

DOI: 10.13382/j.jemi.B2508481

融合表情识别技术与人工智能模型的 书籍推荐智能体*

舒锐 彭挺 傅铭伟

(西南交通大学希望学院机电与轨道车辆工程系 成都 610400)

摘要:随着人工智能模型的飞速发展,合理的开发和利用人工智能模型服务人们生活变得十分重要。鉴于传统推荐系统无法感知用户实时情绪,难以提供情绪适配书籍的问题,针对这一问题,设计了一种融合表情识别技术与人工智能模型的书籍推荐智能体。为了使表情识别结果更加精确,创新性地采用 MTCNN-MobileNet_V2 融合架构进行训练,训练结果显示此方法相较于传统基于 HOG 的识别方法平均准确率提高约 23.6%。该智能体基于树莓派 4B 硬件平台,首先通过 Camera V2 实时采集用户面部图像,采用 MTCNN 算法检测并对齐人脸区域,再使用 MobileNet_V2 轻量化卷积神经网络进行表情分类;随后,系统集成大语言模型 Qwen-3-32B 进行语义推理和生成个性化书籍推荐;最后利用 MQTT 与 Home Assistant 实现推荐结果实时展示。实验结果表明,此系统平均准确率为 94.72%,在标准光照(≈ 800 Lux)下情绪识别准确率为 97.7%;在低照度(≈ 350 Lux)下为 83.1%。5 人场景下识别准确率约 92%,响应时延约 5 s,系统响应延迟适中,可实现实时推荐。该智能体显著提升了推荐准确度和用户体验,同时验证了此方法的有效性,为未来基于情感输入的个性化推荐提供了有效依据。

关键词:表情识别;人工智能;智能体;MTCNN;MobileNet_V2

中图分类号: TN014 **文献标识码:** A **国家标准学科分类代码:** 520.60

Book recommendation agent based on expression recognition technology and artificial intelligence model

Shu Rui Peng Ting Fu Mingwei

(Department of Mechanical, Electrical and Rail Transit Engineering, Hope College of
Southwest Jiaotong University, Chengdu 610400, China)

Abstract: With the rapid development of artificial intelligence models, the reasonable development and utilization of these models to serve people's daily lives have become increasingly important. Given that traditional recommendation systems are unable to perceive users' real-time emotions and therefore struggle to provide emotion-aligned book recommendations, this study proposes an intelligent book recommendation agent that integrates facial expression recognition technology with artificial intelligence models. In order to make the expression recognition results more accurate, this paper innovatively uses MTCNN-MobileNet_V2 fusion architecture for training. The training results show that the average accuracy of this method is about 23.6% higher than that of the traditional HOG based recognition method. The agent is based on the Raspberry Pi 4B hardware platform. Firstly, the user's face image is collected in real time through Camera V2, and the face region is detected and aligned using MTCNN algorithm. Then, the expression classification is performed using MobileNet_V2 lightweight convolutional neural network; Then, the system integrates the large language model Qwen-3-32B to perform semantic reasoning and generate personalized book recommendations; Finally, MQTT and Home Assistant are used to realize the real-time display of recommendation results. The experimental results show that the average accuracy of this system is 94.72%, and the accuracy of emotion recognition is 97.7% under standard light (≈ 800 Lux); 83.1% under low illumination (≈ 350 lux). In the five person scene, the recognition accuracy is about 92%, the response delay is about 5 seconds, and the system response delay is moderate, which can realize real-time recommendation. The agent significantly improves the recommendation accuracy and user experience, and

收稿日期: 2025-06-20 Received Date: 2025-06-20

* 基金项目:四川省第一批现场工程师专项培养计划项目(教职成厅函[2023]6号)、西南交通大学希望学院实践示范课程建设项目(SJSF2025006)资助

verifies the effectiveness of this method, which provides an effective basis for personalized recommendation based on emotional input in the future.

Keywords: expression recognition; artificial intelligence; agent; MTCNN; MobileNet_V2

0 引言

随着互联网技术的迅猛发展,网络信息日益丰富,人们在工作与生活中面临的心理压力不断增长,情绪状态也因此更容易受到外部环境的干扰和波动。尽管人们可以通过一些娱乐方式进行情绪调节,但这些方式大多停留在表层,难以实现深层次的情绪引导与改善。相比之下,通过阅读优质书籍来提升个体精神内核,在情绪调节方面具有更加持久和积极的影响。因此,依据个体当前情绪状态进行针对性书籍推荐,具有一定的研究意义。近年来,微表情识别研究已由传统的手工特征提取逐步转向以深度神经网络为核心的方法,其中卷积神经网络(convolutional neural network, CNN)仍占主导地位。以增强型长记忆卷积网络(enhanced long-recall convolutional network, ELRCN)^[1]与双时间尺度卷积神经网络(dual timescale convolutional neural network, DTS-CNN)^[2]为代表的混合架构,通过将卷积网络与时序模型结合,增强了对表情随时间变化的时域信息捕捉能力。与此同时,图神经网络(graph neural network, GNN)在微表情领域的应用也开始兴起。Lo等^[3]提出基于面部动作单元(action unit, AU)的一种用于微表情识别的图卷积网络(graph convolutional network, GCN)方法,利用3D-CNN提取AU特征,再通过GCN建模AU依赖关系;Xie等^[4]则采用AU辅助的图注意力网络则引入生成对抗网络进行数据增强,这些方法通过复杂网络结构来捕捉更为精细的运动模式。此外,中科院光电研究院赵兆辉团队等提出了基于视觉Transformer的MCCA-VNet模型,该模型将通道注意力和空间注意力融入Vision Transformer中,用以捕捉面部微表情的微妙变化特征,在常用的公开数据集上获得了优异性能^[5]。尽管网络结构日趋复杂,但微表情识别的整体准确率仍面临数据稀缺与细微变化难以建模的挑战。在图书推荐领域,国际研究则延续了协同过滤与基于内容方法,并结合混合推荐、上下文感知与知识图谱等策略以应对冷启动与多样性需求。Wayesa等^[6]提出的基于语义规则挖掘的混合推荐系统,利用关联规则在数字图书馆场景下提升了新用户推荐质量。国内郭佳杰团队首次将宽-深度学习模型应用于Book-Crossing数据集,将稀疏特征与深度网络有机融合,在部分指标上优于传统算法^[7]。其他研究还结合用户显式兴趣建模、聚类优化与贝叶斯网络^[8-9],持续探索更精细的个性化推荐策略。

微表情识别与图书推荐的交叉研究尚处探索阶段。针对实时感知与设备部署的不足,本研究提出基于微表情的情感感知推荐框架:采用轻量化CNN在树莓派等设备上实时分类情绪,并将情绪动态嵌入协同过滤与内容推荐的混合策略中,以动态调整推荐列表,从而提升推荐精度与用户体验。该方法在计算资源与性能平衡的前提下,为个性化推荐提供了新思路。

1 网络架构

1.1 微表情识别技术

微表情识别技术通过分析面部微表情来判断个体的情感状态,是计算机视觉和情感计算领域的一个重要研究方向,其基本流程包括人脸检测^[10]、面部区域提取^[11]、表情分类^[12]和结果可视化。微表情识别的挑战主要在于表情的多样性、不同个体的面部差异以及光照和环境的变化,但是随着深度学习和计算机视觉技术的发展,表情识别技术已逐步成熟并且在智能设备上得到广泛应用。在此设计中,首先利用人脸检测算法多任务级联卷积神经网络(multi-task cascaded convolutional networks, MTCNN)^[13]同时完成定位面部区域和提取面部特征,再通过训练好的深度学习模型MobileNet_V2对面部表情进行分类。MobileNet_V2面部表情基础数据集主要由Emotion-Domestic国内(亚洲)表情识别数据集和MMAFEDB表情识别数据集^[14]提供,其总量超过150 000张人脸图,面部表情包含生气(angry)、厌恶(disgust)、害怕(fear)、快乐(happy)、中性(neutral)、悲伤(sad)、惊奇(surprise)7种表情。为了提高系统处理效率,对MobileNet_V2采用轻量化模型设计,它通过深度可分离卷积减少计算复杂度,以适用于资源受限的设备。

1) 算法选型

在人脸识别应用中,常见的算法包括Eigenfaces^[15]、Fisherfaces^[16]、CNN^[17]、FaceNet^[18]、MTCNN^[19]等。传统算法如Eigenfaces与Fisherfaces计算简单但对光照、姿态等变化适应性较差。相比之下,CNN与FaceNet等深度模型具备更强的特征提取能力,但是对计算资源需求较高,不适合嵌入式平台。然而MTCNN兼顾精度与效率,具备良好的鲁棒性,适合在树莓派等资源受限设备上运行,主流算法性能参数对照如表1所示。

MobileNet_V2作为一种专为移动设备和边缘设备高效运算设计的轻量化卷积神经网络。其相较于MobileNet_V1拥有更高的计算效率,通过引入导致残差

表 1 主流算法性能参数对比

Table 1 Comparison of performance parameters of mainstream algorithms

算法	准确性	鲁棒性	实时性	计算复杂度	适用性	可扩展性	能效
Eigenfaces	70	60	80	70	65	65	85
Fisherfaces	75	75	70	75	70	70	80
CNN	95	90	70	60	90	85	65
FaceNet	95	95	65	85	90	90	60
MTCNN	80	85	75	70	80	80	75

注:各性能指标均采用统一的百分制归一化评分(0~100 分)。该评分基于算法在公开数据集、典型实验场景以及实验结果中的综合表现,用于反映各算法的相对性能趋势,而非绝对实验数值。

结构(inverted residuals, IR)^[20]和线性瓶颈(linear bottleneck, LB)大幅减少了计算冗余,提升了特征表达能力,在低 Alpha 值下表现的更为稳定,MobileNet 系列性能对比如图 1 所示。

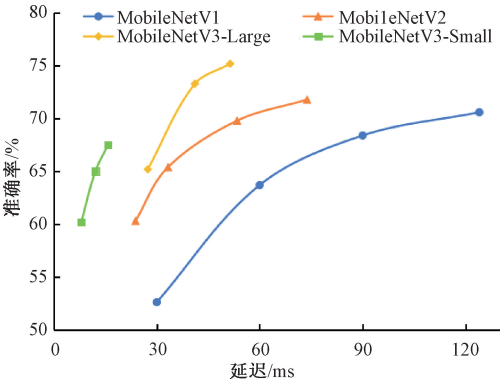


图 1 MobileNet 系列性能对比

Fig. 1 MobileNet series performance comparison

由图 1 的对比结果可以看出,MobileNet_V2 在相似甚至更低的延迟下可获得更高的分类准确率;相比进一步优化后的 MobileNet_V3 系列,尽管 MobileNet_V3-Large 在高算力平台上的准确率略有提升,但其更复杂的结构在资源受限的嵌入式平台上会带来额外的运算与内存开销,导致推理延迟增加,在图 1 中可以观察到,MobileNet_V3-Large 的延迟高于 MobileNet_V2,而 MobileNet_V3-Small 虽具更低延迟,但准确率下降明显。通过对比,

MobileNet_V2 拥有更均衡的权衡、更广泛的兼容性,特别适合于对实时性和准确性均有要求的微表情识别任务。因此,选择 MobileNet_V2 作为骨干网络,既能满足嵌入式设备的实时推理需求,也能在低算力平台上保持良好的分类准确性。

2) 网络架构设计

针对树莓派 4B 等嵌入式设备硬件资源有限以及系统实时性和计算效率的问题,对 MTCNN 和 MobileNet_V2 进行了一定的修改和优化,以确保模型可以在硬件资源有限的设备上高效运行。

在此设计中,网络框架采用 MTCNN 和 MobileNet_V2 融合设计完成微表情识别分类,即先利用 MTCNN 级联子网候选网络(P-Net)、精炼网络(R-Net)和输出网络(O-Net)在原始图像上完成人脸候选窗口的检测与 5 个关键点回归;随后在回归的关键点的基础上对每一张人脸进行仿射对齐、裁剪与归一化预处理;下一步将裁剪后的人脸输入预训练的 MobileNet_V2 模型中提取深度特征,并在其顶层添加微表情情绪检测的分类头;最后,通过替换全连接层输出 512 维特征向量,采用 ArcFace、Softmax 等损失函数进行模型联合微调,以增强特征的类内紧凑性和类间可分性,并在训练中兼顾多任务损失的权重平衡。这一结构具备近似检测与识别的端到端效率,同时通过轻量化骨干与多任务学习提升了系统的实时性与鲁棒性,模型网络结构如图 2 所示。

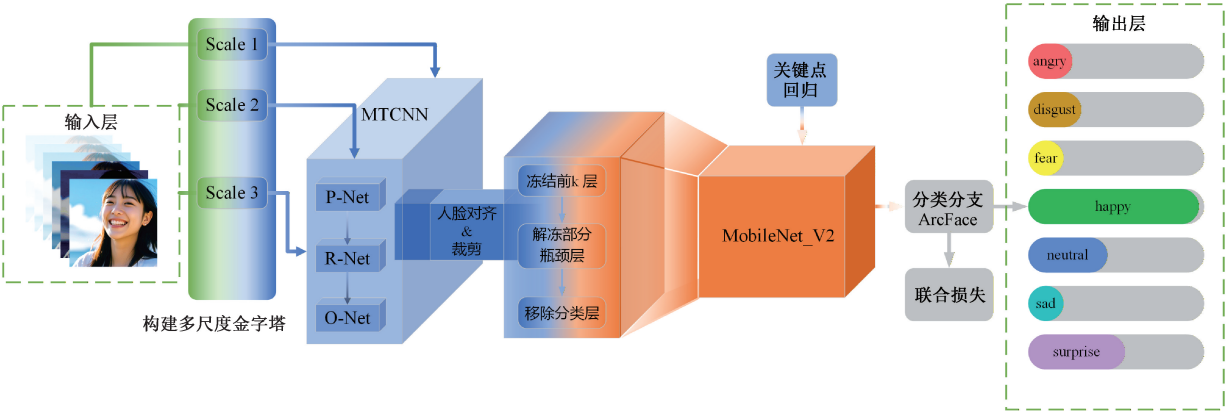


图 2 模型网络结构

Fig. 2 Network architecture of the model

系统以原始输入图像为起点,在多个预设缩放比例下进行连续下采样,搭建不同分辨率的图像金字塔并记录每层的缩放因子方便后续映射。随后各尺度图像依次送入轻量级 P-Net,在每个位置并行输出人脸置信度和边界框微调向量,保留置信度高于 0.65 的候选窗口并根据回归偏移调整框坐标,同时同尺度上以 $\text{IoU} = 0.66$ 的非极大抑制剔除冗余框。之后将 P-Net 筛选出的候选框在原图中统一裁剪缩放至 24×24 pixels 后作为 R-Net 输入;R-Net 则通过深层卷积与全连接网络对每个窗口再次进

行人脸/非人脸二分类,剔除大量假阳性且输出更加精确的边界框回归向量,此阶段采用置信度阈值 0.76 和 $\text{IoU} = 0.78$ 的 NMS 实现更严格的筛选。最后,对 R-Net 输出框裁剪至 48×48 向 O-Net 进行最终精炼,O-Net 除了给出更高置信度(阈值为 0.85)和边界回归外,还同时回归双眼、鼻尖以及双嘴角五点的坐标,在 $\text{IoU} = 0.86$ 的 NMS 之后输出最终的人脸边界框和五点关键点,用于后续的图像对齐与特征分析,MTCNN 网络结构如图 3 所示。

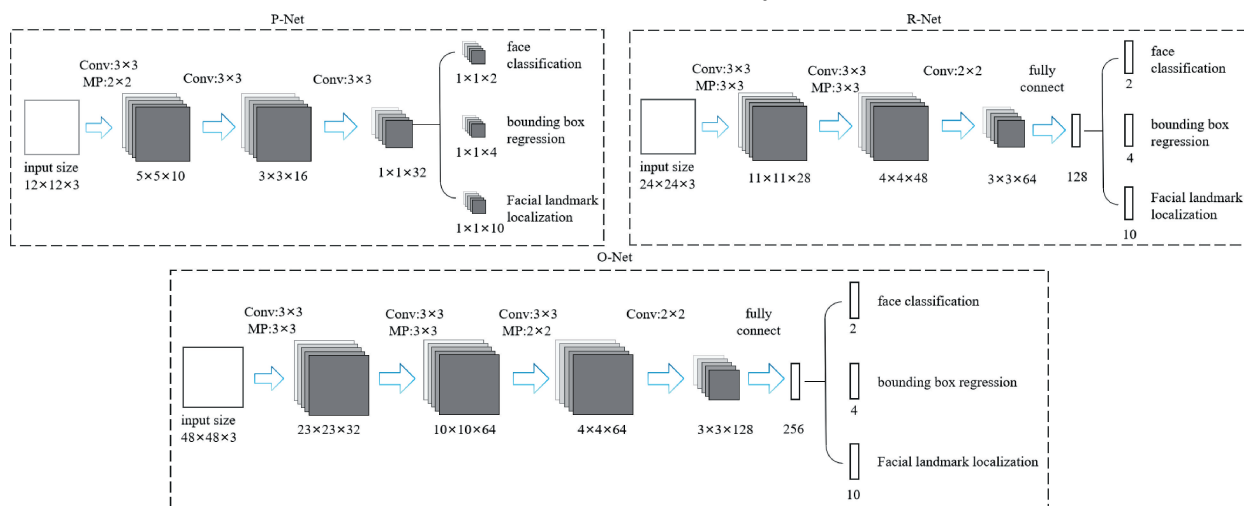


图3 MTCNN 网络结构

Fig. 3 MTCNN architecture

根据 O-Net 回归出的 5 个关键点坐标对人脸进行仿射变换对齐,在此之前需设定一组模板关键点坐标,本模型照片大小为 112×112 ,坐标分别为左眼(35,40)、右眼(77,40)、鼻尖(56,56)、左嘴角(40,80)以及右嘴角(72,80)。对检测到的关键点与模板关键点用 Opencv 接口进行最小二乘法求解仿射矩阵,然后对对齐后的人脸区域周围添加适当边距(15%)进行裁剪。之后将对齐并裁剪好的人脸图像缩放到模型输入尺寸,同时使用插值算法保持图像清晰度,并将像素值除以 255,使其映射至 $[0, 1]$,再根据训练集统计的均值与标准差进行零均值单位方差归一化,进一步减小光照与设备差异对特征提取的影响。通过上述对模型的优化,进一步提高了系统在树莓派 4B 上运行的高效性和准确性。

在对齐并预处理后的单人脸图像送入 MobileNet_V2 骨干网络进行深度特征提取。MobileNet_V2 骨干网络主要由倒残差和线性瓶颈构成,在对模型进行调整过程中为丰富特征的维度,首先使用 1×1 卷积将通道数按扩张系数的 6 倍放大;随后在此高维空间内进行 3×3 深度可分离卷积逐通道提取空间信息,并以 1×1 卷积融合通道间特征信息;最后采用线性激活的 1×1 卷积将通道数投影回低维,并在输入输出维度一致时通过残差连接强化

梯度传递。此设计既保留了残差网络的优化优势,又在参数量和 FLOPs 上实现显著压缩,让模型即使在移动设备上也可以高效运行。在此基础上,移除原有的分类层,添加一个输出 512 维特征向量的全连接层,为下游人脸微表情情绪分类任务提供紧凑表示,MobileNet_V2 结构如图 4 所示。通过加载在 ImageNet 上预训练的权重,既保留了大规模语义特征,又加速了在微表情数据上的收敛速度。

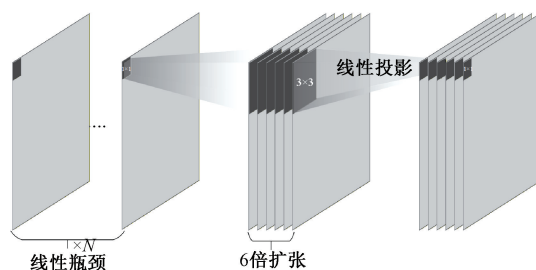


图4 MobileNet_V2 结构

Fig. 4 Architecture of MobileNet_V2

在模型的训练过程中,为更好利用大规模语义特征抽取能力,采用多任务联合微调策略,以充分发挥

MTCNN 与 MobileNet_V2 的优势。在初始化模型时直接加载官方在 ImageNet 上训练好的权重,并将倒残差瓶颈模块前 k 层冻结,仅保留新增的 512 维全连接层与多任务头可训练,同时检查梯度流以确保骨干层冻结生效。随后采用分阶段微调策略:阶段一仅训练全连接层和表情/识别分类头,学习率定义为 1×10^{-3} ,以此来快速适配新任务,直到验证集分类损失在 5~10 个 epoch 内收敛;阶段二解冻最后几层倒残差模块,使用余弦退火或分段衰减调度将初始学习率降至 1×10^{-4} ,并联合优化 Softmax、ArcFace、CosFace 分类损失及关键点回归损失,根据验证表现动态调整各分支权重,确保表情分类不被其他任务干扰。在超参数方面,批量大小取为 48,优化器选用带动量的 SGD,全连接层引入 Dropout (0.2~0.5) 和 L2 权重衰减 (1×10^{-4}),并对训练样本应用随机水平翻转、微小旋转 ($\pm 15^\circ$)、随机裁剪、色彩抖动及高斯噪声等增强。训练时在独立验证集上实时监控表情准确率、识别准确率和人脸检测 AP,当主指标在连续 5 个 epoch 内无提升即触发早停并回滚至最优权重,以防止过拟合并保留最佳模型。通过上述对模型的优化和修改,进一步提高了在树莓派 4B 上的高效性和准确性。

在部署阶段,为满足实时微表情识别需求,采用 OpenCV 和 MTCNN 配合进行视频流逐帧处理。系统以 25 FPS 的速率从摄像头读取视频帧,在每帧中先执行 MTCNN 人脸检测与关键点定位,再进行仿射对齐与预处理,然后将对齐后的图像输入 MobileNet_V2 骨干网络,最后通过分类头输出微表情类别及置信度。为了提高结果可靠性,设定分类置信度阈值 0.85,只有当模型预测置信度 ≥ 0.85 时,才认定为有效表情,否则忽略该检测结果,从而降低误报率。该阈值与帧率设置均经过多轮实验验证,既能实现动态视频流的实时处理,又保证了识别准确性与系统稳定性。训练完成后,训练集的准确度在 95.6% 以上,测试集的准确度在 94.5% 左右。模型在训练集和测试集上的表现如图 5、6 所示。通过上述设计,该微表情识别系统在树莓派 4B 上具有良好的实时性和鲁棒性,可为精准的人机交互与情感计算应用提供技术支持。

MTCNN 在进行人脸检测和对齐时能够高效处理多种光照和姿势变化,且在树莓派 4B 上能够流畅运行。而 MobileNet_V2 的轻量化设计使得它能够在性能受限的设备上进行高效的表情分类。相比于其他如 FaceNet 或大型 CNN 模型,这种组合能够在保证一定识别精度的前提下,大大减少计算开销,并保证实时性,部分情绪识别效果图如图 7 所示。因此,综合考虑到树莓派 4B 的硬件资源限制以及对实时性和计算效率的要求,MTCNN 和 MobileNet_V2 是理想的选择。

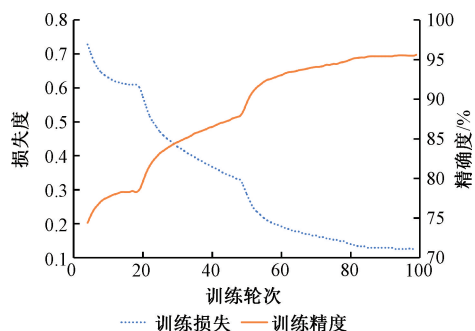


图 5 训练集损失与精度

Fig. 5 Training loss and accuracy curves

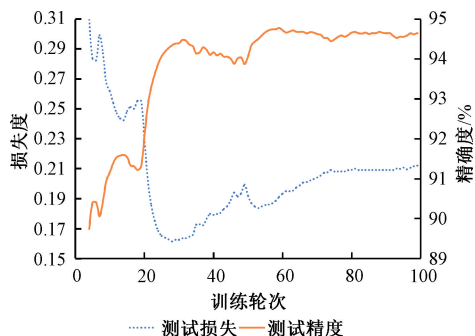


图 6 测试集损失与精度

Fig. 6 Testing loss and accuracy curves



图 7 情绪识别效果

Fig. 7 Effect picture of emotion recognition

1.2 推荐算法设计

推荐算法采用结合微表情识别和大型语言模型 (large language model, LLM) 推理的智能书籍推荐技术。此方法首先利用摄像头实时采集用户面部图像,并采用深度学习模型进行人脸检测与情绪分类,提取用户的即时情绪状态。随后,基于识别的情绪状态,调用大语言模型 (Qwen-3-32B) 进行语义理解与推理,采用 Prompt

工程生成符合当前情绪场景的个性化书单,结合 Home Assistant 平台与 MQTT 协议实现推荐结果可视化。此算法能够实现基于用户情绪的智能推荐,具有实时性、个性化和交互性等特点。

1) 多模态情感识别框架

传统推荐算法存在情感响应迟滞问题,而基于静态内容特征的推荐策略缺乏对用户情感偏好的动态适应能力,导致在用户情绪转换场景下推荐精准度显著降低。针对上述情感适应性缺陷,本研究创新性地构建了基于多模态情感识别的动态推荐框架。系统通过高精度微表情识别模型实时捕获用户的面部动作编码(FACS),结合情绪分类器解析用户的7种基础情感状态。在此基础上,设计情感-内容双通道权重分配机制,将情感向量与协同过滤推荐结果进行张量融合,通过门控循环单元动态调节内容特征的权重系数。此架构不仅保留传统推荐算法的优势,更能依据用户实时的情感反馈建立情绪-兴趣映射模型,最终生成具有情感感知能力的个性化推荐列表,在生成最终推荐分数时,本系统通过动态权重调整,将情绪状态与传统的协同过滤结果结合起来。推荐分数的计算公式如下:

$$S_{final} = \alpha \times S_{emotion} + (1 - \alpha) \times S_{CF} \quad (1)$$

式中: $S_{emotion}$ 表示基于情绪的推荐分数; S_{CF} 为协同过滤的推荐分数; α 为情绪权重系数,控制情绪和协同过滤推荐的融合程度。此公式通过权重调整,动态平衡情绪与历史行为的影响,确保推荐结果既符合用户的情绪需求,也考虑到其历史偏好。

2) 推荐模块实现

本系统的推荐算法基于用户的情绪状态进行个性化书籍推荐。情绪识别是前置步骤,它通过分析用户的面部表情或其他情绪信息为推荐提供基础。在情绪状态的准确识别后,系统将这一情绪标签作为输入,向推荐系统发送查询请求,获取与该情绪状态匹配的书籍。情绪驱动的推荐设计使得推荐系统能够根据用户的当前情绪需求提供更加个性化、情感化的内容,而不是基于简单的用户历史行为或偏好。

系统首先通过人脸检测与情绪识别算法,对输入图像或视频帧进行预处理,利用预训练深度学习模型提取用户面部情感标签。通过分析这些情绪分布,系统精准捕捉用户当下情感状态,为后续推荐提供依据。随后,系统将检测到的情绪信息作为输入,与书籍推荐模型对接,生成一本或多本适合当前情绪的书籍推荐列表。其中,推荐模块利用 Gradio 客户端接口调用在线推荐引擎,通过构造自然语言查询,获得对应情绪下的书籍建议。为了提升响应速度和用户体验,系统设计了缓存机制,将每种情绪对应的推荐结果存储于内存中,避免重复调用外部接口而产生延迟。同时,通过 MQTT 协议将推荐结果

发布到 Home Assistant,实现与智能显示屏、移动设备等终端的实时数据同步,确保整个系统在高并发场景下依然保持稳定高效运行,书籍推荐流程如图8所示。

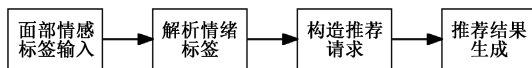


图8 书籍推荐流程

Fig. 8 Flow chart of book recommendation

2 系统设计

2.1 总体设计

本文设计采用树莓派 4B (Raspberry Pi 4 Model B)^[21] 作为主控平台,提供低功耗和足够的计算能力,适用于嵌入式应用。摄像头使用树莓派 Camera V2,能够实时捕捉用户的面部微表情,确保了良好的图像质量和识别效果。软件方面,系统运行基于 Linux 的 Raspberry Pi OS,提供稳定的操作系统支持,兼容 Python 编程语言,便于实现微表情识别和推荐算法,此外,还配备 Home Assistant 显示界面,用于展示书籍推荐结果和与用户的交互。微表情识别使用 MTCNN 和 MobileNet_V2 模型相融合的设计,通过深度学习模型识别用户实时的情绪状态,推荐算法则基于强大的 Qwen-3-32B,能够结合用户情绪与历史数据生成个性化的书籍推荐,其系统流程如图9所示。

2.2 硬件设计

在此设计中,选择树莓派 4B 作为核心硬件平台,结合 Qwen-3-32B 推荐算法进行个性化推荐。其低功耗和良好的硬件扩展性使得该系统能够满足日常应用需求,为用户提供精准的情感驱动书籍推荐,提升整体用户体验。

Camera V2 通过 CSI 与树莓派主板连接,利用移动产业处理器接口 (mobile industry processor interface, MIPI) 协议进行数据传输,确保高速、稳定的数据交互。其基本原理是通过镜头将外部光线聚焦在图像传感器上,图像传感器将光信号转化为电信号后,利用 MIPI 协议通过 CSI 传输到树莓派的处理器,再经过树莓派处理器的数字信号处理 (digital signal processing, DSP),生成可视的数字图像或视频信号。它的设计具有低功耗和小体积的特点,便于在空间受限的项目中集成使用。它广泛支持 Python、OpenCV、Picamera 等开源库,可以轻松地进行图像处理、姿态检测、运动分析等任务^[22],其硬件实物如图10所示。

2.3 软件设计

Home Assistant 作为一款开源家庭自动化平台,通过 Docker 容器化部署在 Raspberry Pi OS 系统上,能够为智能家居设备管理提供灵活的技术解决方案。Home

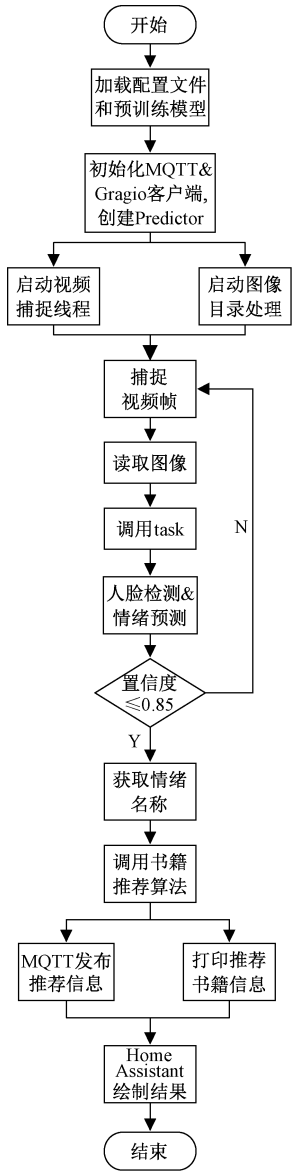


图 9 系统流程图
Fig. 9 System flow chart

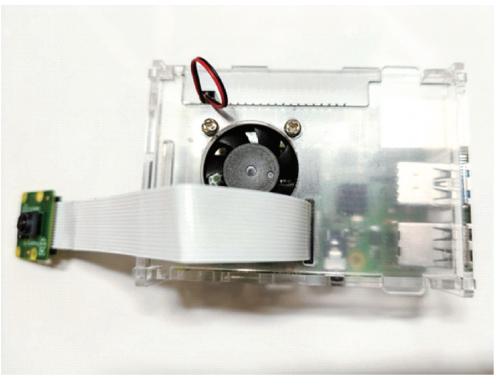


图 10 硬件实物
Fig. 10 Hardware object drawing

Assistant,新的配置便会迅速生效,使整个智能家居系统能够以最新的规则运行;另一方面,Home Assistant 内置的 Lovelace 前端界面为用户提供了一个直观的拖放式编辑功能,允许进行简单的图形化操作添加、修改或删除各种卡片,另外用户可以自由组合和调整布局,从而为客户端定制个性化的交互式 UI 界面如图 11 所示,实现信息展示与设备控制的无缝融合。



图 11 Home assistant UI 界面
Fig. 11 Home assistant UI interface

Assistant 的优势不仅在于对多种设备的广泛兼容性,还在于其注重本地化控制和隐私保护,使得用户数据始终存储在本地网络中,而无需依赖第三方云服务,因而降低了数据泄露的风险。

在安装上可以利用 Docker 引擎^[23]在 Raspberry Pi OS 上部署 Home Assistant,可以将其作为一个独立的服务运行在轻量级硬件上,这种安装方式不仅节省资源,还为其提供较高的运行效率。在安装成功后通过访问 configuration.yaml 配置文件,可以借助 Nano 编辑器对该文件进行直接编辑,添加 MQTT^[24]等集成配置,实现不同设备间的互联互通,并根据实际需求调整自动化规则和场景设置。编辑完成后,保存文件并重启 Home

3 实验测试与分析

3.1 数据集与响应时间

通过一系列实验旨在验证融合表情识别技术与人工智能模型的书籍推荐智能体的可靠性、准确性以及稳定性。在测试过程中,为更加准确测量模型在树莓派 4B 上的推理能力,帧读取频率为 25 fps,且 Qwen-3-32B 推理间隔为 50 帧,即智能体每 2 s 进行一次书籍推荐。并对优化、调整后的 MobileNet_V2 模型进行了 1 轮共 12 000 张人脸图片的测试,以验证模型在长时间运行下的稳定性。

测试结果表明:优化、调整后的 MobileNet_V2 在树莓派 4B 上单帧推理时间稳定在 50~70 ms,情绪分类准确率与原始模型持平,内存占用优化后可在设备资源限

制下持续运行,且能适应不同光照条件下的表情变化。但远程 API 处理长文本请求仍需约 5 s 响应时间,其训练集与响应时间变化趋势如图 12 所示。

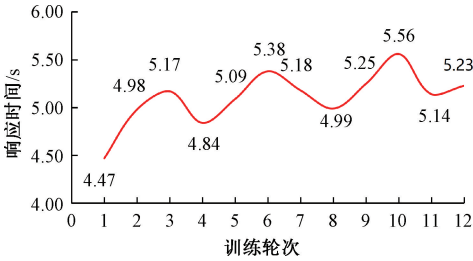


图 12 训练集与响应时间
Fig. 12 Training set and response time

通过分析测试结果,发现尽管情绪识别模块已经优化得较为高效,但由于 Qwen-3-32B 处理长文本的推理时间较长,导致最终的系统响应时间存在约 5 s 的延迟,但在实际应用中,5 s 的响应时间在用户可接受范围内。

3.2 实验结果及分析

智能体的人物设定是一位心理分析师,主要任务是通过分析人物当前的情绪状态,然后推荐适合阅读的书籍并给出推荐理由。测试过程中,主要对智能体从情绪识别准确率和书籍推荐效果两个方面测试。

1) 情绪识别准确率

在情绪识别准确率的测试中,主要围绕不同光照值和多用户并发处理进行,如图 13 所示设置 4 种典型光照场景:分别采用照度值约为 800、650、500 和 350 Lux,详细评估了系统在不同光照条件下的面部微表情识别性能。实验结果如表 2 所示,实验数据表明光照强度与识别准确率呈现非线性负相关关系,但即便在复杂光照条件下,系统综合表现仍明显优于传统基于 HOG 的识别方法,平均提升达到 23.6%。这一结果充分证明了所采用的硬件优化和轻量化卷积神经网络设计在不同光照环境下均具有较强的鲁棒性,为后续基于情感驱动的书籍推荐提供了可靠的基础支持。

表 2 不同光照下识别准确率
Table 2 Recognition accuracy at different illumination levels

照度/Lux	准确率/%
800	97.7
650	95.3
500	92.5
350	83.1

为进一步验证系统在多用户场景下的鲁棒性,测试时设计了 4 组不同测试人数的实验,分别对应单人、双



图 13 照度值不同的场景
Fig. 13 Scenes with different illuminance values

人、4 人和 5 人同时进入摄像头视野的场景,如图 14 所示。测试环境统一设置为标准光照条件(800 Lux),以避免光照因素对结果的干扰。通过分析系统在不同人数下的情绪识别准确率、响应延迟及资源占用情况,全面评估其实际应用潜力。

在单人测试场景中,由于系统处理任务相对单一,图像采集、情绪识别及书籍推荐均能够在较短时间内完成。实验数据显示,单人场景下系统平均准确率为 97.3%,与上一实验的基准结果一致。单人测试结果表明,在无并发负载情况下,系统各模块均能达到预期性能指标,为后续多用户实验提供了良好的基线数据。其次,在少数人测试(2~4 人)情况下,系统需同时处理来自多个用户的视频流及实时情绪数据。结果显示,此时系统的情绪识别准确率略有下降,但仍能保持在 95%左右。与此同时,由于树莓派 4B 在处理多路数据时存在一定的计算资源竞争,系统的响应时间略有增加。在多人测试(5 人及以上)场景中,由于系统同时接入的摄像头数据和情绪信息大幅增加,树莓派处理负荷显著上升。实验观察到,多人同时入镜时,系统在情绪识别过程中受背景干扰和图像重叠的影响较大,准确率降至约 92%。

2) 书籍推荐效果

在书籍推荐性能测试过程中,传统的协同过滤通过用户行为数据发现相似用户的偏好,具有无需依赖书籍内容的优势,但受冷启动和数据稀疏影响较大;内容过滤依靠书籍特征与用户偏好进行匹配,适用于新书场景且具有良好的可解释性,但存在特征提取困难与推荐结果单一的问题;混合推荐则结合了两种方法的优势,通过模



图 14 人数不同的典型场景

Fig. 14 Typical scenarios with different numbers of people

型融合显著提升了推荐的准确性与鲁棒性,是当前应用最广的技术路线。性能评估显示,混合推荐在准确率方面表现最佳,在 Netflix 数据集上融合 KNN 与 SVD 的方法显著降低 RMSE 并提升推荐质量;响应时间方面,传统推荐系统能达到毫秒级,而运行在边缘设备上的混合情境系统响应约为 5 s,仍保持实时性;同时,混合策略有效缓解了协同过滤的冷启动难题和内容过滤的过度专一问题,并在推荐多样性方面具有明显优势。

在智能体测试过程中,重点测试了智能体在不同情绪状态下生成的书籍推荐效果,智能体会根据识别出的情绪状态推荐一本书籍并且说明推荐理由,如图 15 所示。通过对用户面部微表情的实时捕捉和情绪分类,系统能够区分出 angry、disgust、fear 等多种情绪,并根据各情绪状态生成具有针对性的书籍推荐列表。

为验证本文设计能够根据用户即时情绪状态生成个性化、精准的书籍推荐,在采集 7 类情绪状态下的面部图像后,分别对系统的情绪识别稳定性以及推荐模块的匹配一致性进行测试,并对各情绪状态下系统输出的书籍



图 15 书籍推荐效果

Fig. 15 Book recommendation effect

类型及数量进行统计分析,逐类计算推荐输出的书籍类型占比,整理得到情绪与书籍适配表如表3所示,其中括号数据表示投入该情绪类型的照片数量。

表3 情绪与书籍适配表

Table 3 Emotion and book adaptation table		
情绪类型	推荐书籍类型	推荐对应书籍类型 百分比/%
生气(1 820)	自我修复、心理学	92. 58
厌恶(1 684)	心理学、哲学类	88. 66
害怕(1 756)	温馨治愈类	91. 34
快乐(1 827)	幽默、轻松类	95. 89
中性(1 603)	科普、文学类	89. 46
悲伤(1 652)	治愈系、励志类	92. 07
惊奇(1 658)	科幻、冒险类	89. 75

实验结果表明,不同情绪状态下推荐主题与内容呈现出显著差异,这不仅验证了情绪驱动推荐算法的有效性,也反映出用户在情绪变化时对书籍内容的不同偏好。通过与传统协同过滤方法的融合,本设计在保持较高推荐准确率的同时,仍能够对情绪变化做出稳定的实时响应,从而进一步提升整体推荐的可靠性与个性化水平。

4 结 论

本文提出并实现了一种融合表情识别技术与人工智能模型的书籍推荐智能体,成功突破了传统推荐系统在实时响应用户情绪变化方面的不足。设计采用MTCNN实时捕捉用户面部表情,并利用经过优化调整后的轻量化卷积神经网络 MobileNet_V2 进行情绪分类,确保在标准及复杂光照条件下均能保持较高的识别准确率。基于情绪标签与传统协同过滤推荐算法的深度融合,以及引入大语言模型 Qwen-3-32B 进行语义推理,实现了更具人性化、情境适应性的动态推荐,展现出在情绪适配能力上的显著创新与优势。

通过情绪识别准确率和书籍推荐效果两个方面的实验测试,实验结果表明系统在不同场景下均展现出较优异的性能与适应性。系统在光照适应性、多用户场景鲁棒性及情感驱动的推荐精准性方面均达到预期目标,其低功耗、实时性与动态优化机制为嵌入式情感计算与个性化推荐提供了可行的技术路径,为后续智慧图书馆与在线阅读平台的应用奠定了一定基础。虽然当前智能体在实验测试中受限于硬件条件,其稳定性、响应速度及用户交互体验尚未达到最优水平。未来在实际应用中若能进一步提升硬件配置,增强算力与通信速率,则系统在上述方面的表现均有较大提升空间。

参考文献

[1] KHOR H Q, SEE J, PHAN R C W, et al. Enriched

long-term recurrent convolutional network for facial micro-expression recognition[C]. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE, 2018: 667-674.

[2] PENG M, WANG C, CHEN T, et al. Dual temporal scale convolutional neural network for micro-expression recognition [J]. Frontiers in Psychology, 2017, 8: 1745.

[3] LO L, XIE H X, SHUAI H H, et al. MER-GCN: Micro-expression recognition based on relation modeling with graph convolutional networks [C]. 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE, 2020: 79-84.

[4] XIE H X, LO L, SHUAI H H, et al. Au-assisted graph attention convolutional network for micro-expression recognition [C]. Proceedings of the 28th ACM International Conference on Multimedia. 2020: 2871-2880.

[5] ZHANG D, ZHANG T, SUN H, et al. MCCA-VNet: A vit-based deep learning approach for micro-expression recognition based on facial coding[J]. Sensors (Basel, Switzerland), 2024, 24(23): 7549.

[6] WAYESA F, LERANSO M, ASEFA G, et al. Pattern-based hybrid book recommendation system using semantic relationships[J]. Scientific Reports, 2023, 13(1): 3693.

[7] GUO J J. Enhancing book recommendation systems through the application of sparse features in wide and deep learning models [J]. International Journal of Frontiers in Engineering Technology, 2024, 6(5).

[8] LIU B, LIU X, LUO Q, et al. Variational bayesian personalized ranking [J]. ArXiv preprint arXiv: 2503. 11067, 2025.

[9] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback[J]. ArXiv preprint arXiv:1205. 2618, 2012.

[10] 张立军,曹江涛,姬晓飞,等. 基于深度学习的学生课堂学习状态监测系统设计[J]. 电子测量与仪器学报, 2024,38(4):37-45.

ZHANG L J, CAO J T, JI X F, et al. Design of a classroom learning state monitoring system for students based on deep learning [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38 (4): 37-45.

[11] 王建,宋晓宁. 融合多尺度特征的轻量级人脸检测算法[J]. 模式识别与人工智能,2022,35(6):507-515.

WANG J, SONG X N. Lightweight face detection algorithm fusing multi-scale features [J]. Pattern Recognition and Artificial Intelligence, 2022, 35 (6):

- 507-515.
- [12] DINO H I, ABDULRAZZAQ M B. Facial expression classification based on SVM, KNN and MLP classifiers[C]. 2019 International Conference on Advanced Science and Engineering (ICOASE). IEEE, 2019: 70-75.
- [13] KU H, DONG W. Face recognition based on mtcnn and convolutional neural network[J]. Frontiers in Signal Processing, 2020, 4(1): 37-42.
- [14] AGRAWAL I, KUMAR A, SWATHI D G, et al. Emotion recognition from facial expression using cnn[C]. 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC). IEEE, 2021: 01-06.
- [15] HO H T, NGUYEN L V, LE T H T, et al. Face detection using eigenfaces: A Comprehensive review[J]. IEEE Access, 2024.
- [16] ANGGO M, ARAPU L. Face recognition using fisherface method[C]. Journal of Physics: Conference Series. IOP Publishing, 2018, 1028(1): 012119.
- [17] ALZUBAIDI L, ZHANG J, HUMAIDI A J, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions[J]. Journal of Big Data, 2021, 8: 1-74.
- [18] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 815-823.
- [19] ZHANG N, LUO J, GAO W. Research on face detection technology based on MTCNN[C]. 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA). IEEE, 2020: 154-158.
- [20] 陈广秋,温奇璋,尹文卿,等. 用于红外与可见光图像融合的注意力残差密集融合网络[J]. 电子测量与仪器学报, 2023, 37(8): 182-193.
- CHEN G Q, WEN Q ZH, YIN W Q, et al. Attention residual dense fusion network for infrared and visible image fusion[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(8): 182-193.
- [21] LAKI S, STOYAANOV R, KIS D, et al. P4Pi: P4 on Raspberry Pi for networking education[J]. ACM SIGCOMM Computer Communication Review, 2021, 51(3): 17-21.
- [22] 舒锐,傅铭伟,彭挺,等. 基于人工智能模型的婴幼儿行为监护系统[J]. 电子测量技术, 2025, 48(7): 107-116.
- SHU R, FU M W, PENG T, et al. Infant behavior monitoring system based on artificial intelligence models[J]. Electronic Measurement Technology, 2025, 48(7): 107-116.
- [23] COMBE T, MARTIN A, DI P R. To docker or not to docker: A security perspective[J]. IEEE Cloud Computing, 2016, 3(5): 54-62.
- [24] UPADHYAY Y, BOROLE A, DILEEPAN D. MQTT based secured home automation system[C]. 2016 Symposium on Colossal Data Analysis and Networking (CDAN). IEEE, 2016: 1-4.

作者简介



舒锐(通信作者), 2017年于西南交通大学获得学士学位, 2024年于西南石油大学获得硕士学位, 现为西南交通大学希望学院讲师, 主要研究方向为嵌入式系统与电力电子技术。

E-mail: srwby2022@163.com

Shu Rui (Corresponding author) received his B. Sc. degree from Southwest Jiaotong University in 2017 and M. Sc. degree from Southwest Petroleum University in 2024, respectively. Now he is lecturer in Southwest Jiaotong University Hope College. His main research interests include embedded systems and power electronics technology.