

分类号: TP391

密 级: 公开

学校代码: 10363

学 号: 2210320119



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目

基于深度学习的人脸

微表情识别研究

论 文 作 者

唐绍语

校 内 导 师

魏利胜（教授）

校 外 导 师

程武山（教授）

专业学位类别

控制科学与工程

研究方向（领域）

模式识别与智能系统

论文提交日期: 2024 年 5 月 31 日

分类号: TP391

单位代码: 10363

密 级: 公开

学 号: 2210320119

题 目 基于深度学习的人脸微表情识别研究

题 目 Research on Face Micro-Expression Recognition
Based on Deep Learning

学 生 姓 名: 唐绍语
校 内 导 师: 魏利胜 教授
校 外 导 师: 程武山 教授
专 业 学 位 类 别: 控制科学与工程
研 究 方 向 (领 域): 模式识别与智能系统

论文答辩日期: 2024 年 5 月 12 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：唐绍语

日期：2024年5月30日

安徽工程大学学位论文授权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：唐绍语

日期：2024年5月30日

指导教师签名：魏利胜

日期：2024年5月30日

基于深度学习的人脸微表情识别研究

摘 要

微表情是一种发生在人脸局部区域极其短暂、难以察觉的自发肌肉动作，往往难以控制和无法隐藏。微表情具有揭示个体真实情感和意图的作用，在心理学、教育、医疗、刑侦等领域意义重大。然而，人工识别微表情需要特定的学习和研究经验，难以普遍适用。随着计算机视觉技术的发展，采用深度学习方法自动识别微表情成为研究热点。深度学习自动提取图像中的特征表示，减少对专业微表情知识的依赖，使得微表情规模识别及应用成为可能。本文深入研究了基于深度学习的微表情识别，对图像预处理技术和特征提取网络模型展开研究，主要研究内容如下：

首先，研究了图像增强、特征提取等方面的人脸微表情图像预处理技术。在 LLFormer 网络的基础上，引入卷积块注意力机制(Convolutional Block Attention Module, CBAM)增强网络对浅层重要特征的提取能力，并在 LOL 数据集上实验验证了算法对低光照图像增强的有效性；此外，融合全变分 L1 正则化(Total Variation L1-norm, TVL1)光流特征和光流应变特征作为微表情的特征描述符，并采用支持向量机(Support Vector Machine, SVM)作为分类器进行实验，实验证明了融合光流特征描述符的有效性。

其次，探讨了一种融合 MobileViT 和坐标注意力机制(Coordinate Attention, CA)的人脸微表情识别网络。在 MobileViT-V3 轻量化网络的基础上，引入轻量型的 CA 注意力计算特征权重，在关注重要特征的同时抑制不重要特征；采用 SiLU 激活函数替换 ReLU6 激活函数，进一步增强网络的非线性表达能力；利用 Focal Loss 函数作为训练损失函数，缓解类别不平衡的问题；通过在 CASME 和 CASME II 数据

集上的消融实验和对比实验，验证所研究网络对微表情识别的有效性。

然后，研究了一种融合残差网络和两种注意力机制的人脸微表情识别网络。将全局注意力(Global Attention Mechanism, GAM)和双重注意力(Dual Attention, DA)与浅层残差网络融合，在加强局部重要特征关注、提取关键信息的同时，降低过拟合风险。通过公共数据集实验验证了所提模型微表情识别的优越性，经自建数据集实验检验模型识别性能良好。

最后，设计了一款基于深度学习的微表情识别系统。在软件系统需求分析的基础上，采用 PostgreSQL 数据库实现信息存储，利用 Pycharm 软件，添加 PyQt5 库和 QT-Designer 工具，分别实现 UI 界面以及界面逻辑控制。系统包含登陆验证、视频采集、微表情识别、查看历史记录和结果显示等功能，可以直观地展示识别结果。

综上，本文采用深度学习方法对人脸微表情识别进行系统性研究，为微表情识别理论研究和实际应用提供了重要的参考价值。

关键词：微表情识别；深度学习；Transformer；MobileViT；残差网络；注意力机制；微表情识别系统

Research on Facial Micro-Expression Recognition Based on Deep Learning

ABSTRACT

Micro-expression is a kind of spontaneous muscle action that occurs in the localized area of human face, which is extremely transient and hard to detect, and is often difficult to control and cannot be hidden. Micro-expressions have the function of revealing an individual's true emotions and intentions, and are of great significance in the fields of psychology, education, medical care, and criminal investigation. However, manual recognition of micro-expressions requires specific learning and research experience and is difficult to apply universally. With the development of computer vision technology, it has become a research hotspot to automatically recognize micro-expressions using deep learning methods. Deep learning automatically extracts feature representations in images and reduces the dependence on specialized micro-expression knowledge, making micro-expression scale recognition and application possible. In this paper, micro-expression recognition based on deep learning is studied in depth, and the image preprocessing technology and feature extraction network model are studied, and the main research contents are as follows:

Firstly, the image enhancement, feature extraction and other aspects of face micro-expression image preprocessing technology are studied. Based on the LLFormer network, the Convolutional Block Attention Module (CBAM) is introduced to enhance the ability of the network to extract important features in the shallow layer. Experiments on LOL datasets verify the effectiveness of the algorithm for low-light image

enhancement; In addition, the Total Variation L1-norm (TVL1) optical flow features and optical flow strain features are fused as feature representations of micro-expressions, and Support Vector Machine (SVM) is used as a classifier to carry out experiments. Experiments prove the effectiveness of the fused optical flow feature.

Secondly, a facial micro-expression recognition network based on MobileViT and Coordinate Attention (CA) is discussed. On the basis of MobileViT-V3 lightweight network, lightweight CA attention is introduced to calculate feature weights, which pays attention to important features and suppresses unimportant features; ReLU6 activation function is replaced by SiLU activation function to further enhance the nonlinear expression ability of the network; Focus Loss function is used as training loss function to alleviate the problem of class imbalance; Through the ablation experiment and contrast experiment on CASME and CASME II data sets, the effectiveness of the network for micro-expression recognition is verified.

Then, a facial micro-expression recognition network based on residual network and two attention mechanisms is studied. Combining Global Attention Mechanism (GAM) and Dual Attention (DA) with shallow residual network, it can strengthen the attention of local important features, extract key information and reduce the risk of over-fitting. The superiority of the proposed model is verified by experiments on public data sets, and the recognition performance of the model is verified by experiments on self-built data sets.

Finally, a micro-expression recognition system based on deep learning is designed. Based on the analysis of software system

requirements, PostgreSQL database is used to store information, Pycharm software is used to add PyQt5 library and QT-Designer tools to realize UI interface and interface logic control respectively. The system includes login verification, video capture, micro-expression recognition, viewing history records and displaying results, etc., which can display the recognition results intuitively.

In summary, this paper uses deep learning method to systematically study face micro-expression recognition, which provides important reference value for theoretical research and practical application of micro-expression recognition.

Tang Shaoyu(Control Science and Engineering)

Supervised by Wei Lisheng

Key words: Micro-expression recognition; Deep learning; Transformer; MobileViT; Residual network; Attention mechanism; Micro-expression recognition system

目 录

摘 要.....	I
ABSTRACT.....	III
目 录.....	VI
第 1 章 绪 论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 人脸微表情图像预处理技术研究现状.....	2
1.2.2 人脸微表情识别算法研究现状.....	5
1.3 主要研究内容.....	6
1.4 本章小结.....	8
第 2 章 人脸微表情图像预处理技术.....	9
2.1 图像增强.....	9
2.1.1 CBAM-LLFormer 低光照图像增强网络模型.....	10
2.1.2 CBAM 模块.....	11
2.1.3 Transformer 模块.....	11
2.1.4 跨层注意力融合模块.....	13
2.1.5 实验分析.....	14
2.2 关键帧选取.....	15
2.3 尺度变换.....	16
2.4 特征提取.....	19
2.4.1 TVL1 光流特征.....	19
2.4.2 光流应变.....	19
2.4.3 融合光流.....	20
2.4.4 实验分析.....	20
2.5 本章小结.....	22
第 3 章 融合 MobileViT 和注意力机制的人脸微表情识别.....	23
3.1 CA-MobileViT 特征提取网络结构设计.....	23
3.1.1 整体网络模型.....	23
3.1.2 MobileViT-V3 网络模型.....	24
3.1.3 注意力模块.....	26
3.2 实验结果与分析.....	27
3.2.1 人脸微表情数据集.....	27
3.2.2 实验设置和评价指标.....	28
3.2.3 消融实验.....	30
3.2.4 对比实验.....	31

3.3 本章小结.....	33
第 4 章 融合残差网络和注意力机制的人脸微表情识别.....	34
4.1 GAM-Res-DA 特征提取网络结构设计.....	34
4.1.1 整体网络模型.....	34
4.1.2 残差网络模块.....	35
4.1.3 注意力模块.....	36
4.2 实验结果与分析.....	39
4.2.1 实验设置.....	39
4.2.2 消融实验.....	39
4.2.3 对比实验.....	40
4.2.4 自建数据集实验.....	42
4.3 本章小结.....	45
第 5 章 人脸微表情识别系统设计.....	46
5.1 系统需求分析.....	46
5.2 系统功能设计.....	46
5.3 系统实现.....	47
5.3.1 开发平台与数据库.....	47
5.3.2 ID 验证登陆与注册界面的实现.....	47
5.3.3 系统主界面的实现.....	50
5.4 系统测试.....	53
5.4.1 ID 验证登陆与注册功能测试.....	53
5.4.2 系统主界面功能测试.....	55
5.5 本章小结.....	58
第 6 章 总结与展望.....	59
6.1 总结.....	59
6.2 展望.....	59
参考文献.....	61
攻读学位期间发表的学术论文.....	66
攻读学位期间获得的奖励.....	66
致 谢.....	67

插图清单

图 1-1 研究内容框架图	7
图 2-1 数据集预处理流程图	9
图 2-2 CBAM-LLFormer 低光照图像增强网络模型图	10
图 2-3 CBAM 结构图.....	11
图 2-4 Transformer 模块结构图.....	12
图 2-5 深度可分离卷积结构图	12
图 2-6 双门前馈网络结构图	13
图 2-7 跨层注意力融合模块结构图	13
图 2-8 低光照图像增强对比图	14
图 2-9 LBP 计算示例图	16
图 2-10 顶点帧求取示例图	16
图 2-11 MTCNN 人脸检测和关键点检测流程	17
图 2-12 人脸检测与人脸定位示例图	18
图 2-13 人脸旋转对齐图与裁剪示例图	18
图 2-14 融合特征图获取和数据增强示例图.....	21
图 3-1 CA-MobileViT 模型结构图	23
图 3-2 SiLU 和 ReLU6 激活函数图	24
图 3-3 MobileNetV2 模块结构图	25
图 3-4 MobileViT 模块结构图.....	25
图 3-5 Unfold 和 Fold 操作示例图	26
图 3-6 CA 注意力机制结构图和池化计算示例图	27
图 3-7 CASME 和 CASME II 数据集样本.....	28
图 3-8 混淆矩阵结构图	29
图 3-9 在 CASME 数据集上的混淆矩阵.....	31
图 3-10 在 CASME II 数据集上的混淆矩阵	31
图 4-1 GAM-Res-DA 模型结构图.....	35
图 4-2 残差学习构建模块图	36
图 4-3 GAM 模块结构图	37
图 4-4 DA 模块结构图.....	38
图 4-5 微表情识别的类激活图	40
图 4-6 在 CASME 数据集上的混淆矩阵.....	41
图 4-7 在 CASME II 数据集上的混淆矩阵	41
图 4-8 部分数据集样本示例图	43
图 4-9 训练时损失函数收敛曲线和测试集准确率曲线图.....	44
图 4-10 在自建数据集上的混淆矩阵	44
图 5-1 系统整体架构图	47
图 5-2 登陆与注册流程图	48

图 5-3 ID 验证登陆与注册界面图	48
图 5-4 系统首页图	50
图 5-5 微表情识别顺序逻辑图	51
图 5-6 微表情识别初始界面图	51
图 5-7 微表情识别结果历史记录图	52
图 5-8 修改密码顺序逻辑图	52
图 5-9 修改密码界面图	53
图 5-10 系统登陆功能测试图	54
图 5-11 系统注册功能测试图	55
图 5-12 微表情识别功能测试图 1	56
图 5-13 微表情识别功能测试图 2	56
图 5-14 修改密码功能测试图	57

插表清单

表 1-1 部分运动单元编码表	1
表 1-2 微表情类别与运动单元组合对照表.....	2
表 2-1 LOL 数据集低光照增强结果对比表.....	15
表 2-2 不同特征识别结果对比表	21
表 3-1 实验使用的公开数据集简介表	28
表 3-2 使用 Focal Loss 和交叉熵损失函数的结果对比表	30
表 3-3 改进网络前后的识别结果对比表.....	30
表 3-4 不同方法在 CASME 数据集的结果对比表	32
表 3-5 不同方法在 CASME II 数据库的结果对比表	32
表 4-1 不同网络结构实验结果对比表	39
表 4-2 不同方法在 CASME 数据集识别性能对比表	42
表 4-3 不同方法在 CASME II 数据集识别性能对比表	42
表 4-4 自建数据集信息表	43
表 5-1 登录功能测试对照表	53
表 5-2 注册功能测试对照表	54
表 5-3 主界面功能测试对照表	55
表 5-4 微表情识别功能测试对照表	56
表 5-5 修改密码功能测试对照表	57

第 1 章 绪 论

1.1 研究背景及意义

通常情况下，人们的面部表情会反映其真实情绪状态。然而，有时候人类会刻意控制自己的表情，以隐藏或伪装真实情绪，使得其他人难以准确理解他的意图和情感，从而导致沟通不畅或产生误解^[1]。微表情是人类面部表情中持续时间较短、幅度较小且通常难以察觉的表情，持续时间约为 0.04 秒~0.2 秒^[2]。是一种很难通过故意伪装来掩饰真实情感的自发动作，对于人类情感、意图和认知状态的理解具有重要价值^[3]。

在发现微表情的重要价值后，人们开始研究如何识别微表情。最早的微表情研究围绕心理学展开，研究者们关注人类面部表情中瞬息即逝的微小变化，并尝试通过心理学实验来揭示微表情特征及其与情绪、行为的关联^[4]。在上个世纪，心理学家 Ekman 等^[5]通过大量研究和调查，总结面部表情的动作规律，设计出面部动作编码系统，利用不同运动单元(Action Unit, AU)组合来表示不同的表情，提高了表情分类的准确率和说服力。部分面部 AU 编码如表 1-1 所示。微表情作为表情中的一种，也同样适用面部动作编码系统，该编码系统奠定了微表情研究的基础，部分运动单元组合和微表情对应关系如表 1-2 所示。

前期的微表情识别方法主要依靠人工识别微表情，需要训练有素的观察者来捕捉面部微小的变化，还要有耐心专业的知识来区分短暂出现的微表情类别，还要通过大量实践和训练来提高识别微表情的准确性和可靠性，耗时耗力。

表 1-1 部分运动单元编码表

AU	对应动作	AU	对应动作	AU	对应动作
AU1	内侧眉毛抬起	AU10	上嘴唇抬起	AU17	下巴抬起
AU2	外侧眉毛抬起	AU11	鼻唇沟加深	AU18	嘴唇鼓起
AU4	眉毛下压	AU12	嘴角上扬	AU20	嘴唇向后拉扯
AU5	上眼皮抬起	AU13	拉动嘴角向上	AU22	嘴唇呈漏斗形
AU6	脸颊抬起	AU14	嘴角收缩	AU23	双唇收紧成一
AU7	眼睛收缩	AU15	嘴角向下拉动	AU24	双唇收缩
AU9	皱鼻	AU16	嘴唇向上挤动	AU38	鼻孔扩张

表 1-2 微表情类别与运动单元组合对照表

微表情类别	运动单元组合
高兴	AU6+AU12 等
伤心	AU1+AU4+AU5 等
惊讶	AU1+AU2 等
压抑	AU4 +AU14 等
生气	AU4+AU5+AU7+AU23 等
厌恶	AU4+AU9 等

后来随着研究者们对微表情认知的不断加深，开始着手建立微表情数据集，进行相应的 AU 标注工作，以便开展更深入的研究和分析^{[6][7]}。因为计算机技术在其他计算机视觉任务中取得令人满意的成果，研究者们将机器学习和深度学习应用到微表情的检测和识别任务中，探索如何利用计算机技术来自动检测和识别微表情^{[8][9][10]}，成功地利用深度学习方法提高了微表情识别的准确率和鲁棒性，并尝试将微表情识别技术应用到欺骗检测^[11]、侦察审讯^[12]、医疗诊断^[13]实际场景中。随着微表情识别技术不断进步和研究成果的不断丰富，微表情识别肯定能够在更多领域得到应用和推广，因此基于深度学习的自动微表情识别研究有着重要意义^[14]。

1.2 国内外研究现状

微表情识别算法的一般流程为使用现有公共数据集，通过一些数据预处理技术获得理想的特征图，再利用人工或者计算机去总结学习特征图，经过反复训练和实验以达到准确识别微表情目的^[15]。参照微表情识别一般流程，本节将从人脸微表情图像预处理技术和人脸微表情识别算法两个方面出发，阐述国内外的研究现状。

1.2.1 人脸微表情图像预处理技术研究现状

由于微表情特征不明显，导致很难直接进行微表情识别，为帮助识别微表情研究者们使用了众多的数据预处理技术。现有的预处理技术主要包括人脸检测、帧归一化、运动放大和特征提取等，各预处理技术研究现状如下：

1) 人脸检测

人脸检测是检测视频或图像中是否包含人脸，是微表情识别非常重要的一

步。有些方法使用基于局部图像区域提取的手工特征来构建分类器实现人脸检测，如 Haar 级联分类器^[16]、定向梯度直方图^[17]和基于机器学习的 SVM 方法^[18]，但这些方法无法在不受控制的环境中获得高精度图像。随着神经网络的高速发展与成功，研究者尝试将深度学习应用到人脸检测任务中，受级联分类器影响，Zhang 等^[19]设计了基于级联的多任务卷积神经网络(Multi-Task Convolutional Neural Network, MTCNN)进行人脸检测，并将该方法扩展到关键点检测和人脸定位，其利用级联架构精心设计了深度卷积网络的三个阶段，使用图像金字塔技术让设计的 MTCNN 在多个分辨率下运行，在算法每个级联卷积神经网络(Convolutional Neural Networks, CNN)检测阶段之后引入了校准阶段，用以从粗到精的方式检测人脸和关键点位置。此外，受到特征金字塔网络启发，Zhang 等^[20]提出了用于单张人脸检测的特征聚集网络，该模型核心思想是利用 CNN 固有的多尺度特征，通过聚集不同尺度的高级语义特征图作为线索，通过分层聚集的方式增强低级特征图来检测图像中的人脸。

2) 帧归一化

帧归一化是指在整个微表情序列中为微表情识别选择有意义的帧。微表情识别的许多研究人员应用了时间插值法将输入样本均匀地对齐到相同数量的帧中，以此减小数据之间的尺度差异，提高模型收敛速度和稳定性。计算流程是给定一个微表情视频，时间插值法可以将每一帧中提取的高维视觉特征映射到由嵌入在路径图中的一组三角函数确定的低维连续曲线上，并将整个曲线投影回原始域以获得插值视频。除了时间插值法之外，Xu 等^[21]使用线性插值以较小的误差指示面部微表情运动模式；Ngo 等^[22]提出了稀疏促进动态模态分解来更好地挖掘数据的低维结构和特征，并通过实验验证了稀疏性促进动态模态分解在插值少量帧时，可以提高微表情识别性能。

3) 运动放大

运动放大是将微弱的动作放大，以增加不同运动之间的区分能力，适用于微表情识别任务中。Wu 等^[23]提出欧拉视频放大技术，通过空间金字塔分解、时域带通滤波和线性放大滤波结果，再将滤波结果加上原始图像获得经过放大的结果。但欧拉视频放大也会放大图像中的噪音，后面 Oh 等^[24]提出了基于深度学习的视频放大算法，其中训练网络使用的数据集是人工合成的，人为地制造物体微移动，再人为地放大物体的运动，采集前后的视频作为对照组建立运动

放大数据集，并以此数据集训练神经网络，使网络表现出了强大的运动放大能力。

4) 特征提取

特征提取将现有的图像或视频转化为更容易被计算机理解的其他特征图，也是微表情识别非常重要的一步。其中经常使用的特征是局部二进制模式及其变体和各种光流算法。其中 Ojala 等^[25]是最早在微表情识别任务中引入局部二进制模式(Local Binary Pattern, LBP)算子的团队，该算子可用于描述微表情的纹理特征，但 LBP 特征描述符仅包含空间信息。考虑到微表情中还包含有时间信息，Zhao 等^[26]提出了三正交平面上的局部二进制模式(Local Binary Pattern on Three-Orthogonal Planes, LBP-TOP)，根据空间 X, Y 和时间 T 方向，计算 XY、XT 和 YT 平面的 LBP 特征表示微表情，但 LBP-TOP 计算成本高。为降低计算成本，Wang 等^[27]提出了仅考虑位于三个正交平面重叠线上六个交点的六交点局部二进制模式。Wang 等^[28]提出局部立方体二进制模式，计算感兴趣区域的 LBP 立方体块描述纹理特征，目的是充分利用时间和空间特征。此外光流特征通过利用亮度守恒原理计算位移，估计运动的大小。由于微表情识别高度依赖于面部成分的运动，Liong 等^[29]提出了光流应变作为特征描述符，根据光流应变特征和应变量计算出光流应变图，利用图中位移矢量表示微表情。Happy 等^[30]提出了光流方向模糊直方图特征描述符，使用直方图模糊化从光流矢量方向构建合适的角度直方图表示微表情。Liong 等^[31]提出了双加权定向光流作为特征描述符，使用单个顶点帧的双加权定向光流表示微表情。Liu 等^[32]引入了主向平均光流(Main Directional Mean Optical Flow, MDMO)作为特征描述符，从感兴趣区域提取基于局部的光流，计算在主方向上的分量用于表示微表情。

综上，通过对人脸微表情识别图像预处理技术国内外研究现状分析可知，目前大多数研究内容都是对现有数据集的数据预处理技术，加强微表情的时空信息表达，但在微表情识别的实际应用过程中会出现低光照环境，拍摄的图像通常会出现明显退化会影响识别准确率，是微表情识别系统应用过程中需要面对的问题。本课题针对真实环境下的低光照情况，将研究一种基于 Transformer 的低光照图像增强算法，拟融合 Transformer 和注意力机制，加强对重要特征的关注，以此提高网络对低光照图像增强的能力；拟探究一种融合 TVL1 光流特征、光流应变特征的微表情特征描述符，以此提高光流估计精度，

增强面部运动特征表达。

1.2.2 人脸微表情识别算法研究现状

当 Ekman 总结出微表情的特性和表现方式并发表论文时，标志着系统性微表情研究开始。最早的微表情识别方法是基于人工的方法，研究各种特征来描述微表情，通过统计归纳的方式帮助识别微表情。随着计算机技术高速发展，将各种特征输入到分类器完成微表情识别，常使用的分类器有支持向量机、k 最近邻、随机森林和基于深度学习的 Softmax 等。其中深度学习的 Softmax 方法表现优异，成为现在研究的主流。

Patel 等^[33]将 CNN 模型首次应用于微表情识别任务中，但当时缺乏微表情数据无法训练出可靠的 CNN。为应对 CNN 训练中缺乏数据的挑战，Takalkar 等^[34]通过数据增强技术扩展了可用的数据集，从这些数据集中生成额外的合成图像，这些数据集中的图像被水平翻转以生成合成数据，随后被裁剪和对齐作为 CNN 的输入进行微表情识别。陈汤慧等^[35]将在宏表情数据集上预训练的 Xception 网络，通过迁移学习方式运用到微表情识别上，并使用基于深度学习的运动放大技术放大微表情，结合压缩与激励注意力机制提高微表情识别准确率。Quang 等^[36]基于胶囊网络提出了一个完整的微表情识别框架，该框架检测并预处理微表情序列中的顶点帧，再使用 ImageNet 数据集预训练胶囊网络的第一个卷积层，最后将顶点帧的面部区域输入到胶囊网络中进行微表情识别。考虑更好地提取微表情的时间和空间特征，Peng 等^[37]提出了一种双时间尺度卷积神经网络(Dual Temporal Scale Convolution Neural Network, DTSCNN)来识别自发的面部微表情，使用光流序列作为网络的输入，网络使用不同的 DTSCNN 流来处理微表情视频的光流序列，DTSCNN 的每个流中都存在一个独立的浅层网络，可以有效应对过拟合问题，并通过浅层网络生成高级特征，最后使用线性 SVM 对输出特征进行微表情识别。Kim 等^[38]提出了 CNN 和长短期记忆网络的组合模型来处理空间和时间信息，在空间域表示中，由 CNN 学习微表情从开始到结束的每个表达阶段图像的空间特征；再使用长短期记忆网络学习从 CNN 中提取的空间信息之间的时间特征，考虑不同微表情类别的状态和状态连续性的变化，使得提取的特征对光照变化具有鲁棒性，可以更有效地识别微表情。Fan 等^[39]根据面部表情光流的对称性，使用自监督学习技术开发了一种对称对比视觉转换器来加强面部左右部分相似面部动作特征的学习，从而有效地识别微表情。赵

明华等^[40]提出了一种结合注意力机制的三流 CNN，模型引入 CBAM 注意力机制关注重要的时间和空间特征并忽略不相关特征，从而实现精确微表情识别。Wang 等^[28]提出了一种能够处理非欧几里得结构数据的局部立方体二值模式时空图卷积网络(Local Cube Binary Pattern Spatial-Temporal Graph Convolutional Network, LCBP-STGCN)，该网络以感兴趣区域的局部立方体二进制模式为输入，通过空间图卷积网络和时空卷积网络交替卷积获得高层次的时空特征，提高微表情识别准确率。

综上，通过对人脸微表情识别算法研究现状分析可知，基于深度学习的微表情识别最重要的是提取样本中的时间和空间特征。考虑到微表情发在人脸局部区域的特性，这使得不同位置的特征重要程度不一样。为有效识别微表情，本课题将以融合光流特征为输入，研究结合注意力机制模块的神经网络模型，加强模型捕捉人脸关键信息的能力，提高对微表情识别的准确率。

1.3 主要研究内容

微表情是一种短暂、微弱且难以察觉的情绪表达形式，通常持续时间很短，仅有几分之一秒，又因其不可控制或隐藏的特性，可以反映个人的真实情感。因此，微表情识别在人际交流、情绪领域以及安全监控等方面都具有重要的应用意义。本文将围绕微表情预处理技术以及微表情识别展开研究，首先，使用基于 Transformer 的低光照图像增强网络获取高质量图像，再使用其他预处理技术获得融合光流特征；其次，使用融合光流作为网络输入，探究融合注意力机制的神经网络模型的微表情识别效果；最后，利用研究的网络模型设计微表情识别系统，尝试微表情识别的实际应用。全文的主要研究内容框架以及各章节安排如图 1-1 所示，具体章节安排如下：

第 1 章：阐述本课题的研究背景和意义，分析人脸微表情图像预处理技术和微表情识别算法的研究现状；

第 2 章：研究人脸微表情预图像处理方法。探究基于 Transformer 的图像增强方法、融合 TVL1 光流和光流应变的特征提取方法，并给出相应实验和结果分析；阐述基于 LBP 的关键帧选取方法、基于 MTCNN 的尺度变换方法，为后续微表情识别的研究提供理论基础；

第 3 章：探讨一种融合 MobileViT 和 CA 注意力机制的轻量化微表情识别模

型。阐述网络模型结构，包括倒残差结构、深度可分离卷积、局部特征提取模块、全局特征提取模块、特征融合模块以及注意力模块，之后求取 CASME 和 CASME II 两个数据集的融合光流作为输入对网络模型进行微表情识别实验，证明了网络模型对微表情识别的有效性；

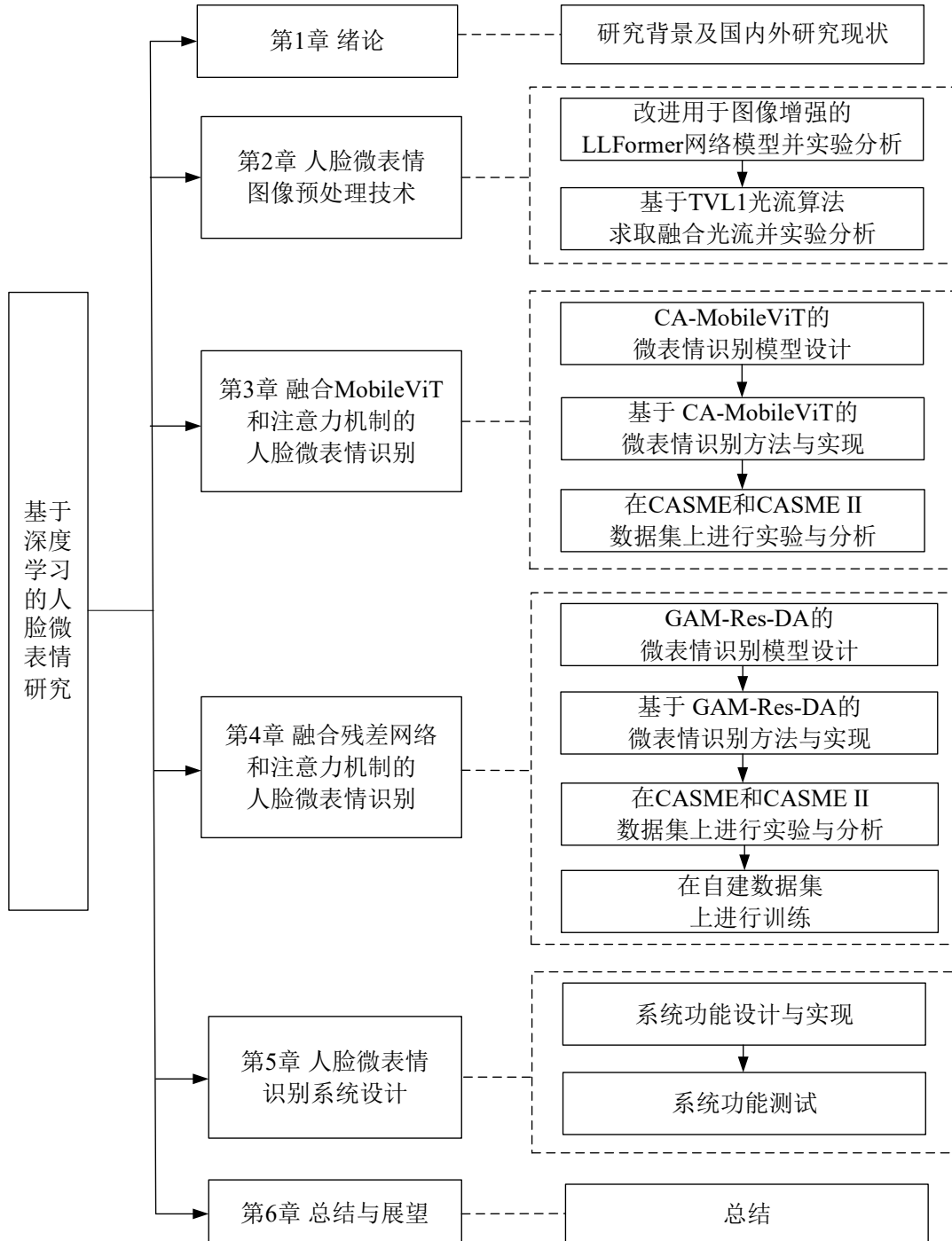


图 1-1 研究内容框架图

第 4 章：研究一种融合 GAM 和 DA 两种注意力机制的残差网络微表情识别

模型。阐述模型各个模块结构，包括残差网络模块、GAM 模块以及 DA 模块，之后求取 CASME 和 CASME II 两个数据集的融合光流作为输入对网络模型进行微表情识别实验，实验证明所研究模型微表情识别准确率更高，并在自建数据集实验检验模型识别性能；

第 5 章：设计一款基于深度学习的微表情识别系统。首先分析系统需求，其次确定系统功能，然后分模块完善各个功能，最后进行整体系统功能测试，实现微表情识别的应用；

第 6 章：总结与展望。对本文的研究内容进行概括总结，分析研究的不足之处，并提出后面的研究方向。

1.4 本章小结

本章首先阐述了微表情识别的研究背景和重要的现实意义；然后按照人脸微表情预处理技术和识别算法两个方向来阐述国内外的研究现状；最后对全文研究框架进行概括介绍，简要阐述了各个章节的主要内容，为后续的详细研究做铺垫。

第2章 人脸微表情图像预处理技术

微表情是一个持续时间短、运动幅度小的面部运动，头部的移动和环境背景信息会影响微表情识别，所以为减少外界干扰通常会裁剪出视频序列中的人脸图像并通过标定点旋转对齐。此外还需选择合适的特征图像作为深度神经网络的输入，微表情是发生在人脸的连续动作，所以选取的输入特征最好包括时间信息和空间信息，这样可以最大程度帮助识别微表情以此提高微表情识别准确率。本章会阐述为提高微表情识别准确率使用的预处理技术，包含图像增强、选取关键帧、尺寸变换、特征提取 4 个步骤。微表情数据集预处理流程如图 2-1 所示。

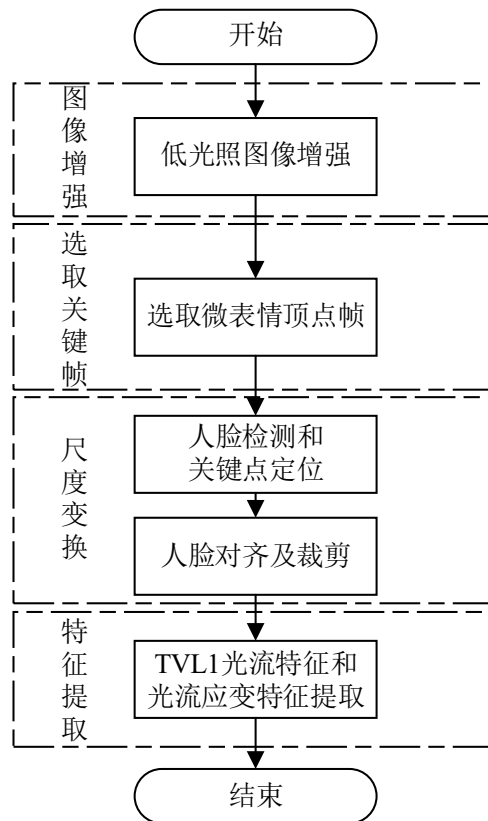


图 2-1 数据集预处理流程图

2.1 图像增强

考虑到在微表情识别的实际应用过程中，真实的外部环境无法像采集现有数据集时的实验室环境一样可控。会出现弱光环境下的微表情，在弱光条件下拍摄的图像通常会出现明显退化，例如能见度低、对比度低和噪声高等问题^[41]，

这些会极大地影响微表情识别准确率。出于减弱这些影响的目的，本章研究了一种基于 CBAM-LLFormer 的深度学习低光照增强方法。

2.1.1 CBAM-LLFormer 低光照图像增强网络模型

CBAM-LLFormer 网络模型由 CBAM 注意力模块、Transformer 模块和跨层注意力融合模块组成。在 LLFormer 网络^[42]采用分层结构缓解低光照图像增强任务的计算瓶颈的基础之上，引入已有的 CBAM 注意力加强对浅层重要特征的关注度。CBAM-LLFormer 低光图像增强网络模型如图 2-2 所示。

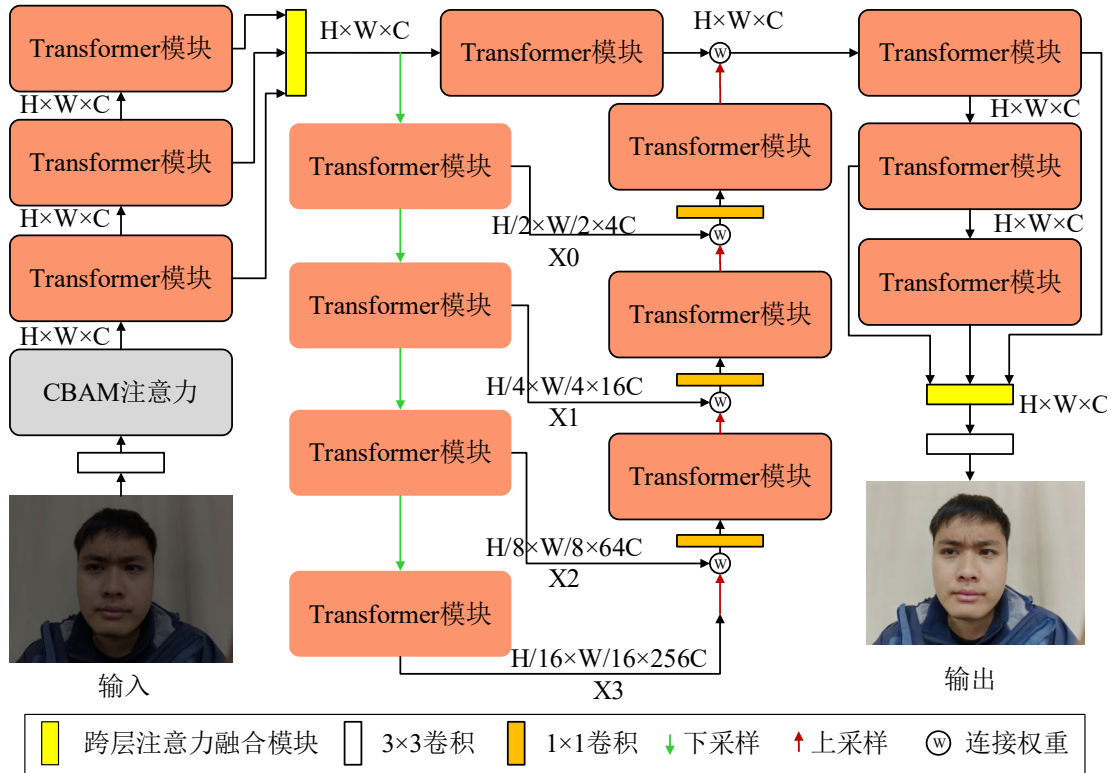


图 2-2 CBAM-LLFormer 低光照图像增强网络模型图

首先，通过 3×3 卷积操作提取低光照图像的浅层特征，再引入 CBAM 注意力提取重要位置信息；其次，通过三个 Transformer 模块进行特征提取，将三次提取的特征图利用跨层特征融合模块形成增强特征图；然后，对增强特征图使用 unshuffle 操作进行四次下采样，再使用 shuffle 操作进行四次上采样，每上采样后，进行注意力特征融合，并使用 1×1 卷积操作减少特征的损失，逐步恢复到原始输入特征的尺寸；最后，经过三个 Transformer 模块进行特征提取，再利用跨层特征融合模块生成增强特征用于图像重建，并用 3×3 卷积操作获得增强后的图像。

2.1.2 CBAM 模块

CBAM 是一种在空间和通道两个维度上进行注意力操作的注意力机制^[43]。具体结构如图 2-3 所示。首先，通道注意力模块将输入特征 F 执行并行的最大池化和平均池化操作，将特征 F 从 $\mathbb{R}^{C \times H \times W}$ 大小变为两个 $\mathbb{R}^{C \times 1 \times 1}$ 大小的特征；其次，使用多层感知器进行特征提取，再将两个输出结果相加；然后，使用 Sigmoid 函数求得通道注意权重 M_c 。空间注意力模块会对经过通道注意力模块优化后的输出特征矩阵 F_1 ，先执行最大池化和平均池化操作将特征 F_1 从 $\mathbb{R}^{C \times H \times W}$ 大小变为两个 $\mathbb{R}^{C \times 1 \times 1}$ 大小的特征，经过 Concat 操作拼接两个特征，再使用卷积进行特征提取；最后，使用 Sigmoid 函数求得空间注意权重 M_s 。由此可以得出如下的 CBAM 注意力计算公式：

$$F_1 = M_c(F) \otimes F \quad (2-1)$$

$$F_2 = M_s(F_1) \otimes F_1 \quad (2-2)$$

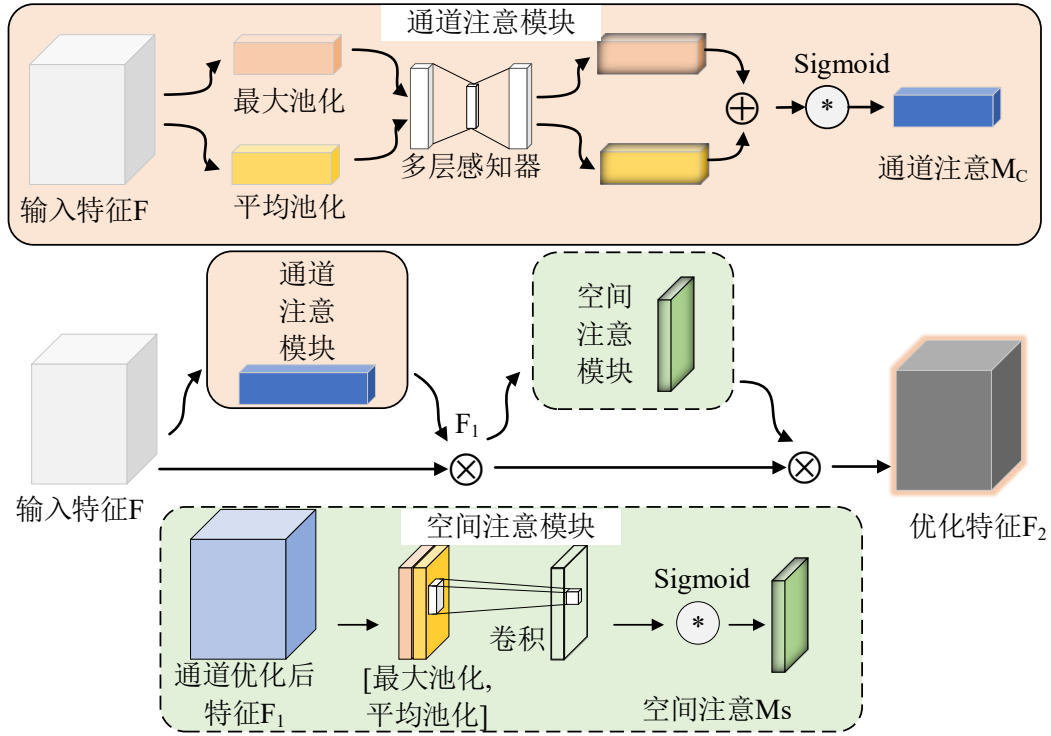


图 2-3 CBAM 结构图

2.1.3 Transformer 模块

Transformer 模块会在特征的高和宽上执行自注意力机制，以较少的计算复杂度捕获非局部自相似度和特征的远程依赖关系。此外，在自注意力机制后，设计了一种新的双门控前馈网络，该网络采用双门控机制来关注有用的特征。

Transformer 模块结构如图 2-4 所示。

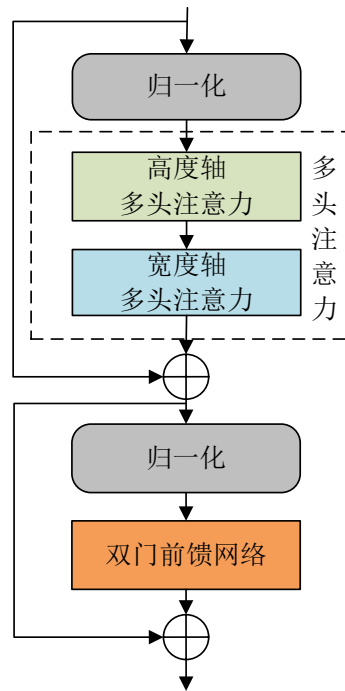


图 2-4 Transformer 模块结构图

高度轴多头注意力首先会对给定大小为 $\mathbb{R}^{C \times H \times W}$ 特征 X ，并行使用三个 1×1 卷积操作加强特征 X 的特征表达和 3 个 3×3 的深度可分离卷积提取局部特征获得三个特征 Q 、 K 、 V ；其次对 Q 、 K 进行 reshape 操作和点积操作获得权重矩阵；然后对 V 进行 reshape 操作和权重矩阵相乘，再使用 reshape 和 1×1 卷积操作获得输出特征。宽度轴的计算流程和高度轴类似。使用的深度可分离卷积结构如图 2-5 所示，将正常卷积分为深度卷积(Depthwise Convolution, DW)和逐点卷积两个部分。

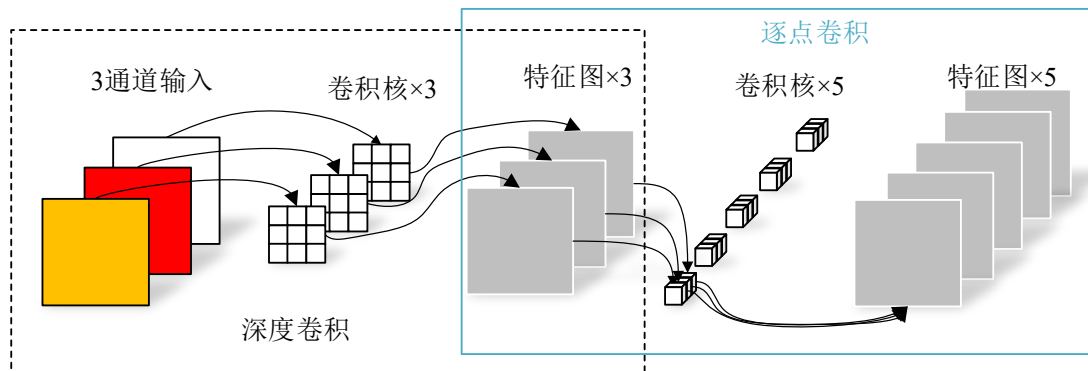


图 2-5 深度可分离卷积结构图

双门前馈网络结构如图 2-6 所示。为实现高效的特征转换，引入双门控机制

和局部信息增强，并提出了一种新的双门控前馈网络。对于双门控机制，首先，在每条路径上应用 1×1 卷积和 3×3 深度卷积操作来丰富局部信息；然后，应用双 GELU 函数和元素乘积来过滤信息较少的特征；最后，使用元素求和融合两条路径中的有用信息。

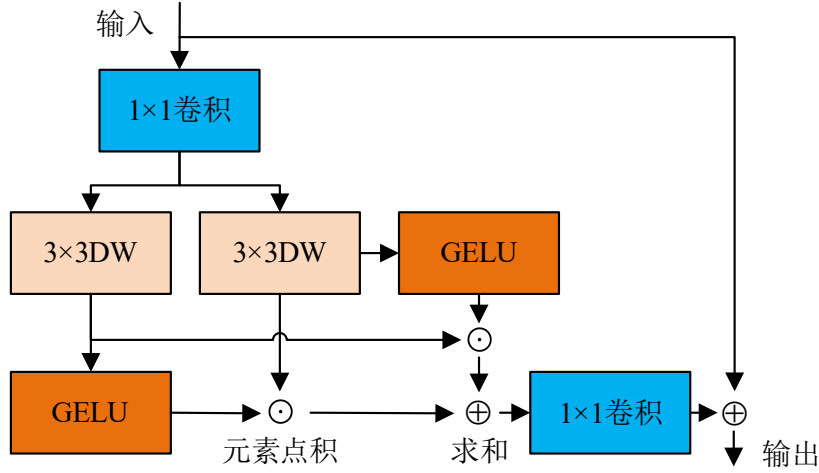


图 2-6 双门前馈网络结构图

2.1.4 跨层注意力融合模块

跨层注意力融合模块会将三个 Transformer 的输出根据注意力权重融合为加强特征图。跨层注意力融合模块结构如图 2-7 所示。首先，将 3 个串联特征 reshape 为 $\mathbb{R}^{H\times W\times 3C}$ 尺寸的特征 F_{in} ，使用 1×1 卷积操作来聚合逐像素的跨通道上下文；其次，使用 3×3 深度卷积操作来生成 Q ， K 和 V ；然后，使用 reshape 将 Q 和 K 分别转为 $\mathbb{R}^{3\times HWC}$ 和 $\mathbb{R}^{HWC\times 3}$ 的二维矩阵，再通过矩阵乘法计算大小为 3×3 的注意力权重矩阵；最后，将 V 乘以注意力权重矩阵。并添加输入特征 F_{in} 再进行 1×1 卷积操作获得最终输出。

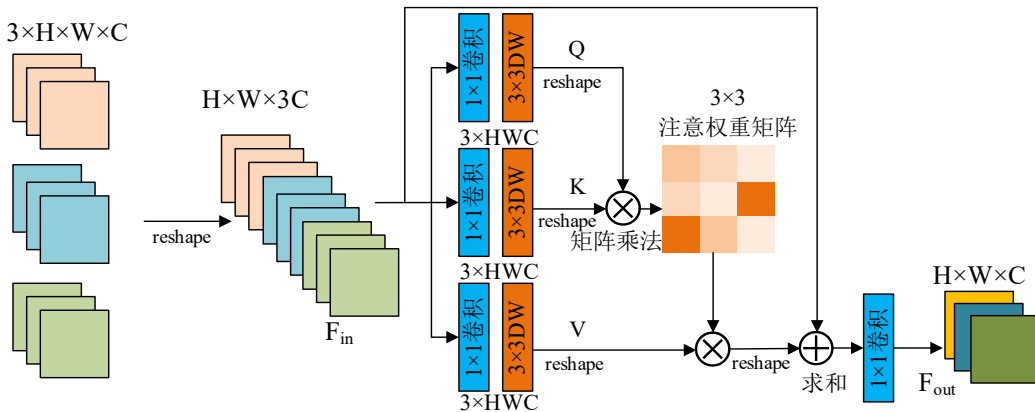


图 2-7 跨层注意力融合模块结构图

2.1.5 实验分析

为验证研究的深度学习低光照增强方法的可行性，采集了真实环境下作者本人的低光照人脸微表情图像，再使用本章的低光照图像恢复算法进行图像增强效果测试，真实环境下作者本人的低光照图像处理前后如图2-8所示。由图可知，经过图像增强后，能够有效地提高图像的亮度且可以保留人脸图像的色彩信息和细节信息，有助于后续微表情识别。



图 2-8 低光照图像增强对比图

为更加客观地分析图像增强效果，选用峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似度(Structural Similarity index, SSIM)作为图像质量评价指标。PSNR经常用作图像压缩等领域信号重建质量的评价，PSNR的单位是dB，数值越大表示失真越小，求取 PSNR 常用的方式是通过均方误差(Mean Squared Error, MSE)来计算。每张图像都具有特定的结构，同一张图像上的像素间有很强的从属关系，这些从属关系反映了视觉场景中的结构信息。由此可以根据两张图像的结构相似性来评价图像恢复程度，称为结构相似法。结构相似性取值范围从 0 到 1，两张图像结构越相似，则 SSIM 的值越接近 1。两个评价指标的计算公式如下：

$$\begin{cases} MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - Y(i, j))^2 \\ PSNR = 10 \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right) \end{cases} \quad (2-3)$$

$$SSIM(X, Y) = \frac{(2\mu_X \mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \quad (2-4)$$

其中, X 和 Y 分别表示当前图像和参考图像, H 和 W 分别表示图像的高和宽, $X(i, j)$ 和 $Y(i, j)$ 分别表示图像 X 和图像 Y 上像素点, n 表示每个像素的比特数, μ 表示图像像素平均值, σ 表示图像像素方差, σ_{xy} 表示协方差, c_1 和 c_2 表示用来维持稳定的常数。

为验证研究的深度学习低光照增强方法的可靠性, 在公开的 LOL 数据集^[44]上进行对比试验。LOL 数据集由 500 对低光照图像和正常图像组成, 训练集有 485 对图像, 测试集有 15 对图像, 其图像大都是室内场景, 图像分辨率为 600×400 。为获取更多的上下文信息, 选取 128×128 的图像 patch 在 LOL 数据集上进行 100 epochs 的实验; 为更快地训练模型和提高网络性能, 使用余弦退火法来确定训练过程学习率并使用鲁棒性较强的 SmoothL1Loss 作为网络训练时的损失函数。并将本章方法和其他专门为低光照图像增强设计的 5 种算法进行比较, 具体结果如下表 2-1 所示, 表中最优结果加粗显示。

表 2-1 LOL 数据集低光照增强结果对比表

方法	PSNR	SSIM
LIME ^[41]	16.7586	0.4449
BIMEF ^[45]	13.8752	0.5950
KinD ^[46]	17.6476	0.7715
Uformer ^[47]	18.5479	0.7212
LLFormer ^[42]	20.7429	0.8035
本章方法	21.3476	0.8323

从表 2-1 可以看出本章研究的基于 Transformer 的深度学习方法在 LOL 数据集上获得了最高的 PSNR 和 SSIM, 取得优异的性能。实验证明了本章方法优于其他算法, 可以更好地提高图像质量, 保留更多细节信息。

2.2 关键帧选取

微表情是一个连续的变化过程, 在微表情视频序列中微表情开始的那一帧被称为起始帧, 面部肌肉运动强度最高的那一帧被称为顶点帧, 面部肌肉从收缩恢复到松弛状态的那一帧被称为结束帧。其中顶点和起始帧, 两者时间间隔短且面部差异最为明显, 所以基于两者的光流特征能够帮助识别微表情。那么要求两者的光流就需要定位顶点帧, 除人工标注外, 常用的自动顶点帧定位方

法是计算每帧的 LBP 纹理特征，再计算每帧与视频第一帧之间 LBP 值的差值，差距最大的那一帧作为顶点帧。

其中 LBP 是一种简单且高效的局部纹理特征描述算子，具有旋转不变性和灰度不变性，可以在一定程度上消除光照变化带来的影响。其计算过程是先求取图像的灰度图像，在灰度图像中划出一个 3×3 的像素区域，周围的 8 个像素值和中心像素值作比较，当周围像素值大于中心像素值，则该像素点的位置被标记为 1，相反则标记为 0。中心像素的 LBP 值反映了该像素周围区域的纹理信息。计算示例如图 2-9 所示。

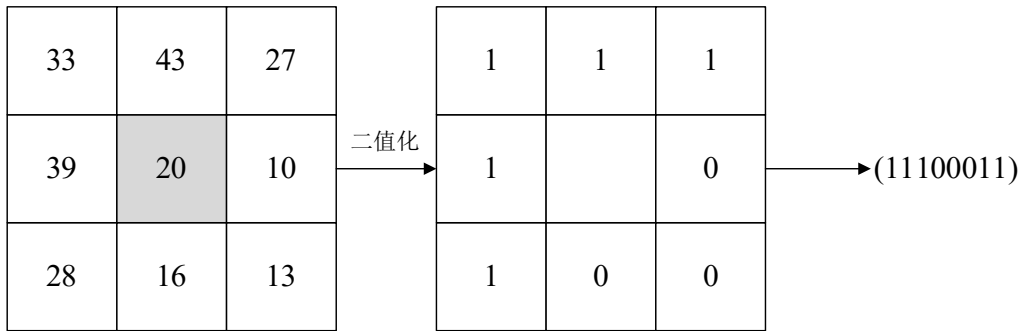


图 2-9 LBP 计算示例图

视频样本顶点帧的求取示例如图 2-10 所示，图中为作者本人。为降低计算成本，后面在公共数据集的识别算法实验中，使用公共数据集已标注的起始帧和顶点帧信息。



图 2-10 顶点帧求取示例图

2.3 尺度变换

真实环境下每个微表情样本中人脸的位置各不相同，环境背景也千变万化，但背景信息对微表情识别是无用的，不仅会消耗大量计算机算力，还会使神经网络学习到无用信息，从而降低微表情识别准确率。为了可以集中学习到微表

情视频中关键的人脸信息，在获取到顶点帧和起始帧位置后，先使用MTCNN^[19]完成两帧图像中的人脸区域检测和人脸关键点检测，再根据人脸关键点的信息进行人脸对齐并裁剪出人脸图像，用以除去多余的无用背景信息。MTCNN可以同时完成人脸检测和关键点检测两个任务，输出人脸框的长度、宽度以及左上角坐标，此外还有左眼、右眼、鼻子、左嘴角以及右嘴角五个关键点的坐标信息。网络人脸检测和关键点检测流程如图 2-11 所示。

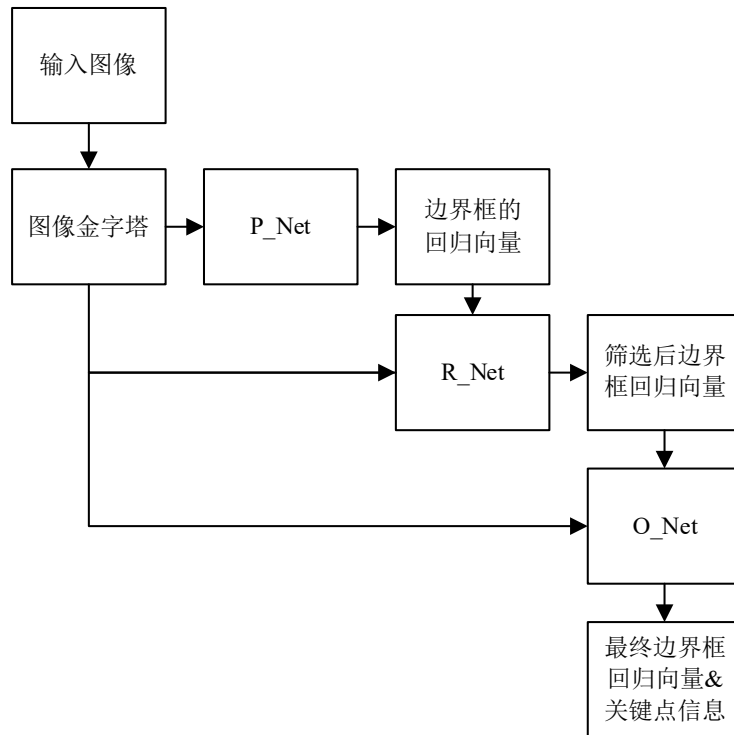


图 2-11 MTCNN 人脸检测和关键点检测流程

具体操作流程如下：第一步对输入的微表情单帧图像进行金字塔操作(进行一系列尺度变换)；第二步将得到的图像金字塔输入到建议网络(Proposal Network, P-Net)，该层网络会将输入进行三次卷积和一次最大池化操作，再通过人脸检测器判断是否有人脸，如果有人脸会获取人脸区域的候选窗口和边界回归向量，并使用非极大值抑制合并高度重叠的候选框，过滤掉那些效果差的候选框；第三步使用优化网络(Refine Network, R-Net)去除大量的非人脸框，其结构在 P-Net 的基础上增加了池化层和全连接层，从而筛除在 P-Net 层生成的效果差的人脸区域窗口候选框；最后一步使用输出网络(Output Network, O-Net)在 R-Net 网络的基础上进行更精细的处理，与 R-Net 的结构相比多了卷积层和池化层，加强提取特征，进行人脸判别和人脸区域判定，输出人脸区域的关键点坐

标和人脸框位置信息。人脸检测和关键点检测结果如图 2-12 所示，左侧两列分别为 CASME II 和 CASME 数据集样本，右侧为作者本人图像。

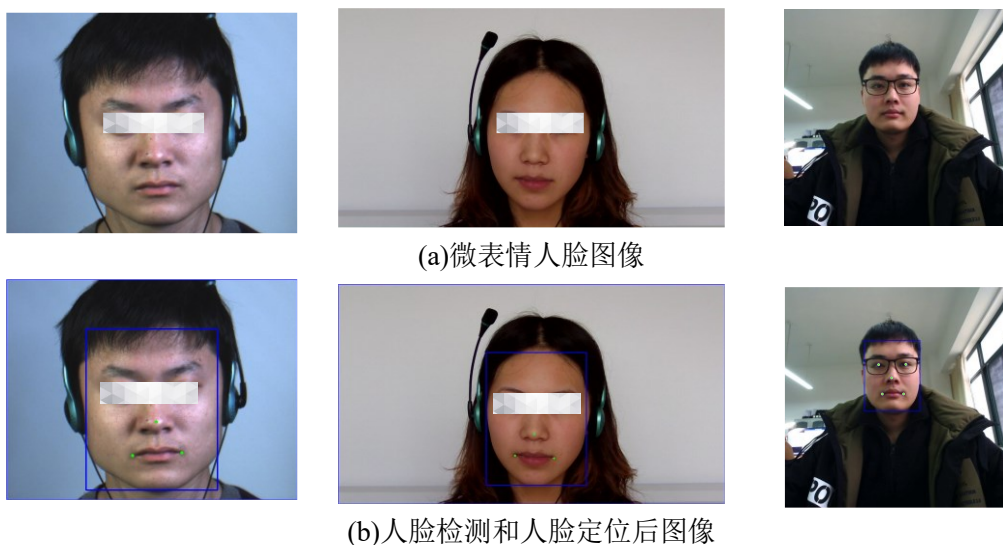


图 2-12 人脸检测与人脸定位示例图

其次根据左眼和右眼的位置信息，使用反正切函数计算出人脸对齐需要的旋转角度，然后以头部框中点作为原点对人脸进行旋转保证左右眼睛位置在同一个水平线上，旋转角度计算公式如公式(2-5)所示。然后根据人脸位置和尺寸裁剪出人脸区域，最后调整为指定大小并保存，对齐及裁剪示例如图 2-13。

$$\theta = \arctan[(y_2 - y_1)/(x_2 - x_1)] \quad (2-5)$$

其中， (x_1, y_1) 和 (x_2, y_2) 分别代表人脸关键点检测得到的左眼和右眼的坐标。 θ 代表待求取的图像旋转角度。

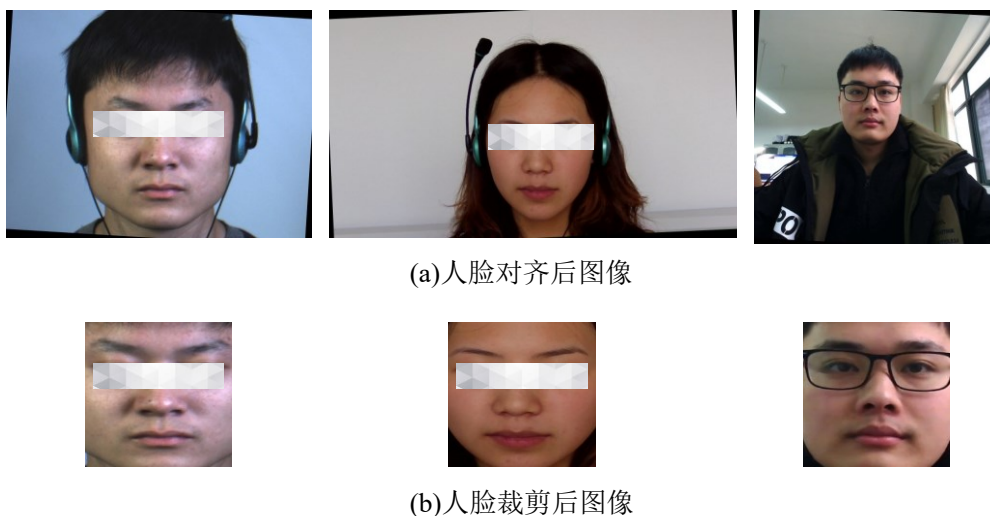


图 2-13 人脸旋转对齐图与裁剪示例图

2.4 特征提取

光流估计常被用于视觉领域内的目标检测、分类和跟踪等任务。表达的是空间运动物体在成像平面上像素运动的瞬时速度，可以表达相邻两帧图像的变化，包含了运动目标的时间和空间信息，因此可以作为网络输入特征帮助识别微表情。又因为微表情中包含较多微弱运动，光流应变可以近似的描述面部运动强度，所以还计算光流应变作为补充特征。

2.4.1 TVL1 光流特征

光流的求取要符合基本的条件。条件一是每一帧图像中的亮度恒定不变，意味着运动物体的像素值不变只有空间位置变化，这样就可以将像素的位移作为物体的运动矢量；条件二是运动连贯且运动幅度小，意味着可以将图像之间的灰度变化视为由光流引起的。常见的光流估计方法有许多种，其中 TVL1 光流算法^[48]是一种基于变分原理的光流估计算法，通过最小化一个能量函数来计算两帧图像间像素点的运动，计算方法如公式(2-6)所示。算法使用变分原理保证两帧之间灰度值变化光滑性，使用 L1 范数保证相邻像素点之间的光流的相似性，保证对光照变化和噪声有较好的鲁棒性，本章利用 OpenCV 库 DualTVL1Optical-Flow 光流计算函数获取图像光流特征图。

$$E(F(X)) = \int_{\Omega} \{ \lambda |I_0(X) - I_1(X + F(X))| + |\nabla F(X)| \} d_x \quad (2-6)$$

其中， E 表示待求取的能量函数， Ω 表示整个图像面， I_0 和 I_1 表示前后两帧图像， X 等于 (x, y) 表示图像上的一个像素点， λ 是自由参数保证光流场变化平滑， $F(X)$ 表示在 X 处的光流场包含垂直光流矢量和水平光流矢量， $\nabla F(X)$ 表示光流场在 X 处的微分。

2.4.2 光流应变

光流应变^[29]是指光流场 $F(X)$ 的空间变化率，描述的是在水平和垂直方向光流在不同位置的变化情况。通过求取梯度来检测图像边缘的强度信息，减少光流估计过程中产生的误差。公式(2-7)是光流应变的 Hessian 矩阵。

$$\varepsilon = \begin{bmatrix} \varepsilon_{xx} = \frac{\partial u}{\partial x} & \varepsilon_{xy} = \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \\ \varepsilon_{yx} = \frac{1}{2} \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) & \varepsilon_{yy} = \frac{\partial v}{\partial y} \end{bmatrix} \quad (2-7)$$

其中, ε 表示光流应变, ε_{xx} 和 ε_{yy} 表示光流矢量法向应变分量, ε_{xy} 和 ε_{yx} 表示光流矢量切应变分量, 并用各分量的平方和 $S = \sqrt{\varepsilon_{xx}^2 + \varepsilon_{yy}^2 + \varepsilon_{xy}^2 + \varepsilon_{yx}^2}$ 计算像素点的应变大小。

2.4.3 融合光流

在计算出样本视频起始帧和顶点帧的垂直光流特征图、水平光流特征图和光流应变图之后, 使用 Python 中 numpy 库中的 stack 函数堆叠三种特征图, 获取三通道的融合特征图。具体步骤如下: 先根据图像地址读取三种光流特征图的灰度图像; 再将图像转化为数组形式; 最后将三个数组堆叠获得融合光流图像。获取融合特征图的相关代码如下:

代码: 获取融合光流特征图

```
import numpy as np

from PIL import Image

#获取水平光流 U、垂直光流 V 和光流应变 S 图像

image1 = Image.open('U-path').convert("L")
image2 = Image.open('V-path').convert("L")
image3 = Image.open('S-path').convert("L")

#转换为数组格式

array1 = np.array(image1)
array2 = np.array(image2)
array3 = np.array(image3)

#堆叠为三通道数组

combined_array = np.stack((array1, array2, array3), axis=2)
```

2.4.4 实验分析

为验证融合光流特征的有效性, 求取 CASME II 数据集中各样本的水平光流、垂直光流、光流应变和融合光流特征图, 并将数据集样本按照 3:1 的比例划分为训练集和测试集, 再使用 SVM 分类器进行识别实验, 采用准确率(Accuracy, ACC)作为识别实验的评价指标, ACC 计算公式如(2-8)所示。输入不同特征识别结果如表 2-2 所示。

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2-8)$$

其中, TP 是真阳性的数量, TN 是假阴性的数量, FP 是假阳性的数量, FN 是真阴性的数量。

表 2-2 不同特征识别结果对比表

输入特征	ACC
水平光流	0.4032
垂直光流	0.5645
光流应变	0.5000
融合光流	0.5968

根据表 2-2 中实验结果, 四种光流特征在 SVM 分类器中融合光流的识别效果最优, 较其他特征准确率有所提升。但使用 SVM 分类器的识别方法的总体识别准确率有待提高, 所以需要进一步改进人脸微表情识别算法。

在输入到深度神经网络模型前, 会使用随机水平翻转来增加网络训练时的数据量, 降低过训练拟合风险, 提高网络性能。融合特征图获取和数据增强示例如图 2-14 所示。训练集数据增强代码如下:

代码: 实现数据增强与归一化

```
from torchvision import transforms

transform_train = transforms.Compose([
    transforms.RandomHorizontalFlip(),
    transforms.ToTensor(),
    transforms.Normalize((0.4914, 0.4822, 0.4465), (0.2023, 0.1994, 0.2010)),])

transform_test = transforms.Compose([
    transforms.ToTensor(),
    transforms.Normalize((0.4914, 0.4822, 0.4465), (0.2023, 0.1994, 0.2010)),])
```

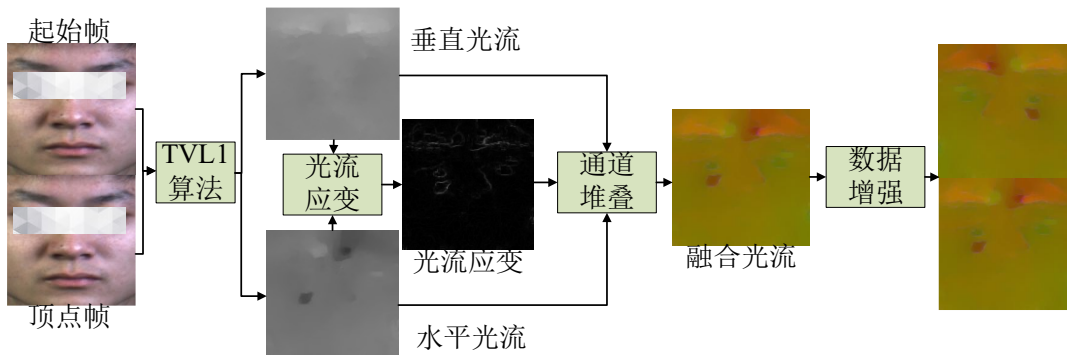


图 2-14 融合特征图获取和数据增强示例图

2.5 本章小结

本章阐述了用于提高微表情识别准确率的图像预处理技术，包括图像增强技术、关键帧选取技术、尺度变换技术和特征提取技术。主要阐述图像增强和特征提取的部分优化，考虑到微表情识别系统的实际应用情况，研究了一种基于 Transformer 的低光照微表情图像增强方法，在 LLFormer 基础之上，引入 CBAM 注意力，加强对浅层特征中重要信息关注度，在 LOL 数据集上验证了方法的有效性；特征提取时在 TVL1 光流特征的基础上引入光流应变特征作为补充，并通过实验对比四种光流特征对微表情识别的影响，发现融合光流可以提高微表情识别准确率，为后续识别算法研究提供帮助。

第3章 融合 MobileViT 和注意力机制的人脸微表情识别

针对以融合光流特征为输入，SVM 分类器的微表情识别准确率表现不佳的问题，本章研究了一种融合 CA 注意力机制和 MobileViT-V3 的 CA-MobileViT 特征提取网络用以完成特征提取和微表情识别。CA-MobileViT 在 MobileViT-V3 网络的基础上进行优化。首先，针对微表情发生在人脸局部区域的特性，引入 CA 注意力增强网络特征表达能力；其次，使用更加平滑的 SiLU 替换 ReLU6 激活函数，改善网络优化能力；然后，考虑到微表情数据集样本不平衡，使用 Focal loss 损失函数来降低类别不平衡带来影响；最后，在 CASME 和 CASME II 数据集上进行实验并和其他经典方法对比，结果证实了该方法的有效性和可行性。

3.1 CA-MobileViT 特征提取网络结构设计

3.1.1 整体网络模型

为了可以更有效的提取微表情光流图中关键的信息，在 MobileViT-V3-small 网络第一次下采样后引入现有的 CA 注意力机制，根据注意力权重关注重要的微表情浅层特征。如图 3-1 所示是 CA-MobileViT 模型结构。

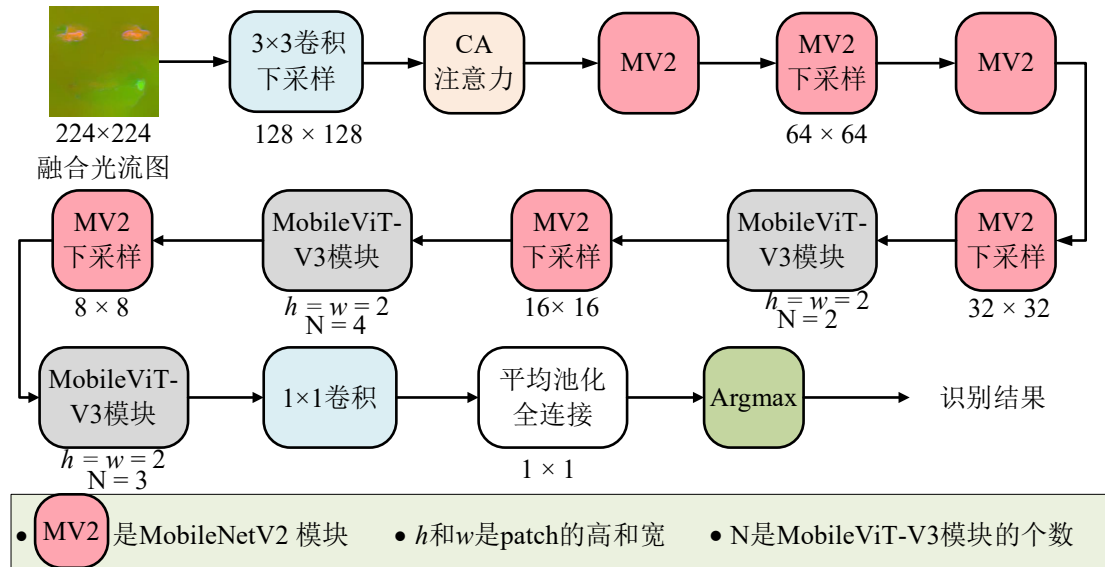


图 3-1 CA-MobileViT 模型结构图

将原 MobileViT-V3 模型中的 ReLU6 激活函数换为 SiLU 激活函数，以此缓解梯度消失问题，并且 SiLU 函数比 ReLU6 函数更加平滑在整个定义域都有导数，这会提供更好的梯度传播有利于网络优化。函数表达式如公式(3-1)和公式

(3-2)。在定义域的曲线如图 3-2 所示。

$$\text{ReLU6}(x) = \min(6, \max(0, x)) \quad (3-1)$$

$$\text{SiLU}(x) = \frac{x}{1 + e^{-x}} \quad (3-2)$$

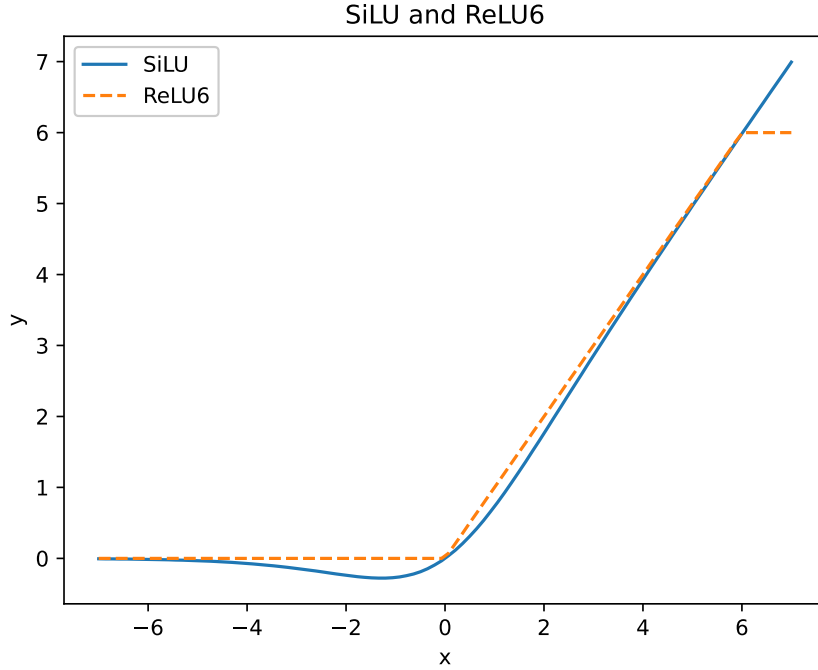


图 3-2 SiLU 和 ReLU6 激活函数图

考虑到微表情数据集中各个类别样本数量不均衡，使用 Focal Loss 函数代替交叉熵损失函数，其在交叉熵损失函数的基础上引入一个可调参数 γ 来降低易分类样本的权重，从而使模型更加关注分类困难的小样本。这样可以有效地解决类别不平衡问题，提高模型在少数类别上的关注度，计算公式如下：

$$\text{Focalloss} = -\alpha(1 - p)^\gamma * \log(p) \quad (3-3)$$

其中， p 是模型预测的概率， α 是样本类别间的权重参数， γ 是可调参数，且取值要大于等于零。参考 Lin 等^[49]实验结果将 α 设为 0.25，将 γ 设为 2。

3.1.2 MobileViT-V3 网络模型

考虑到 MobileViT-V3^[50]是结合 CNN 和 Transformer 的混合架构，网络模型具有两种架构的优势，可以通过 CNN 获得空间归纳偏置和局部像素之间的相关性，也可以通过 Transformer 的自注意力机制获取面部微表情全局信息，所以选用 MobileViT-V3 作为微表情识别的主干网络。MobileViT-V3 是苹果公司发布一

个轻量型主干网络，由一系列 MobileNetV2 模块^[51]和 MobileViT 模块串联而成。其中 MobileNetV2 模块具体结构如图 3-3 所示，左边是步长为 2 时的结构，右边是当步长为 1 时引入了跳跃连接的结构。

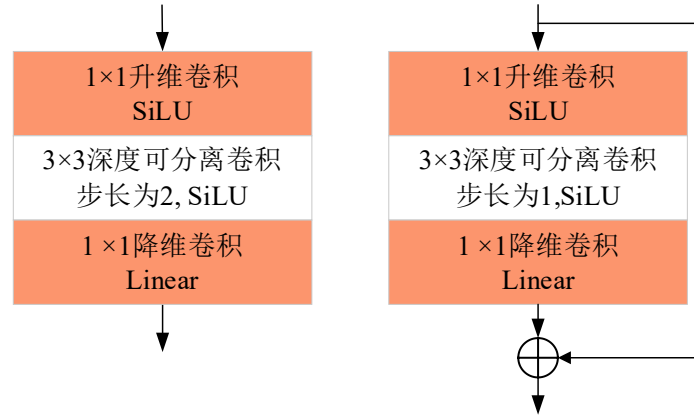


图 3-3 MobileNetV2 模块结构图

MobileNetV2 模块采用的是倒残差结构。首先，使用 1×1 卷积操作提升特征通道维度，后接 SiLU 激活函数，以减少后续卷积操作造成特征信息丢失；然后，使用 3×3 深度可分离卷积操作，后接 SiLU 激活函数，以实现特征提取的同时减少网络模型计算量；最后，使用 1×1 卷积操作降低特征图通道维度，后接 Linear bottleneck 层，输出所需通道数的特征图同时也可以减少特征提取过程中信息丢失。

MobileViT 模块是由局部表达、全局表达和特征融合三个部分组成。MobileViT-V3 结构如图 3-4 所示。

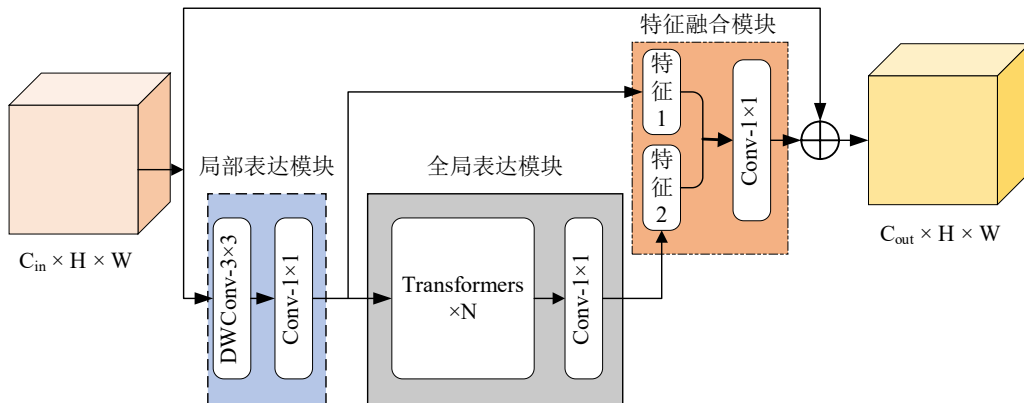


图 3-4 MobileViT 模块结构图

在 MobileViT 模块中，首先使用 3×3 深度可分离卷积操作，在降低计算量的情况下实现局部特征建模，配合 1×1 卷积操作将特征图调整为合适的通道数；

其次使用 MobileViT 特有的 Unfold→Transformer→Fold 结构进行全局特征建模，配合 1×1 卷积操作调整为与局部输出特征图一样的通道数；然后沿通道方向将局部特征图和全局特征图拼接，配合 1×1 卷积操作输出和初始输入相同通道数的特征图；最后使用跳跃连接将初始输入和融合模块的输出特征图相加作为输入。自此实现了输入、局部特征和全局特征的融合。其中 MobileViT 特有的 Unfold→Transformer→Fold 结构是区别于其他 Transformer 的部分，特征图在展开为小的特征 patches 后，MobileViT 的自注意力机制只选择相同位置的像素点进行点积操作，以此避免特征冗余，这样操作大大降低了模型的数量和计算量，这也是模型轻量化的原因。具体 Unfold 和 Fold 操作如图 3-5 所示。

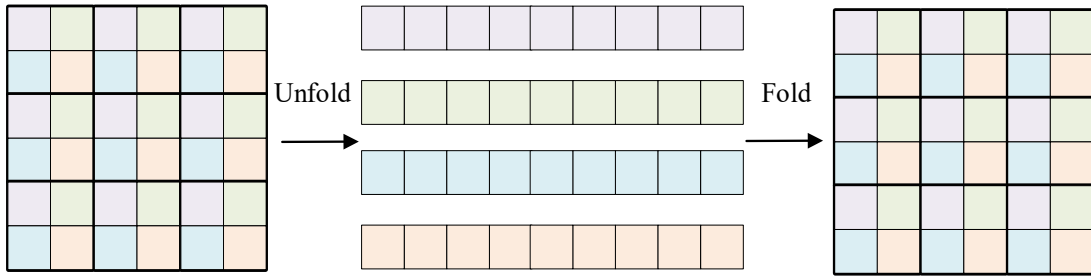


图 3-5 Unfold 和 Fold 操作示例图

3.1.3 注意力模块

为提高轻量型网络特征提取性能引入 CA 注意力模块^[52]，不仅考虑了通道之间的相互依赖关系，还考虑了空间相关的位置信息，并且该模块有很强的灵活性和轻量化，可以很简单的插入到网络模型中，是一款即插即用的注意力模块。CA 注意力机制结构如图 3-6 中(a)图所示。

首先，当输入特征尺寸为 $\mathbb{R}^{C \times H \times W}$ 时，分别对 H 方向和 W 方向进行平均池化，用以捕获特征的空间位置信息和各个特征之间的空间关系，同时还可以避免将空间信息全部压缩到通道中，获得尺寸为 $\mathbb{R}^{C \times H \times 1}$ 和 $\mathbb{R}^{C \times 1 \times W}$ 的特征图，如图 3-6 中图(b)是一个简单池化计算示例；然后，将经过 W 池化后特征图进行矩阵变换获得尺寸为 $\mathbb{R}^{C \times 1 \times H}$ 的特征图，因为 H 和 W 相等，直接将矩阵变换后的特征和 W 池化特征图拼接获得尺寸为 $\mathbb{R}^{C \times 1 \times (H+W)}$ 的特征图，再使用大小为 1×1 的卷积操作降低通道维度获得尺寸为 $\mathbb{R}^{C' \times 1 \times (H+W)}$ 的特征图；最后，沿着空间维度进行分割操作恢复到拼接前的尺寸，再使用 Sigmoid 激活函数得到最后的注意力权重，将输入和注意力权重相乘就得到 CA 注意力机制的输出。

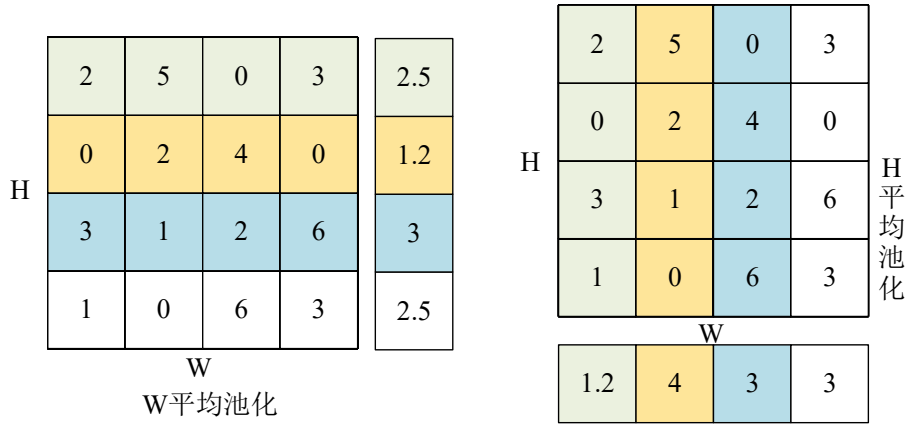
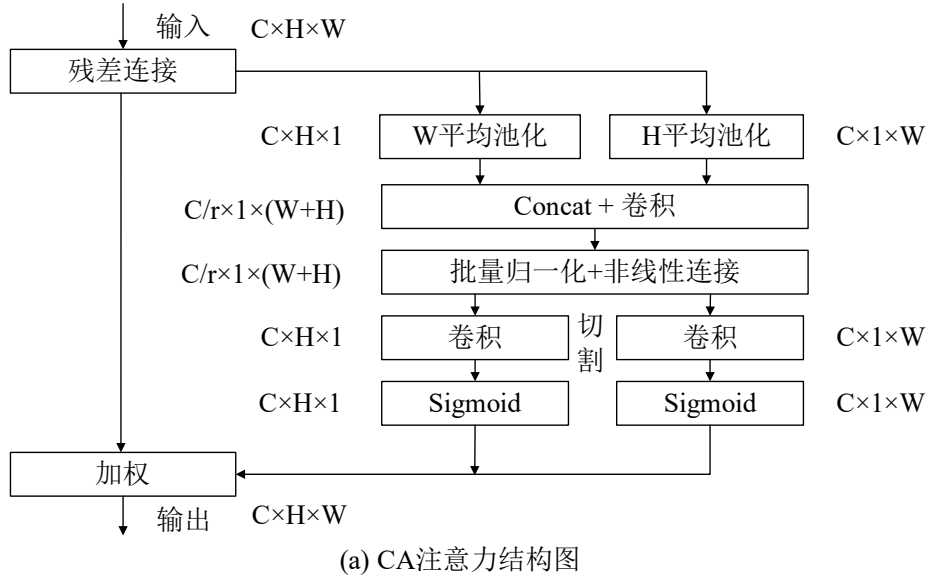


图 3-6 CA 注意力机制结构图和池化计算示例图

3.2 实验结果与分析

3.2.1 人脸微表情数据集

本章实验使用的是 CASME^[53]和 CASME II^[54]两个自发式微表情数据集，数据集简介如表 3-1 所示。数据集是由中国科学院提供，拍摄地点是完全控制的实验室环境。实验室使用 4 个补光灯进行照明，保证良好光照以减少产生面部阴影，保证获取高质量微表情视频。数据集采集期间给受试者观看高情感视频，以此引发受试者的面部动作，期间受试者需尽量维持自然表情。后期会交由微表情研究专家从大量面部运动视频中挑选出包含微表情的视频段，筛选出微表情并标注运动单元和视频中的重要帧的位置。

表 3-1 实验使用的公开数据集简介表

名称	类别数目	样本个数	受试者人数	图像分辨率	视频帧率/fps
CASME	8	195	35	640×480	60
CASME II	7	256	26	640×480	200

表中 CASME 数据集在 2013 年发布。视频采集时共招募 35 名参与者，平均年龄为 22.03 岁。从拍摄的 1500 个面部动作中，经过专业人员编码分析选出 195 个微表情样本，并将样本分为 8 类，类别名称分别是愉悦、厌恶、伤心、蔑视、害怕、紧张、惊讶、压抑。CASME II 数据集在 2014 年发布。视频采集时共招募 22 名参与者，最终筛选出 255 个样本，被分为 7 类，类别名称分别为其他、高兴、压抑、伤心、害怕、惊讶、厌恶。数据集中部分样本顶点帧图像如图 3-7 所示。



图 3-7 CASME 和 CASME II 数据集样本

3.2.2 实验设置和评价指标

实验会先进行如下操作：对微表情数据集样本中起始帧和顶点帧使用第 2 章中 MTCNN 实现尺度变换获得大小为 224×224 的正面人脸图像，再求取两者之间 TVL1 光流和光流应变堆叠成 3 通道融合光流特征作为网络输入。

实验所使用硬件配置为 64 位 Windows 11 系统，16GB 内存的 Intel Core i7-12700H CPU，6GB 显存的 NVIDIA GEFORCE RTX 3060 GPU，11.2 版本的 CUDA。网络训练参数设置如下：使用 Pycharm 软件 PyTorch 架构搭建网络模型，实验采用 Adam 优化器，损失函数为 Focal Loss 损失函数，每次批处理大小为 8，迭代次数为 100，学习率为 0.0001。

实验使用微表情识别最常用的留一受试者(Leave-One-Subject-Out, LOSO)交

叉验证法，具体操作是每次将一位受试者的样本集作为测试集，其他受试者的样本集作为训练集，训练结束后更换另一位受试者的作为样本集作为测试集，重复实验直到所有受试者样本都做过测试集^[55]。LOSO 验证方法的优势是可在数据较少的情况下充分利用数据，对不平衡数据集更具鲁棒性，更准确地评估模型的泛化能力，适用于小样本数据集。

为评估网络性能，引入混淆矩阵来直观的展示实验对各个类别分类结果，更好地分析算法对于各个类别微表情的识别准确度，混淆矩阵结构如图 3-8 所示，图中参数 TP 是真阳性表示实际值为正，预测结果也为正； TN 是真阴性表示实际值为负，预测结果也为负； FP 是假阳性表示实际值为负，预测结果为正； FN 是假阴性表示实际值为正，预测结果为负。

预测值 真实值	正类(P)	负类(N)
正类(T)	TP	FN
负类(F)	FP	TN

图 3-8 混淆矩阵结构图

实验采用 ACC 作为识别算法的评价指标，其数值是模型分类正确样本占总样本的比例，所以准确率数值越高表明模型效果越好。考虑到数据集样本类别不均衡，同时还选用更加公平的未加权 F1 分数(Unweighted F1-score, UF1)作为评价指标，最大取值为 1，最小取值为 0，其数值越大意味着模型效果越好。实验中使用 sklearn.metrics 库中的 f1_score 函数求取 UF1。两个评价指标计算公式如下：

$$ACC = \sum_{i=1}^k \sum_{j=1}^C \frac{TP_i^j + FP_i^j}{TP_i^j + TN_i^j + FP_i^j + FN_i^j} \quad (3-4)$$

$$\begin{cases} F1_j = \sum_{i=1}^k \frac{2TP_i^j}{2TP_i^j + FP_i^j + FN_i^j} \\ UF1 = \frac{1}{C} \sum_{j=1}^C F1_j \end{cases} \quad (3-5)$$

其中， C 表示类别数， k 表示被试者人数， j 表示第 j 类样本， i 表示第 i 受试者

样本作测试集， TP 、 TN 、 FP 、 FN 等参数定义同混淆矩阵参数定义。

3.2.3 消融实验

1) 验证 Focal Loss 函数的有效性

为验证本章算法中 Focal Loss 损失函数的有效性，使用控制变量法进行两组实验。两组实验只有使用的损失函数不同，其他硬件和参数设置与 3.2.2 节的相同。通过两组实验的对比，比较 Focal Loss 损失函数和交叉熵损失函数两者对本章算法的影响，实验结果如表 3-3 所示。

表 3-2 使用 Focal Loss 和交叉熵损失函数的结果对比表

损失函数类型	CASME		CASME II	
	ACC (%)	F1 得分	ACC (%)	F1 得分
交叉熵	75.75	0.7571	78.05	0.7442
Focal Loss	76.36	0.7710	78.86	0.7633

根据表 3-3 中结果可知，使用交叉熵损失函数的实验结果没有使用 Focal Loss 函数作为损失函数的实验结果好。使用后者作为损失函数在 CASME 数据集上 ACC 和 F1 得分分别提高 0.61%和 0.0139；在 CASME II 数据集上 ACC 和 F1 得分分别提高 0.81%和 0.0191。证明了 Focal Loss 损失函数可以在微表情识别的任务中减少微表情数据集分布不均衡所带来的负面影响。

2) 验证引入注意力机制的有效性

为验证本章算法引入 CA 注意力机制的有效性，使用 LOSO 交叉验证法在 CASME 和 CASME II 两个数据集上进行实验，实验将融合光流作为网络模型输入，并使用 Focal Loss 损失函数训练模型。最终不同网络结构的识别结果如表 3-4 所示。

表 3-3 改进网络前后的识别结果对比表

网络结构	CASME		CASME II	
	ACC (%)	F1 得分	ACC (%)	F1 得分
MobileViT-V3	75.58	0.7572	77.23	0.7463
本章模型	76.36	0.7710	78.86	0.7633

根据表 3-4 的结果可以看出，在加入 CA 注意力模块后网络性能相较于原 MobileViT-V3 网络有所提高。在 CASME 数据集上 ACC 提高 0.78%，F1 得分提高了 0.0138；在 CASME II 数据集上 ACC 提高了 1.63%，F1 得分提高 0.017。

验证了引入 CA 注意力模块对提高网络性能的有效性。

3.2.4 对比实验

为更加直观地分析网络对各个类别的识别结果，将研究的 CA-MobileViT 人脸微表情识别算法分别在 CASME 和 CASME II 两个数据集上进行试验并绘制分类混淆矩阵。在 CASME 数据集上使用样本数较多的厌恶、紧张、压抑、惊讶四个类别进行四分类实验；在 CASME II 数据集上使用样本数较多的高兴、压抑、厌恶、其他、惊讶五个类别进行五分类实验。

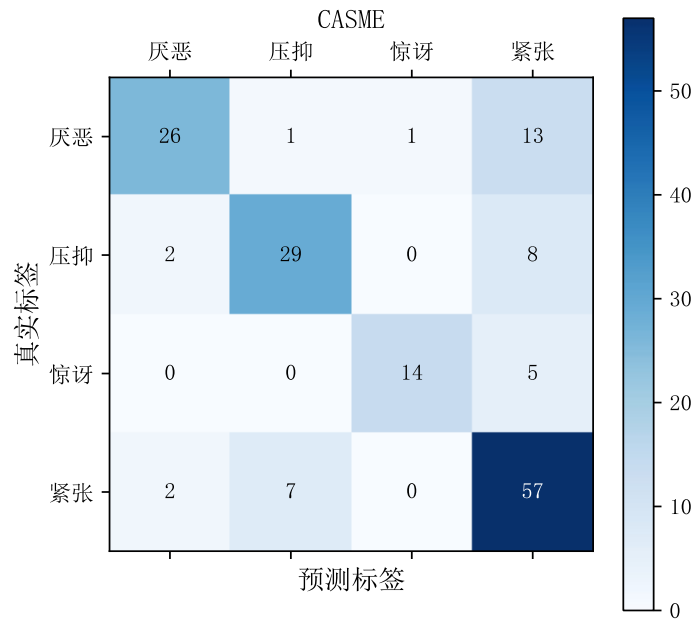


图 3-9 在 CASME 数据集上的混淆矩阵

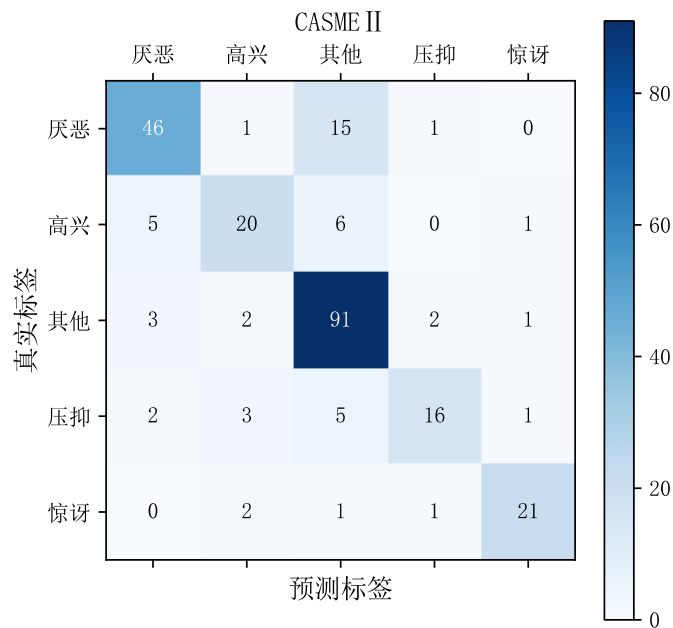


图 3-10 在 CASME II 数据集上的混淆矩阵

为验证本章研究算法的有效性，与多种识别算法在数据集 CASME 和 CASME II 上的识别结果进行对比。在两个数据集上不同算法的实验分类结果如表 3-5 和表 3-6 所示。表中算法均采用 F1 分数和 ACC 作为微表情识别性能优劣的评价指标，表中 LBP-SIP^[27]、LCBP^[28]是基于 LBP 变种的手工特征识别算法，MDMO^[32]和 Sparse MDMO^[56]算法是基于光流的手工特征识别算法；3D-FCNN^[57]，TSCNN-II^[58]，Dual-ResNet^[60]是基于卷积神经网络的深度学习方法，LCBP-STGCN^[28]，Graph-TCN^[61]，AU-GCN^[62]是使用图卷积神经网络的深度学习方法。

表 3-4 不同方法在 CASME 数据集的结果对比表

方法	ACC (%)	F1 得分
LBP-SIP ^[27]	43.56	0.3300
LCBP ^[28]	76.39	0.6603
MDMO ^[32]	56.29	0.5551
Sparse MDMO ^[56]	74.83	0.7490
3D-FCNN ^[57]	54.44	--
TSCNN-II ^[58]	73.88	0.7270
本章算法	76.36	0.7710

表 3-5 不同方法在 CASME II 数据库的结果对比表

方法	ACC (%)	F1 得分
LCBP ^[28]	70.10	0.7004
Sparse MDMO ^[56]	66.95	0.6911
Dual-ResNet ^[60]	69.19	0.6624
LCBP-STGCN ^[28]	76.14	0.7535
Graph-TCN ^[61]	73.98	0.7201
AU-GCN ^[62]	74.26	0.7003
本章算法	78.86	0.7633

在表 3-5 和表 3-6 中将表现最优异的识别结果加粗显示。根据上述两个表所展示的实验结果，本章算法在两个数据集中都取得了较好的识别结果。在 CASME 数据集上的识别准确率为 76.36%，略低于 LCBP 算法的 76.39%，F1 得分是所比算法中的最优值；在 CASME II 数据集上的识别准确率和 F1 得分为

78.86%和 0.7633 是所比算法中最优的。在 CASME 中表现较差的原因是 CASME 数据集人脸较小，求取融合光流会引入较多噪声。

3.3 本章小结

本章结合 MobileViT-V3 和 CA 注意力机制研究了一种 CA-MobileViT 网络模型。在使用 MobileViT-V3 网络基础上，引入 CA 注意力机制加强对感兴趣区域特征的关注度。通过消融实验，证实了引入的 CA 可以提高 MobileViT-V3 网络对微表情识别的 ACC 和 F1 得分；也证实了 Focal Loss 损失函数可以减少微表情数据集分布不均衡所带来的负面影响。又和其他算法在 CASME 和 CASME II 两个数据集上的实验结果作对比，验证了 CA-MobileViT 网络模型对微表情识别的有效性。

第4章 融合残差网络和注意力机制的人脸微表情识别

针对以融合光流特征为网络输入，轻量型 CA-MobileViT 特征提取网络对微表情识别准确率并不是特别理想的问题。本章研究了一种融合残差网络与两种注意力机制的 GAM-Res-DA 微表情识别算法，以此进一步提高基于深度学习的微表情识别准确率。首先，考虑到本章的微表情识别算法属于小样本图像分类的问题，采用浅层残差神经网络以此避免出现过拟合现象；然后，引入 GAM 注意力机制和 DA 注意力机制改进残差神经网络结构，利用 GAM 注意力机制在减少信息弥散的情况下放大各维度的交互作用，利用 DA 注意力机制来捕捉空间和通道两个维度上特征的全局依赖关系，以此提取更多融合光流的关键运动信息；最后，在 CASME 和 CASME II 数据集上进行实验并和其他经典方法对比，结果证实了该方法的有效性和可行性，并在自建数据集上进行实验检验模型微表情识别性能。

4.1 GAM-Res-DA 特征提取网络结构设计

4.1.1 整体网络模型

GAM-Res-DA 网络模型引入现有的 GAM 注意力模块和 DA 注意力模块改进残差模块。残差网络的快捷连接可以为网络补充浅层特征信息，有效避免梯度爆炸和梯度消失的发生。引入的 GAM 注意力模块可以帮助模型更好地获取图像全维度上的特征信息，DA 注意力模块可以加强各个特征间的联系，帮助模型更有效地捕捉重要信息，从而提高网络模型的性能。并且为了更好更快地训练出可靠的微表情识别网络，设计的特征提取网络为浅层的神经网络。因此网络模型没有繁重的计算负担，可以很快的达到收敛，有效地学习微表情的面部特征表示，有助于提高网络模型对面部特征的建模能力，从而提高网络模型对微表情识别的准确率。

网络模型结构如图 4-1 所示。首先，特征提取网络前段采用和残差网络相同的输入和处理方式，初始输入 $224 \times 224 \times 3$ 大小的融合光流图像，对其使用 7×7 卷积和最大池化进行 2 倍下采样操作和初步的特征提取，以减小图像尺寸和提取低级特征，获得 112×112 大小的特征图；其次，特征提取网络中段堆叠 4 个融

合了 GAM 模块和 DA 模块的残差模块进一步特征提取，每个残差块包含一个带有跳跃链接的卷积层，在每个卷积层后会进行批量归一化和激活函数 ReLU，获得 7×7 大小的特征图；然后，特征提取网络尾段经过 2×2 的平均池化层对特征图进行降维处理，将每个特征图转化为一个固定长度的特征向量；最后，使用全连接层对提取到的特征进行展平，将展平后的特征通过 Softmax 函数生成最终的类别预测结果。

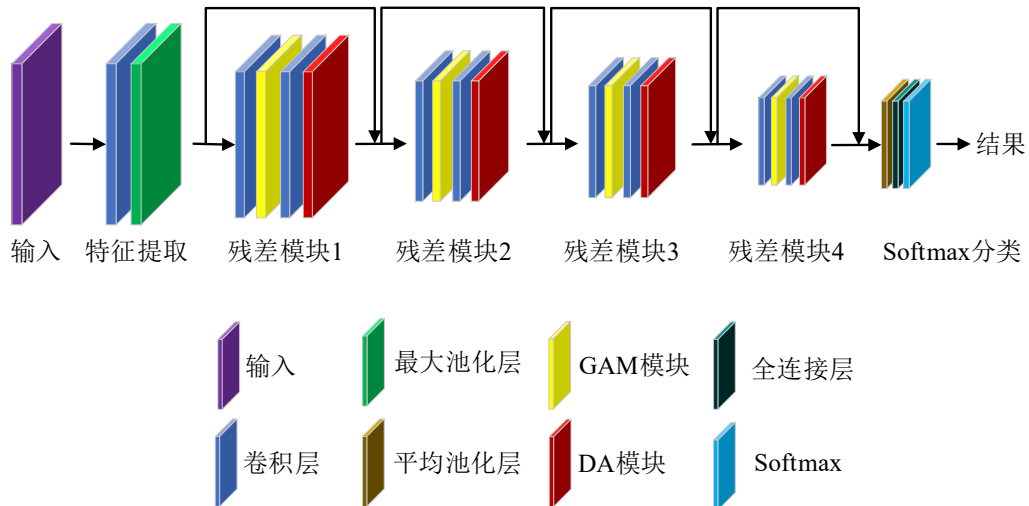


图 4-1 GAM-Res-DA 模型结构图

4.1.2 残差网络模块

本章选择残差网络模型进行改进并应用于微表情识别任务的主要原因是现有的微表情数据集中样本较少，难以训练出强壮的网络。残差网络让原始特征可以直接从一层传递到另一层，避免了信息逐层丢失，有助于保留更多有用信息，有助于梯度的快速传播，减少训练复杂度，使得网络更加容易收敛，训练出更强壮的网络。

在 2015 年，He 等^[63]针对较深的神经网络难以训练的问题，提出了残差学习框架来简化深度神经网络的训练，残差结构有快捷连接和恒等映射两个优秀设计。快捷连接跳过中间网络直接将输入添加到输出中，从而使网络可以学习关于输入和输出两者的权重关系，避免训练过程中出现梯度消失和梯度爆炸的问题；普通的网络随着网络层数的增加，会出现训练误差先减少后增多即网络退化的问题，残差网络让网络在下层误差变化不大时，将学习目标转变为学习残差函数，也就是使输入近似于输出，以保持在今后的层次中不会造成严重的

精度下降，让网络在增加层数的同时还保持良好性能。因此，残差网络在各种计算机视觉任务中表现出色，并被广泛应用于深度学习领域，并对后面如何建立深层网络产生深远影响。

残差网络在求取层的映射表达过程中，让网络学习残差函数取代学习未知函数。其将多个非线性层的映射看作一个复杂的函数计算，如果网络的输入为 x ，设网络求取的底层映射为 $H(x)$ ，那么网络堆叠层的非线性映射为 $F(x)=H(x)-x$ ，这样就可以得到底层映射新表达式的 $H(x)=F(x)+x$ 。这个新的映射表达式可以通过具有快捷链接的前馈神经网络来实现，残差学习和残差块结构如图 4-2 所示。

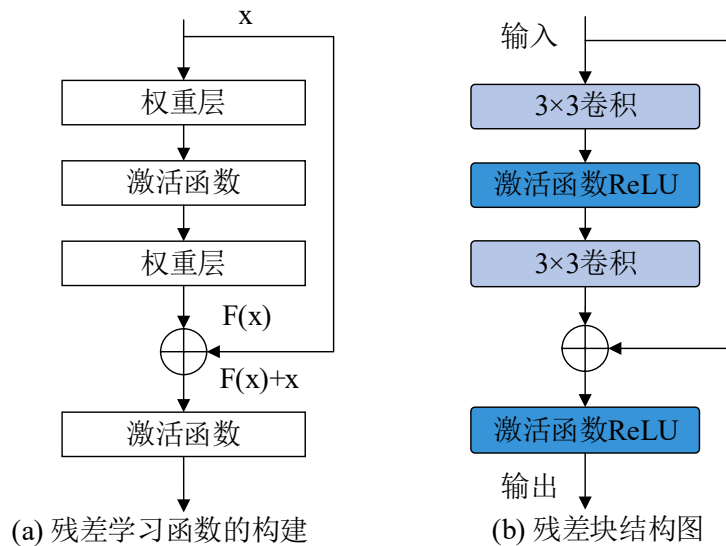


图 4-2 残差学习构建模块图

残差块的输出由两部分构成。一部分输出是将原输入经过第一个权重层和 ReLU 激活函数，再经过第二个权重层运算获得；另一部分输出是将原输入通过快捷连接方式获得，且快捷连接方式既不增加额外的参数也不增加计算复杂度。残差块会利用原始输入作为输出元素之一，以此提高网络的表达能力和训练效率。残差网络模型是通过一系列残差块串联而成，常见的残差网络有 ResNet18, ResNet34, ResNet50……，通过不断增加残差块使网络更具表现力。本章作为研究的网络是由四个残差块串联组成。

4.1.3 注意力模块

1) GAM 模块

现有的很多注意力机制都忽视了通道和空间两个维度之间信息交互的重要

性。GAM^[64]是同时考虑空间维度和通道维度的注意力机制，先计算通道注意力，保留三个维度上的信息同时放大通道-空间依赖性；再计算空间注意力，实现空间信息融合。通过通道、空间高度与空间宽度之间的注意力权重来确定各个特征的重要程度。使用 GAM 可以让网络在减少面部运动信息弥散的同时放大不同光流特征维度间的相互作用。GAM 模块输出特征计算公式如下：

$$F_2 = M_c(F_1) \otimes F_1 \quad (4-1)$$

$$F_3 = M_s(F_2) \otimes F_2 \quad (4-2)$$

其中， F_1 表示给定输入特征图， F_2 表示中间特征图， F_3 表示输出特征图， M_c 和 M_s 分别表示通道注意力和空间注意力特征图， $\otimes$ 表示按元素进行乘法操作。

GAM 模块结构如图 4-3 所示。通道注意力子模块可以提取融合光流特征三个维度上的信息，过程中使用三维排列来保留跨三维的信息，之后又使用一个多层感知器对通道和维度的相关性进行放大提取，再经过 Sigmoid 函数生成通道注意特征图，关注重要的跨维度信息。在空间注意力子模块中，由于最大池化操作会减少特征信息的利用，因此使用两个 7×7 卷积层进行空间特征融合从而保留更多信息的特征映射。空间注意子模块为避免显著增加参数量，使用带有 Channel Shuffle 的分组卷积，可以减少计算量和计算参数，提高模型计算效率，再经过 Sigmoid 函数生成空间注意特征图，提高感兴趣区域的权重，获取更多面部关键信息。

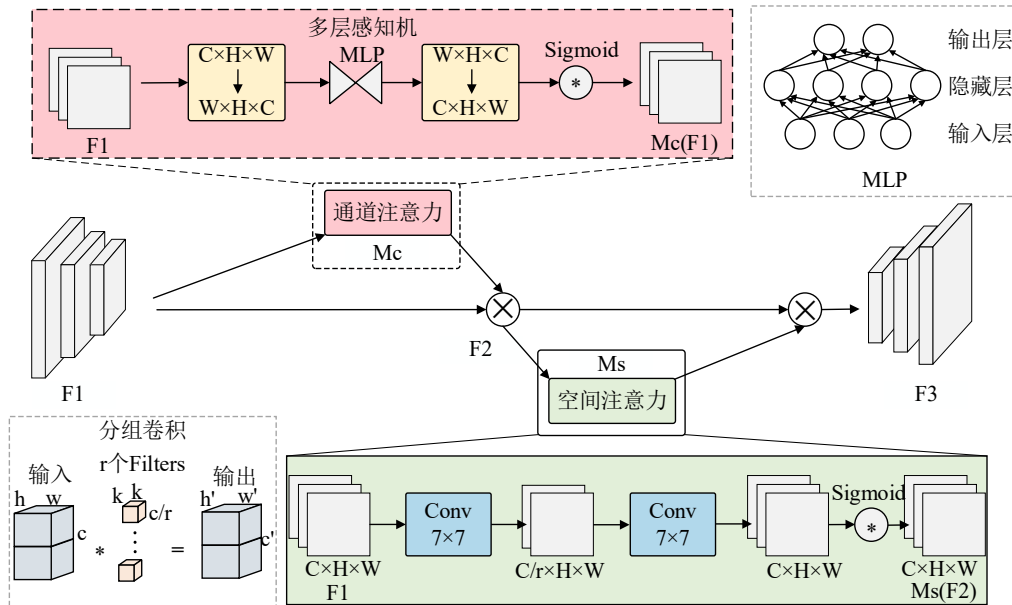


图 4-3 GAM 模块结构图

2) DA 模块

由于卷积操作存在局部感受野的特点，无法准确提取微表情重要区域间的联系，进而影响微表情识别准确率。为捕获面部特征中任意两个位置的空间依赖关系，在特征提取网络的残差块中引入 DA 注意力^[65]。DA 模块是由并行的位置注意力机制和通道注意力机制两条支路组成，其中位置注意力模块通过所有位置特征的加权选择性地聚合每个位置的特征，不管各个特征之间的距离如何将相似的特征关联起来；另外通道注意力模块通过整合所有通道映射之间的关联特征，选择性地突出相互依赖特征的表示。之后对两个分支的输出进行求和，自适应地整合局部特征及其全局依赖关系进一步强化特征表示。DA 模块结构如图 4-4 所示。

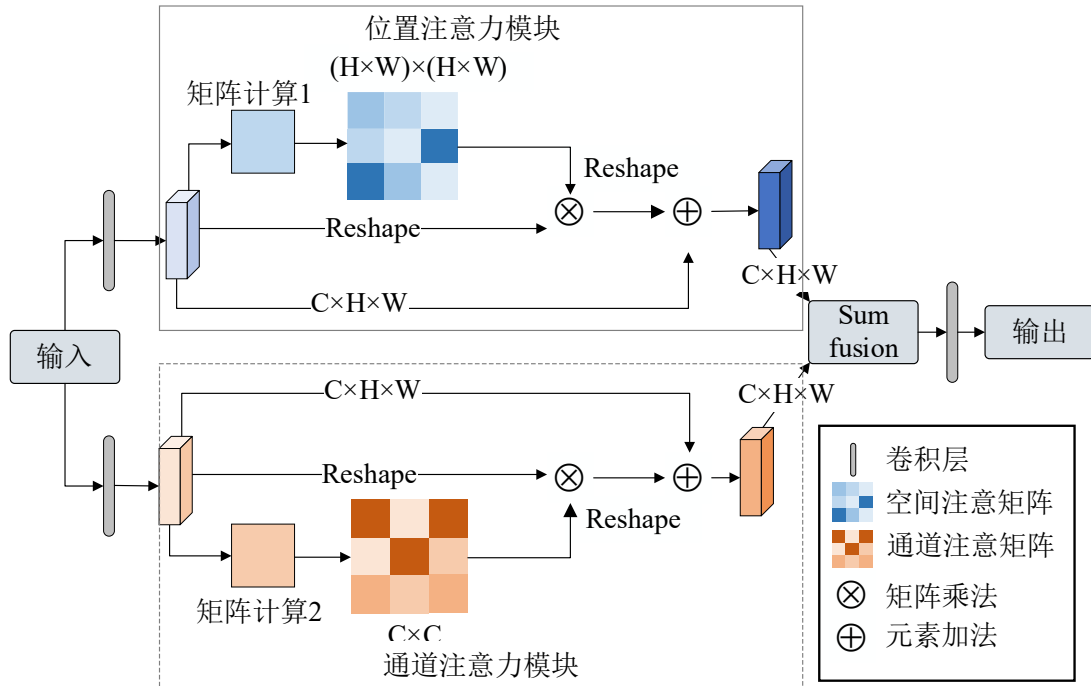


图 4-4 DA 模块结构图

DA 模块拥有位置注意力和位置注意力两条支路，其中位置注意力模块使用卷积操作来降低特征的维度。然后将特征输入到位置注意模块，通过以下三个步骤生成新的特征：第一步是生成空间注意矩阵，该矩阵可以获得同一个通道中任意两个像素之间的关系；第二步是将 Reshape 操作后的原始特征矩阵和空间注意矩阵进行矩阵乘法；第三步，对矩阵乘法的结果使用 Reshape 操作与原始

始特征矩阵进行元素求和操作，以获得位置加权后的特征矩阵；另一路通道注意力与位置注意力的计算流程相似，区别在于第一步是在通道维度上计算通道注意矩阵，获得通道之间的相互依赖关系。最后，将两个注意力分支的输出聚合起来，获得更好的特征表示。

4.2 实验结果与分析

4.2.1 实验设置

在 CASME 和 CASME II 数据集上使用 LOSO 交叉验证法进行实验，数据集的基本信息如 3.2.1 节。网络训练时使用 Focal Loss 损失函数代替交叉熵损失函数，其他参数设置与 3.2.2 节一致。实验使用 ACC 和 F1 得分作为微表情识别算法性能优劣的评价指标，绘制在 CASME 和 CASME II 数据集上的混淆矩阵来分析对各类别的分类结果，求取类激活图观察不同面部区域对微表情分类的贡献度，并在自建数据集上测试本章算法的性能。

4.2.2 消融实验

为验证两个注意力机制对微表情识别的影响，使用不同网络结构在 CASME II 数据集上进行四组消融实验：只使用初始的残差网络；使用 DA 模块不使用 GAM 模块的残差网络；使用 GAM 模块不使用 DA 模块的残差网络；本章算法。具体实验结果如表 4-2 所示。

表 4-1 不同网络结构实验结果对比表

网络模型	ACC (%)	F1
ResNet10	79.77	0.7290
ResNet10+DA 模块	80.42	0.7324
ResNet10+GAM 模块	82.44	0.7758
本章算法	83.33	0.8115

根据表中结果可知两种注意力机制的都有效提升了网络性能，且同时引入 GAM 和 DA 模块的网络性能最好，微表情识别准确率达到了 83.33%，F1 达到了 0.8115，对比 ResNet10 分别提高了 3.56%和 0.0825，实验证实引入的两个注意力机制可以提高 ResNet10 对微表情识别准确率。为更加直观的观测分类的依

据，使用类激活图可视化面部区域，观察对微表情识别贡献程度高的部位，结果如图 4-5 所示。

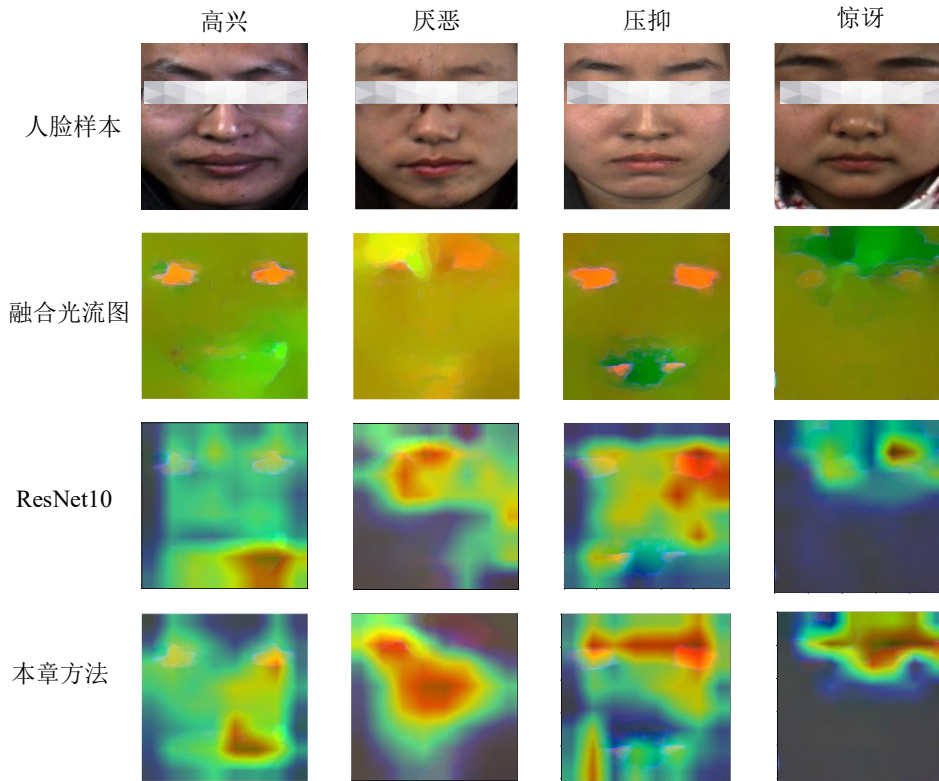


图 4-5 微表情识别的类激活图

从图 4-5 可以看出对于高兴类别的微表情，嘴角和脸颊部分对识别的贡献度最高，符合该类微表情的运动单元 AU6 脸颊抬起和 AU20 嘴唇向后拉扯组合；对于厌恶类别的微表情，眉毛和鼻子部位贡献最大，符合该类微表情的运动单元 AU4 眉毛下压和 AU9 皱鼻组合；对于压抑类别的微表情，眉毛和嘴角贡献最大，符合该类微表情的运动单元 AU4 眉毛下压和 AU14 嘴角下压组合；对于惊讶类别的微表情，眉毛贡献度最大，符合该类微表情的运动单元 AU1 内侧眉毛抬起和 AU2 眉毛外侧抬起组合。

4.2.3 对比实验

将研究的融合残差网络和注意力机制人脸微表情识别算法分别在 CASME 和 CASME II 数据集上进行试验。在 CASME 数据集中使用厌恶、紧张、压抑、惊讶四个样本数较多的类别进行四分类实验；在 CASME II 数据集中使用高兴、压抑、厌恶、其他、惊讶五个样本数较多的类别进行五分类实验。在两个数据集识别结果的混淆矩阵如图 4-6 和图 4-7 所示。

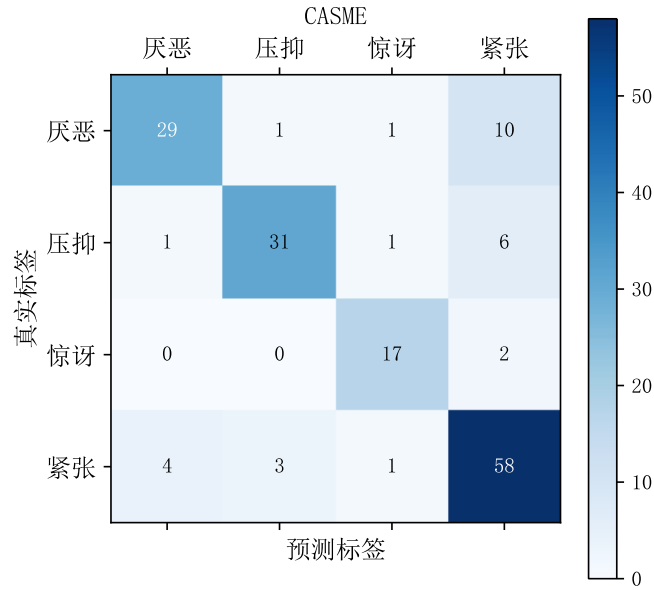


图 4-6 在 CASME 数据集上的混淆矩阵

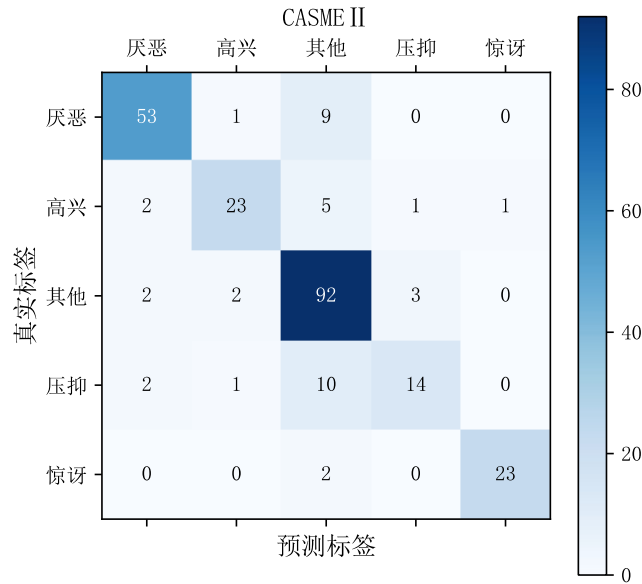


图 4-7 在 CASME II 数据集上的混淆矩阵

从图 4-6 和图 4-7 可以看出本章算法在 CASME 数据集的实验中容易将厌恶、压抑、惊讶类别的微表情误判为紧张；在 CASME II 数据集上容易将压抑、高兴、厌恶三类的部分样本错判为其他。分析原因是因为 CASME 数据集中大部分样本为紧张使训练出的模型性能不平衡，忽略了其他类别的样本，导致发生类别判定错误。

此外为验证本章所研究的微表情识别方法的有效性。用 ACC 和 F1 作为微表情识别算法的评价指标，将本章算法和其他的微表情识别方法在 CASME 和 CASME II 两个数据集上进行性能比较。不同算法在两个数据集上的识别结果如

表 4-3 和表 4-4 所示，并将数据集中取得最高评价指标的结果加粗显示。

表 4-2 不同方法在 CASME 数据集识别性能对比表

算法类别	方法	ACC (%)	F1 得分
基于手工特征	LBP-TOP ^[26]	42.83	0.3706
	LBP-SIP ^[27]	43.56	0.3300
	MDMO ^[32]	56.29	0.5551
	LCBP ^[28]	76.39	0.6603
基于深度学习	TSCNN-II ^[58]	73.88	0.7270
	LCBP-STGCN ^[28]	81.26	0.7701
	CA-MobileViT	76.36	0.7710
	本章方法	81.81	0.8249

表 4-3 不同方法在 CASME II 数据集识别性能对比表

算法类别	方法	ACC (%)	F1 得分
基于手工特征	LBP-TOP ^[26]	40.49	0.3542
	LBP-SIP ^[27]	42.07	0.4501
	Bi-WOOF ^[31]	57.89	0.6100
	LCBP ^[28]	70.10	0.7000
基于深度学习	SE-ResNet50 ^[59]	59.26	0.5853
	Dual-ResNet ^[60]	69.19	0.6624
	LCBP-STGCN ^[28]	76.14	0.7535
	TSCNN-II ^[58]	80.97	0.8070
	CA-MobileViT	78.86	0.7633
	本章方法	83.33	0.8115

从表 4-3 和表 4-4 可以看出基于深度学习的微表情识别算法通常比基于手工特征的微表情识别算法更加具有优势，本章研究的算法在 CASME 数据集和 CASME II 数据集上的准确率和 F1 评分，分别取得 81.81%、0.8249 和 83.33%和 0.8115，在两个数据集上都表现出了相较于其他方法更加优秀的识别性能。实验结果证明了融合残差网络和注意力机制的微表情识别算法的有效性。

4.2.4 自建数据集实验

考虑到后续会设计微表情识别系统，为提高系统对微表情的识别准确率，

仿照公开数据集的采集流程，使用华为 mate50 手机相机建立了作者的微表情数据集，由于无法完美的区分微表情种类，参考常用的微表情三分类实验中的类别，将所采集的微表情分为积极(高兴)，消极(压抑、厌恶、伤心、害怕…)和惊讶三个类别。自建数据集的微表情样本类别与数量如表 4-5 所示，部分微表情样本的顶点帧示例如图 4-8 所示。

表 4-4 自建数据集信息表

类别	自建数据集
积极	167
消极	174
惊讶	160



图 4-8 部分数据集样本示例图

将自建数据集按照 8:2 的比例来划分训练集与测试集，考虑到样本较少使用旋转 5 度，10 度，15 度以及水平翻转进行数据增强，以此训练融合残差网络和注意力机制的网络模型。训练硬件配置如 3.2.2 小节，训练参数设置如下：训练轮次为 100 epochs，损失函数为 Focal loss 损失函数，网络优化器为 Adam，学习率设置为 0.00002，weight_decay 设置为 1，amsgrad 设置为 True。训练时损失函数收敛曲线、测试集准确率曲线如图 4-9 所示。

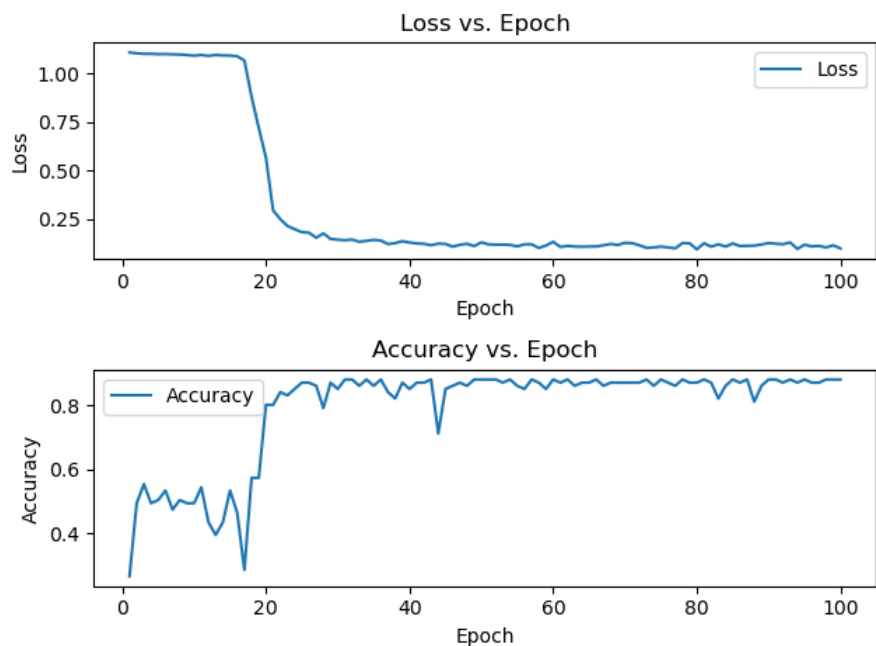


图 4-9 训练时损失函数收敛曲线和测试集准确率曲线图

从图 4-9 可以看出训练损失函数在 20 epoch 之后接近收敛，测试集的准确率还会出现波动最终准确率收敛于 88.12%附近，取得了较好的识别结果。为更加了解网络模型对自建数据集各个类别的分类结果绘制了混淆矩阵，结果如图 4-10 所示，可以看出对三种类别都表现出了良好的识别准确率。

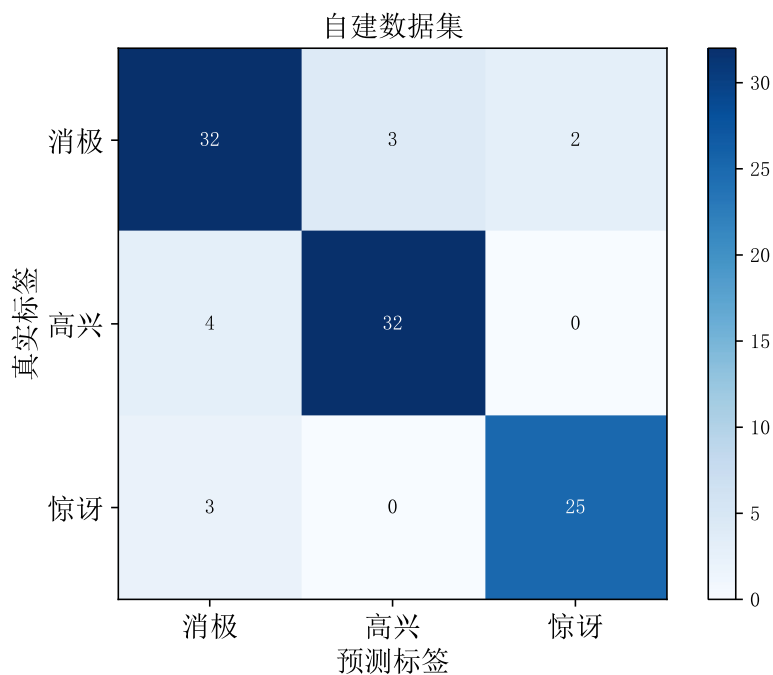


图 4-10 在自建数据集上的混淆矩阵

4.3 本章小结

本章研究了一种融合残差网络和注意力机制的人脸微表情识别算法。首先，阐述了残差网络、GAM 和 DA 注意力机制的结构，并结合三者的优势构建了改进后的残差网络；其次，使用 CASME II 数据集进行消融实验验证了改进后残差网络中 GAM 和 DA 模块对网络性能的影响；然后，对本章算法与其他算法进行了对比分析，画出混淆矩阵直观展现算法对各类别的分类结果，根据与多个模型的对比结果，发现本章方法在 CASME 和 CASME II 数据集上的微表情识别性能相当有竞争力；最后，在自建数据集上进行实验，取得较好的识别准确率，为后续识别系统研究提供技术支持。

第 5 章 人脸微表情识别系统设计

随着情感识别技术的发展，微表情作为一种重要的非言语行为信号，对于理解个体内心状态和意图具有重要意义。然而，由于微表情的瞬时性和微小幅度特性，其识别一直是一个具有挑战性的问题。因此，针对微表情的准确识别与分析成为当前人机交互、虚拟现实、安全监控等领域的热点研究之一。由于建立微表情数据集时需要耗费大量的时间和相对专业的知识，公开数据集样本较少，导致现在微表情识别的准确率不是特别的高。现有的微表情识别系统都是一些学术论文中实验性的产品，作为辅助判断的工具，比如研究生情感分析系统、基于微表情的线上课堂学生注意力分析系统和微表情预测系统等。本章基于前面几章的研究基础，尝试运用微表情识别系统识别用户微表情，提供当前用户的情绪置信度，为分析用户情感需求提供参考。

5.1 系统需求分析

近年来，随着科技和医疗技术的进步，中国的人口老龄化成为主要社会现象。伴随人口老龄化出现了小部分失能弱智的老人，他们无法通过语言表达自己的想法，需要专业的护理人员帮助实现需求，但现在社会并没有足够从事康复护理的人员，这是一个巨大的社会痛点问题。针对这一问题，有些科技公司已经开发出专门的康复护理智能机器人，产品通过如生命体征测量系统和微表情识别系统实现对目标群体的需求分析，再使用多自由度的机器人来满足目标群体的需求，提高目标人群护理质量，同时也可以在一定程度上减轻护理人员的劳心程度。本章准备设计一款基于深度学习的人脸微表情识别系统，由于前面几章研究的数据预处理技术和算法模型都是基于视频中关键帧的微表情识别，所以设计的微表情识别系统需要有视频采集；数据预处理；微表情识别；数据传输等功能。

5.2 系统功能设计

因为系统会采集人脸信息，为防止用户信息泄露，本章将会设计 ID 登录界面保证信息安全。结合系统需求分析，系统整体结构如图 5-1 所示。开始是 ID

验证登录与注册，通过比对输入的账号和密码是否是数据库中存储的信息，来判定是否为合法的访问者，验证正确才可以进入系统的内部界面。在开放注册权限的情况下，新用户可通过注册账户，获得合法登陆系统的权限。

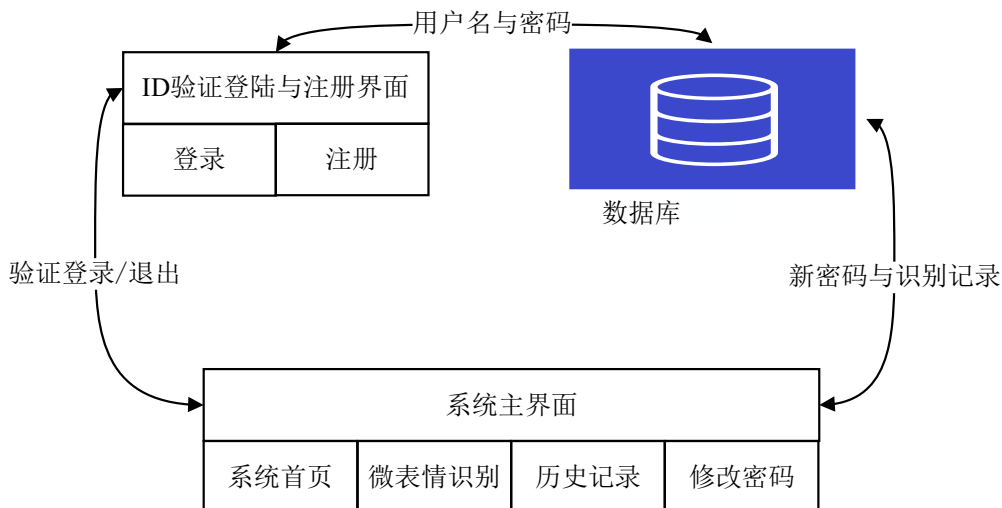


图 5-1 系统整体架构图

进入系统内部会有四个功能界面分别为介绍系统和微表情知识背景的系统首页；实现微表情识别功能的识别模块，该模块又包含视频采集、数据预处理、一键识别等，模块会将识别结果传给数据库，为他人对用户的需求分析提供参考；记录日期、时间和识别结果的历史记录界面；还有可以修改账号和密码的修改密码界面，该模块会将修改后的信息传输给数据库，下次登陆需使用新的账号和密码才可以进入系统主界面。

5.3 系统实现

5.3.1 开发平台与数据库

本章使用的开发硬件是 64 位 Windows11 操作系统，处理器是因特尔的 14 核 20 线程 i7-12700H，使用编程软件为 PyCharm 2023 社区版，使用 PostgreSQL 作为后端数据库来存储重要信息和数据中转站。UI 界面的设计使用的是 PyCharm 中的 PyQt5 库和 QT Designer 软件。

5.3.2 ID 验证登陆与注册界面的实现

为方便系统的设计开发，根据功能需求，分析内在逻辑，绘制了登陆与注册顺序逻辑图，具体如下图 5-2 所示。

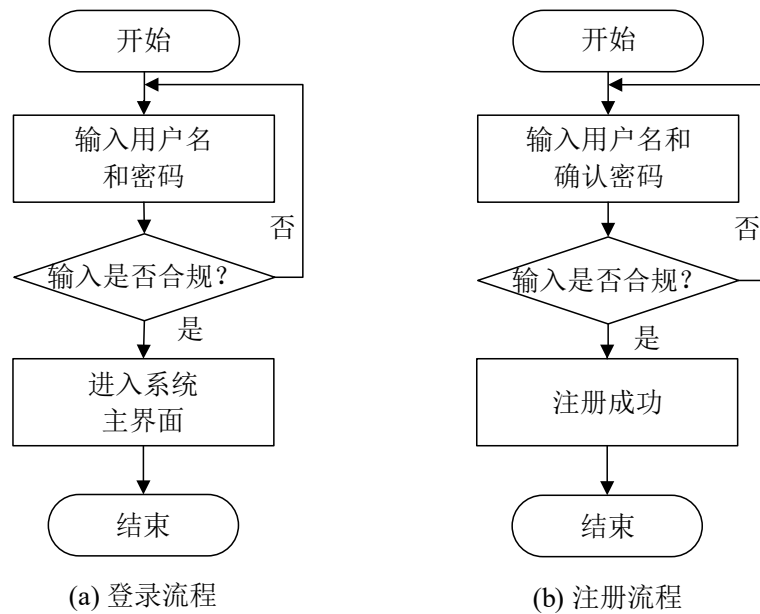


图 5-2 登陆与注册流程图

使用者需要输入正确的账号和与账号相对应的密码才能完成登录，进入微表情识别系统主界面和使用系统功能，否则会提示“账号或密码错误！”。在开放注册权限时需要键入账号、密码和确认密码，设置账号和密码的最大长度分别是 10 和 20，只有密码和确认密码相同才可以完成注册并会返回登陆界面，否则会提示“两次密码不一致！”。本章根据流程图设计了如下图所示的 UI 界面，此界面共有两个子界面分别实现登录和注册两个功能。



图 5-3 ID 验证登陆与注册界面图

下面是数据库建立的功能代码和实现登陆功能的部分代码展示，注册功能的实现和登陆界面类似。

数据库建立的部分代码：

代码：使用 PostgreSQL 建立用户数据库

```
/*建立表格存放账号和密码*/  
  
create table users (accounts varchar (10),passwords varchar(20));  
  
/*插入数据*/  
  
insert into users values('data1','data2');  
  
/*查看数据*/  
  
select * from users;
```

登陆界面 Python 代码：

代码：登陆界面

```
import psycpg2  
  
from LoginUi import * # 登陆与注册界面  
  
from interMERUi import * # 系统主界面  
  
from PyQt5.QtWidgets import QApplication, QMainWindow  
  
#定义登陆函数  
  
def login_in(self):  
  
    account = self.ui.lineEdit_L_account.text()  
  
    password = self.ui.lineEdit_L_password.text()# 获界面输入的账号和密码  
  
    account_list = []  
  
    password_list = []  
  
    # 访问数据库  
  
    conn =psycpg2.connect(database='数据库名', user='postgres', password= '123456',  
host='127.0.0.1', port='5432')  
  
    cur = conn.cursor()  
  
    cur.execute("select * from users") #查询 users 表格  
  
    rows = cur.fetchall()  
  
    for row in rows:  
  
        account_list.append(row[0])  
  
        password_list.append(row[1])  
  
    conn.commit()  
  
    conn.close()
```

```

for i in range(len(account_list)):
    if len(account) == 0 or len(password) == 0:
        self.ui.stackedWidget.setCurrentIndex(1) # 提示 “账号密码不能为空”
    elif account == account_list[i] and password == password_list[i]:
        self.win = MainWindow()
        self.close() # 关闭登陆界面，打开主界面
    else:
        self.ui.stackedWidget.setCurrentIndex(2) # 提示账号或密码不正确
    
```

5.3.3 系统主界面的实现

当账户和密码验证通过后，会进入系统的主界面，主界面包含系统首页、微表情识别、历史记录和修改密码四个界面。

第一个页面是系统首页，是用户访问平台时的第一印象，承载着对平台定位、功能特色和使用方法的直观展示。所以简明扼要的介绍系统的定位和使用说明，也简单介绍了微表情的特性和重要性。系统首页界面图如下图所示。

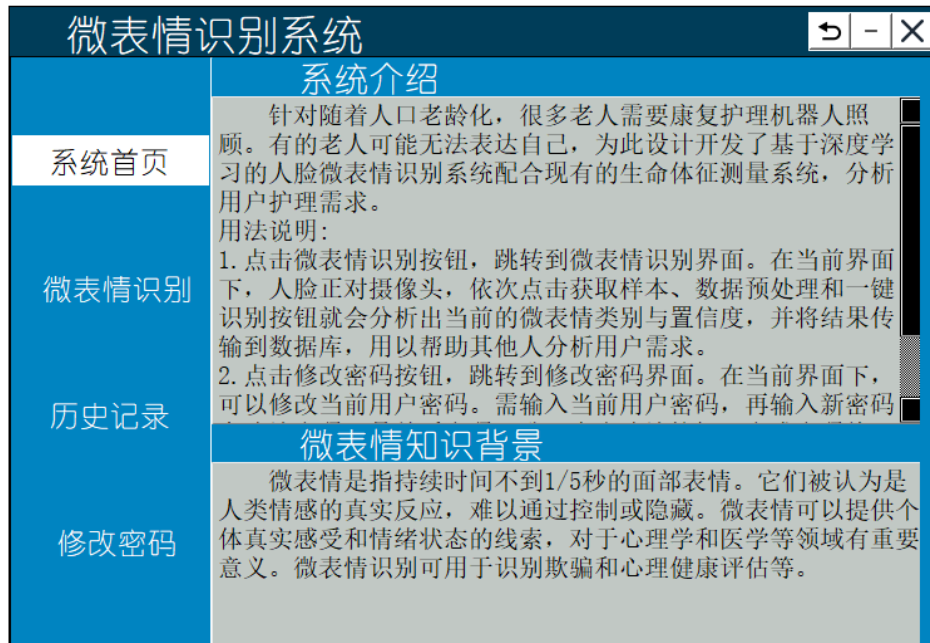


图 5-4 系统首页图

第二个页面是微表情识别页面，是本系统最重要的功能界面。点击微表情识别会进入如图 5-6 所示的界面首页。首先，设计了获取样本按钮，点击按钮就可以通过 OpenCV 调用摄像头获取 1 秒钟的微表情视频；然后，设计了数据预处理按钮，点击按钮就可以使用第 2 章所研究的图像增强、关键帧获取、尺度

变换和特征提取四个功能函数，自动获得基于起始帧和顶点帧的融合光流特征图，其中尺度变换使用 MTCNN 网络完成，该网络带有人脸检测功能，采集的视频中没有人脸会报错并跳转回首页；最后，通过点击一键识别按钮，将获取到的特征图输入到第 4 章预先训练好的深度学习网络中获取结果并传送到数据库储存，微表情识别界面顺序逻辑图如图 5-5 所示。

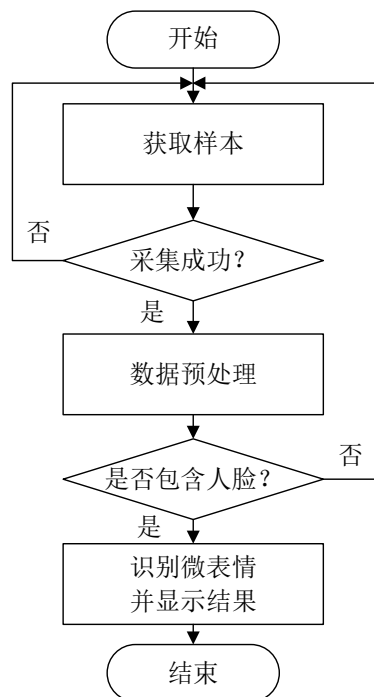


图 5-5 微表情识别顺序逻辑图



图 5-6 微表情识别初始界面图

第三个页面是历史记录页面，为了更好地管理用户的微表情识别记录，本章设计了如图 5-7 所示的历史记录表格界面，会记录当前的日期、时间、微表情识别结果和结果的置信度。这样就可以记录用户的情绪，有助于其他人更好的了解心态变化。

微表情识别系统				
系统首页 微表情识别 历史记录 修改密码	序号	日期 (年/月/日)	时间 (时/分)	结果 置信度
	1	2024/3/9	9:02	高兴 94.06%
	2	2024/3/10	10:35	惊讶 96.99%
	3	2024/3/11	16:12	高兴 91.34%
	4			
	5			
	6			
	7			
	8			
	9			
	10			
	11			

图 5-7 微表情识别结果历史记录图

第四个界面是修改密码界面，界面如图 5-9 所示。用户需要在旧密码框里输入当前密码，再通过键盘在新密码框和确认密码框中输入相同的字符串，此时会更新数据库中当前用户的登录密码，否者会报告“前后密码不一致!”或“密码错误!”。修改密码界面顺序逻辑图如图 5-8 所示。

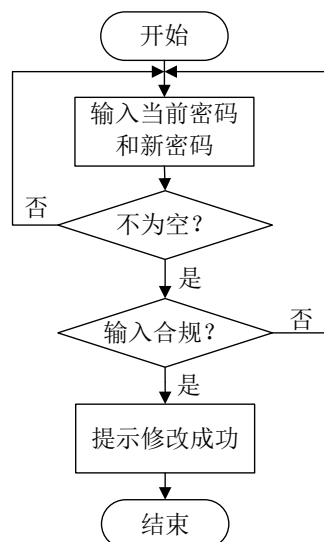


图 5-8 修改密码顺序逻辑图



图 5-9 修改密码界面图

5.4 系统测试

为测试微表情识别系统的可靠性，作者本人对系统的登录、注册和主界面功能进行了测试。

5.4.1 ID 验证登陆与注册功能测试

1) 登录功能测试

登陆界面包括“关闭”、“确认”、“注册”、“登录”四个按钮和“账号”和“密码”两个输入框。操作与正常反馈对照关系如表 5-1 所示。

表 5-1 登录功能测试对照表

操作	正常反馈
点击×按钮	关闭登录界面
点击注册按钮	转到注册界面
点击账号框	可以输入字符
点击密码框	可以输入字符，且无法看见输入值
输入为空，点击确认按钮	红色文字提示“账号或密码不能为空！”
输入错误的密码或账号，点击确认按钮	红色文字提示“账号或密码错误”
输入正确的账号和密码，点击确认按钮	转到系统主界面

输入错误密码或账号时，登录界面如图 5-10 所示。



图 5-10 系统登陆功能测试图

2) 注册功能测试

登陆界面包括“×”、“确认”、“注册”、“登录”四个按钮；“账号”、“密码”和“确认密码”三个输入框，还有一个文本提示框。操作与正常反馈对照关系如表 5-2 所示。

表 5-2 注册功能测试对照表

操作	正常反馈
点击×按钮	关闭登录界面
点击登录按钮	转到登录界面
点击账号框	可以输入字符
点击密码框	可以输入字符，且无法看见输入值
点击确认密码框	可以输入字符，且无法看见输入值
输入为空，点击确认	红色文字提示“账号或密码不能为空!”
密码输入两次不一致，点击确认按钮	红色文字提示“前后密码不一致!”
输入的账号或密码过长，点击确认按钮	红色文字提示“账号或密码过长!”
输入账号和前后密码一致，点击确认按钮	绿色文字提示“注册成功!”

输入空密码或账号时，注册界面如图 5-11 所示。



图 5-11 系统注册功能测试图

5.4.2 系统主界面功能测试

主界面包含状态栏三个按钮，导航栏四个按钮。操作与正常反馈对照关系如表 5-3 所示。

表 5-3 主界面功能测试对照表

操作	正常反馈
点击状态栏左边按钮	可以返回系统首页
点击状态栏中间按钮	可以最小化窗口
点击状态栏右边按钮	可以关闭窗口
点击左侧导航栏系统首页按钮	可以跳转到系统首页界面
点击左侧导航栏微表情识别按钮	可以跳转到微表情识别界面
点击左侧导航栏历史记录按钮	可以跳转到历史记录界面
点击左侧导航栏修改密码按钮	可以跳转到修改密码界面

1) 系统首页功能测试

系统首页包含系统介绍、操作说明和微表情知识背景，主要检测文字说明是否完整，是否可以查看完整的使用说明。

2) 微表情识别功能测试

微表情识别包括“获取样本”、“数据预处理”和“一键识别”三个按钮；图像显示区和文本提示区。其中文本显示区有：绿色字体：“采集成功!”、“预处理成功”；红色字体：“采集失败!”、“无人脸图像!”、显示识别结果的名称与置

信度。操作与正常反馈对照关系如表 5-4 所示。

表 5-4 微表情识别功能测试对照表

操作	正常反馈
点击获取样本按钮	打开摄像头获取视频，在显示区展示；文本提示采集是否成功
点击数据预处理按钮	文本提示有无人脸；有则显示区展示处理结果，无则返回主页
点击一键识别按钮	显示区展示顶点帧图像；文本区提示微表情类别置信度

如图 5-12 和图 5-13 是作者本人的识别示例。

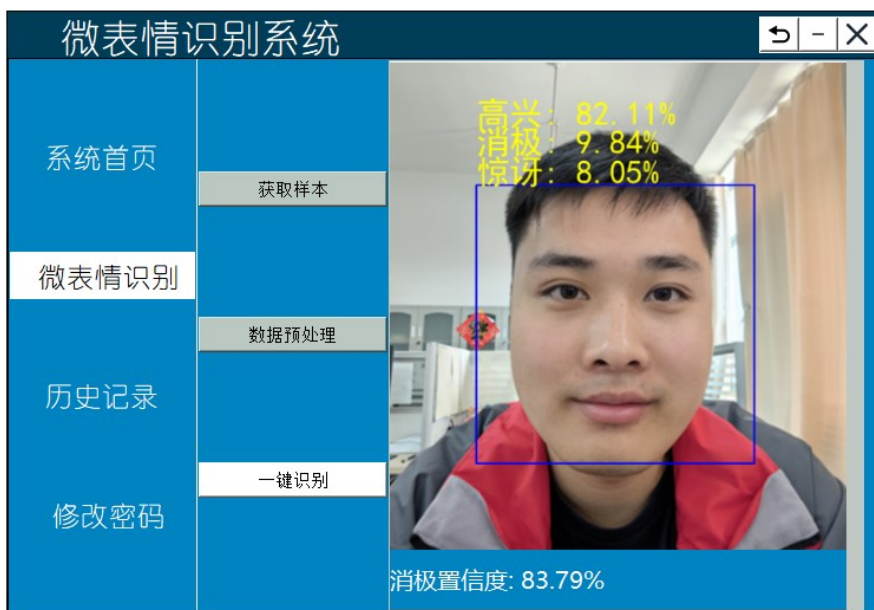


图 5-12 微表情识别功能测试图 1

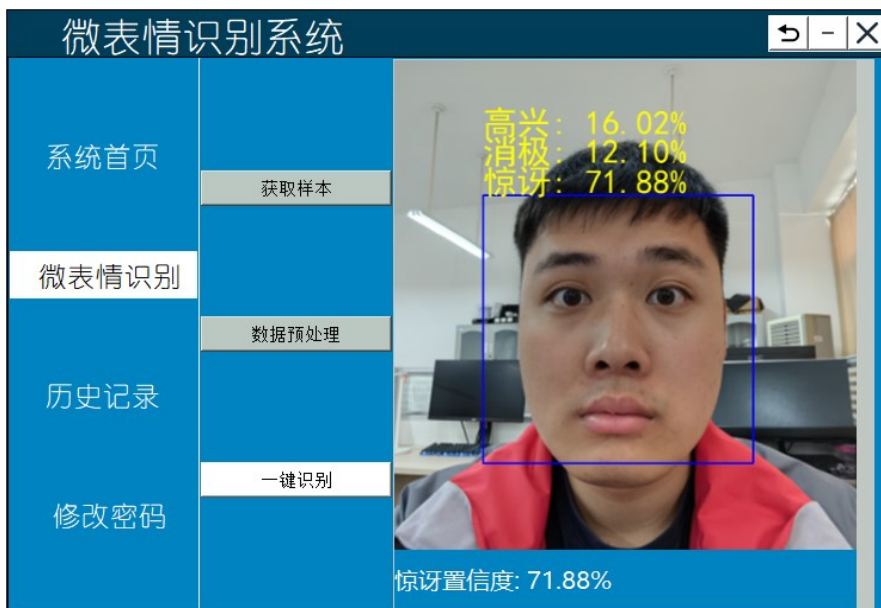


图 5-13 微表情识别功能测试图 2

3) 历史记录功能测试

历史记录会记录日期、时间、识别结果以及置信度。主要检测是否可以查看历史记录和记录内容是否正确，

4) 修改密码功能测试

修改密码界面包括“确认”按钮和“账号”、“密码”两个输入框。操作与正常反馈对照关系如表 5-5 所示。

表 5-5 修改密码功能测试对照表

操作	正常反馈
点击旧密码框	可以输入字符
点击新密码框	可以输入字符
点击确认新密码框	可以输入字符
输入为空，点击确认按钮	红色文字提示“输入不能为空!”
输入错误的旧密码，点击确认按钮	红色文字提示“密码错误!”
新密码和确认密码不一致，点击确认按钮	红色字体提示“前后密码不一致!”
输入正确新旧密码，点击确认按钮	绿色字体显示“修改成功!”

如图 5-14 是输入正确的旧密码且输入的新密码和确认新密码一致时的测试结果图。



图 5-14 修改密码功能测试图

5.5 本章小结

本章通过在 CASME II 数据集上训练好的 GAM-Res-DA 模型为基础, 结合第 2 章研究的数据预处理技术, 设计开发了一款微表情识别系统。首先, 介绍了研究设计本系统的意义; 其次, 研究了人脸微表情识别系统的功能需求; 然后, 根据系统的功能需求进行功能逻辑分析, 以此设计系统 UI 界面和程序; 最后, 进行了系统功能测试, 验证了系统的可行性。为微表情识别的信息化、智能化提供了思路, 也是对微表情识别算法的应用做出尝试, 推动微表情识别系统的实际应用。

第 6 章 总结与展望

6.1 总结

面部表情是人类表达情绪的途经之一，在人类交流和人机交互中起着至关重要的作用。微表情是一种轻微的面部表情，是强度很低且持续时间短暂的自发动作，其难以控制的特性使其可以揭示一个人的真实情绪，被应用于司法审问、智慧教育和医疗保健等多个领域，因此基于深度学习的自动人脸微表情识别研究具有重大意义。本文首先研究了融合注意力机制和 LLFormer 图像增强网络和提取融合光流特征的图像预处理方法；其次探讨了融合 MobileViT 和注意力机制的网络模型，并用于微表情的识别；然后探讨了融合两种注意力机制的残差网络模型；最后，设计了一款基于深度学习的微表情识别系统。具体的研究内容如下：

1) 研究了融合 CBAM 和 LLFormer 的低光照图像增强网络，并在 LOL 数据集上进行试验，验证了该算法可以有效提高图像质量；还研究了基于 TVL1 光流的融合光流特征提取方法，并使用 SVM 分类器进行分类实验，验证了融合光流的有效性。

2) 探讨了 CA-MobileViT 和 GAM-Res-DA 特征提取网络模型。通过引入注意力机制提高网络对重要信息的关注度，使用 Focal Loss 损失函数减弱数据不平衡的影响，在 CASME 和 CASME II 数据集上进行 LOSO 实验评估模型性能，验证了两种模型对微表情识别的有效性。并将在自建数据集上训练 GAM-Res-DA 模型，取得了较好的识别准确率。

3) 设计了一款微表情识别系统。系统运用第 2 章的预处理技术和第 4 章中训练好的神经网络模型来识别微表情，尝试基于深度学习的微表情识别算法实际应用。

6.2 展望

本文研究的人脸微表情预处理方法和识别模型，虽然可以识别微表情，但还有很多的不足之处，未来研究的方向有以下几个：

1) 研究更好的图像预处理技术。因为微表情属于面部动作，平面 2D 图像

无法完全展示全部信息，所以考虑使用 3 维重建技术建造 3D 人脸模型，让微表情面部运动单元更加明显。再根据不同面部动作单元组合分类微表情，提高微表情识别的可解释性和可靠性。

2) 设计多模态网络模型。微表情发生是个人情绪变化的体现，情绪变化会引起体温，血压、心跳、呼吸等生命体征的变化，也会伴随其他的肢体动作。未来可以通过结合更多特征信息训练多模态模型，提高微表情识别的准确性。

3) 建立专业的微表情识别系统。本文设计的系统使用自建数据集训练的网络模型，泛化能力较差。可以考虑根据系统的应用场景，分析场景需求和难点，建立专门的大型数据集训练特定场景的网络模型，这样可以提高系统的泛化能力和可靠性，推进微表情识别系统的应用进程。

参考文献

- [1] 贾晓烨,杨明强,张鹏等.微表情自动识别综述[J].计算机辅助设计与图形学学报,2014,26(09):1385-1395.
- [2] 周伟航,肖正清,钱育蓉等.微表情自动分析方法研究综述[J].计算机应用研究,2022,39(07):1921-1932.
- [3] Shen X B, Qi W, Fu X L. Effects of the duration of expressions on the recognition of micro expressions[J]. Journal of Zhejiang University SCIENCE B, 2012, 13(03): 221-230.
- [4] Darwin C. The Expression of the Emotions in Man and Animals[M]. Chicago: University of Chicago Press, 1965.
- [5] Ekman P, Friesen W V. Facial action coding system (FACS): a technique for the measurement of facial actions[J]. Rivista Di Psichiatria, 1978, 47(02): 126-138.
- [6] Davison A K, Lansley C, Costen N, et al. SAMM: A Spontaneous Micro-Facial Movement Dataset[J]. IEEE Transactions on Affective Computing, 2018, 9(01): 116-129.
- [7] Li X, Pfister T, Huang X, et al. A spontaneous micro-expression database: Inducement, collection and baseline[C]//2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. IEEE, Shanghai, China, 22-26 April, 2013: 1-6.
- [8] Zhi R, Hu J, Wan F. Micro-expression recognition with supervised contrastive learning[J]. Pattern Recognition Letters, 2022, 163: 25-31
- [9] Adegun I P, Vadapalli H B. Facial micro-expression recognition: A machine learning approach[J]. Scientific African, 2020, 8: e00465.
- [10] Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(01): 221-231.
- [11] 申寻兵,隋华杰,傅小兰.微表情在欺骗检测中的应用[J].心理科学进展,2017,25(02):211-220.
- [12] 彭玉伟.微表情分析技术在侦查讯问工作中的应用研究[J].中国刑事法杂志,2015(02):95-103.
- [13] 李雯,李豪喆,陈琛等.面部微表情分析技术在法医精神病学领域的研究现状及应用展望[J].法医学杂志,2023,39(05):493-500.
- [14] 徐峰,张军平.人脸微表情识别综述[J].自动化学报,2017,43(03):333-348.
- [15] Goh K M, Ng C H, Lim L L, et al. Micro-expression recognition: an updated review of current trends, challenges and solutions[J]. The Visual Computer, 2020, 36(03): 445-468.
- [16] Cuimei L, Zhiliang Q, Nan J, et al. Human face detection algorithm via Haar cascade classifier combined with three additional classifiers[C]//2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI). IEEE, Yangzhou, China, 20-22 October, 2017: 483-487.
- [17] Nazir M, Jan Z, Sajjad M. Facial expression recognition using histogram of oriented gradients based transformed features[J]. Cluster Computing, 2018, 21(01): 539-548.
- [18] Xi D, Lee S-W. Face Detection based on support vector machines[C]//Pattern Recognition with

- Support Vector Machines. Springer Berlin, Niagara Falls, Canada, 10 August, 2002: 370-387.
- [19] Zhang K, Zhang Z, Li Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.
- [20] Zhang J, Wu X, Hoi S C H, et al. Feature agglomeration networks for single stage face detection[J]. Neurocomputing, 2020, 380: 180-189.
- [21] Xu F, Zhang J, Wang J Z. Microexpression identification and categorization using a facial dynamics map[J]. IEEE Transactions on Affective Computing, 2017, 8(02): 254-267.
- [22] Ngo A C L, See J, Phan R C W. Sparsity in Dynamics of Spontaneous Subtle Emotions: Analysis and Application[J]. IEEE Transactions on Affective Computing, 2017, 8(03): 396-411.
- [23] Wu H Y, Rubinstein M, Shih E, et al. Eulerian video magnification for revealing subtle changes in the world[J]. ACM Trans. Graph, 2012, 31(04): 1-8.
- [24] Oh T H, Jaroensri R, Kim C, et al. Learning-based video motion magnification[C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2018: 663-679.
- [25] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern Recognition, 1996, 29(01): 51-59.
- [26] Zhao G, Pietikainen M. Dynamic texture Recognition using local binary patterns with an application to facial expressions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(06): 915-928.
- [27] Wang Y, See J, Phan R C W, et al. Lbp with six intersection points: reducing redundant information in lbp-top for micro-expression recognition[C]//Computer Vision-ACCV 2014: 12th Asian Conference on Computer Vision, Springer International Publishing, Singapore, 1-5 November, 2014, Revised Selected Papers, Part I 12, 2015: 525-537.
- [28] Wang Y, Han J, Guo Z. LCBP-STGCN: A local cube binary pattern spatial temporal graph convolutional network for micro-expression recognition[J]. Journal of Intelligent & Fuzzy Systems, 2023, 44(02): 1601-1611.
- [29] Liong S T, Phan R C W, See J, et al. Optical strain based recognition of subtle emotions[C]//2014 International Symposium on Intelligent Signal Processing and Communication Systems, IEEE, Kuching, Malaysia, 01-04 December, 2014: 180-184.
- [30] Happy S L, Routray A. Fuzzy histogram of optical flow orientations for micro-expression recognition[J]. IEEE Transactions on Affective Computing, 2019, 10(03): 394-406.
- [31] Liong S T, See J, Wong K, et al. Less is more: micro-expression recognition from video using apex frame[J]. Signal Processing: Image Communication, 2018, 62: 82-92.
- [32] Liu Y J, Zhang J K, Yan W J, et al. A main directional mean optical flow feature for spontaneous micro-expression recognition[J]. IEEE Transactions on Affective Computing, 2016, 7(04): 299-310.
- [33] Patel D, Hong X, Zhao G. Selective deep features for micro-expression recognition[C]//2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, Cancun, Mexico, 04-08

- December, 2016: 2258-2263.
- [34] Takalkar M A, Xu M. Image based facial micro-expression recognition using deep learning on small datasets[C]//2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, Sydney, NSW, Australia, 29 November-01 December, 2017: 1-7.
- [35] 陈汤慧,高美凤.基于 ME-Xception 卷积神经网络的微表情识别[J].信号处理,2022,38(05):992-1000.
- [36] Quang N V, Chun J, Tokuyama T. CapsuleNet for micro-expression recognition[C]//2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, Lille, France, 14-18 May, 2019: 1-7.
- [37] Peng M, Wang C, Chen T, et al. Dual temporal scale convolutional neural network for micro-expression recognition[J]. Frontiers in Psychology, 2017, 8: 273835.
- [38] Kim D H, Baddar W J, Jang J, et al. Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition[J]. IEEE Transactions on Affective Computing, 2019, 10 (02): 223-236
- [39] Fan X, Chen X, Jiang M, et al. SelfME: self-supervised motion learning for micro-expression recognition[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Vancouver, BC, Canada, 17-24 June, 2023: 13834-13843.
- [40] 赵明华,董爽爽,胡静等.注意力引导的三流卷积神经网络用于微表情识别[J].中国图象图形学报,2024,29(01):111-122.
- [41] Guo X, Li Y, Ling H. LIME: low-light image enhancement via illumination map estimation[J]. IEEE Transactions on image processing, 2016, 26(02): 982-993.
- [42] Wang T, Zhang K, Shen T, et al. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method[C]//Proceedings of the AAAI Conference on Artificial Intelligence. AAAI Press, Washington, DC, USA, 7-14 February, 2023, 37(03): 2654-2662.
- [43] Woo S, Park J, Lee J-Y, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision (ECCV). Springer Cham, Munich Germany, 8-14 September, 2018: 3-19.
- [44] Wei C, Wang W, Yang W, et al. Deep retinex decomposition for low-light enhancement[J]. arXiv preprint arXiv: 1808.04560, 2018: 1-12.
- [45] Ying Z, Li G, Gao W. A bio-inspired multi-exposure fusion framework for low-light image enhancement[J]. arXiv preprint arXiv: 1711.00591, 2017: 1-10.
- [46] Zhang Y, Zhang J, Guo X. Kindling the darkness: A practical low-light image enhancer[C]//Proceedings of the 27th ACM international conference on multimedia. Association for Computing Machinery, Nice France, 21-25 October, 2019: 1632-1640.
- [47] Wang Z, Cun X, Bao J, et al. Uformer: A general u-shaped transformer for image restoration[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition

- (CVPR). IEEE, New Orleans, LA, USA, 18-24 June, 2022: 17662-17672.
- [48] Zach C, Pock T, Bischof H. A duality based approach for realtime tv-l1 optical flow[C]//Pattern Recognition: 29th DAGM Symposium. Springer, Berlin Heidelberg, Germany, 12-14 September, 2007: 214-223.
- [49] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). IEEE, Venice, Italy, 22-29 October, 2017: 2999-3007.
- [50] Liu J, Chen H, Zhou W. Improved MobileViT: A more efficient light-weight convolution and vision transformer hybrid model[J]. Journal of Physics: Conference Series, 2023, 2562(01): 012012.
- [51] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). IEEE, Salt Lake City, UT, USA, 18-23 June, 2018: 4510-4520.
- [52] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Nashville, TN, USA, 20-25 June, 2021:13708-13717.
- [53] Wen J Y, Wu Q, Yong J L, et al. CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces[C]//2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE, Shanghai, China, 22-26 April, 2013: 1-7.
- [54] Yan W J, Li X, Wang S J, et al. CASME II: An improved spontaneous micro-expression database and the baseline evaluation[J]. Plos One, 2014, 9 (01): e86041
- [55] 牛瑞华,杨俊,邢斓馨等.基于卷积注意力模块和双通道网络的微表情识别算法[J].计算机应用,2021,41(09):2552-2559.
- [56] Liu Y J, Li B J, Lai Y K. Sparse mdmo: Learning a discriminative feature for micro-expression recognition[J]. IEEE Transactions on Affective computing,2018, 12 (01): 254-261.
- [57] Li J, Wang Y, See J, et al. Micro-expression recognition based on 3D flow convolutional neural network[J]. Pattern Analysis and Applications, 2019, 22: 1331-1339.
- [58] Song B, Li K, Zong Y, et al. Recognizing spontaneous micro-expression using a three-stream convolutional neural network[J]. IEEE Access, 2019, 7: 184537-184551.
- [59] 任宇,陈新泉,王岱嵘等.改进的残差网络与峰值帧的微表情识别[J].重庆工商大学学报(自然科学版),2023,41(01):1-11.
- [60] 王晔,王峰,贾海蓉等.结合人脸关键点与光流特征的微表情识别[J].激光杂志,2023,44(05):72-77.
- [61] Lei L, Li J, Chen T, et al. A novel graph-tcn with a graph structured representation for micro-expression recognition[C]//Proceedings of the 28th ACM International Conference on Multimedia. Seattle, WA, USA, 12-16 October, 2020: 2237-2245.
- [62] Lei L, Chen T, Li S, et al. Micro-expression recognition based on facial graph representation

- learning and facial action unit fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Nashville, TN, USA, 19-25 June, 2021: 1571-1580.
- [63] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, 27-30 June, 2016: 770-778.
- [64] Jiang S, Campbell D, Lu Y, et al. Learning to estimate hidden motions with global motion aggregation[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Montreal, QC, Canada, 10-17 October, 2021: 9752-9761.
- [65] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA, 15-20 June, 2019: 3141-3149.

攻读学位期间发表的学术论文

- [1] 唐绍语,魏利胜.融合光流与多注意力机制的微表情识别[J].安徽工程大学学报(三类,已录用:2023.11.07.0001)
- [2] **Shaoyu Tang**, Lisheng Wei. Micro-expression recognition based on improved MobileViT[C]//2024 the 4th International Conference on Computer, Control and Robotics. (EI,已录用: SC059)
- [3] **Tang Shaoyu**, Wei Lisheng, Wang Rui and Zhang Pinggai. Detection with thermal imaging for packaging bag sealing based on knowledge transfer[J]. International Journal of Modelling, Identification and Control. (EI,已录用: IJMIC-164322)
- [4] 高港,魏利胜,**唐绍语**. 一种基于改进 YOLOv5 单目测距方法研究 [J]. 安徽工程大学学报, 2023, 38 (01): 42-48.
- [5] Xiang Shiyu, Wei Lisheng, **Tang Shaoyu**. Fusion of UNet and Criminisi algorithms for endoscopic specular highlight removal[C]//2023 38th Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE, 2023: 1093-1097. (EI:20240815614624)
- [6] Shen Xin, Wei Lisheng, **Tang Shaoyu**. Dermoscopic image classification method using an ensemble of fine-tuned convolutional neural networks[J]. Sensors, 2022, 22(11): 4147. (SCI:000808672100001)
- [7] Tao xinjie, Wei Lisheng, **Tang Shaoyu**, Lu huacai. Image Recognition of Alzheimer's disease Based on Improved MobileVit Network [C]// the 43rd Chinese Control Conference. (EI,已投稿:1902)

攻读学位期间获得的奖励

- [1] 获得 2023-2024 学年安徽工程大学研究生**二等奖学金**
- [2] 获得 2022-2023 学年安徽工程大学研究生**二等奖学金**
- [3] 获得 2021-2022 学年安徽工程大学研究生**二等奖学金**
- [4] 获得 2022 年安徽工程大学第八届“互联网+”大学生创新创业大赛研究生创意组**三等奖**

致 谢

文章写到这里代表我的研究生求学之路即将结束，回首这三年，我深切感受到自己的成长与进步，这要感谢所有给予我支持和帮助的人，他们的陪伴、鼓励、教导对我产生了深远的影响，让我得以完成这篇文章。借此，向他们表达诚挚谢意。

首先，衷心感谢我的导师，感谢尊敬的魏利胜老师给予我的教诲，在学术上为我指引方向、提供专业的建议，在生活 and 思想上传授给我诸多的人生道理；老师的敬业精神、学识渊博以及老师对学术的执着追求深深激励着我，让我更加坚定地走在学术道路上。

然后，诚挚的感谢诸位同门和朋友，感谢他们在我迷茫之时提供观点和建议，在懈怠之时给予适当的激励，让我可以更加清晰且积极的完成学业；也感谢他们的陪伴让我感受到属于研究生的美好生活；同时也非常感谢校外指导老师程武山教授对我学习的支持和指导，在学习过程中，为我提供了宝贵的知识和经验，使我能够更加深入地理解学科内容。

最后，感谢我的父母和姐姐，他们用辛勤的劳动和无私的奉献，为我创造了温馨的家庭环境，提供了全面的情感、物质、精神等多方面的支持，让我可以更有信心面对学习和生活中的挑战。

再次感谢所有人，祝福你们生活幸福、万事如意！

唐绍语

安徽工程大学电气工程学院

二〇二四年四月

尚德敬学 唯实惟新

