宇. 基于深度学习的广告联盟作弊行为智能识别系统[P]. 广州扬悦博众信息科技有限公司,ZL202312156236.6.

[10] 左海强,郑宇博,黄启洲,王荣迪,任伟强 . 基于深度学习算法的考试作弊行为检测方法[P]. 中国石油大学(华东),ZL202210166556.6.