双流网络信息交互机制下的微表情识别

朱伟杰, 陈莹*

(江南大学轻工过程先进控制教育部重点实验室 无锡 214122)
(chenying@jiangnan.edu.cn)

摘要: 针对深度学习的方法用于微表情识别时微表情识别的实验数据库非常稀缺, 导致神经网络在学习的过程中知识获取有限而难以提高精度及泛化能力的问题, 提出基于双流网络信息交互的微表情识别方法. 通过改进的深度互学习策略引导图像序列不同模态之间的交互训练, 提高网络的识别率. 方法基于 RGB 图像序列建立主体网络, 基于光流建立辅助网络; 在训练阶段, 通过设计互学习损失中的有监督学习损失和拟态损失, 优化训练过程, 使得每一种模态都能学习正确地预测训练样本的真实标识, 同时能与其他模态的预测相匹配; 在测试阶段, 由于互学习机制增强了 RGB 分支的判别能力, 因此可对光流分支进行剪裁, 在保证精度的前提下提高识别速度. 在 CASME, CASME II 和 SMIC 数据库上的实验结果表明, 该方法有效地提高了识别精度, 整体性能优于已有方法.

关键词: 微表情; 双流网络; 信息交互; 互学习机制

中图分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2021.18456

Micro-Expression Recognition Based on Dual-Stream Networks Information Interaction

Zhu Weijie and Chen Ying*

(Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122)

Abstract: Scarcity of experimental databases for deep learning based micro-expression recognition leads to limited knowledge acquisition in learning process and increasing difficulty for improving accuracy and generalization capabilities. In view of the problem, a micro-expression algorithm based on dual-stream network information interaction is proposed. An improved deep mutual learning strategy is designed to guide the interactive training between different modalities of the image sequence to improve the recognition rate of the network, in which the main network is built based on the RGB image sequence, and the auxiliary network is built based on the optical flow. In the training phase, the training process is optimized via a mutual learning loss which includes the supervised learning loss and mimic loss, which ensures the prediction of each mode is consistent with the true identity of the training sample and the predictions of other modalities. In the testing phase, since the mutual learning mechanism enhances the discrimination ability of RGB branches, the optical flow branches can be tailored to improve the recognition speed while ensuring accuracy. The experimental results on the CASME, CASME II and SMIC data sets show that the algorithm effectively improves the recognition accuracy and the overall performance is better than existing algorithms.

Key words: micro-expression; dual-stream networks; information interaction; mutual learning mechanism

通常观察一个人的状态时,首先想到的是观察人脸的表情,同时也会根据对方的表情进行交流.在商务谈判、刑事审讯和心理治疗等场合中,往往人们会有意无意地隐藏自己的真实情绪,以此来达到他们的目的.微表情是人类的肌肉语言,在此类情况下,使用微表情识别的结果作为判断依据,识别的成功率则会大大提升.相关研究表明^[1-2],微表情相比常见的宏表情有持续时间短和肌肉运动幅度小的显著特点.

计算机视觉技术在微表情识别中的主要方法分成传统方法和深度学习方法 2 大类.通过对近年来微表情识别方法研究发现,目前用于微表情识别仍以传统方法居多.微表情识别传统方法主要有局部二值模式(local binary pattern, LBP)^[3]及其改良的方法、基于光流场特征^[4]及其改良的方法,和基于张量变化分析^[5]及其改良的方法 3 类.2014 年, Wang 等^[5]提出基于判别式张量子空间分析(deterministic tensor subspace analysis, DTSA)的方法,通过对判别式张量的子空间进行最优的特征表示学习,实现微表情的识别分类.2016 年, Huang 等^[6]在时间-空间域特征(local binary patterns from three orthogonal planes, LBP-TOP)的基础上,提出一种新时空域特征(spatial temporal completed local quantized patterns, STCLQP),通过提取帧序列指示、幅值和方向 3 种特征进行微表情识别,相比 LBP-TOP 方法,在主流数据库实验效果都有所提升.2016 年, Liu 等^[7]提出主方向平均光流(main directional mean optical-flow, MDMO)特征,将一种鲁棒的光流方法应用到微表情视频序列中,并将面部基于面部运动单元(action units, AUs)的兴趣区域(region of interest, ROI)划分,更加精确地对各区域进行处理;同年,张轩阁等^[8]提出分区的特征统计方法,将光流图像划分为若干矩形区域,在局部区域内归纳各点光流运动情况,增强特征的有效性.

传统方法在为微表情识别打下了很好的基础同时也做出了很大的贡献,但是仍然有很大的提升空间:(1) 识别过程中计算量庞大、耗时长并且泛化能力差;(2) 过多地依赖于手工特征,使得研究者必须掌握很多心理学的知识,对于非专业人员使用此类方法非常困难;(3) 目前微表情的识别率离能够真正的应用距离还很远.因此研究人员迫切地需要探索能够提取更高层特征的方法.

近年来,深度学习在计算机视觉中得到了广泛的应用,并在图像分类^[9]、目标检测^[10]、语义分

割^[11]以及超分辨率重建^[12]等多方面都取得了不俗的成绩.一般来说,深度学习的方法所提取的特征在计算机视觉中表现大多优于手工特征.2019 年, Gan 等^[13]提出计算出每个微表情序列中的顶帧,即运动幅度最大的帧,然后分别求出顶帧与起始帧之间竖直方向和水平方向光流图,输入卷积神经网络(convolutional neural networks, CNN)中提取特征,最终得到的分类效果比仅使用光流方法有很大提升.2019 年,刘汝涵等^[14]针对微表情动作微弱不利于识别,提出一种基于眼部干扰消除的视频放大方法,结合深度神经网络进行微表情识别,与仅使用深度神经网络相比结果有所提升.微表情识别中的深度学习常用方法一般是 CNN 或循环神经网络(recurrent neural networks, RNN),或者将两者结合.但是这些方法不能够同时编码视频序列时间-空间域的关系.近年来,3D-CNN 在基于视频分类中得到广泛的应用,它能够同时提取时间-空间域特征,而时间域特征对于微表情识别也非常重要.使用 3D-CNN 提取特征^[15-16]比仅使用 CNN 或者 RNN 识别效果提升明显.然而,目前微表情识别中使用深度学习方法也受到挑战:由于微表情数据库的稀缺造成在搭建神经网络时不能构建大网络,因此网络的参数会过拟合,造成微表情识别精度的下降.

为了解决这一问题,本文在识别过程中引入了深度互学习机制,提出双流网络信息交互网络(information interaction in dual stream networks, I2DSN)进行微表情识别.2018 年, Zhang 等^[17]提出,深度互学习的方法能够有效地缓解网络过拟合的情况,并且网络之间可以互相学习新的知识,进而有效地提高网络的性能.本文在微表情识别任务中融入深度互学习的策略,提出了 2 种训练方法 M-3D 和 M-3D(S),同时加入光流序列来辅助灰度图主网络的训练,通过常规的有监督学习损失和面向其他模态的拟态损失,使得 RGB 流与光流子网络之间通过不断分享学习经验,提高判别性能.此外,由于互学习机制增强了各模态分支的判别能力,测试阶段只需要输入 RGB 流即可完成识别任务.

1 基于 I2DSN 的微表情识别

1.1 双流深度互学习网络

由于本文针对微表情视频序列处理,数据中含有的时间域信息也是至关重要的,因此采用 3D-CNN 为基础网络来提取微表情序列的时空特

征. 3D-CNN 卷积层以数帧为单位处理图像后得到一幅特征图, 其既包含了空间信息, 也包含了时间信息, 特征信息更加丰富.

与 DML^[17]网络训练不同, 本文 M-3D 将双流网络分别输入微表情序列的 2 种不同模式, 灰度序列以及光流序列; M-3D(S)则在 M-3D 的基础上增加了分阶段训练的策略. 图 1 中主网络的输入微表情灰度序列蕴藏丰富的空间域信息, 用肉眼可以看出这位女性的表情大致是消极的, 而辅助网络输入的光流序列蕴藏丰富的时间域信息, 可以在光流序列中观察得出面部肌肉运动的强度和方向. 本文网络用信息交互的方式学习到两者所包含的信息, 结合两者所包含的信息预测微表情类别将

更加精确. 图 1 所示为双流信息交互微表情识别网络架构, 本文方法 M-3D 使用了双流网络来训练样本, 主网络输入微表情的灰度序列, 辅助网络的输入为微表情光流序列. 双流网络同时训练, 本文采用 3D-CNN 作为特征提取器, 不同颜色的箭头代表微表情序列中的不同帧, 网络在提取特征的过程中, 会将多帧图像的特征关联, 同时提取出空间域信息和时间域信息形成特征图. 特征图经过多层 3D 卷积、池化, 最终在决策层输出预测结果. 在训练时, 主网络与辅助网络都是组合损失函数监督进行训练, 组合损失函数由交叉熵 L_K^h 和 L_A^h , 以及散度损失 L_K^s 和 L_A^s 组成. 在训练的过程中相互学习、相互促进, 最终双流网络的预测性能都得到提升.

图 1 I2DSN

M-3D(S)改进了 M-3D 的训练策略, 在 M-3D 的损失函数上赋予组合损失函数不同的权重, 将训练分为 2 个阶段. 在第 1 个阶段, 在网络开始训练时, 双流网络的参数随机初始化, 网络预测结果接近均匀分布; 在这一阶段交叉熵损失大, 而双流网络的交互损失(本文交互损失即 KL 散度损失)小, 预测的正确率都不高, 可能会对另一流网络的预测结果有负面影响, 所以此时由标签的交叉熵损失作为主导, 监督网络训练. 训练一段时间之后进入第 2 个阶段, 2 个网络预测能力得到提升, 每个网络在训练的过程中学习到的知识不同, 所呈现的预测能力也不同; 这一阶段标签已经不能带来更多有价值的信息, 而与辅助网络的交互损失可以给主网络带来更多有价值的信息, 此时设定交互损失占主导.

1.2 损失函数

传统的监督训练损失函数为交叉熵损失, Szegedy 等^[18]认为训练时使用 label 编码会使得网络对预测结果太过自信, 容易过拟合. 本文网络的损失函数由 2 部分组成, 以缓解过拟合问题. 第 1 部分是主网络预测值与真实值的交叉熵损失, 记作 L^h . 第 2 部分是主网络与辅助网络交互 KL 散度损失, 用来度量双流网络预测结果之间距离, 记作 L^s . 以下标 K 表示主网络, A 表示辅助网络, N 为样本类别数, M 为样本数, 将相应的标签定义为 $Y = \{y_i\}_{i=1}^M$, 且 $y_i \in \{1, 2, \dots, N\}$, 则主网络交叉熵为

$$L_K = -\sum_{i=1}^M \sum_{n=1}^N I_K(y_i, n) \ln(p_K^n(x_i)) \quad (1)$$

$$\text{其中, } I_K(y_i, n) = \begin{cases} 0, & \text{if } y_i = n \\ 1, & \text{otherwise} \end{cases}$$

主网络交互损失为

$$L_K^s = D_{KL}(p_A \parallel p_K) = \sum_{i=1}^M \sum_{n=1}^N p_A^n(x_i) \ln \frac{p_A^n(x_i)}{p_K^n(x_i)} \quad (2)$$

M-3D 训练方法主网络损失函数为

$$L_K = L_K^h + L_K^s \quad (3)$$

M-3D(S)训练方法改良了 M-3D 的训练过程. 考虑 L_K^h 和 L_K^s 在训练过程中不同阶段对网络的监督作用大小不同, 将训练过程分为 2 个阶段, 对应不同的损失函数. 第 1 阶段为训练起步阶段, 此时双流预测的正确率较低, 因此由标签的交叉熵损失为主导监督网络训练, M-3D(S)损失函数为

$$L_K = \alpha L_K^h + (1 - \alpha) L_K^s \quad (4)$$

第 2 阶段为训练逐稳阶段, 双流预测正确率提高, 各分支通过相互学习中不断增强网络的识别能力, 此时由交互 KL 散度损失为主导监督网络训练, M-3D(S)损失函数为

$$L_K = (1 - \alpha) L_K^h + \alpha L_K^s \quad (5)$$

类似地, 辅助网络损失函数为

$$L_A^h = - \sum_{i=1}^M \sum_{n=1}^N I_A(y_i, n) \ln(p_A^n(x_i)) \quad (6)$$

其中, $I_A(y_i, n) = \begin{cases} 0, & \text{if } y_i = n \\ 1, & \text{otherwise} \end{cases}$.

辅助网络的交互损失为

$$L_A^s = D_{KL}(p_K \parallel p_A) = \sum_{i=1}^M \sum_{n=1}^N p_K^n(x_i) \ln \frac{p_K^n(x_i)}{p_A^n(x_i)} \quad (7)$$

辅助网络 M-3D 训练方法损失函数为

$$L_A = L_A^h + L_A^s \quad (8)$$

辅助网络 M-3D(S)训练方法与主网络相同, 将训练过程分为 2 个阶段, 对应不同的损失函数. 第 1 阶段由标签的交叉熵损失主导监督网络训练, 其损失函数为

$$L_A = \alpha L_A^h + (1 - \alpha) L_A^s \quad (9)$$

第 2 阶段由交互 KL 散度损失主导监督网络训练, 其损失函数为

$$L_A = (1 - \alpha) L_A^h + \alpha L_A^s \quad (10)$$

其中, $\alpha > 0.5$.

1.3 网络训练与测试

M-3D 方法中双流网络同时训练, 主网络输入灰度序列, 辅助网络输入光流序列. 双流网络都是由 L^h 和 L^s 共同监督同时训练, 此时式(3)(7)为损失函数. M-3D(S)方法对 M-3D 训练过程进行改进,

此时使用式(4)(8)作为损失函数来监督网络训练, 训练分为 2 个阶段, 在网络的训练精度小于 60% 时, 属于网络训练的第 1 个阶段, L^h 在损失函数中占主导作用. 当训练精度大于 60%, 训练进入第 2 阶段, L^s 在损失函数中占主导位置, 故设置 $\alpha > 0.5$, 测试方式与 M-3D 方法相同.

由于互学习的机制在训练过程中增强了主网络的预测能力, 因此测试阶段可对光流分支进行剪裁, 仅用原来的主网络即可完成对灰度图像的预测. 图 2 所示为测试阶段本文网络对微表情序列预测的网络结构图.

图 2 测试网络结构

2 实验与分析

2.1 数据库

本文实验使用了 CASME, CASME II 和 SMIC 数据库, 图 3 所示样本分别来自 CASME 数据库、CASME II 数据库和 SMIC 数据库中 SMIC-HS.

2.1.1 CASME 数据库

CASME 数据库^[19]由中国科学院心理研究所傅小兰团队建立, 使用 60 帧/s 记录了总共包含了 19 个主题 195 个自发的微表情序列. 在他们的研究中共招募了 35 位参与人员, 13 位女性, 22 位男性, 平均年龄在 22 岁. CASME 将表情分为愉悦(amusement)、害怕(fear)、悲伤(sadness)、厌恶(disgust)、惊喜(surprise)、蔑视(contempt)、压抑(repression)和紧张(tense)共 8 个类别.

2.1.2 CASME II 数据库

CASME II^[20]也是由傅小兰团队建立并公开可供科研使用. CASME II 由 255 个微表情序列组成, 采用 200 帧/s 拍摄, 参与人员平均年龄 22 岁, 分为 26 个主题 9 个类别, 分别为高兴(happiness)、惊喜(surprise)、厌恶(disgust)、害怕(fear)、悲伤(sadness)、生气(anger)、压抑(repression)、紧张(tense)和消极(negative). 实验大部分在样本数比较多的 5 大类展开: 紧张(tense)、厌恶(disgust)、高兴(happiness)、惊喜(surprise)和压抑(repression).

2.1.3 SMIC 数据库

SMIC^[21]由芬兰奥卢大学赵国英团队建立. SMIC 包含了 3 个子数据集, SMIC-HS, SMIC-VIS 和 SMIC-NIR. 本文实验选用 SMIC-HS 作为数据样本, 在其中含有 164 个微表情视频序列, 分为 16 个主题. 使用 100 帧/s 录制而成, 分为积极(positive)、消极(negative)和惊喜(surprise)3 类.

a. CASME 数据库

b. CASME II 数据库

c. SMIC 数据库

图 3 3 个数据库样本

2.2 预处理

对微表情数据库进行预处理, 分为空间域和时间域部分. 空间域部分需要对人脸进行定位、配准和剪裁, 在输入网络之前去除图像序列中无关的信息, 将人脸序列作为网络的输入将有利于网络提取出更加有用的特征. 由于 CASME II, CASME 数据库在原始的数据中包含了已经预处理过的图像序列, 因此本文只针对 SMIC 库进行预处理. 首先使用 Fac++API 对图像序列进行人脸的 landmark 定位, 然后使用 Dlib 对图像序列进行人脸的对齐, 最后剪裁出图像序列中的人脸 ROI; 同时将所有图像序列都进行尺寸归一化到 112×112. 时间域部分需要对图像序列的时间尺度进行归一化, 本文使用时间插值模型(temporal interpolation model, TIM)^[21]将所有的图像序列的时间长度都归一化为 32 作为主网络的输入; 同时将所有图像序列时间长度归一化为 33, 对其进行光流计算, 得到时间长度为 32 的微表情光流序列作为辅助网络的输入.

本文使用 TV- L^1 光流方法^[22]计算光流估计, 它有 2 个主要的优点: (1) 对于噪声更加的鲁棒; (2) 能够有效地阻止光流图的不连续. 光流用向量对物体的运动进行编码, 以此来表示图像像素运动

流的方向和大小. 光流的水平和垂直分量定义为

$$\mathbf{v} = \left[p = \frac{dx}{dy}, q = \frac{dy}{dt} \right]^T. \text{ 其中, } dx \text{ 和 } dy \text{ 表示像素沿}$$

着在 x 轴维度和 y 轴维度像素的变化; 而 dt 表示时间的变化. 为了形成三维流图像, 本文将水平和垂直流图像 p 和 q , 以及光流大小 $m = |\mathbf{v}|$ 连接起来.

2.3 实验环境设置

本文实验环境为 64 位的 Ubuntu16.04 操作系统, 32GB 内存, 以及 2 块 Geforce GTX 1080TI GPU. 使用 Pytorch 框架搭建神经网络, 使用 SGD 作为优化器训练网络, 初始学习率设置为 10^{-5} , 同时设置了每 100 个 epoch 衰减到原来的 1/2.

2.4 实验结果及分析

2.4.1 SMIC 数据库实验

本部分实验针对 SMIC 数据库进行实验. 本实验划分数据集所使用的首先采用 3D-Resnext101^[23]作为基础网络进行实验. 表 1 所示为实验所取得结果对比.

为了验证本文方法的先进性, 对 SMIC 进行预处理, 将数据集 5 等份, 将其中一份作为测试集, 其余 4 份作为训练集. 将本文方法与 MircoExpSTCNN^[23]的方法结果进行对比, 其利用微表情的时序特征构建三维网络 MicroExpSTCNN 对微表情序列进行识别, 其结果如表 1 所示, 第 2 列所示数据为灰度序列结果, 第 3 列所示为光流序列结果. 可以看出, 表 1 中的深度学习方法相较于传统方法取得了较大的进步.

表 1 不同方法在 SMIC 数据库识别准确率 %

方法	数据库	
	SMIC	SMIC(optical-flow)
LBP-TOP ^[24]	42.72	
STCLQP ^[6]	64.04	
3D-FCNN ^[25]	55.49	
MicroExpSTCNN ^[23]	68.75	
3D-Resnext101	66.67	60.61
M-3D	72.73	63.64
M-3D(S)	75.76	66.67

通过观察表 1 发现, 3D-Resnext101 方法结果比 MircoExpSTCNN 方法准确率低, 当加入互学习策略之后, M-3D 的效果较之前提升明显, 灰度图序列准确率提升了 6.06%, 光流图序列准确率提升了 3.03%. 在加入分阶段训练的策略之后 M-3D(S) 在原来 M-3D 的基础上也取得了较大提升, 灰度图序列准确率提升了 3.03%, 光流图序列准确率提升了 3.03%.

2.4.2 跨数据库实验

在本实验中, 本文选取了 3 个数据库 CASME, CASME II 和 SMIC 中的数据, 混合数据库有助于模拟真实环境中样本的多样性, 以检验方法的泛化能力. 实验分别在 3 个数据库中选取了 181, 246 和 164 微表情序列作为实验样本进行训练以及测试. 本实验选用数据集情况如表 2 所示, 总共 591 个样本被分为积极(positive, 92 个样本)、消极(negative, 179 个样本)、惊喜(surprise, 88 个样本)和其他(others, 232 个样本)4 类.

本实验采用的 5 折交叉验证进行对比, 数据库均等分 5 份, 其中一份作为测试集, 剩下作为训

练集的每一份将会被作为测试集, 4 份作为训练集, 5 次实验取平均值. MicroExpSTCNN 方法按照 Teja Reddy 等^[26]方法进行预处理, 将输入序列归一化为 $64 \times 64 \times 96$, DTSCNN 方法也按照 Peng 等^[27]的方法进行预处理, 所取得的 5 折实验结果如表 3 所示.

表 2 CASEM, CASME II, SMIC 数据集样本分布

微表情类别	CASME	CASME II	SMIC	包含类别
negative	46	63	70	厌恶、悲伤、害怕
positive	9	32	51	高兴
surprise	20	25	43	惊喜
others	106	126		紧张、压抑
总计	181	246	164	

表 3 不同方法跨数据库实验准确率

方法	5 折交叉验证					平均
	fold1	fold2	fold3	fold4	fold5	
MicroExpSTCNN ^[23]	51.30	49.50	47.83	46.09	56.52	50.26
DTSCNN TIM64	51.30	53.10	54.78	61.47	61.47	56.55
DTSCNN TIM128	50.43	54.70	57.39	60.00	65.22	57.57
DTSCNN(fusion) ^[23]	51.30	53.10	60.00	59.13	66.09	57.94
3D-Resnet101	51.30	53.10	55.65	61.47	64.35	57.17
M-3D	52.17	57.30	58.26	62.61	66.09	59.30
M-3D(S)	53.91	57.30	59.13	64.35	67.83	60.52

从表 3 中可以看到, 3D-Resnet101 方法 5 折交叉验证法所取得的结果是略低 DTSCNN 方法的最好结果, 但是使用了本文的 M-3D 方法之后, 结果得到了明显提高, 超过了 DTSCNN 的最好结果, 改进训练过程的 M-3D(S)方法取得了目前最好效果.

图 4 所示为跨数据库光流序列实验结果, 光流序列通过损失函数获得灰度序列的空间域信息, 从而加强了自身的空间信息. 根据图 4 所示结果看出, 3D-Resnet101 作为 baseline, 在使用了 M-3D, 和 M-3D(S)方法之后, 模型对于每一折的预测效果得到了提升, 进一步证明了方法的普适性. M-3D 方法比仅 3D-Resnet101 平均准确率提升了 2.13%, M-3D(S)相对于 M-3D 平均准确率提升了 1.22%.

图 4 跨数据库光流序列 5 折交叉验证实验结果

2.5 设定 α 值

图 5 所示为主网络和辅助网络训练过程交叉熵损失变化. 从中可以观察到在训练初期, 交叉熵损失较大, 主辅网络预测精度较低; 为减少交互监督对识别精度的影响, 网络由预测值与真实值的交叉熵损失作为主导, 监督网络训练, 此时交叉熵损失系数 $\alpha > 0.5$. 在训练过程中, 交叉熵损失呈下降趋势, 当其下降到一定程度时, 即网络的预测值与真实值之间差距较小, 二者交叉熵损失对网络的监督能力下降, 而此时由于主辅网络的预测值仍有较大差距, 交互损失对网络的监督作用将增强. 基于此, 本文 M-3D(S)方法将网络的训练分为了 2 个阶段, 并在不同的阶段赋予了 2 部分损失不同的权重值. 为确定权重 α , 以 0.25 为梯度进行实验. 图 6 中所展示的数据为第 2.4.2 节实验中的 fold1, fold3, fold5 取不同 α 值的结果. 从图 5 中可以看出, α 值在 0.55~0.70 选取实验会得到较好结果.

3 结 语

本文提出了一种基于双流网络信息交互的微表情识别方法, 实验采用 3 个微表情数据库 SMIC,

图5 交叉熵损失变化

图6 不同 α 值下 fold1, fold3, fold5 准确率

CASME 和 CASME II 评估方法效果, 结果表明, 其与目前的方法相比有一定的优势. 为了能够更好地获得时间域的特征, 本文使用 3D-CNN 作为基础网络, 并使用光流序列图来辅助训练; 同时使用了互学习的策略, 旨在进行微表情图像序列不同模态之间互学习, 互相补充, 以此来达到提高模型性能及其泛化能力的效果. 实验证实了本文方法的有效性和先进性, 其在识别精度及泛化能力上均优于现有方法. 未来希望在互学习交互的策略下进行更多的改进, 网络不仅局限于对空间域的不同模态, 也考虑对时间域的不同状态之间进行实验, 以此来得到更加优异的模型.

参考文献(References):

- [1] Porter S, Brinke L T. Reading between the lies: identifying concealed and falsified emotions in universal facial expressions[J]. *Psychological Science*, 2008, 19(5): 508-514
- [2] Yan W J, Wu Q, Liang J, *et al.* How fast are the leaked facial expressions: the duration of micro-expressions[J]. *Journal of Nonverbal Behavior*, 2013, 37(4): 217-230
- [3] Pfister T, Li X B, Zhao G Y, *et al.* Recognising spontaneous facial micro-expressions[C] // *Proceedings of the International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2011: 1449-1456
- [4] Xu F, Zhang J P, Wang J Z. Microexpression identification and categorization using a facial dynamics map[J]. *IEEE Transactions on Affective Computing*, 2017, 8(2): 254-267
- [5] Wang S J, Chen H L, Yan W J, *et al.* Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine[J]. *Neural Processing Letters*, 2014, 39(1): 25-43
- [6] Huang X H, Zhao G Y, Hong X P, *et al.* Spontaneous facial micro expression analysis using spatiotemporal completed local quantized patterns[J]. *Neurocomputing*, 2016, 175(A): 564-578
- [7] 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(4): 299-310
- [8] Zhang Xuange, Tian Yantao, Yan Fei, *et al.* Micro-expression recognition based on global optical flow feature[J]. *Pattern Recognition and Artificial Intelligence*, 2016, 29(8): 760-768(in Chinese)
(张轩阁, 田彦涛, 颜飞, 等. 基于全局光流特征的微表情识别[J]. *模式识别与人工智能*, 2016, 29(8): 760-768)
- [9] Sun M, Yuan Y C, Zhou F, *et al.* Multi-attention multi-class constraint for fine-grained image recognition[C] // *Proceedings of the European Conference on Computer Vision*. Heidelberg: Springer, 2018: 805-821
- [10] Lu Q S, Jiang Z Q, Men A D, *et al.* Object detection using convolutional networks with adaptively adjusting receptive field of convolutional filter[J]. *IET Computer Vision*, 2019, 13(6): 562-568
- [11] Luc P, Couprie C, Chintala S, *et al.* Semantic segmentation using adversarial networks[OL]. [2020-06-29]. <https://arxiv.org/abs/1611.08408>
- [12] Ren S, Jain D K, Guo K H, *et al.* Towards efficient medical lesion image super-resolution based on deep residual networks[J]. *Signal Processing: Image Communication*, 2019, 75: 1-10
- [13] Gan Y S, Liong S T, Yau W C, *et al.* Off-apexnet on micro-expression recognition system[J]. *Signal Processing: Image Communication*, 2019, 74: 129-139
- [14] Liu Ruhan, Xu Dan. Video amplification and deep learning in micro-expression recognition[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2019, 31(9): 1535-1541(in Chinese)
(刘汝涵, 徐丹. 视频放大和深度学习在微表情识别任务上的应用[J]. *计算机辅助设计与图形学学报*, 2019, 31(9): 1535-1541)
- [15] Wang S J, Li B J, Liu Y J, *et al.* Micro-expression recognition with small sample size by transferring long-term convolutional neural network[J]. *Neurocomputing*, 2018, 312: 251-262
- [16] Li J, Wang Y D, See J, *et al.* Micro-expression recognition based on 3D flow convolutional neural network[J]. *Pattern Analysis and Applications*, 2019, 22(4): 1331-1339
- [17] Zhang Y, Xiang T, Hospedales T M, *et al.* Deep mutual learning[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2018: 4320-4328
- [18] Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the incept-

- tion architecture for computer vision[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 2818-2826
- [19] Yan W J, Wu Q, Liu Y J, *et al.* CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces[C] //Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Los Alamitos: IEEE Computer Society Press, 2013: 1-7
- [20] Yan W J, Li X B, Wang S J, *et al.* CASME II: an improved spontaneous micro-expression database and the baseline evaluation[J]. PLoS One, 2014, 9(1): e86041
- [21] Li X B, Pfister T, Huang X H, *et al.* A spontaneous micro-expression database: Inducement, collection and baseline[C] //Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Los Alamitos: IEEE Computer Society Press, 2013: 1-6
- [22] Zach C, Pock T, Bischof H, *et al.* A duality based approach for realtime TV- L^1 optical flow[C] //Proceedings of Joint Pattern Recognition Symposium. Heidelberg: Springer, 2007: 214-223
- [23] Hara K, Kataoka H, Satoh Y. Learning spatio-temporal features with 3D residual networks for action recognition[C] //Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 3154-3160
- [24] Guo Y J, Tian Y T, Gao X, *et al.* Micro-expression recognition based on local binary patterns from three orthogonal planes and nearest neighbor method[C] //Proceedings of the IEEE International Joint Conference on Neural Networks. Los Alamitos: IEEE Computer Society Press, 2014: 3473-3479
- [25] Li J, Wang Y D, See J, *et al.* Micro-expression recognition based on 3D flow convolutional neural network[J]. Pattern Analysis and Applications, 2019, 22(4): 1331-1339
- [26] Teja Reddy S P, Teja Karri S, Dubey S R, *et al.* Spontaneous facial micro-expression recognition using 3D spatiotemporal convolutional neural networks[C] //Proceedings of the IEEE International Joint Conference on Neural Networks. Los Alamitos: IEEE Computer Society Press, 2019: 1-8
- [27] Peng M, Wang C Y, Chen T, *et al.* Dual temporal scale convolutional neural network for micro-expression recognition[J]. Frontiers in Psychology, 2017, (8): Article No.1745