

文章编号:1007-2780(2022)03-0386-09

基于视频放大和双分支网络的微表情识别

李召峰^{1,2}, 朱 明^{1*}

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049)

摘要:微表情是指当人们想要隐藏其真实情绪时,所自发产生的表情,持续时间在 1/5 s 之内,其面部肌肉运动短暂、幅度微小,有限的数据集使得特征提取变得困难,给其识别带来巨大的挑战。针对这些问题,本文提出一种基于图像预处理技术和双分支网络的识别方法,首先利用先进的人脸对齐网络获取有效的面部表情区域,再采取欧拉视频放大技术捕捉面部动作的微弱变化,以及提取光流信息作为视频序列的特征,然后将光流信息不同组合的特征图输入到双分支分类网络中得到表情标签输出。在 SMIC 和 CASME II 两个公开的数据集上进行实验,采用留一法交叉验证,准确率分别达到 0.545 和 0.584,实验结果的定量分析和定性分析均验证了所提出识别模型的有效性。

关键词:微表情; 视频放大; 光流; 双分支网络

中图分类号:TP391.4;TH691.9 文献标识码:A doi:10.37188/CJLCD.2021-0320

Micro-expression recognition based on video magnification and dual-branch network

LI Zhao-feng^{1,2}, ZHU Ming^{1*}

(1. *Changchun Institute of Optics, Fine Mechanics and Physics,*
Chinese Academy of Sciences, Changchun 130033, China;

2. *University of Chinese Academy of Sciences, Beijing 100049, China)*

Abstract: Micro-expression refers to the expression that people spontaneously generate when they want to hide their true emotions. The duration is within 1/5 s. Their facial muscles are short and small in amplitude. The limited data set makes feature extraction difficult. These factors bring huge challenges to its recognition. In response to these problems, this paper proposes a recognition method based on image preprocessing technology and dual-branch network. Firstly, an advanced face alignment network is used to obtain effective facial regions, and then Euler video magnification technology is used to capture the subtle changes in facial movements. Optical flow information is used as the feature of the video sequence, and the feature maps of different combinations of optical flow information are input into the two-branch classification network to obtain the expression label output. Experiments are performed on the two public data sets of SMIC and CASME II, and the leave-one-subject-out (LOSO)

收稿日期:2021-12-19; 修订日期:2021-12-24.

基金项目:国家自然科学基金青年基金(No.61401425)

Supported by Youth Program of National Natural Science Foundation of China (No. 61401425)

* 通信联系人, E-mail: zhu_mingca@163.com

is employed for cross-validation. The accuracy can reach 0.545 and 0.584, respectively. The quantitative and qualitative analysis of the experimental results verify the effectiveness of the proposed method.

Key words: micro-expression; video magnification; optical flow; dual-branch network

1 引言

面部表情是重要的非语言性情绪表达,日常生活中最常见和熟知的类型是宏观表情,还存在一种不同于宏观表情的人脸表情称作微表情,微表情既不能隐瞒也不能主观压制,能揭示人所潜藏的感情表达和心理活动,可以作为人类主观感受的重要依据^[1]。通常情况下,微表情的持续时间为 $1/25\sim 1/5$ s^[1-2]。微表情分析在临床诊断、国家安全和刑事侦查等领域有很大的应用潜力^[3-5]。作为一个新领域,因为不同微表情序列之间较大的歧义性,直接分析存在很大的挑战^[6]。

微表情因其快速和细微的脸部自发运动使其难以被观察到,即使是专业人士也难以准确识别微表情。早期的一些工作提出了一些辅助性的工具,比如微表情训练工具(METT)^[7],但因其成本过高和较低的人工识别率,使得这些方法很难在实际中应用。近些年,微表情数据库的建立和发布^[8-10]使微表情的识别和应用取得了很大进展。基于公开的数据集,研究者们提出了大量实验方法来自动识别微表情。Liu等提出的光学流动(MDMO)方法^[11]利用时域中像素的变化和相关性在视频帧之间挖掘对应关系,通过定位主要的面部运动方向来计算平均光流脸块,从而识别微表情。有研究利用已经成熟的图像预处理技术,如人脸特征点检测^[12-13]来通过感兴趣区域的选择对特定面部区域进行分析^[14]。这类基于纹理和基于运动的方法在一定程度上解决了微表情识别问题,但这些手工方法得到的特征需要人工去设计,具有一定的复杂度。

近些年,深度学习在表情分类、人脸检测等视觉领域发挥了巨大的优势。已经有许多基于深度学习的方法被用于微表情识别,这些方法避免了繁琐的手工特征设计,可以自动捕捉微表情的细微变化。Kim等采用两步模型CNN-LSTM来识别微表情^[15],先利用卷积神经网络(CNN)来提取

空间帧的特征,然后将特征放入长短期记忆网络(LSTM)中学习微表情的时间信息。Wang等提出了3D-FCNN提取深度时空特征来识别微表情^[16]。Khor等使用光流来学习脸部区域空间关系,再利用VGG-16和预训练的VGGFace模型来增加输入通道和深度特征的堆叠^[17]。Lei等使用了一个图时间卷积网络(Graph-TCN),使用迁移学习将基于学习的视频运动放大应用于放大微表情并提取形状信息,旨在解决脸部肌肉运动强度低的问题^[18]。Liu等结合EVM和光流来提取和放大细微的运动变化以获得更准确的结果^[19]。

受这些工作的启发,本文提出了一种基于视频动作放大和双分支网络的微表情识别框架。已有工作展示了光流在微表情识别中的有效性,脸部运动信息可以被光流所捕获,光流特征包含微表情动态的特征细节,然后将它们输入到卷积神经网络,可以构建和选择更多通过监督学习的重要脸部运动特征。本文进一步探索了光流信息的组合应用,使用双分支网络结构对预处理之后的图像特征图进行识别分类,在该网络中融合了改进的ResNet-18中的残差模块,最后使用全连接层预测微表情的类别。

2 算法流程

本文所提出的微表情识别算法主要由3部分组成。视频预处理部分,将最新的人脸关键点检测及人脸对齐技术应用用于微表情序列,旨在获取主要肌肉运动的人脸区域,然后采用欧拉视频放大技术对视频序列进行微表情中微弱运动的放大,同时在视频序列中提取出对应的初始Oneset帧和强度最大的Apex帧;光流特征处理部分,获取Apex帧的光流特征,进而获得光流强度和光流应变,将其作为对应视频序列的特征;双分支分类网络部分,用于表情的识别和分类,其中融合了改进的ResNet残差模块。本文的实验算法流程如图1所示。

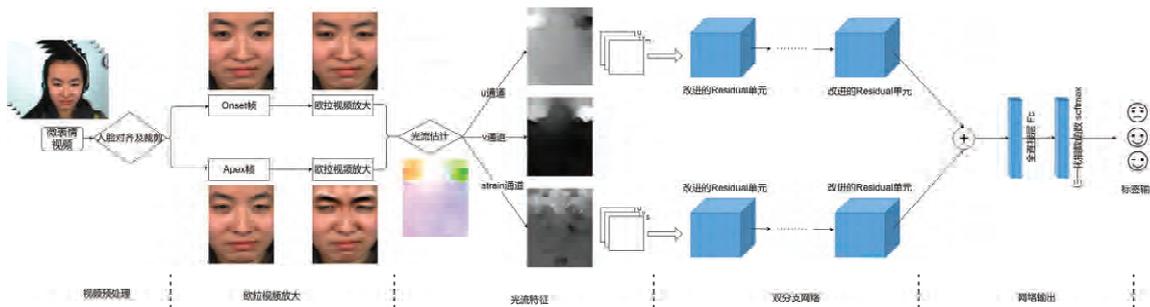


图 1 本文算法流程

Fig.1 Algorithm flow chart of this paper

2.1 视频预处理

人脸的一些部位是不包含微表情信息的。为了避免模型训练中过大的类间差异,进行数据预处理是非常有必要的,有助于减小脸部大小和脸部形状的差异,可以使要识别的人脸进行空间的归一

化,使后续模型提取到的人脸特征与五官的位置无关。视频预处理分为空间和时序两个方面:空间上完成从原始数据到裁剪后的人脸图像序列;时序上对视频帧进行了欧拉视频放大,有助于突显和放大微表情的脸部动作,处理步骤如图 2 所示。



图 2 视频预处理

Fig.2 Video image preprocessing

本文采用了文献[20]中的 2D 人脸对齐网络 (2D-FAN)来生成 68 个人脸关键点以及完成人脸对齐。2D-FAN 是目前已知最强的 2D/3D 的人脸对齐网络,在超过 220 000 张图像上做了验证,是结合了已有最优人脸关键点检测的网络结构。它的网络结构基于 Hour-Glass^[21]网络并采取堆叠的方式。Hour-Glass 网络采用瓶颈块 (Bottleneck block) 作为主要的构建模块,2D-FAN 中则更进一步使用最新提出的分级并行和多尺度的网络块对瓶颈块进行提升,在大规模的综合 2D 人脸关键点数据集上进行了训练。本文采用 2D-FAN 直接检测和提取二维面部关键点,利用该先进的残差网络架构,在整个面部范围内,将整个面部区域的图像块作为回归 CNN 的输入,面部的 2D 坐标向量作为输出。裁剪人脸的过程是依据人脸关键点中眼睛的位置而确定的矩形,以图像序列中 Onset 帧为参考裁剪图像。

用于人脸对齐的传统度量是两眼距离归一化之后,点对点的欧几里得距离,但如文献[22]所指出,对于双眼间距很小的轮廓人脸,该误差度量是有偏差的。本文通过边界框大小进行归一化,使用公式(1)中的平均误差定义作为人脸对齐的误差评测。其中 x 表示给定人脸的真实特征点; y 是对应的预测特征点; d 由公式(2)得到,是真实边界框高(h)和宽(w)乘积的平方根。

$$AE = \frac{1}{N} \sum_{k=1}^N \frac{\|x_k - y_k\|_2}{d} \tag{1}$$

$$d = \sqrt{w * h} \tag{2}$$

微表情最大的挑战是脸部运动的幅度过小而难以区分。对齐的脸部区域在时序上的变化细节依然很难被人眼所观察到,也很难被机器学习技术捕捉到这些微小变化。

为了增大不同微表情之间的区别,采用欧拉视频放大方法^[23]对微小的脸部运动进行放大处

理,使得在时序上有明显的肌肉运动变化。运动放大步骤如下:视频帧分解为不同的拉普拉斯金字塔频带,然后用巴特沃斯滤波器提取感兴趣的频带,最后不同的带通信号乘以放大系数 α 添加到原始图像。

基于运动的时序放大可由式(3)得到, $I(x, t)$ 是图像在 x 位置和 t 时间的强度值, $f(x)$ 是视频帧起始时刻在 x 位置的强度值,图像中的区域块经历了平移运动, $\delta(t)$ 作为位移函数的观测强度。假设一阶泰勒展开中只考虑放大后的最强扰动,式(3)可以简化为式(4),表明空间分布局部图像 $f(x)$ 在时间 t 的位置 $\delta(t)$ 已被放大到 $(1 + \alpha)$ 的量级。

$$I(x, t) \approx f(x) + (1 + \alpha)\delta(t) \frac{\partial f(x)}{\partial x}, \quad (3)$$

$$I(x, t) \approx f(x) + (1 + \alpha)\delta(t), \quad (4)$$

$$(1 + \alpha)\delta(t) < \frac{\lambda}{8}. \quad (5)$$

文献[23]中视频放大包含动作放大和颜色放大两种类型,本文采用动作放大,参数 α 是控制运动放大程度的参数,若太大会引入较大的脸部错位和噪声。根据公式(5)将放大系数固定在一个固定区间范围内,空间波长 λ 取值为 16,选取 α 值为 20,选择无限脉冲响应(IIR)滤波器作为时序滤波器。

2.2 光流信息提取

本文采用微表情强度最强的 Apex 帧和初始 Oneset 帧来代表对应的微表情序列[24]。CASME II 数据库[10]提供了 Apex 帧的序号,SMIC 数据库[8]未提供 Apex 帧序号,采用 Liong 的方法[13]来自动获取对应的 Apex 帧。本文应用 FlowNet 2.0 来生成图像对的光流特征,FlowNet 2.0 的结果有着更清晰的物体边缘,对运动伪影也具有鲁棒性[25],在许多视觉任务中得到了广泛的应用。从 Oneset 帧和 Apex 帧获得的光流场中提取出水平和垂直方向的信息,分别用 u 和 v 表示。再用 u 和 v 分别计算出光流强度和光流应变的大小,公式(6)中采用了 L2 距离计算光流强度 m 。

$$m = d_{L2}(u, v) = \sqrt{u^2 + v^2}. \quad (6)$$

光流应变能计算面部的微小变化,能表达出运动变形的相对量[26],假设进行了很小的运动,光流应变强度可由公式(7)中的海森矩阵计算

得到。

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

可将其简化为

$$\epsilon = \frac{1}{2} [\nabla \mathbf{u} + (\nabla \mathbf{u})^T], \quad (8)$$

其中 \mathbf{u} 为二维位移向量, $\mathbf{u} = [u + v]^T$, $(\epsilon_{xx}, \epsilon_{yy})$ 是法向应变分量, $(\epsilon_{xy}, \epsilon_{yx})$ 是切向应变分量。光流应变强度推导最终可由公式(9)得到:

$$\epsilon = \sqrt{\epsilon_{xx}^2 + \epsilon_{yy}^2 + \epsilon_{xy}^2 + \epsilon_{yx}^2}. \quad (9)$$

2.3 双分支网络模型

深度残差网络[27]在图像分类、语义分割等领域的应用中具有先进的性能,在不增加网络复杂性的情况下,残差网络通过堆叠残差模块的方式来构建网络,经典的残差块如图 3 所示,以两个 3×3 的卷积层为主,外加 BN 层和 ReLU 层,通过 1×1 卷积跨层直接连接,使得输入 x 无损失地传递到更深层的网络当中,将上一层或多层之前的输出与本层的计算输出相加,解决了传统神经网络中梯度弥散和网络性能退化的问题。

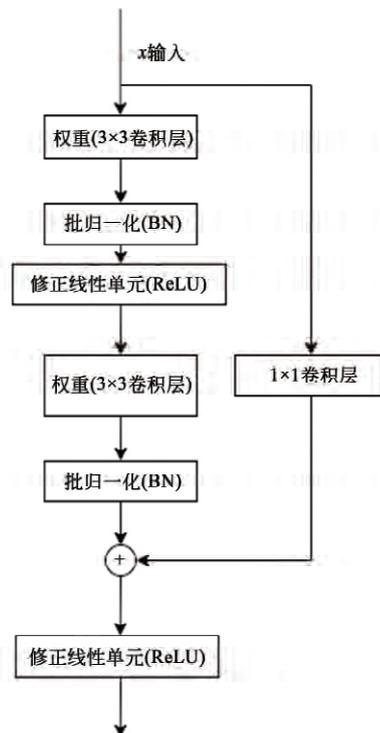


图 3 ResNet 残差模块

Fig.3 Residual module

为了适应不同任务的特性,ResNet 的结构也可以进行调整^[28],本文对 ResNet 残差模块进行改进,融入了 CBAM^[29]的空间注意模块,如图 4 所示,来让网络从空间尺度学习目标“在哪里”。因卷积运算是通过跨通道和空间信息来提取特征的,为了计算空间注意特性,沿着残差模块的通道轴应用平均池化和最大池化操作,连接生成有效的特征描述符。再沿通道轴应用池化的操作能突出有效的信息区域^[30]。该空间注意模块可由式(10)计算,Sigmoid 为 Sigmoid 激活函数。

$$C(x) = \text{Sigmoid}(\text{Conv}(\text{avgPool}(x), \text{maxPool}(x))) \quad (10)$$

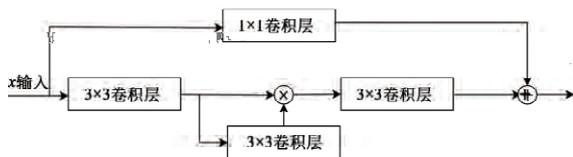


图 4 改进的 ResNet 残差模块
Fig.4 Improved residual module

有不少研究^[31-32]显示了在浅层网络一定条件下小数据集和不平衡的类别上的分类性能,与手工特征相比,显示出很大的优越性,且减少网络深度可以减少可学习的参数。受文献^[33]的启发,本文选取两个独立的分支网络 N_a 和 N_b ,每个分支 N 由 16 个改进的 ResNet 残差模块 c 串行叠加组成,即

$$N = \{c_1, c_2, \dots, c_{16}\} \quad (11)$$

通过训练,这样的网络能够关注到面部的微表情区域。将以下两组的光流特征组合图各送入网络 N_a 和 N_b 中:水平光流场、垂直光流场和光流强度的组合 (u, v, m) ,水平光流场、垂直光流场和和光流应变的组合 (u, v, ϵ) 。两分支网络输出后的合并采用公式(12)的特征图映射加法,来对时间-空间特征进行融合,网络输出之后连接的是全连接层和归一化层。

$$C = C_n a \oplus C_n b \quad (12)$$

3 实验结果与分析

3.1 微表情数据库

自发微表情数据库 SMIC 是最常见的基准数据库^[8],使用高速摄像机以 100 fps 拍摄,最短的视频序列只有 11 帧长(持续时间约为

0.11 s),平均帧长为 29 帧(约 0.30 s),SMIC 数据集包含 16 名参与者,共 164 个视频,包含“负面”(70 个)、“正面”(51 个)和“惊喜”(43 个)3 类。CASME II^[10]是一个综合性自发微表情数据库,包含 247 个视频样本,包含来自亚洲的 26 名参与者。所有视频均以 200 fps 拍摄,分辨率为 640×480 ,该数据库中的视频包含厌恶、压抑、惊讶、高兴、其他、害怕和悲伤 7 类微表情,因最后两种表情样本过少,本文只对前 5 种表情进行评测。

3.2 实验设置及定量结果

本文在公开的微表情数据库 SMIC 和 CASME II 上评测了所提出的识别算法,采用留一法交叉验证。本文的实验部分在 NVIDIA GeForce RTX 2080Ti 上进行,显存为 11 G。采用 PyTorch 深度学习框架和 Adam 优化器($\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1e-8$),最初的学习率设置为 $1e-4$,使用大小为 16 的 batch size,共训练了 30 个 epochs。由于微表情数据集样本量小且不平衡,在训练中还进行了数据增强,包括水平翻转和高斯模糊(5×5)。混淆矩阵结果如图 5 和图 6 所示。

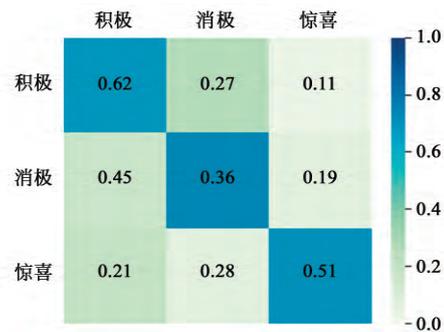


图 5 SMIC 的混淆矩阵
Fig.5 Confusion matrix of SMIC



图 6 CASME II 的混淆矩阵
图 6 Confusion matrix of CASME II

表1是已有研究工作与本文算法的对比,包含了经典的传统算法LBP-TOP和手工提取特征算法,以及最新的深度学习算法。除最常使用的

准确率以外,还增加了同时考虑查准率和查全率的F1指数对算法进行评价。基于SMIC和CASME II两个数据集,本文方法的准确率和F1指

表1 不同算法识别准确率和F1指数比较

Tab.1 Comparison of recognition accuracy and F1-score of different algorithms

方法	SMIC		CASME II	
	准确率	F1 指数	准确率	F1 指数
LBP-TOP ^[34]	0.434	0.342	0.465	0.424
FDM ^[35]	0.549	0.532	0.417	0.297
LBP-STP ^[36]	0.445	0.449	0.466	0.448
Bi-WOOF ^[24]	0.593	0.620	0.579	0.610
ELRCN ^[17]	——	——	0.524	0.500
Image-based CNN ^[37]	0.531	0.476	0.564	0.507
本文方法	0.545	0.563	0.584	0.528

数均优于大多数算法。

3.3 定性结果

为了探索本文模型在“看哪里”及如何进行识别分类的,本文应用梯度加权类激活映射 Grad-CAM 技术^[38]对双分支网络最后卷积层合并后的激活强度进行可视化显示,模型映射后以热图的方式展现,热图有助于直观地解释和分析网络所激活的脸部位置以及网络的运行机制,图7是CASME II所激发的与具体的表情类别相关的面部区域,通过最后双分支各自的激活值相加得到。表明本文模型是能够从表情密切对应的区域学习特征的。根据Ekman于1978年开发的面部动作编码系统^[39],可以看到热图的脸部运动单元(action unit, AU)与表情面部动作编码是对应和一致的,见表2。特别能看到在压抑表情中,嘴唇两侧的凹陷具有很强的网络激活;在厌恶表情中,

眉毛的小幅度提升运动也被本文识别算法检测到。

表2 表情类别与对应的AU组合

Tab.2 Emotion category and corresponding AU combinations

表情类别	运动单元(AU)组合
厌恶	AU4(眉毛压低并使眉毛聚拢), AU5(抬起上眼皮,瞪大眼睛), AU10(抬起上嘴唇)
压抑	AU15(拉动嘴角肌肉向下拉)
惊讶	AU1(眼睛周围一圈肌肉往上拉), AU2(眉毛外侧向上拉)
高兴	AU6(眼睛外圈收缩), AU12(嘴角肌肉向耳窝方向拉)

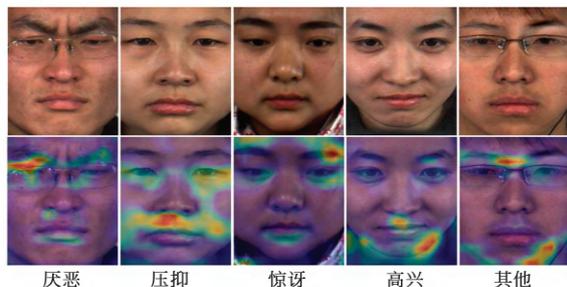


图7 双分支网络激活图可视化

Fig.7 Visualization of two-branch network activations

4 结 论

本文提出了一种微表情的识别方法,利用先进的人脸对齐网络和欧拉视频放大技术对视频序列进行预处理,之后抽取Oneset帧和Apex帧得到对应视频的光流特征,相同结构的分支网络对光流特征图的输入进行进一步的特征提取之后,输出表情的分类标签。所提出算法在公开数据库中的性能优于大多数现有的方法,

最后激活热力图提供了定性验证,也表明所提出网络架构的合理性。已有数据库的情绪标注是自我评估所产生的,与真实表情标签有一定的误差性。本文下一步工作可以在已有模型基础上验证标签误差对性能的影响,此外,数据库

中不同的微表情可能共享相同的一组脸部运动单元,其如何定性决定表情类别需进行进一步的研究。最后,微表情数据集的大小影响分类性能,可以在本文数据增强的基础上进行更有效的数据增广。

参 考 文 献:

- [1] EKMAN P, FRIESEN W V. Nonverbal leakage and clues to deception [J]. *Psychiatry*, 1969, 32(1): 88-106.
- [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] WANG S J, YAN W J, LI X B, *et al.* Micro-expression recognition using dynamic textures on tensor independent color space [C]//*Proceedings of the 22nd International Conference on Pattern Recognition*. Stockholm: IEEE, 2014: 4678-4683.
- [4] ENDRES J, LAIDLAW A. Micro-expression recognition training in medical students; a pilot study [J]. *BMC Medical Education*, 2009, 9(1): 47.
- [5] FRANK M G, KIM D J, KANG S, *et al.* Improving the ability to detect micro expressions in law enforcement officers. Manuscript in preparation, 2014.
- [6] DAVISON A K, MERGHANI W, YAP M H. Objective classes for micro-facial expression recognition [J]. *Journal of Imaging*, 2018, 4(10): 119.
- [7] EKMAN M P. Micro expression training tool[R]. Oakland: CD-ROM, 2003.
- [8] 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*. Shanghai, China: IEEE, 2013: 1-6.
- [9] 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*. Shanghai, China: IEEE, 2013: 1-7.
- [10] 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.
- [11] 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.
- [12] CRISTINACCE D, COOTES T F. Feature detection and tracking with constrained local models [C]//*Proceedings of the British Machine Vision Conference 2006*. Edinburgh, 2006: 929-938.
- [13] LIONG S T, SEE J, WONG K, *et al.* Automatic micro-expression recognition from long video using a single spotted apex [C]//*Proceedings of the Asian Conference on Computer Vision*. Taipei, China: Springer, 2016: 345-360.
- [14] LIONG S T, SEE J, WONG K, *et al.* Automatic apex frame spotting in micro-expression database [C]//*Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. Kuala Lumpur: IEEE, 2015: 665-669.
- [15] KIM D H, BADDAR W J, RO Y M. Micro-expression recognition with expression-state constrained spatio-temporal feature representations [C]//*Proceedings of the 24th ACM International Conference on Multimedia*. Amsterdam, Netherlands: ACM, 2016: 382-386.
- [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] KHOR H Q, SEE J, PHAN R C W, *et al.* Enriched long-term recurrent convolutional network for facial micro-expression recognition [C]//*Proceedings of the 13th IEEE International Conference on Automatic Face & Ges-*

- ture Recognition (FG 2018). Xi'an, China; IEEE, 2018.
- [18] LEI L, LI J F, 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; ACM, 2020.
- [19] LIU J M, LI K, SONG B L, *et al.* A multi-stream convolutional neural network for micro-expression recognition using optical flow and EVM [J]. *arXiv:2011.03756*, 2020.
- [20] BULAT A, TZIMIROPOULOS G. How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks) [C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice; IEEE, 2017: 1021-1030.
- [21] NEWELL A, YANG K Y, DENG J. Stacked hourglass networks for human pose estimation [C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam; Springer, 2016.
- [22] ZHU X X, RAMANAN D. Face detection, pose estimation, and landmark localization in the wild [C]//*Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence; IEEE, 2012.
- [23] WU Y H, RUBINSTEIN M, SHIH E, *et al.* Eulerian video magnification for revealing subtle changes in the world [J]. *ACM Transactions on Graphics*, 2012, 31(4): 65.
- [24] 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.
- [25] ILG E, MAYER N, SAIKIA T, *et al.* FlowNet 2.0: evolution of optical flow estimation with deep networks [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu; IEEE, 2017.
- [26] GODAVARTHY S. Microexpression spotting in video using optical strain [D]. Florida; University of South Florida, 2010.
- [27] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas; IEEE, 2016: 770-778.
- [28] HE K M, ZHANG X Y, REN S Q, *et al.* Identity mappings in deep residual networks [C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam; Springer, 2016.
- [29] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich; Springer, 2018: 3-19.
- [30] ZAGORUYKO S, KOMODAKIS N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer [C]//*Proceedings of the 5th International Conference on Learning Representations*. Toulon; OpenReview.net, 2017.
- [31] PENG M, WU Z, ZHANG Z H, *et al.* From macro to micro expression recognition: deep learning on small datasets using transfer learning [C]//*Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition*. Xi'an, China; IEEE, 2018: 657-661.
- [32] 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.
- [33] KHOR H Q, SEE J, LIONG S T, *et al.* Dual-stream shallow networks for facial micro-expression recognition [C]//*Proceedings of 2019 IEEE International Conference on Image Processing (ICIP)*. Taipei, China; IEEE, 2019: 36-40.
- [34] ZHAO G, PIETIKALNEN M. Dynamic texture recognition using local binary patterns with an application to facial expressions [J]. *IEEE Trcms. Pattern Anal. Mach. Intell.*, 2007, 29(6):915-928.
- [35] XU F, ZHANG J P, WANG J Z. Micro expression identification and categorization using a facial dynamics map [J]. *IEEE Transactions on Affective Computing*, 2017, 8(2): 254-267.
- [36] WANG Y, SEE J, PHAN R C, *et al.* LBP with six intersection points: Reducing redundant information in LBP-TOP for micro-expression recognition [C]//*Proc. Asian Conf. Comput. Vis.*(Beijing; IEEE), 2014: 525-537.
- [37] TAKALKAR M A, XU M. Image based facial micro-expression recognition using deep learning on small datasets [C]//*Proceedings of 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. Sydney; IEEE, 2017: 1-7.

- [38] SELVARAJU R R, COGSWELL M, DAS A, *et al.* Grad-CAM: visual explanations from deep networks via gradient-based localization [C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017: 618-626.
- [39] EKMAN P, FRIESEN W V. *Facial Action Coding System: Investigator's Guide* [M]. Palo Alto: Consulting Psychologists Press, 1978.

作者简介:



李召峰(1991—),男,甘肃天水人,硕士研究生,2013年于西北工业大学获得学士学位,主要从事图像处理、机器视觉方面的研究。E-mail: lzf0215@163.com



朱明(1964—),男,江西南昌人,硕士,研究员,1991年于中国科学院长春光学精密机械研究所获得硕士学位,主要从事视频图像处理,自动目标识别技术及成像目标跟踪等方面的研究。E-mail: zhu_mingca@163.com