

中图法分类号: 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-15

论文引用格式: Chang Heyou, Yang Jiazheng, Gao Guangwei, Zhang Jian, Zheng Hao. XXXX. Micro-expression Recognition Method Based on Multi-Optical Flow Fusion and KAN. Journal of Image and Graphics, XX(XX):0001-0015(常合友, 杨佳铮, 高广谓, 张健, 郑豪. XXXX. 融合多光流与KAN的微表情识别方法. 中国图象图形学报, XX(XX):0001-0015)[DOI:10.11834/jig.240572]

融合多光流与KAN的微表情识别方法

常合友², 杨佳铮^{1,2}, 高广谓^{3,4}, 张健¹, 郑豪^{2*}

1. 江苏海洋大学计算机工程学院, 江苏省连云港市 222005; 2. 南京晓庄学院信息工程学院, 江苏省南京市 211171; 3. 南京邮电大学先进技术研究院, 江苏省南京市 211112; 4. 人工智能教育部重点实验室, 上海, 200240

摘要: 目的 微表情是由个体的内在情感反应引发的面部肌肉活动, 在心理诊断、医学、刑侦测谎等领域有着广泛的应用场景。现有微表情识别方法大都利用单一光流获取面部运动差异, 无法有效应对光照变化或表情强度不一致等问题。为了解决上述问题, 本文提出一种融合多光流与KAN的微表情识别方法(multiple optical flow feature fusion, MOFFFN), 通过捕获多层次、多角度的面部运动差异, 提高微表情识别性能。**方法** 首先, 本文提取三种不同的光流特征, 并构造光流融合模块以捕获这些光流特征水平和垂直方向的信息。其次, 本文构造了一个新颖的特征提取模型, 利用柯尔莫哥洛夫-阿诺尔德网络(Kolmogorov-Arnold networks, KAN)与卷积注意力机制捕捉微表情的细微变化, 提取更具鉴别能力的特征。最后, 本文还设计了一个高效的注意力下采样自注意力特征融合模块, 能够在融合多光流特征的同时突出微表情变化的关键区域特征。**结果** 本文使用主流的留一交叉验证法(leave-one-subject-out-cross-validation, LOSOCV), 分别在中国科学院微表情数据库第二版(Chinese Academy of Sciences micro-expression II, CASME II)、自发性动作与微动作数据库(spontaneous actions and micro-movements, SAMM)和高速自发性微表情数据库(spontaneous micro-expression corpus-high speed, SMIC-HS)三个公开数据集以及复合数据集(composite dataset, CD)上进行了验证。本方法在四个数据集上分别取得了91.79%, 85.69%, 86.56%和85.03%的未加权平均召回率(unweighted average recall, UAR), 以及92.95%, 89.1%, 91.78%和87.63%的未加权F1分数(unweighted F1-score, UF1), 性能优于主流的微表情识别方法。**结论** 本文所提出的方法通过融合多种光流特征, 利用KAN和注意力机制提取更具鉴别能力和鲁棒性的特征, 显著提高了微表情识别的结果。本文代码将在论文被接收后公开: <https://github.com/useless12138/mofffn>。

关键词: 微表情识别; 光流; 特征融合; KAN; 自注意力机制

Micro-expression Recognition Method Based on Multi-Optical Flow Fusion and KAN

Chang Heyou², Yang Jiazheng^{1,2}, Gao Guangwei^{3,4}, Zhang Jian¹, Zheng Hao^{2*}

1. School of Computer Engineering, Jiangsu Ocean University, Lianyungang 222005, China; 2. School of Information Engineering, Nanjing Xiaozhuang University, Nanjing 211171, China; 3. Institute of Advanced Technology, Nanjing University of Post and Telecommunications, Nanjing 211112, China; 4. Key Laboratory of Artificial Intelligence, Ministry of Education, Shanghai, 200240, China

收稿日期: 2024-09-23; 修回日期: 2025-04-24

* 通信作者: 郑豪, zhh710@163.com

基金项目: 国家自然科学基金项目(61976118, 61806098), 高维信息智能感知与系统教育部重点实验室开放基金(No. 2023-3, 2023-10), 人工智能教育部重点实验室开放基金(No. AI202404)

Supported by: National Natural Science Foundation of China (61976118, 61806098), Open Fund of Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information (No. 2023-3, 2023-10), Open Fund of Key Laboratory of Artificial Intelligence (No. AI202404)

©中国图象图形学报版权所有

Abstract: Objective Micro-expressions are rapid, involuntary facial muscle movements that occur when individuals attempt to conceal their genuine emotions. These fleeting facial cues typically last less than half a second and are often imperceptible to the naked eye. Despite their subtle nature, micro-expressions carry significant emotional information and have demonstrated great potential in various real-world applications such as psychological evaluation, medical diagnosis, law enforcement, and deception detection. However, accurate recognition of micro-expressions remains a highly challenging task due to their low intensity, short duration, and vulnerability to environmental conditions such as lighting variation and inconsistent facial expression strength. Most existing micro-expression recognition (MER) approaches rely on a single optical flow algorithm to capture motion information, which often results in inadequate performance when faced with real-world complexities. These methods typically fail to capture the full spectrum of motion dynamics due to the limitations of individual flow models in handling subtle and diverse motion patterns. To overcome these limitations, this paper proposes a novel approach named Multiple Optical Flow Feature Fusion Network (MOFFFN), which leverages the advantages of multiple optical flow types and advanced attention mechanisms to improve the recognition performance of micro-expressions.

Method We propose a novel Micro-Expression Recognition Framework based on multiple optical flow feature fusion networks (MOFFFN), designed to capture the subtle spatiotemporal variations inherent in micro-expressions. The proposed framework consists of three key components: an Optical Flow Fusion Module (OFFM), a mobile residual KAN CBAM block net (MRKCBN), and an Attention Pooling Self-Attention Block (APSB). Firstly, the OFFM module integrates multiple types of optical flow features, including Total Variation L1 (TVL1), Dense Inverse Search (DIS), and Principal Component Analysis (PCA) optical flow. These methods capture different aspects of motion across pixels. To preserve directional cues, we extract both horizontal and vertical components from each flow type. By fusing these components, we construct three composite flow images: global fusion, horizontal fusion, and vertical fusion. This multi-flow fusion strategy captures more comprehensive motion information and enhances robustness to lighting and expression variance, thereby improving recognition performance. Secondly, to extract more discriminative features from the fused flow images, we introduce a novel MRKCBN architecture that integrates Kolmogorov–Arnold Networks (KAN) into a lightweight residual backbone. KAN has recently gained attention for its strong generalization and interpretability, particularly in scientific computing and image analysis. We embed KAN within both the channel and spatial attention submodules of the classic Convolutional Block Attention Module (CBAM), replacing the original MLP and convolution layers with KANLinear and KAN_Conv2d respectively. These KAN-based components offer greater adaptability by leveraging B-spline approximations, group convolution, and dropout, thereby improving the network's ability to model fine-grained spatiotemporal variations in facial micro-expressions. Lastly, the APSB module is designed to fuse the multi-source features extracted from TVL1, DIS, and PCA flows. In contrast to conventional self-attention mechanisms that operate solely on a single feature sequence, the APSB module is designed to learn an effective fusion strategy across diverse optical flow representations. The entire feature fusion process is conducted in parallel across multiple attention heads, enabling the capture of feature relationships at different dimensions and semantic levels. By embedding an Attention Pooling layer, the APSB module is able to learn the importance weights of different facial regions. Emphasizing key regions enhances the model's recognition performance and robustness. Collectively, these modules work synergistically to deliver state-of-the-art performance in micro-expression recognition, particularly under challenging conditions of subtle motion and limited data. **Results** To evaluate the effectiveness and generalizability of MOFFFN, we conduct experiments on three publicly available benchmark datasets: CASME II, SAMM, SMIC-HS, and a composite dataset (CD) that combines all three. We follow the leave-one-subject-out cross-validation (LOSOCV) protocol, which is widely adopted in micro-expression research to ensure robust subject-independent evaluation. Our proposed method achieves unweighted average recall (UAR) scores of 91.79%, 85.69%, 86.56%, and 85.03% on CASME II, SAMM, SMIC-HS, and CD, respectively. Furthermore, the corresponding unweighted F1-scores (UF1) are 92.95%, 89.10%, 91.78%, and 87.63%. These results consistently outperform existing state-of-the-art approaches, validating the superior performance of the MOFFFN framework across multiple datasets with different characteristics and challenges. In addition to the quantitative results, qualitative visualizations of attention maps further confirm that our model successfully highlights the most discriminative facial regions and captures the most relevant motion patterns. **Conclusion** This paper presents a novel micro-expression recognition method that fuses multiple

optical flow features and leverages the representational power of Kolmogorov–Arnold Networks and attention mechanisms to improve recognition performance. By incorporating diverse motion representations, learning discriminative features through KAN, and using attention-guided fusion, our method effectively addresses the challenges posed by the subtle and complex nature of micro-expressions. The incorporation of Kolmogorov–Arnold Networks (KAN) into the attention mechanism—replacing traditional MLP and convolutional layers—further improves the network’s capacity to model subtle spatiotemporal facial dynamics. This integration allows the model to extract more discriminative and interpretable features from complex micro-expression patterns. In addition, the Attention Pooling Self-Attention Block (APSB) facilitates cross-stream feature fusion, learning to emphasize critical facial regions across different motion representations. Collectively, these innovations contribute to significant performance gains on benchmark datasets, especially under the challenges of subtle motion and limited data. The proposed framework demonstrates strong generalization and robustness, making it a promising approach for real-world micro-expression recognition. Our code will be available when the paper is accepted: <https://github.com/less12138/mofffn>.

Key words: Micro-expression recognition; Optical flow; Feature fusion; KAN; Self-attention mechanism

0 引言

微表情指的是在非常短暂的时间内(通常为 1/25 秒至 1/5 秒)闪现在面部的微小表情,这些表情往往会被人们忽略,但它们却包含着丰富的情感信息(Li 等, 2024; Zheng 等, 2016; Zhang 等, 2020)。微表情涉及到人类情感与认知的深层次理解,在心理学、医学、安全监控等领域有着至关重要的作用。比如,在儿童心理健康领域,由于心智尚未成熟,医生往往难以通过正常交流获得准确的信息,容易导致误诊。基于深度学习的微表情识别可以作为心理诊断的辅助工具,有效缓解这一问题。

微表情领域的研究主要包含两个任务:微表情定位和微表情识别。微表情定位旨在在一个微表情视频序列中找到微表情发生的峰值帧。峰值帧(Apex frame)是微表情发生过程中面部抖动最为剧烈的一帧,包含了最为丰富的信息。微表情识别的目标是对于微表情数据进行分类,识别出当前微表情所对应的情感。

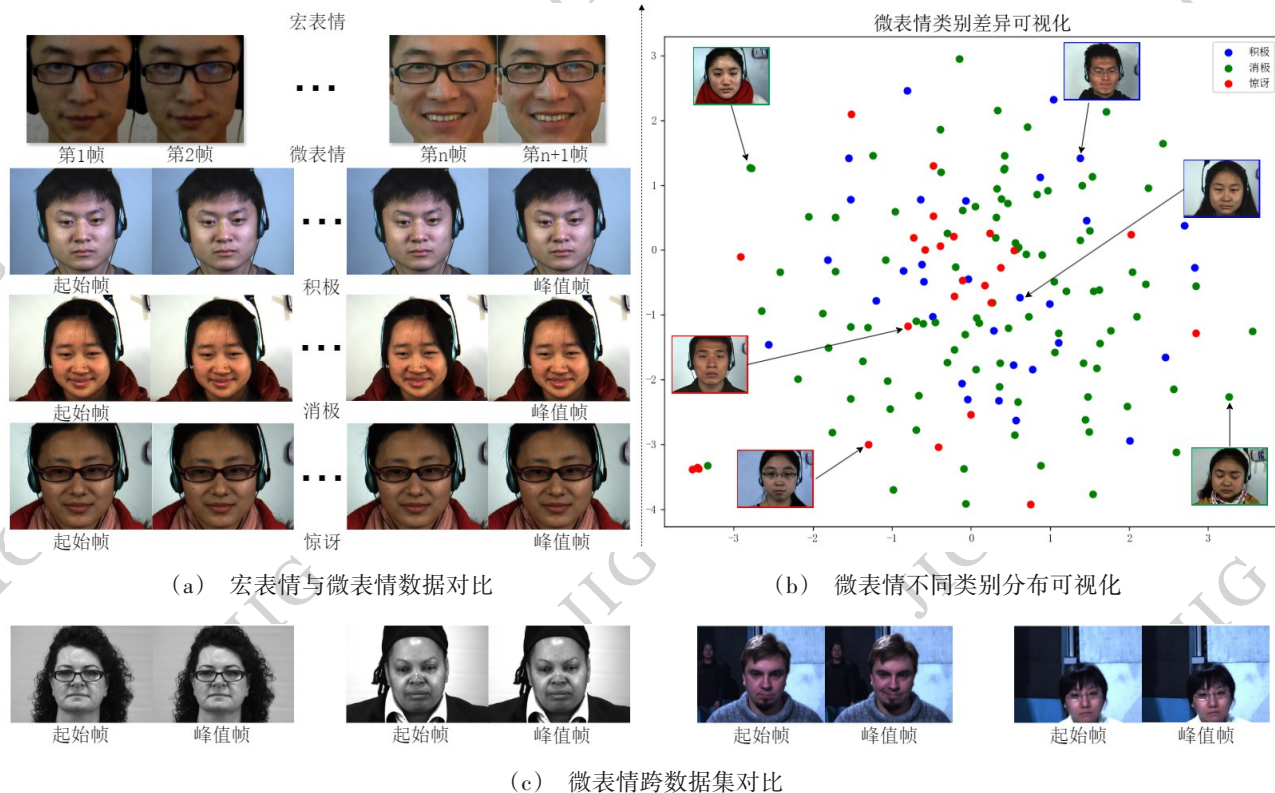
目前,微表情识别方法可以分为传统方法和基于深度学习方法两大类。在传统方法中,Zhao 提出一种动态纹理分析的方法,使用三正交面局部二值模式(local binary patterns from three orthogonal planes, LBP-TOP)处理视频中的动作和表情识别(Zhao 等, 2007)。Liu 提出一种主方向平均光流法(main directional mean optical flow, MDMO)来提取微表情时序中的面部信息(Liu 等, 2016)。在图像序列中,每个像素点的运动可以被视为一个向量,这些向

量的方向和大小代表了物体的运动。MDMO 方法试图找到这些向量的一个主要方向,这个方向可以代表整个图像序列中物体运动的总体趋势。然而这些传统方法提取的特征表达能力有限,不仅无法提取微表情图像的高维特征,也无法有效利用图像序列的时间和空间信息。近几年来,深度学习被广泛应用于各个领域并取得了显著性的优势(Chang 等, 2021; Li 等, 2020; Chang 等, 2024)。在微表情领域, Kim 使用卷积神经网络(convolutional neural network, CNN)和长短期记忆网络(long Short-Term memory, LSTM)依次从关键帧序列中学习空间和时间的微表情特征(Kim 等, 2016)。Zhou 采用初始网络从峰值帧的光流信息中提取空间特征(Zhou 等, 2022)。Gan 提出一种基于 CNN 的微表情识别方法,构造了一个从峰值帧获取光流特征的模型(OFF-ApexNet on micro-expression recognition system, OFF-ApexNet)(Gan 等, 2019)。Nguyen 提出一种基于 bert 的面部微表情识别(BERT-based Facial Micro-Expression Recognition, Micron-bert),通过对角微注意力机制(diagonal Micro-Attention, DMA)和关注区域块(patch of interest, POI)在端到端深度网络中实现无监督捕捉微表情(Nguyen 等, 2023)。Zhao 提出一种注意力引导的三流卷积神经网络用于微表情识别(Zhao 等, 2024)。然而,由于大都利用单一光流特征,在采集环境和图像质量较差情况下,上述方法的性能大幅下降。

当前微表情识别研究面临最大问题在数据集本身,原因有以下三点:首先,面部肌肉在微表情发生时的细微变化,单一光流无法有效提取面部运动特

征。其次,由于存在受试者不同人种和不同的拍摄条件等问题,导致现有方法在执行跨数据集任务时性能大幅下降。此外,由于微表情识别数据集样本

稀少,基于深度学习的微表情识别方法往往存在过拟合现象。



((a)comparison between macro-expression and micro-expression data;(b)visualization of micro-expression category distribution;(c) cross-dataset comparison of micro-expressions)

图1 微表情数据集分析

Fig. 1 Analysis of micro expression datasets

如图1所示。图1(a)展示了宏表情图像序列与不同类别微表情图像序列的对比。可以看出,宏表情(例如高兴)的表现比微表情更为显著,微表情从起始帧到峰值帧的变化极其微小。此外,利用t-分布随机邻居嵌入(t-distributed stochastic neighbor embedding, t-SNE)算法对三类微表情数据进行可视化处理,如图1(b)所示,可以看出三类微表情的数据分布过于随机,类内差异大于类间差异,进一步增加了微表情识别的难度。本文展示了不同微表情数据集之间的巨大差异。如图1(c)所示,不同微表情数据集的拍摄光线、角度,以及受试者的年龄、人种均有巨大差异。

为了解决上述问题,本文提出了融合多光流与KAN的微表情识别方法(multiple optical flow feature fusion, MOFFFN)。该方法的主要创新点如下:1)为

了解决单一光流无法有效捕获面部肌肉细微变化的问题,本文提出了一种多光流融合模块(optical flow fusion module, OFFM)。通过融合多种光流信息,OFFM在缓解不同拍摄角度、光照带来影响的同时,可以多层次、多角度地捕获面部肌肉运动差异,提取更丰富的微表情特征。2)本文首次将KAN(Liu等, 2024)引入到微表情识别领域,构造了基于残差KAN的特征提取模块(mobile residual KAN CBAM block net, MRKCBN),利用KAN可学习的边激活函数提取更具表示能力的特征,有效缓解模型过拟合问题。3)本文还设计了一个改进的多头注意力机制模块(attention pooling self-attention block, APSB)用以特征融合,通过嵌入注意力下采样提高模型对关键区域特征的权重。

1 相关工作

1.1 基于光流的微表情识别方法

最初, 研究人员大都使用基于局部二值模式 (local binary pattern, LBP) 特征的方法识别微表情。虽然 LBP 特征可以较好地表征微表情纹理变化的局部信息, 但由于缺乏全局信息而且无法表征面部纹理信息之间的差异, 这些方法的性能仍有较大提升空间。微表情通常持续时间非常短, 并且涉及面部肌肉的细微运动, 传统的静态图像处理方法难以准确捕捉和分析这些动态特征。光流法通过分析图像序列中像素点的移动, 能够有效计算出图像中的运动信息。2015 年, Liong 提出将光流法运用于微表情识别 (Liong 等, 2015), 该方法提出了一种新的基于光应变的加权特征提取方案, 通过时空池化运动信息为图像平面上的块区域分配权重。然而, 该方法在建模高维非线性特征方面仍存在不足, 难以适应复杂背景或头部轻微运动的干扰。Huang 结合了光流和局部二值模式来提取微表情的时空特征, 提出了一种基于新的时空面部表示框架 (Huang 等, 2015)。2019 年, Liu 提出了利用光流法 (optical flow) 提取的光流特征表示面部肌肉运动 (Liu 等, 2019)。该方法利用迁移学习技术和光流算法, 取得了较好的识别性能, 验证了光流特征的有效性。尽管取得了较好的性能, 这些方法大多依赖单一光流形式, 在面对复杂场景 (如光照波动、表情不对称、视频压缩) 时仍缺乏鲁棒性。如图 2 所示, 对起始帧和峰值帧进行光流估计, 并进行可视化。光流可视化使用颜色空间的色调和明度表示光流特征的方向和强度。

1.2 基于深度学习的方法

随着深度学习和卷积神经网络的兴起, 很多深度学习与光流相结合的方法被提出。Peng 提出一种利用三维光流特征结合三维卷积的网络模型捕捉微表情的时空动态变化 (Peng 等, 2017)。Liong 构造了经典的浅层三流三维卷积神经网络模型 (shallow triple stream three-dimensional CNN, STSTNet), 该模型利用三个光流特征 (光应变、水平和垂直光流场) 从视频的起始和顶峰帧中提取区分性高阶特征和微表情细节, 在当时取得了最佳性能 (Liong 等, 2019)。Nguyen 提出一种新型微表情识别方法, 利用 DMA 和

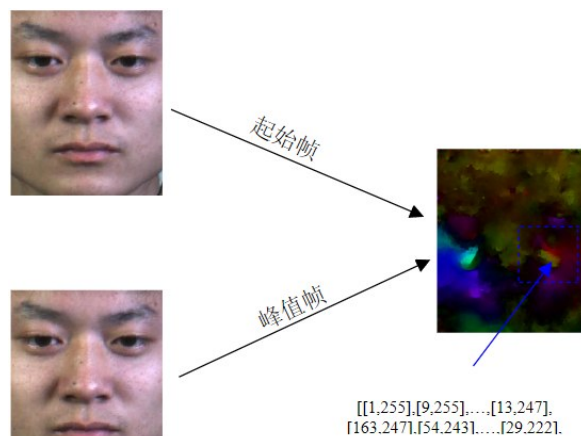


图2 光流图像

Fig. 2 Optical flow image

POI 模块实现自动将注意力集中在面部最能体现微表情变化的部分, 进一步提高了对微表情的检测和识别精度。特征融合通过结合多种特征类型和信息源, 实现了更加全面和精准的表情识别。但该方法计算成本较高, 且在小样本学习场景中稳定性有待进一步验证。Liong 提出了一种基于场景流注意力的微表情网络 (scene flow attention-based micro-expression network, SFAMNet) (Liong 等, 2024), 利用 RGB 图像和深度信息的方法 (RGB-Depth, RGB-D) 计算的场景流作为输入, 并将输入数据中不同维度的特征进行融合, 提高了模型微表情识别和定位的能力。

总体而言, 主流的微表情识别方法依靠单一光流特征难以提取微表情细微的运动变化, 对光照变化、背景变化等较为敏感。考虑到微表情数据采集困难, 样本稀少等情况, 本文所提方法通过提取并融合多种光流特征刻画面部多元运动信息, 利用 KAN 捕获微表情局部运动细节, 进而提升模型的识别性能。

2 融合多光流与 KAN 的微表情识别方法

本文提出的 MOFFFN 结构如图 3 所示, 主要包括光流融合模块 (optical flow fusion module, OFFM) (2.1 节)、基于残差 KAN 的特征提取模块 (mobile residual KAN CBAM block net, MRKCBN) (2.2 节) 以及注意力下采样自注意力特征融合模块 (attention

pooling self-attention block, APSB)(2.3节)。

2.1 OFFM 模块

由于微表情发生过程面部肌肉运动十分细微,主流方法大都先提取图像的光流特征,如总变差 L1 光流算法 (total variation - L1, TVL1) (Zach 等, 2007)、稠密逆向搜索算法 (dense inverse search, DIS) (Kroeger 等, 2016) 或主成分分析算法 (principal component analysis, PCA) (Wulff 等, 2015) 等。不同

的光流特征能够刻画每个像素点不同维度的变化信息。相较于单一光流,利用多种光流特征可以有效提升微表情的识别率 (Du 等, 2025)。与 (Du 等, 2025) 不同,本文构造的多光流融合模块新颖的提出将三种光流信息进行融合,并提取水平和垂直分量的光流信息进行融合,能够提取微表情变化的多层次信息。OFFM 结构如图 4 所示。

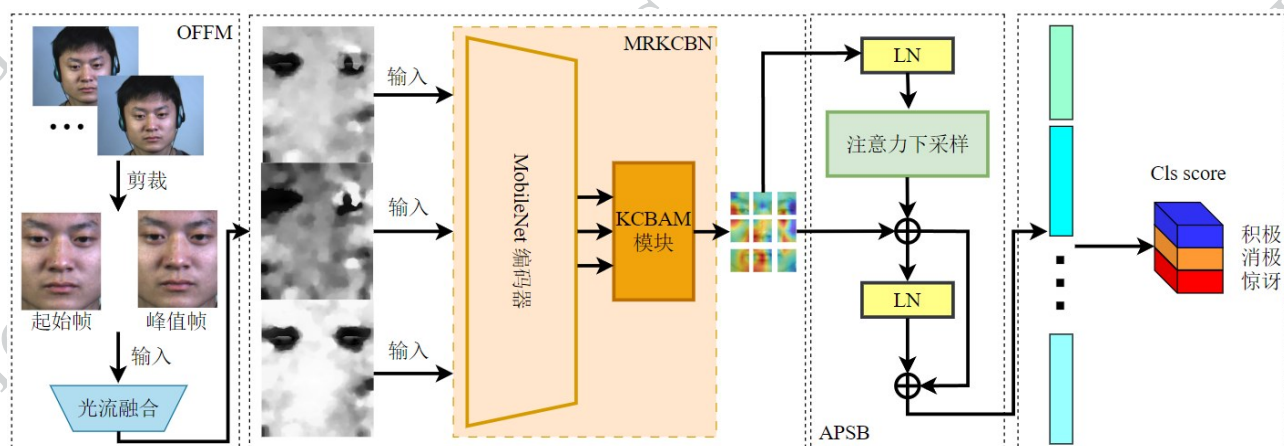


图 3 多光流特征融合网络模型示意图

Fig. 3 The diagram of multiple optical flow feature fusion network

首先,将每一个微表情视频序列裁剪成图片帧的格式,使用 dlib 库提供的 68 个人脸关键点检测的模型来实现面部对齐和剪裁操作。然后,根据起始帧和峰值帧分别计算 TVL1、DIS 和 PCA 三种光流特征,得到三种光流图像,即 TVL1 光流图像、DIS 光流图像和 PCA 光流图像。接着,分别提取三种光流图像的水平 and 垂直光流信息,并将三种光流图像的光流信息按照一定的比例进行融合,得到三种融合光流图像:整体融合光流图像、水平融合光流图像和垂直融合光流图像。

通过融合多种光流算法提取的光流信息,OFFM 可以获得微表情变化的多维度信息,提高模型对光照和表情变化等场景的鲁棒性,有助于提高微表情识别的准确率。此外,水平融合光流和垂直融合光流可以提供额外的线索信息,在复杂的面部表情中,有助于区分真实的表情和伪装的表情,减少识别误差。

2.2 MRKCBN 模块

由于微表情识别数据稀少且微表情变化极其微弱,传统的多层感知机 (Multi-Layer Perception, MLP)

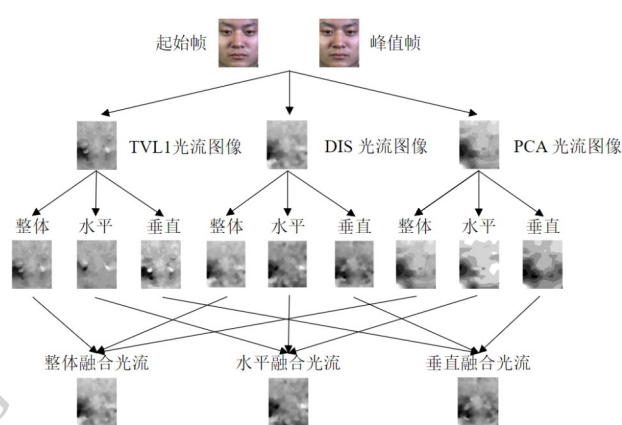


图 4 光流融合模块示意图

Fig. 4 Diagram of optical flow fusion module

不仅容易发生过拟合,而且难以捕捉复杂的表情边缘或肌肉变化。近两年,KAN 因其自适应性及可解释性受到越来越多关注,并在科学计算、图像分割、时间序列分析等领域获得验证。受此启发,本文首次将 KAN 引入到微表情识别领域,构造了一个新颖的 MRKCBN 特征提取网络。MRKCBN 主要包括三个模块,即 MobileNetV3, KAN 卷积注意力模块 (KAN convolutional attention module, KCBAM) 和自

适应平均池化(AdaptiveAvgPool),其结构图如图5所示。

输入光流图像,MRKCBN先对其进行卷积(卷积核大小为 3×3 ,数量为16)、归一化和Hardwish非线性变换操作,然后将特征图输入到MobileNetV3模型中,利用MobileNetV3提取光流图像的特征。然后,利用本文构造的KCBAM模块提取图像的时空特征,KCBAM结构图如图6所示。

卷积注意力模块(convolutional block attention module,CBAM)是一种结合通道注意力和空间注意力的深度学习模块,被广泛应用于各种卷积神经网络。

CBAM 通常使用 MLP 来压缩通道(Woo 等, 2018),然而 MLP 容易陷入局部最优解。此外,CBAM 中的空间注意力机制模块使用普通 2D 卷积对拼接后的特征图进行特征提取,2D 卷积只捕获局部领域内信息的局限性,这导致 CBAM 难以捕获光流图像的全局信息。最近,Liu 证明 KAN 通过卷积操作进行局部特征提取(Liu 等, 2024),能够更好地捕捉图像中像素之间的空间关系。受此启发,本文将 KAN 分别嵌入到 CBAM 中的通道注意力模块和空间注意力模块中,利用 KAN 的自适应性提高特征的时空信息表征能力。

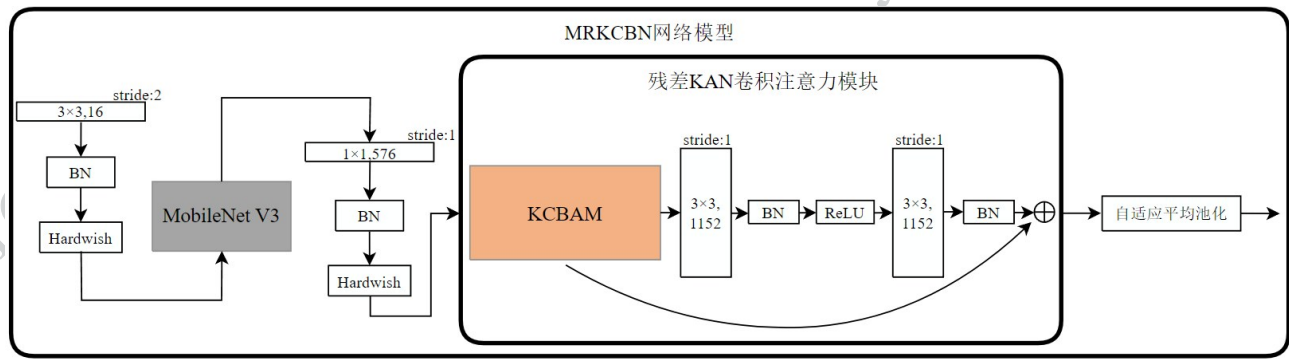


图5 MRKCBN网络模型结构图

Fig. 5 MRKCBN (mobile residual KAN CBAM block net)

KAN是一种基于Kolmogorov-Arnold定理的神经网络结构。Kolmogorov-Arnold定理提出如果是一个有界域上的多连续函数,那么可以表示为单个变量的连续函数和加法运算的有限合成,其数学表达式为:

$$f(\mathbf{x}) = f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \Phi_{q,p}(x_p) \right) \quad (1)$$

式中, Φ_q 和 $\Phi_{q,p}$ 都是单变量连续函数, \mathbf{x} 是输入变量。

如图6(b)所示,KAN通道注意力模块(KAN channel attention module, KCAM)使用KANLinear模块优化通道注意力机制中原始的Linear模块,KAN-Linear通过引入B样条 Φ 函数(B-splines)来近似传统的激活函数,提供了一种自适应的线性层(Liu等, 2024)。它能够根据输入数据动态调整其行为,并且通过正则化损失来控制模型的复杂度。KCAM模型的核心运算表示如下:

$$M_c(F_1) = \sigma(\text{KAN}(\text{AvgPool}(F_1) + \text{KAN}(\text{MaxPool}(F_1)))) \quad (2)$$

式中, F_1 表示KCAM的输入序列, AvgPool表示平均池化操作, MaxPool表示最大池化操作, σ 表示激活函数, $M_c(F_1)$ 表示KCAM的输出序列。KCAM模型使用KAN_Conv2d替换掉空间注意力机制CBAM中的卷积模块,KAN_Conv2d通过B样条来近似传统的激活函数,通过分组卷积、归一化和dropout等特性,提供了一种更加灵活和强大的卷积层实现。这些特性使得KCAM能够更好地适应数据的特征,提高模型的表达能力和泛化性能。

类似的,在空间注意力模块中引入KAN,得到KAN空间注意力模块(KAN spatial attention module, KSAM),如图6(c)所示。KSAM的核心运算表示如下:

$$M_s(F_2) = \sigma(\text{KANConv}([\text{AvgPool}(F_2); \text{MaxPool}(F_2)])) \quad (3)$$

式中, F_2 表示KSAM的输入序列, KANConv表示卷积操作, $M_s(F_2)$ 表示KSAM的输出序列。

为了缓解梯度消失和模型收敛困难的问题,本
© 中国图象图形学报版权所有

表 1 MRKCBN 模块具体配置

Table 1 MRKCBN module configuration

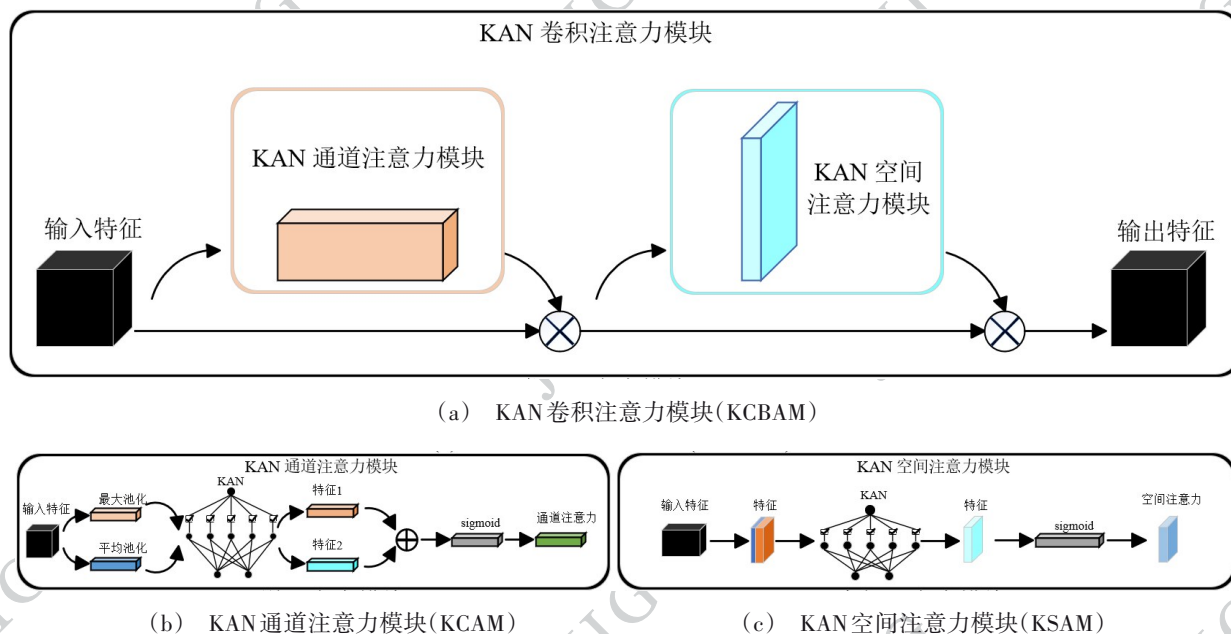
层	输入形状	输出形状	内核
Conv2d	3×224×224	16×112×112	3
BatchNorm2d	16×112×112	16×112×112	
Hardswish	16×112×112	16×112×112	
MobileNetV3	16×112×112	96×7×7	
Conv2d	96×7×7	576×7×7	1
BatchNorm2d	576×7×7	576×7×7	
Hardswish	576×7×7	576×7×7	
RKCB	576×7×7	1152×7×7	

文在 KCBAM 中加入残差模块,包括两个 3×3 卷积层、两个归一化层和一个 ReLU 非线性变换层。最后对特征图进行自适应平均池化操作,通过减少特

征图的空间维度(如高度和宽度)来提取更有意义的特征,同时减小计算量,提升模型的性能。与传统的最大池化和平均池化不同,自适应平均池化可以根据给定的目标输出尺寸动态调整池化操作,使输出尺寸与目标匹配。KCBAM 的具体配置如表 1 所示。

2.3 APSB 模块

在特征融合阶段,为了强化模型对面部关键区域的关注度,本文对自注意力机制(self-Attention)进行改进,提出了一种注意力下采样自注意力特征融合模块(attention pooling self-attention block, APSB)。不同于传统的自注意力机制仅处理单一特征序列,APSB 能够在多个光流特征中学习融合策略,其结构图如图 7 所示。APSB 接收三个光流特征: F_{TVLI} 、 F_{DIS} 和 F_{PCA} 。这三组特征通过通道维度拼接组成新的特征张量 F_{concat} 。具体而言:



((a)KAN convolutional block attention module; (b)KAN channel attention module; (c)KAN spatial attention module)

图 6 KAN 卷积注意力模块(KCBAM)

Fig. 6 KAN convolutional attention module

APSB 首先按照公式(4)得到每个向量 k_i 的注意力分数:

(4)

式中, k_i 是键值 K 的第 i 个向量, W 和 b 分别表示线性层的权重和偏置。接着对序列进行 softmax 操作得到归一化后的注意力分数 α_i , 表示在计算输出结果

时第 i 个元素对整体的贡献权重。然后,使用注意力分数对输入序列进行加权聚合,并通过线性层进行生成的最终输出,计算公式如(5)所示:

$$K' = \text{Attention} F \quad (5)$$

式中, AttentionPooling 表示对于键值 K 使用上述注意力下采样,得到聚合后的 K' , 并使用矩阵乘法计

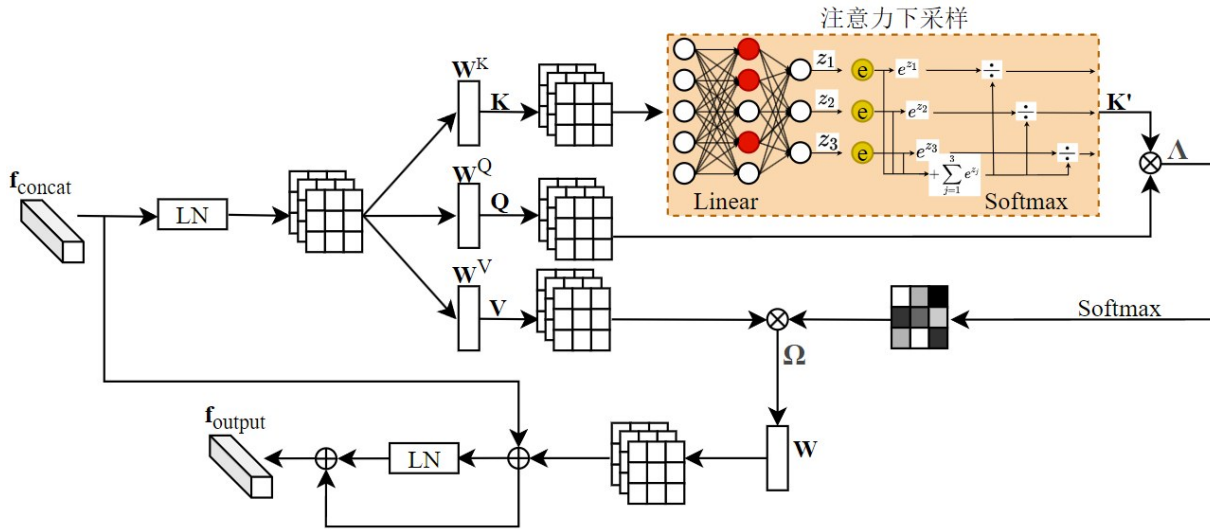


图7 APSB模块示意图

Fig. 7 APSB(attention pooling self-attention block)

算 $Q \cdot K'$:

$$\Lambda = \frac{Q \cdot K'}{\sqrt{d}} + b \quad (6)$$

式中, d 表示注意力头的维度大小, 是用于稳定数值计算的缩放因子。 b 为偏置项, 帮助模型更好的适应数据的分布。最后使用归一化的注意力分数 Λ 与键值 V 进行加权求和, 得到的得到注意力矩阵 Ω , 计算公式如(7)所示:

$$\Omega(Q, K, V) = \text{Softmax}(\Lambda) \cdot V \quad (7)$$

将经过注意力加权后的特征进行残差连接, 再经过 Layer Normalization 归一化, 得到最终输出特征 F_{output} 。

整个特征融合过程在多个注意力头上并行进行, 能够捕捉不同维度和语义层次的特征关系。通过嵌入 Attention Pooling 层, APSB 能够学习面部不同区域的权重。通过提高关键区域的权重提升模型识别的性能和鲁棒性。本文所提出的 MOFFFN 算法如表 2 所示。

3 本文实验

3.1 实验数据集

为了验证所提方法的有效性, 本文分别和 LBP-TOP(Zhao 等, 2007), 基于顶点帧的视频微表情识别 (micro-expression recognition from video using apex frame, Bi-WOOF)(Liong 等, 2018), 用于微表情识别的胶囊网络 (CapsuleNet for micro-expression recogni-

tion, CapsuleNet)(Van 等, 2019), STSTNet (Liong 等, 2019), OFF-ApexNet(Gan 等, 2019), 神经微表情识别器 (neural micro-expression recognizer, EMR)(Liu 等, 2019), 循环卷积网络 (recurrent convolutional network, RCN)(Xia 等, 2020), 特征细化模型 (feature refinement, FeatRef)(Zhou 等, 2022), 结合光流和动态图像卷积神经网络双流微表情识别新算法 (dual-stream combining optical flow and dynamic image convolutional neural networks, FDCN)(Tang 等, 2023) 和注意力引导的三流卷积神经网络 (attention-guided three-stream convolutional neural network, ATSCNN)(Zhao 等, 2024) 在 CASME II(Yan 等, 2014), SAMM(Davison 等, 2016) 和 SMIC-HS(Davison 等, 2023) 三个主流公开数据集上进行对比。CASME II 数据集由中国科学院心理研究所发布, 包含 35 名中国受试者在观看情感刺激视频时的微表情视频。该数据集包括五类情感: 愤怒、厌恶、恐惧、快乐和悲伤。SAMM 数据集由英国曼彻斯特大学发布。该数据集包含 32 名来自不同种族和性别的受试者在观看情感刺激视频时的微表情视频, 微表情涵盖了包括愤怒、厌恶、恐惧、快乐、悲伤、惊讶和其它类别。SMIC-HS 由芬兰奥卢大学发布, 包含三种主要情感: 正面、负面和惊讶。按照 (Liong 等, 2019) 中的设置, 本文将情绪类别统一设置为 3 类, 即“正面”情绪类别包括“快乐”情绪类别, “负面”情绪类别包括“悲伤”、“厌恶”、“蔑视”、“恐惧”和“愤怒”情绪类别, “惊喜”情绪类别仅包括“惊喜”类别。此外, 本文还将三

表 2 MOFFN 算法描述

Table 2 The proposed algorithm of MOFFN	
Algorithm 1: MOFFN	
Input: 微表情图像序列 (I_{start}, I_{apex})	
Output: 微表情类别 y_{pred}	
//光流特征融合模块 OFFM	
步骤 1:	$O_{TV} \leftarrow \text{OpticalFlow}_{TVLI}(I_{start}, I_{apex})$
步骤 2:	$O_{DIS} \leftarrow \text{OpticalFlow}_{DIS}(I_{start}, I_{apex})$
步骤 3:	$O_{PCA} \leftarrow \text{OpticalFlow}_{PCA}(I_{start}, I_{apex})$
步骤 4:	$O_{fused} \leftarrow 2 * O_{TVLI} + 4 * O_{DIS} + 4 * O_{PCA}$
//特征提取模块 MRKCBN	
步骤 5:	$F \leftarrow \text{MobileNetV3}(O_{fused})$
步骤 6:	$F \leftarrow \text{KCBAM}(F)$
步骤 7:	$F \leftarrow \text{Residual}(F)$
步骤 8:	$F \leftarrow \text{AdaptiveAvgPool}(F)$
步骤 9:	$f_{concat} \leftarrow \text{Concat}(F)$
//特征融合模块 APSB	
步骤 10:	$F \leftarrow \text{LayerNorm}(f_{concat})$
步骤 11:	$Q \leftarrow F \cdot W^Q$
步骤 12:	$K \leftarrow F \cdot W^K$
步骤 13:	$V \leftarrow F \cdot W^V$
步骤 14:	$K' \leftarrow \text{AttentionPooling}(K)$
步骤 15:	$\Lambda \leftarrow \text{Softmax}((Q \cdot K') / \text{sqrt}(d))$
步骤 16:	$Z \leftarrow \Lambda \cdot V + f_{concat}$
步骤 17:	$f_{output} \leftarrow \text{LayerNorm}(Z) + Z$
//分类与输出	
步骤 18:	$f_{output} \leftarrow \text{Linear}(f_{output})$
步骤 19:	$y_{pred} \leftarrow \text{Softmax}(\text{Classifier}(f_{output}))$
步骤 20:	return y_{pred}

注:加粗字体为矩阵。

个数据集整合成一个复合数据集(CD)进行验证。图 8 展示了三个数据集部分样本的起始帧和峰值帧图像,可以看出三个数据集的样本差异显著: CASME II 数据集的受试者均为中国人、背景干净, SAMM 数据集的受试者跨年龄段大、光照变化明显、人种差异大,而 SMIC-HS 数据集的样本则分辨率低、背景更复杂。四个数据集的详细信息如表 3 所示。

3.2 实验设置

实验采用微表情识别领域主流的留一交叉验证

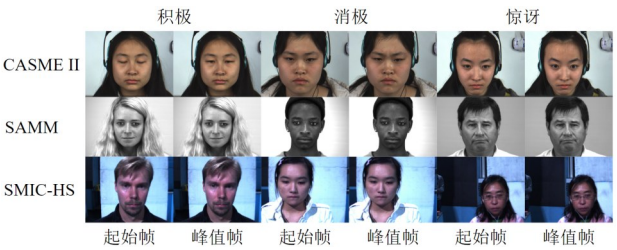


图 8 微表情数据集部分样本

Fig. 8 Some samples of the four datasets

表 3 数据集统计信息

Table 3 Datasets statistics				
分类	CASME II	SAMM	SMIC-HS	CD
消极	88	92	70	250
积极	32	26	51	109
惊讶	25	15	43	83
总和	145	133	163	442

法 (leave-one-subject-out-cross-validation, LOSO CV)。该方法分别将每个受试者作为验证集,所有剩余的样本作为训练集。模型初始学习率(learning_rate)设置为 0.0001,批大小(batch_size)设置为 128,训练轮次(epoch)设置为 300 轮,训练过程需要 15 小时左右。实验环境为 Windows 11 操作系统、PyTorch2.3.0、Intel Core i7-12700、64 GB 内存和 NVIDIA GeForce RTX 4080。

实验同样使用主流的 UF1 和 UAR 作为衡量指标。UF1 也称为宏观平均 F1 分数,是一种通常用于评估具有不平衡类分布的多类分类任务的性能的指标。其公式如下:

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

$$F1 = \frac{2 * \sum_{c=1}^C \text{Precision}_c * \text{Recall}_c}{\sum_{c=1}^C (\text{Precision}_c + \text{Recall}_c)} \quad (9)$$

式中, FP_c 为第 c 类假阳性, FN_c 为第 c 类假阴性, TP_c 为第 c 类真阳性, C 为类别数。

UAR 是一种在存在不平衡类比率的情况下评估模型有效性时特别有用的指标。其公式如下:

$$UAR = \frac{\sum_{c=1}^C \text{Recall}_c}{\sum_{c=1}^C n_c} \quad (10)$$

式中, n_c 为每个类别的样本总数。

3.3 实验结果

表 4 列出了本文方法和对比方法的实验结果。
© 中国图象图形学报版权所有

可以看出,本文方法在四个数据集上都取得了最高的结果。具体而言,在 CASMEII 数据集上,本方法的 UF1 和 UAR 分别达到了 91.79% 和 92.95%,比第二好的 FeatRef 分别提升 2.64% 和 4.22%,比 STST-Net 分别提升了 7.97% 和 6.09%。本文方法在 CASME II 上取得的结果体现了本文所提模型优异的特征提取能力。在 SAMP 和 SMIC-HS 数据集上,性能较好的是 EMR 和 FeatRef, UF1 和 UAR 都超过了 70%。本方法分别提升 8.15% 和 17.55%,由于 SAMP 和 SMIC-HS 数据集受拍摄环境、受试者面部

差异等问题,大多方法的性能出现大幅下降。但本文方法的性能下降幅度较少,体现了本文方法面对数据集图像本身干扰问题更具鲁棒性。在 SMIC-HS 数据集上,本方法的性能较次优方法 EMR 提升了 11.95%UF1 和 16.48%UAR,证明本文提出的光流融合模块相较于其他方法能够提取更丰富的光流特征,提高模型性能。在复合数据集 CD 上,本方法的 UF1 和 UAR 分别较次优方法提升了 6.18% 和 9.31%。说明本方法能够在多种族多年龄段的复合微表情识别任务中具有更强的鲁棒性。

表 4 实验结果对比
Table 4 Performance of different methods on the four datasets

Method	CASME II		SAMP		SMIC		CD	
	UF1(%) (↑)	UAR(%) (↑)	UF1(%) (↑)	UAR(%) (↑)	UF1(%) (↑)	UAR(%) (↑)	UF1(%) (↑)	UAR(%) (↑)
LBP-TOP	70.26	74.29	39.54	41.02	20.00	52.80	58.82	57.85
Bi-WOOF	78.05	80.26	52.11	51.39	57.27	58.29	62.96	62.27
CapsuleNet	70.68	70.18	62.09	59.89	58.20	58.77	65.20	65.06
ATSCNN	--	--	--	--	--	--	73.51	72.05
FDCN	73.09	72.00	58.07	57.00	--	--	--	--
STSTNet	83.82	86.86	65.88	68.10	68.01	70.13	73.53	76.05
OFFApexNet	87.64	86.81	54.09	53.92	68.17	66.95	71.96	70.96
EMR	82.93	82.09	<u>77.54</u>	<u>71.52</u>	<u>74.61</u>	<u>75.30</u>	<u>78.85</u>	78.24
RCN	85.12	81.23	76.01	67.15	63.26	64.41	74.32	71.90
FeatRef	<u>89.15</u>	<u>88.73</u>	73.72	<u>71.55</u>	<u>70.11</u>	70.83	78.38	<u>78.32</u>
MOFFFN (ours)	91.79	92.95	85.69	89.10	86.56	91.78	85.03	87.63

注:加粗字体为最优结果,下划线字体为次优结果,倾斜字体为第三结果。↑表示数值越高越优。

图 9 展示了本方法在四个数据集上的混淆矩阵。图中的“0,1,2”分别表示积极、消极和惊讶三类微表情类别。可以看出,本文所提出的方法在 CASME II 数据集上取得最好的性能,对于积极、消极和惊讶都有着较高的准确率,其中,识别消极类别的准确率到达了 97%。然而,本文方法在识别 SAMP 和 SMIC-HS 数据集上的“积极”和“惊讶”类别时仍存在问题。在 SAMP 数据集上,识别“积极”类别准确率最低仅有 56%。在 SMIC-HS 数据集上识别“惊讶”类别时得到了最低的准确率 70%。从混合数据集的混淆矩阵可以看出,本方法能够较好的处理三个类别,“惊讶”情感的正确率最低,“消极”情感的正确率最高,达到了 85%。不论“惊讶”还是“积极”,

被错误识别为“消极”的比例均超过 20%,说明通过整合数据集虽然在整体上减小了类别不均带来的影响,但“消极”类别对于识别其他两类的结果依旧有着不小的干扰。

导致这种不一致性能有三个主要原因:首先,由于 SAMP 数据集是从具有更多种族和更广泛的年龄属性分布的不同参与者收集的,这导致模型识别更具挑战性。其次,在 SAMP 数据集上,由于 SAMP 存在大量的消极样本,导致类别不平衡的问题。因此微表情类别为积极时,识别准确率不高。最后,与其他两个数据集相比,SMIC-HS 数据集以更低的帧速率和人脸分辨率捕获,丢失了一些关键的表达信息。

表5列出了本文所提出网络模型和对比方法的每秒浮点运算次数(Floating Point Operations Per Second, FLOPs)和总参数量两个指标。与STSTNet相比,在总参数量和FLOPs相近的情况下,本文方法的UF1和UAR评估指标分别提高了11.5%和11.58%。与Dual-Inception(Zhou等,2019)相比,本文方法不仅总参数量和FLOPs各低1个数量级,UF1和UAR均有较大提升。与ATSCNN相比,尽管本文方法的总参数量和FLOPs较大,但UF1和UAR评估指标提升显著。

3.4 消融实验

消融实验结果如表6所示。第一行表示基础模型MobileNet_v3在不提取多元光流特征、不进行特征融合情况下取得的结果。加入光流融合模块后,

表5 模型复杂度对比

Table 5 Model complexity comparison

方法	UF1(%)	UAR (%)	FLOPs/M (↓)	总参数量 (↓)
STSTNet	73.53	76.05	57.45	4.98×10^7
Dual-Inception	73.22	72.78	264.15	1.34×10^8
ATSCNN	73.51	72.05	19.01	4.21×10^6
MOFFFN(ours)	85.03	87.63	50.58	5.06 10⁷

注:加粗字体为本文方法结果。↓表示数值越低越优。

UF1和UAR分别大幅提升了17.7%和13.69%,表明光流融合算法相较于原始的微表情数据集,多光流方法能够提取更加丰富的面部运动特征,可以提高模型的识别性能。继续加入MRKCBN模块后,

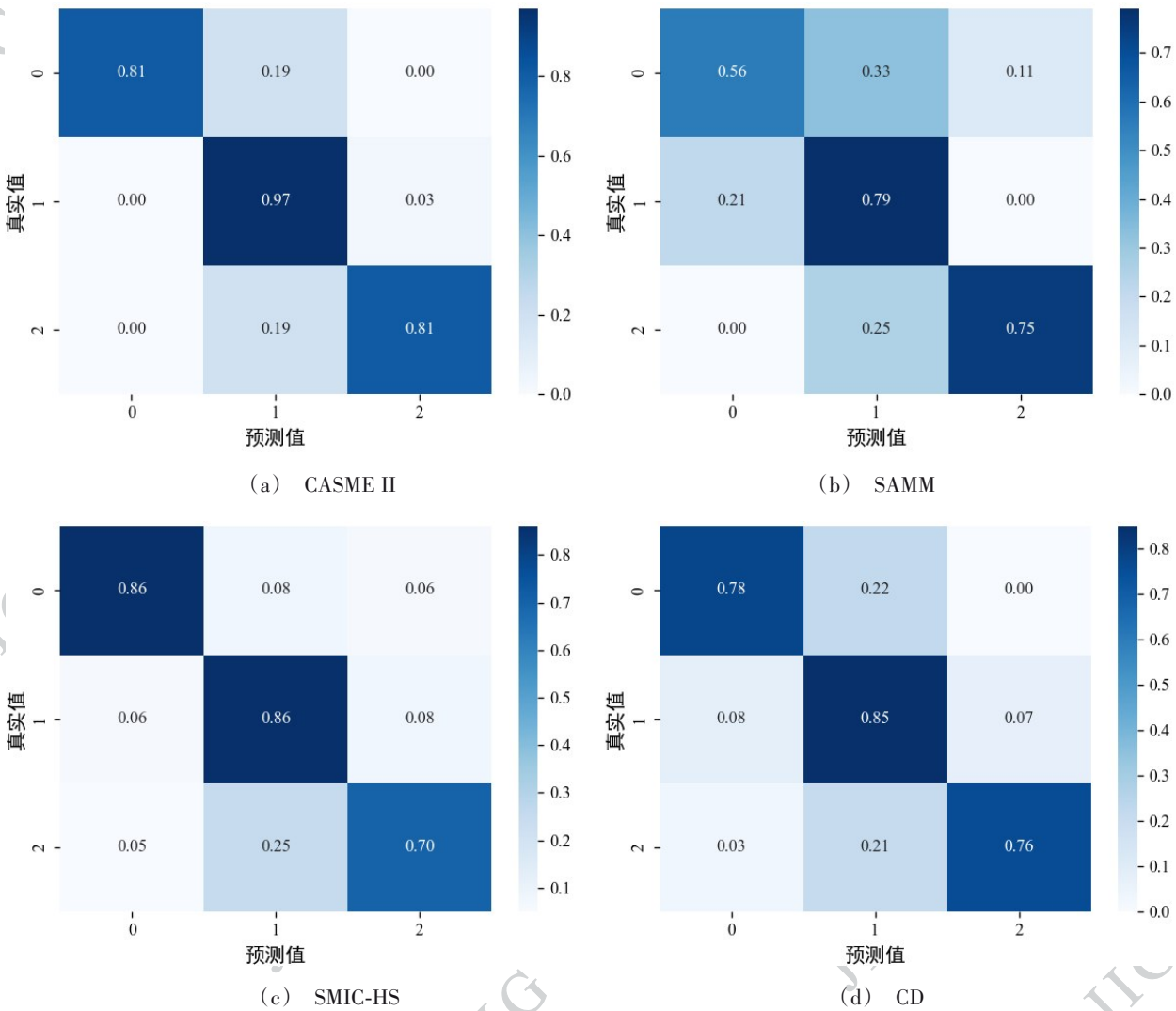


图9 混淆矩阵

Fig. 9 Confusion matrix

UF1 和 UAR 进一步提升了 2.09% 和 1.4%, 说明 KAN 能够凭借其独特的自适应性提高特征的时空信息表征能力。嵌入 APSB 模块后, 方法的 UF1 和 UAR 分别达到 91.79% 和 92.95%, 进一步提升了 2.06% 和 4.74%。

表 6 消融实验
Table 6 Ablation experiments

光流融合模块	MRKCBN 模块	APSB 模块	UF1(%) (↑)	UAR(%) (↑)
			69.94	73.12
√			87.64	86.81
√	√		89.73	88.21
√	√	√	91.79	92.95

注:加粗字体为最优结果。↑表示数值越高越优。

图 10 对本方法的部分中间结果进行可视化。前两列分别表示微表情样本的起始帧和峰值帧, 中间两列表示 TVL1 光流和本文提出的融合光流图像, 最后两列表示未使用 KCBAM 模块和使用 KCBAM 模块的注意力可视化图像。可以看出, 相较于单一的 TVL1 光流, 多光流融合后的图像包含更多的面部运动信息。未使用 KCBAM 时, 模型的注意力关注区域大而散, 并没有集中到微表情发生的关键部位。使用 KCBAM 后, 模型的注意力更加集中在面部运动区域。

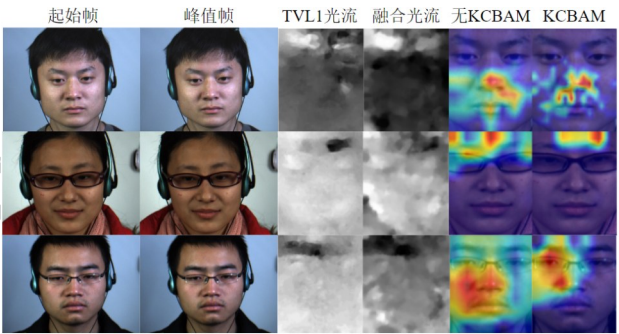


图 10 注意力可视化对比图像
Fig. 10 Comparison images of attention visualization

4 结 论

本文提出一种融合多光流与KAN的微表情识别方法(MOFFN)。MOFFN 主要包括三个模块, 分

别是多光流融合模块(OFFM)、基于残差KAN的特征提取模块(MRKCBN)和注意力下采样自注意力特征融合模块(APSB)。通过融合三种光流信息, OFFM 能够捕获面部肌肉变化的多维度信息; 借助 KAN 和 MobelNetv3, MRKCBN 能够在微表情数据较少的情况下提取更鲁棒和具有鉴别能力的特征; APSB 利用多头注意力机制学习特征权重, 进一步提升模型性能。本文所提的 MOFFN 在四个主流微表情识别数据集上进行了验证, 均取得了最好的性能。

然而, 本方法仍存在以下局限性: 首先, 由于融合了多种光流信息, 模型的时间复杂度较高, 制约着本方法在实际场景中的应用。其次, 微表情数据集较少且存在显著的类别不平衡问题, 虽然数据增广和类别整合能够一定程度上缓解上述问题, 但是本方法对“积极”和“惊讶”的误识率仍然较高。在后续的研究中, 我们将尝试构建更高效的微表情识别模型, 同时融入多模态特征, 将微表情识别应用到实际场景中, 包括谎言检测、情感识别和心理健康评估。

致谢: 本课题得到国家自然科学基金(No.61976118, 61806098), 高维信息智能感知与系统教育部重点实验室开放基金(No.2023-3, 2023-10)以及人工智能教育部重点实验室开放基金(No. AI202404)的资助。

参考文献(References)

Chang H Y, Zhang F L, Ma S, Gao G W, Zheng H and Chen Y. 2021. Unsupervised domain adaptation based on cluster matching and Fisher criterion for image classification. Computers & Electrical Engineering, 91[DOI: 10.1016/j.compeleceng.2021.107041]

Chang H Y, Gao G W, Chen Y, and Zheng H. 2024. Multi-task Contexture Learning Network for automated vertebrae segmentation and tumor diagnosis from MRI. Computers and Electrical Engineering, 113[DOI: 10.1016/j.compeleceng.2023.109032]

Du H Y, Zhang P, Lin Q, Li X T, Xu S and Ben X Y. 2025. Micro-expression recognition algorithm based on motion feature selection and fusion with visual Transformer. Signal processing, 41(2): 267-278 (杜含月, 张鹏, 林强, 李晓桐, 徐森和贾晔烨. 2025. 基于视觉Transformer的运动特征选择融合微表情识别算法. 信号处理, 41(2): 267-278)

Davison A K, Lansley C, Costen N, Tan K and Yap M H. 2016. Samm: A spontaneous micro-facial movement dataset [J]. IEEE transactions on affective computing, 9(1): 116-129. [DOI: 10.1109/TAFFC.2016.2573832]

Davison A K, Li J, Yap M H, See J, Cheng M H, Li X, Hong X and

- Wang S J. 2023. MEGC2023: ACM Multimedia 2023 ME Grand Challenge [C]//Proceedings of the 31st ACM International Conference on Multimedia. New York, NY, USA: ACM: 9625-9629. [DOI:10.1145/3581783.3612833]
- Gan Y S, Liong S T, Yau W C, Huang Y C and Tan L K. 2019. OFF-ApexNet on micro-expression recognition system. *Signal Processing: Image Communication*, 74: 129-139 [DOI: 10.1016/j.image.2019.02.005]
- Huang X, Wang S J, Zhao G and Piteikainen M. 2015. Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection// 2015 IEEE International Conference on Computer Vision Workshop, Santiago, Chile: ICCVW: 1-9. [DOI: 10.1109/ICCVW.2015.10]
- Kroeger T, Timofte R, Dai D and Van G L. 2016. Fast optical flow using dense inverse search//Proceedings of the 14th European Conference on Computer Vision (ECCV 2016). Amsterdam, The Netherlands: Springer: 471-488 [DOI:10.1007/978-3-319-46493-0_29]
- Kim D H, Baddar W J, and Ro Y M. 2016. Micro-Expression Recognition with Expression-State Constrained Spatio-Temporal Feature Representations//Proceedings of the 24th ACM International Conference on Multimedia. New York, USA: ACM: 382-386 [DOI:10.1145/2964284.2967247]
- Li B K, Wu C Z, Xiang B Y, Zang H J, Ren Y S and Zhan S. 2024. Micro-expression classification algorithm guided by peak frame localization. *Journal of Image and Graphics*, 29 (5): 1447-1459 (李博凯, 吴从中, 项柏杨, 臧怀娟, 任永生, 詹曙. 2024. 微表情峰值帧定位引导的分类算法. *中国图象图形学报*, 29(5): 1447-1459)
- Liu Y J, Zhang J K, Yan W J, Wang S J, Zhao G and Fu X. 2016. A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Transactions on Affective Computing*, 7(4): 299 - 310 [DOI:10.1109/TAFFC.2015.2485205]
- Li G H, Yuan Y F, Ben X Y, and Zhang J P. 2020. Facial micro-expression recognition using spatial-temporal attention mechanism. *Journal of Image and Graphics*, 25(11): 2380-2390 (李国豪, 袁一帆, 袁晔, 张军平. 2020. 采用时空注意力机制的人脸微表情识别. *中国图象图形学报*, 25(11): 2380-2390)
- Liu Y, Du H, Zheng L and Gedeon T. 2019. A neural micro-expression recognizer//Proceedings of the 14th IEEE International Conference on Automatic Face & Gesture Recognition. Lille, France: IEEE: 1-4 [DOI:10.1109/FG.2019.8756583]
- Liong S T, See J, Phan R C W, Le Ngo A C, Oh Y H and Wong K. 2015. Subtle expression recognition using optical strain weighted features//Computer Vision- ACCV 2014 Workshops: Singapore, Singapore, Springer: 47-61 [DOI:10.1007/978-3-319-16631-5_47]
- Liong S T, Gan Y S, See J, Khor H Q and Huang Y C. 2019. Shallow triple stream three-dimensional CNN (STSTNet) for micro-expression recognition// Proceedings of the 14th IEEE International Conference on Automatic Face & Gesture Recognition. Lille, France: IEEE: 1-5 [DOI:10.1109/FG.2019.8756567]
- Liong G B, Liong S T, Chan C S and See J. 2024. SFAMNet: A scene flow attention-based micro-expression network. *Neurocomputing*, 566 [DOI:10.1016/j.neucom.2023.126998]
- Liu Z. 2024. Kan: Kolmogorov-arnold networks. *Machine-learning*, 1. [DOI:10.48550/arXiv.2404.19756]
- Liong S T, See J, Wong K S, and Phan R C W. 2018. Less is more: Micro-expression recognition from video using apex frame [J]. *Signal Processing: Image Communication*, 62: 82-92. [DOI: 10.1016/j.image.2017.11.006]
- Liu Y, Du H, Zheng L, and Gedeon T. 2019. A neural micro-expression recognizer [C]//2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019). Lille, France: IEEE: 2019:1-4. [DOI:10.1109/FG.2019.8756583]
- Nguyen X B, Duong C N, Li X, Gauch S, Seo H S and Luu K. 2023. Micron-BERT: BERT-Based Facial Micro-Expression Recognition// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, CVPR: 1482-1492. [DOI:10.1109/CVPR52729.2023.00149]
- Peng M, Wang C, Chen T, Liu G and Fu X. 2017. Dual temporal scale convolutional neural network for micro-expression recognition. *Frontiers in Psychology*, 8 [DOI:10.3389/fpsyg.2017.01745]
- Tang J, Li L, Tang M and Xie J. 2023. A novel micro-expression recognition algorithm using dual-stream combining optical flow and dynamic image convolutional neural networks [J]. *Signal, Image and Video Processing*, 17 (3): 769-776. [DOI: 10.1007/s11760-022-02286-0]
- Van Quang N, Chun J and Tokuyama T. 2019. CapsuleNet for micro-expression recognition [C]//2019 14th IEEE International Conference on Automatic Face & Gesture Recognition. Lille, France: IEEE: 1-7. [DOI: 10.1109/FG.2019.8756544]
- Wulff J and Black M J. 2015. Efficient sparse-to-dense optical flow estimation using a learned basis and layers//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 120-130 [DOI:10.1109/CVPR.2015.7298607]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. Cbam: Convolutional block attention module // Proceedings of the European Conference on Computer Vision, Berlin, Heidelberg: ECCV: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Xia Z, Peng W, Feng X and Zhao G. 2020. Revealing the invisible with model and data shrinking for composite-database micro-expression recognition [J]. *IEEE Transactions on Image Processing*, 2020, 29: 8590-8605. [DOI: 10.1109/TIP.2020.3018222]
- Yan W J, Li X, Wang S J, Zhao G, Liu Y J, Chen Y H and Fu X. 2014. CASME II: An improved spontaneous micro-expression database and the baseline evaluation [J]. *PloS one*, 2014, 9(1): e86041. [DOI:10.1371/journal.pone.0086041]
- Zheng H, Geng X, Tao D C and Jin Z. 2016. A multi-task model for simultaneous face identification and facial expression recognition.

- Neurocomputing, 171: 515 – 523 [DOI: 10.1016/j.neucom.2015.06.079]
- Zhang J, Zhang H, Bo L, Li H R, Xu S and Yuan D. 2020. Subspace transform induced robust similarity measure for facial images. Frontiers of Information Technology & Electronic Engineering, 21(9): 1334 – 1345 [DOI: 10.1631/FITEE.1900552]
- Zhao G and Pietikainen M. 2007. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6): 915 – 928 [DOI: 10.1109/TPAMI.2007.1110]
- Zhou L, Mao Q R, Huang X H, Zhang F F and Zhang Z H. 2022. Feature refinement: An expression-specific feature learning and fusion method for micro-expression recognition. Pattern Recognition, 122 [DOI: 10.1016/j.patcog.2021.108275]
- Zhao M H, Dong S S, Hu J, Du S L, Shi C, Li P and Shi Z H. 2024. Attention-guided triple-stream convolutional neural network for micro-expression recognition. Journal of Image and Graphics. 29(01): 111-122. (赵明华, 董爽爽, 胡静, 都双丽, 石程, 李鹏, 石争浩. 2024. 注意力引导的三流卷积神经网络用于微表情识别. 中国图象图形学报, 29(01): 111-122).
- Zach C, Pock T and Bischof H. 2007. A duality based approach for real-time TV-L 1 optical flow// Proceedings of the 29th DAGM Conference on Pattern Recognition. Berlin, Heidelberg: Springer: 214-223 [DOI: 10.1007/978-3-540-74936-3_22]
- Zhou L, Mao Q, Huang X, Zhang F and Zhang Z. 2022. Feature refinement: An expression-specific feature learning and fusion method for micro-expression recognition [J]. Pattern Recognition, 122: 108275. [DOI: 10.1016/j.patcog.2021.108275]
- Zhou L, Mao Q and Xue L. 2019. Dual-inception network for cross-database micro-expression recognition [C]//2019 14th IEEE international conference on automatic face & gesture recognition .Lille, France: IEEE: 1-5. [DOI: 10.1109 /FG.2019.8756579]

作者简介

常合友, 1989年生, 男, 副教授, 研究方向为: 特征提取、图像识别和医学影像分割。Email: hychang@njxzc.edu.cn.

杨佳铮, 2001年生, 男, 硕士研究生, 研究方向为: 微表情识别。Email: 1435213094@qq.com.

高广谓, 1986年生, 男, 研究员, CCF 高级会员, 研究方向为: 计算机视觉和模式识别, 包括图像处理、图像和视频分析、鲁棒模型拟合等。Email: csgwgao@njupt.edu.cn.

张键, 1978年生, 男, 教授, 研究方向为: 模式识别与智能系统、机器学习、人工智能及其应用、数据挖掘与大数据分析。Email: zhangjian@jou.edu.cn.

郑豪, 1976年生, 男, 教授, 研究方向为: 人工智能、计算机视觉、图像处理、模式识别、大数据。Email: zhh710@163.com.