

文章编号: 2095-2163(2022)10-0144-06

中图分类号: TP183

文献标志码: A

基于面部多个局部特征的人脸表情识别算法

贾茜伟, 闫娟, 杨慧斌, 刘向前

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

摘要: 人脸面部表情通常来说可以显露出人的内心活动变化, 目前现有的表情识别方法一般依靠面部的整体特征进行处理, 没有考虑面部的一些局部特征, 导致面部表情识别的准确度不理想。人的面部表情进行变化时, 面部局部肌肉会随之变化, 基于此, 提出一种基于面部多个局部特征的人脸表情识别算法。本文首先对检测到的人脸进行面部分区, 分为23个子区域, 再将分好的区域输入到卷积神经网络中进行局部特征的提取。最终使用 $AM - softmax$ 函数将表情分为中性、愤怒、厌恶、惊讶、高兴、悲伤和恐惧七类。评估实验在 CK+ 和 JAFFE 数据集上对本文算法进行验证, 得到的平均准确率分别是 99.87% 和 96.72%, 均超过 S-Patches 算法, 该结果表明本文算法对表情识别性能有所提高。

关键词: 表情识别; 面部分区; 卷积神经网络; $AM - softmax$ 函数

Facial expression recognition technology based on multiple local facial features

JIA Qianwei, YAN Juan, YANG Huibin, LIU Xiangqian

(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

[Abstract] Facial expressions can usually reveal changes in human inner activities. Current facial expression recognition methods generally rely on the overall features of the face for processing, and do not consider some local features of the face, resulting in a decrease in the accuracy of facial expression recognition. When a person's facial expression changes, local muscles will change accordingly. Based on this, a facial expression recognition algorithm based on multiple local features of the face is proposed. This paper first partitions the detected face into 23 sub-regions, and then inputs the divided regions into the convolutional neural network to extract local features. Finally, the $AM - softmax$ function is used to divide the expressions into seven categories: neutral, angry, disgusted, surprised, happy, sad and fearful. The evaluation experiment verifies the average accuracy of the algorithm in this paper on the CK+ and JAFFE data sets, are 99.87% and 96.72%, respectively, which both exceed the S-Patches algorithm. The results show that the algorithm has improved the performance of facial expression recognition.

[Key words] expression recognition; face partition; convolutional neural network; $AM - softmax$ function

0 引言

人的面部表情识别可以识别出中性、愤怒、厌恶、惊讶、高兴、悲伤和恐惧七类表情^[1-5], 该表情识别技术有许多应用场景, 例如: 在汽车司机的疲劳驾驶监测中, 通过司机的表情变化来推测司机的驾驶状态, 从而对司机发出提醒。在智能安防中, 通过监测公众场合人们表情的变化来推测其心理想法, 并及时阻止潜在危险行为, 从而避免危险事故的发生。在医疗单位中, 通过监测病人表情的变化来了解病人的身体状况, 一旦发现异样及时采取措施等。

Mehrabian^[6] 研究指出, 人脸面部表情可以传递人们的内心活动和情感变化, 如果计算机可以获取到人脸信息, 人机交互将会有很大的发展空间。经过 Ekman 等人^[7] 的一系列研究, 人脸的面部表情识别技术已经迅速发展起来。一般来说, 将表情识别

分为特征提取和特征分类这两步, 特征提取主要是解析人脸面部图像, 并将局部特征提取出来, 分类器再根据提取到的局部特征进行最优分类^[7-10]。近年来, 部分学者提出依据人脸的特征点发生的形变程度来推测人脸面部表情, 然而仅仅依靠部分特征点的形变并不能代表整个人脸表情, 从而造成表情识别准确率较低。伴随着卷积神经网络 (Convolutional Neural Networks, CNN) 的研发问世, 一些学者提出可以在表情识别中融入 CNN 算法, 经过划分好的区域提取人脸的全局特征迭代回归, 取得了不错的识别效果, 但是只有面部全局特征, 未能突显出局部特征, 因此会造成相似表情在一定程度上的区分困难, 例如惊讶和恐惧的区分^[11-12]。Kaleekal 等人基于面部 68 个特征点周围提取 8 个面部突出块 (Salient Facial Patches, SFP) 的 Hahn 特征, 并采用 SVM 进行分类, 经过试验表明, 在 CK+ 与 JAFFE 数据集的

作者简介: 贾茜伟 (1997-), 男, 硕士研究生, 主要研究方向: 智能控制。

通讯作者: 闫娟 Email: aliceyan_shu@126.com

收稿日期: 2021-11-26

识别率分别为 91.33% 与 93.16%。SFP 方法虽然突出了面部的局部特征, 但是选取时并未设置一个固定的标准, 也忽视了不同表情中局部特征比例上的不同。另外, 根据 Li 等人研究, 面部的 6 类基本表情均有 FACS 编码中对应的面部肌肉运动单元 AU, 但是该分区方法需要大量识别 AU, 而单个 AU 的识别准确率并不高^[13-15]。

因此, 本文提出基于面部多重局部特征的人脸表情识别技术, 论文的创新点可表述为:

(1) 本算法基于人脸的 68 个特征点, 再将额头部分加上, 总共将面部分为 23 个子区域。面部表情也会在额头上体现出来, 例如额头的皱纹和表面舒张等。本文统计了 6 类基本表情 AU 产生域, 再依据其产生域、面部肌肉单元以及面部器官将 23 个子区域归类到 8 个候选区域中, 为 SFP 的选取提供了依据, 而且可以不用重复就能提取不同表情中包含的相似的局部特征, 从而解决了需要重复识别 AU 和识别单个 AU 效率低的问题。

(2) 本文算法采用 8 个并行的分支进行提取, 并根据候选区域所占的面积比例支配不同维全连接层, 解决了局部候选区域提取不够均衡的问题。每个分支的输出按照注意力自适应地连接, 经过 Softmax 函数将人脸面部表情基本分为 7 类: 中性、愤怒、厌恶、恐惧、高兴、悲伤、惊讶。在 CK+ 和 JAFFE 数据集得到的平均准确率分别为 99.87% 与 96.72%, 相比 S-Patches 方法分别增加了 6.03% 与 10.28%。

1 本文人脸表情识别算法

1.1 面部分区

根据人脸的构造以及产生表情时一些运动的局部肌肉单元, 将面部肌肉运动分为 45 种, 现选取部分 FACS 展示, 详见表 1。这些局部肌肉运动单元可以组成多个人脸表情, 其中包含 6 类人脸基本表情。继而研究得到的表情相应 AUs 结果见表 2。分析表 1、表 2 可知, 眉毛紧在一起、眉毛间出现直形皱纹等, 对应着表 1 的面部肌肉运动单元中的 AU_4 、 AU_5 、 AU_7 、 AU_{23} 、 AU_{24} 的一种或几种, 表情表现为愤怒; 眉毛向下压、上眼皮向上抬、下眼皮下面出现皱纹, 对应着表 1 的面部肌肉运动单元中的 AU_9 、 AU_{17} , 表情表现为厌恶; 眉毛紧皱并抬起、额头显现皱纹、上眼睑向上抬起、下眼皮向下拉伸、嘴微微变化或张大等, 对应着表 1 的面部肌肉运动单元中的 AU_4 、 $AU_1 + AU_5$ 、 $AU_5 + AU_7$, 表情表现为恐惧; 牙齿露出、嘴角向

外拉并微微抬起、眉毛可能会向下弯一点、眼皮下面凸起等, 对应着表 1 的面部肌肉运动单元中的 AU_6 、 AU_{12} 、 AU_{25} , 表情表现为高兴; 眉毛向内收缩、嘴角向下拉伸、上眼皮下拉、额头舒张等, 对应着表 1 的面部肌肉运动单元中的 AU_1 、 AU_4 、 AU_{15} 、 AU_{17} , 表情表现为悲伤; 眼睛睁大、上眼皮向上抬起、下眼皮向下拉伸、嘴巴张大等, 对应着表 1 的面部肌肉运动单元中的 AU_5 、 AU_{26} 、 AU_{27} 、 $AU_1 + AU_2$, 表情表现为惊讶。

表 1 部分面部肌肉运动单元 AU

Tab. 1 Part of facial muscles movement unit AU

AU	描述	样图
1	内眉抬起	
2	外眉抬起	
5	上眼睑抬起	
7	眼皮拉紧	
12	嘴角拉紧	
17	下巴抬起	
20	嘴唇拉长	
24	嘴唇下拉	
26	下颚下拉	

表 2 表情相应 AUs

Tab. 2 AUs corresponding to the expression

表情	相应 AUs
愤怒	$AU_4, AU_5, AU_7, AU_{23}, AU_{24}$
厌恶	AU_9, AU_{17}
恐惧	$AU_4, AU_1 + AU_5, AU_5 + AU_7$
高兴	AU_6, AU_{12}, AU_{25}
悲伤	$AU_1, AU_4, AU_{15}, AU_{17}$
惊讶	$AU_5, AU_{26}, AU_{27}, AU_1 + AU_2$

6 类基本表情产生时, 大部分面部肌肉运动单元存在明显动作, 基于此, 将面部分为 23 个局部特征区域, 如图 1 所示, 再将这 23 个面部区域按照人脸的构造分为 8 个局部区域 AU_{O_i} , i 为 1 到 8。每个候选区域都包含一定的特征区域, 且只提取本候选区域内的特征, 在候选区域内产生的表情见表 3。人脸表情的局部区域示意如图 2 所示。由图 2 可看

出,图2(a)中 AU_{o_1} 包含额头和眉毛部分,是 AU_1 、 AU_2 、 AU_5 和 AU_7 的产生域,包含1、2、3、6、7面部的局部特征区域,属于愤怒、恐惧、悲伤、惊讶等表情发生时的肌肉运动单元活动区域;图2(b)中 AU_{o_2} 包含眉毛和眼睛部分,是 AU_4 产生域,包含2、3、6、7、8、11面部的局部特征区域,属于愤怒、恐惧、悲伤表情产生区域;图2(c)中 AU_{o_3} 包含眼睛、鼻梁及其鼻子附近部分,是 AU_6 产生域,包含8、9、10、11、13、15面部的局部特征区域,属于高兴表情产生区域;图2(d)中 AU_{o_4} 包含纵向鼻子及鼻梁部分,是 AU_9 产生域,包含4、5、9、10、14面部的局部特征区域,属于厌恶表情产生区域;图2(e)中 AU_{o_5} 包含鼻子和嘴巴部分,是 AU_{12} 、 AU_{15} 产生域,包含14、17、18、19面部的局部特征区域,属于高兴、惊讶、悲伤表情产生区域;图2(f)中 AU_{o_6} 包含嘴巴、下巴及附近肌肉部分,是 AU_{25} 、 AU_{26} 、 AU_{27} 产生域,包含17、18、19、

20、21、22、23面部的局部特征区域,属于高兴、惊讶表情产生区域;图2(g)中 AU_{o_7} 包含下巴部分,是 AU_{17} 产生域,包含20、21、22、23面部的局部特征区域,属于厌恶、悲伤表情产生区域;图2(h)中 AU_{o_8} 包含嘴巴部分,是 AU_{23} 、 AU_{24} 产生域,包含18面部的局部特征区域,属于愤怒表情产生区域。



图1 面部区域分区

Fig. 1 Facial area division

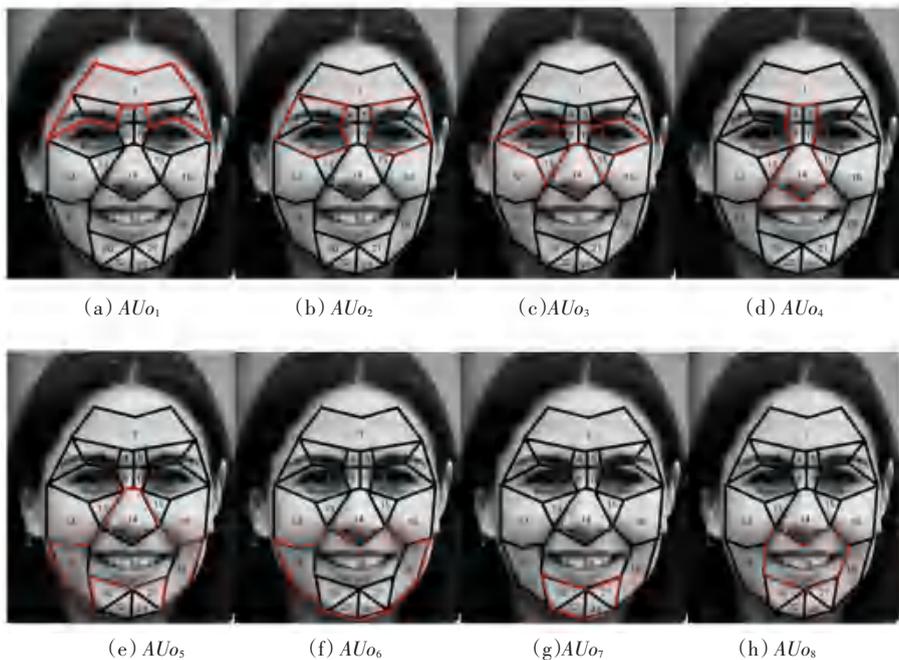


图2 局部区域示意图

Fig. 2 Local area diagram

1.2 与 CNN 算法融合

本文算法基于面部区域的68个特征点和23个分好的局部特征区域,构成8个局部候选区域,并将候选区所占的面积最小的矩形框输入卷积神经网络中。8个局部候选区选取的矩形框如图3所示。

本文算法的整体流程如图4所示。每个 AU_{o_i} 图

像中选取的局部特征区域大小不一,并且都需要经过 CNN 网络提取特征,再将提取得到的特征拼接为4096大小的全连接层。为了得到更好的识别效果,就需要突出不同表情的局部特征,因此再将拼接的全连接层乘以不同表情所占的权重值,后续又经过特征提取与 *Softmax* 函数将表情基本分为7类:中性、愤怒、厌恶、恐惧、惊讶、高兴、悲伤。

表 3 AU 组包含区域
Tab. 3 AU group contains area

AU 候选区	AU 编号	特征区域	产生表情
AU_{o_1}	AU_1, AU_2, AU_5, AU_7	1,2,3,6,7	愤怒、恐惧、悲伤、惊讶
AU_{o_2}	AU_4	2,3,6,7,8,11	愤怒、恐惧、悲伤
AU_{o_3}	AU_6	8,9,10,11,13,15	高兴
AU_{o_4}	AU_9	4,5,9,10,14	厌恶
AU_{o_5}	AU_{12}, AU_{15}	14,17,18,19	高兴、悲伤、惊讶
AU_{o_6}	$AU_{25}, AU_{26}, AU_{27}$	17,18,19,20,21,22,23	高兴、惊讶
AU_{o_7}	AU_{17}	20,21,22,23	厌恶、悲伤
AU_{o_8}	AU_{23}, AU_{24}	18	愤怒



图 3 局部候选区最小矩形示意图

Fig. 3 Schematic diagram of the smallest rectangle of the local candidate area

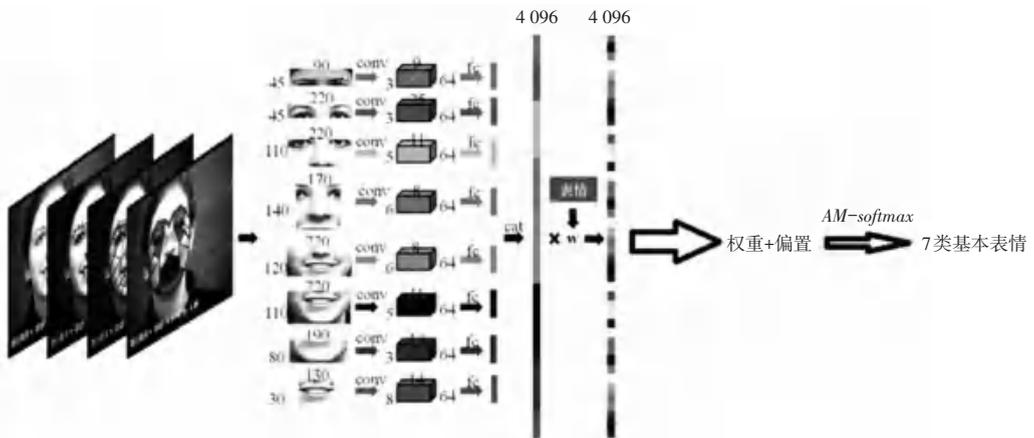


图 4 本文算法流程图

Fig. 4 Flow chart of the algorithm in this paper

1.3 损失函数

每一类基本表情都有不同的局部特征,将其局

部特征突显出来,损失函数是必不可少的。损失函数 $AM - softmax$ 在许多视觉工作中被经常使用,主

要通过最大化真实样本类别的后验概率将特征从不同的类中分离出来,给定输入特征为 x_i ,对应的标记 y_i ,其计算公式如下:

$$L_{softmax} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{w_j^T x_i + b_j}}{\sum_{j=1}^M e^{w_j^T x_i + b_j}} \quad (1)$$

其中, N 为批次的大小; M 为类的数量; $w_j \in R^d$ 是第 j 类最后一个完全连接层的权重; d 是特征维度; $b_j \in R$ 是偏置项。在 w_j 和 x_j 上应用 L_2 归一化来优化球面上的特征,特征距离与特征角的计算关系如下:

$$w_j^T x_i = \|w_j\| \|x_i\| \cos \theta_{ji} \quad (2)$$

其中, θ_{ji} 是 w_j 和 x_i 之间的角度。

2 实验与分析

2.1 实验环境

为验证本文算法的有效性,实验选用 Win10 系统、内存 8 G、CPU 为 i5-9500、硬盘为 512 G 固态硬盘,实验中所需要的软件包括: Ubuntu16.04、Pytorch1.9.0、Opencv4.5.1、Cudnn7.5、CUDA10.1。

2.2 数据集

本文算法的模型选用 CK+数据集和 JAFFE 数据集进行实验,实验前先将图像进行预处理,统一大小为 48×48 的灰度图,训练集和测试集大致按照 4:1 随机进行划分,由此得到的 CK+和 JAFFE 数据集中 7 类基本表情的训练集和测试集的数量分布见表 4 和表 5。

表 4 CK+数据集中各表情分布情况

Tab. 4 Distribution of expressions in the CK+ dataset

表情类别	训练集/张	测试集/张	共计
中性(normal)	1 040	260	1 300
愤怒(angry)	480	120	600
厌恶(disgust)	700	175	875
惊讶(surprised)	880	220	1 100
高兴(happy)	640	160	800
悲伤(sad)	224	56	280
恐惧(fear)	240	60	300

表 5 JAFFE 数据集中各表情分布情况

Tab. 5 The distribution of expressions in the JAFFE dataset

表情类别	训练集/张	测试集/张	共计
中性(normal)	24	6	30
愤怒(angry)	24	6	30
厌恶(disgust)	23	6	29
惊讶(surprised)	24	6	30
高兴(happy)	24	7	31
悲伤(sad)	24	7	31
恐惧(fear)	25	7	32

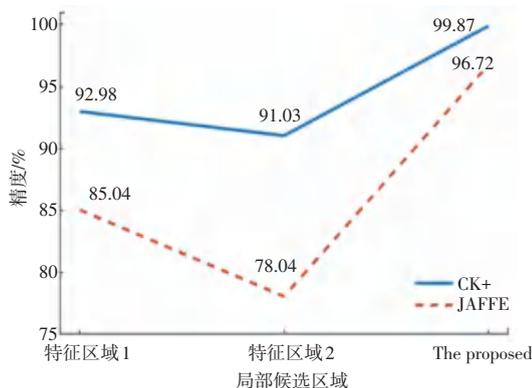
2.3 实验结果与分析

现选取 2 组不同的局部候选区域与 AU_o_i 进行实验验证,对比结果如图 5 所示。选取的 2 组不同局部候选区域组成见表 6。由图 5(a) 可看到,本文算法选取的区域在 CK+数据集的平均准确率达到 99.87%,相比于选取的特征区域 1 与特征区域 2 高出 6.89%与 8.84%。在 JAFFE 数据集的平均准确率达到 96.72%,相比于选取的特征区域 1 与特征区域 2 高出 11.68%与 18.68%。实验表明,精确选取局部特征区域比简单选取面部的器官区域得到的识别效果要高得多,局部特征区域对人脸面部表情识别的准确率起到重要的作用。

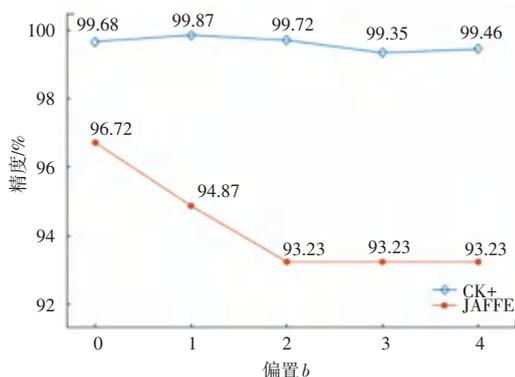
表 6 对比局部候选区域组成

Tab. 6 Comparing the composition of local candidate regions

AU	特征区域 1	特征区域 2
AU_{o_1}	1,2,3,6,7,8,9	3,6,7,8,9
AU_{o_2}	2,3,6,7,8,11,12	6,7,8,11,12,13
AU_{o_3}	8,9,10,11,13,14,15	9,10,11,12,13,15
AU_{o_4}	4,5,9,10,11,12,14	9,10,11,12,14,15
AU_{o_5}	14,15,16,17,18,19	17,18,19,20,21
AU_{o_6}	16,17,18,19,20,21	18,19,20,21,22,23
AU_{o_7}	19,20,21,22,23	21,22,23
AU_{o_8}	17,18	18,19



(a) 不同候选区域精度图



(b) 不同偏置值精度图

图 5 对比图

Fig. 5 Comparison chart

不同表情的权重值 W_j 是不同的, 也会对面部表情的识别造成一定的影响, 现从偏置 b 的值为 0、1、2、3、4 这五种来进行实验验证不同偏置值对表情识别准确率的影响。由图 5(b) 可知, 当偏置值 b 为 1 时, CK+ 数据集中的准确率最高为 99.87%; 当偏置值 b 为 0 时, JAFFE 数据集中的准确率为 96.72%。实验表明, 偏置值不是越大、或者越小越好。

算法对比结果见表 7。将本文算法与 CNN 算法在 CK+ 和 JAFFE 的混合数据集下对 7 类基本表情识别的准确率进行对比后, 从表 7 中可以看出, 本文算法除在厌恶表情识别上略低于 S-Patches 算法外, 其余表情识别和平均准确率均高于 S-Patches 算法。对于 CNN 算法而言, 本文算法只在悲伤表情识别中略低, 其余方面均与其相等或胜过 CNN 算法。总地来说, 实验通过与两大经典算法做对比, 可以证明本文算法在对面部特征划分成有效的局部特征区域后, 具有有效性, 并且提高了表情识别性能。

表 7 算法对比

Tab. 7 Comparison results of the algorithms

算法	中性	愤怒	厌恶	恐惧	高兴	悲伤	惊讶	平均
S-Patches	84.21	88.76	99.24	98.34	89.96	90.65	96.06	92.46
CNN	98.76	92.10	94.13	97.05	100.00	98.60	96.28	96.70
本文算法	98.78	98.35	98.76	98.20	100.00	97.52	100.00	98.80

3 结束语

本文基于 68 个特征点和面部肌肉运动单元, 提出基于面部多个局部特征的人脸表情识别算法。算法将包含额头的人脸面部区域划分成 23 个子区域, 再经卷积神经网络提取特征, 最后经 $AM - softmax$ 函数进行表情分类。本文算法在 CK+ 和 JAFFE 数据集的平均准确率分别为 99.87% 和 96.72%, 相对于 S-Patches 则分别提升了 6.03% 和 10.28%。本文算法也存在一定的不足, 比如大量提取局部特征, 必然会增加训练时长, 还有就是提取局部特征后的降

维问题, 这些都需要后续进一步地去研究。

参考文献

- [1] 党宏社, 王森, 张选德. 基于深度学习的面部表情识别方法综述 [J]. 科学技术与工程, 2020, 20(24): 9724-9732.
- [2] 张智恒. 基于深度学习的人脸面部表情识别 [D]. 西安: 西安石油大学, 2021.
- [3] OWUSU E, KUMI J A, JUSTICE K A, et al. On facial expression recognition benchmarks [J]. Applied Computational Intelligence and Soft Computing, 2021, 2021: 9917246.
- [4] SUN Yudao, WU Chunhua, ZHENG Kangfeng, et al. Adv-emotion: The facial expression adversarial attack [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2021, 35(11): 3-7.
- [5] 章东平, 陈思瑶, 李建超, 等. 基于改进型加性余弦间隔损失函数的深度学习人脸识别 [J]. 传感技术学报, 2019, 32(12): 1830-1835.
- [6] MEHRABIAN A. Communication without Words [J]. Psychology Today, 1968, 2(4): 53-56.
- [7] EKMAN P, FRIESEN E. Facial action coding system: A technique for the measurement of facial movement [M]. Palo Alto: Consulting Psychologists Press, 1978.
- [8] 李清霞. 基于改进的 Gabor 小波特征提取的人脸表情识别算法研究 [D]. 秦皇岛: 燕山大学, 2018.
- [9] 张勇育. 基于单帧图像的超分辨率重建 [D]. 厦门: 厦门大学, 2013.
- [10] 刘元震. 基于深度学习的人脸面部情感识别的研究 [D]. 哈尔滨: 哈尔滨工业大学, 2018.
- [11] YANG Biao, CAO Jingmeng, NI Rongrong, et al. Facial expression recognition using weighted mixture deep Neural Network based on double-channel facial images [J]. IEEE Access, 2018, 6: 4630-4640.
- [12] SAMIRA E K, CHRISTOPHER P, XAVIER B, et al. Combining modality specific Deep Neural Networks for emotion recognition in video [C] // Proceedings of the 15th ACM on International Conference on Multimodal Interaction. Sydney, Australia: ACM, 2013: 543-550.
- [13] 谭小慧, 李昭伟, 樊亚春. 基于多尺度细节增强的面部表情识别方法 [J]. 电子与信息学报, 2019, 41(11): 2752-2759.
- [14] 刘旷. 基于卷积网络集成的面部表情识别方法 [D]. 杭州: 浙江大学, 2016.
- [15] 沈利迪. 融合深度学习和稀疏表示的人脸表情识别方法研究 [J]. 自动化技术与应用, 2019, 38(10): 93-97.