

# 基于自发表情数据集预训练的多任务 深度网络表情识别方法

彭先霖<sup>1</sup>, 张海曦<sup>1</sup>, 胡琦瑶<sup>2</sup>, 温超<sup>3</sup>

(1. 西北工业大学 电子信息学院, 西安 710129;

2. 西北大学 信息科学与技术学院, 西安 710127;

3. 西北大学 智能交互与信息艺术研究中心, 西安 710127)

**摘要:** 为了更好地将现有深度卷积神经网络应用于表情识别, 提出将构建自然表情图像集预训练和多任务深度学习相结合的方法; 首先, 利用社交网络图像构建一个自发人脸表情数据集, 对现有深度卷积神经网络进行预训练; 然后, 以双层树分类器替换输出层的平面 softmax 分类器, 构建深度多任务人脸表情识别模型; 在 CK+数据集和 Oulu-Caisa 数据集上的验证实验表明, 提出方法的准确率较深度卷积神经网络 vgg-face 和 Inception V3 方法平均提高了 3.1%, 较基于多任务深度卷积神经网络方法平均提高了 0.7%。

**关键词:** 表情识别; 自然表情数据集; 预训练; 深度卷积神经网络; 多任务学习

## Multi-task Deep Network Expression Recognition Based on Pre-training of Natural Emotional Data Sets

Peng Xianlin<sup>1</sup>, Zhang Haixi<sup>1</sup>, Hu Qiyao<sup>2</sup>, Wen Chao<sup>3</sup>

(1. School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China;

2. School of Information Science and Technology, Northwest University, Xi'an 710127, China;

3. Intelligent Interaction and Information Arts Research Center, Northwest University, Xi'an 710127, China)

**Abstract:** In order to better apply the existing depth convolution neural network to facial expression recognition, a method combining pre-training of natural facial expression image set and multi-task depth learning is proposed. Firstly, a data set of spontaneous facial expressions is constructed from social network images, and the existing deep convolution neural network is pre-trained. Then, a deep multi-task facial expression model is constructed by replacing the planar softmax classifier of the output layer with a double-tree classifier. The experimental results on CK+ and Oulu-Caisa datasets show that the accuracy of the proposed method is 3.1% higher than that of the deep convolution neural network such as vgg-face and Inception V3, and 0.7% higher than that of the multi-task deep convolution neural network.

**Keywords:** facial expression recognition; natural expression dataset; pre-training; deep convolution neural network; multitask learning

## 0 引言

人脸表情在许多领域中发挥着重要作用<sup>[1]</sup>, 如图像理解、新闻分析、个性化推荐、人机交互等。因此, 近三十年来, 自动人脸表情识别引起了众多学者的关注<sup>[2]</sup>。人脸表情通常被划分为有限类型, 每种类型的特征又因种族、文化、个性等的不同而有很大的差异, 因此, 人脸表情识别仍然是一个具有很大挑战性的问题。

经典的人脸表情识别方法主要分为三类: (1) 主元分

析 (PCA)<sup>[3]</sup>和线性判别分析 LDA (Linear Discriminant Analysis)<sup>[4]</sup>; (2) 主动形状模型 ASM (Active Shape Model) 和主动表观模型 AAM (Active Appearance Model)<sup>[5]</sup>; (3) 基于纹理特征的 Gabor 小波<sup>[6]</sup>和局部二进制模式 LBP<sup>[7]</sup>等。经典方法的不足之处在于, 所用的特征提取方法都是以手工特征为基准进行提取, 损失了原图像的很多细节特征。近年来, 随着深度学习领域的崛起, 卷积神经网络 (Convolution Neural Network, CNN) 逐渐被运用于图像处理与分析领域<sup>[8]</sup>。CNN 避免了对图像的复杂前期预处理, 可以直接输入原始图像, 逐层构建一个多层的深度神经网络, 让机器自主地从样本数据中学习特征。特别是 VGG face<sup>[9]</sup>、Googlenet<sup>[10]</sup>、Resnet34<sup>[11]</sup>等的表现突出。但是, 现有深度神经网络方法大多专注于单个识别任务, 然而, 现实世界中人脸表情与不同个体的面部形态、头部姿势、外部光照等多种因素交织在一起, 为了减弱面部形态对表

收稿日期: 2019-05-09; 修回日期: 2019-05-21。

基金项目: 国家重点研发计划资助课题(2017YFB1402103)。

作者简介: 彭先霖(1989-), 男, 湖南娄底人, 博士研究生, 主要从事图像处理、模式识别方向的研究。

通讯作者: 温超(1978-), 男, 陕西三原人, 硕士生导师, 副教授, 主要从事机器学习、信息交互设计方向的研究。

情识别的影响, 文献 [12] 构建了深度多任务学习框架, 该框架以现有深度卷积神经网络模型为基础, 用双层树分类器替换输出层的平面 softmax 分类器, 通过利用人脸表情标签和人脸标签共同学习更具辨识力的深度特征, 将知识从相关人脸识别任务中迁移至表情识别任务中来, 从而减弱面部形态对表情识别的影响, 有效提高了人脸表情识别率。然而 VGG face<sup>[9]</sup>、Googlenet<sup>[10]</sup>、Resnet34<sup>[11]</sup> 等这些模型都不是针对表情识别任务而训练的, 如果能用大量表情数据对这些模型进行再次训练, 预期能取得更高的识别率。

现有的人脸表情数据集可分为两类。一类专注于六种基本的情绪, 即快乐、悲伤、惊讶、愤怒、厌恶和恐惧。例如, MMI 人脸表情数据集<sup>[13]</sup> 包含 2 000 余幅表情图像或帧, 50 人的表情图像 500 余幅。日本女性人脸表情 (JAFFE) 数据集<sup>[14]</sup> 提供了 213 幅日本女性人脸表情图像, 每个对象分别为 7 个基本表情 (6 个情绪表情加上中性脸) 中的 3 个或 4 个例子。另一个类关注于提取人脸表情的细粒度描述。例如, Cohn-Kanade 数据集<sup>[15]</sup> 包含单动作单元和基于 Ekman 提出的面部动作编码系统 (FACS) 的动作单元的组合。除此之外, 还有一些其他的包括一些常见人脸表情的数据集<sup>[16-18]</sup>。所有这些数据都面临着两个问题: 首先, 每个数据集中图像的数量很少; 第二, 由于表情图像是在人为设置的环境中获得的, 而不是自发表情图像。因此, 很难学习丰富的表情特征并有效地对表情进行分类。为了克服上述两方面的缺陷, 文献 [19] 基于大型网络图像, 结合不同的种族和文化, 提出了一种自发人脸表情数据集构建方法。

本文提出基于自发表情数据集预训练的多任务深度网络表情识别方法, 首先通过自建自然人脸表情数据集对现有成熟深度网络模型进行预训练, 再采用多任务深度学习框架进行表情识别, 有效提高了表情识别准确率。

## 1 自发表情数据集的建立

为了构建表情数据集, 参考文献 [19] 得到如图 1 所示的框图, 在百度和谷歌等搜索引擎中, 总共使用 7 个关键词 (即愤怒、蔑视、厌恶、恐惧、幸福、悲伤、惊讶, 及其对应英文词语) 作为查询词进行搜索。搜索引擎通过每个关键词返回大量图像, 得到初步的表情图像数据集。然后通过人脸检测、特征提取、自动过滤和手动过滤四个步骤, 得到自然表情图像数据集。下面简要介绍这四个步骤。

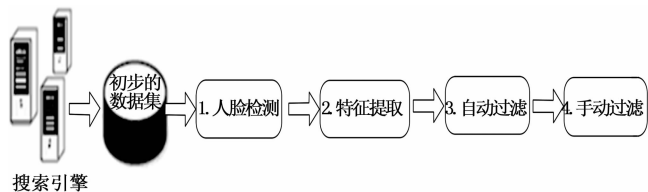


图 1 构建自然表情数据集基本过程

### 1.1 人脸检测

目前, Viola-Jones<sup>[20]</sup> 人脸检测算法与其他算法相比,

表现出较好的性能 (准确性和速度)。因此, 在本文中, 我们利用该算法通过提取 haar 特征和设计 Adaboost 分类器来检测原始数据集中的候选人脸。

### 1.2 特征提取

局部二值模式 (local Binary Patterns, LBP)<sup>[21]</sup>、局部相位量化 (local Phase Quantization, LPQ)<sup>[22]</sup> 和二值统计图像特征 (binarized statistical Image Features, BSIF)<sup>[23]</sup> 等三类特征在人脸表情描述上表现较好, 所以在这里选择这三类特征作为表情特征。

#### 1.2.1 LBP

LBP 描述符是一个高度区分灰度纹理的描述符。对于图像中的每个像素, 二进制代码是通过用中心像素的值对一个圆对称邻域进行阈值计算得到的。

#### 1.2.2 LPQ

LPQ 特性是一个模糊的鲁棒图像描述符。LPQ 描述符的低频相位分量对中心对称模糊不敏感。因此, LPQ 可以使用短期傅里叶变换的相位信息, 在图像的每个像素周围的窗口上进行局部计算得到。

#### 1.2.3 BSIF

BSIF 特征是属于 LBP 和 LPQ 同一家族的二值纹理描述符。但是, BSIF 不使用手工设计的滤波器, 而是使用一组自然图像来自动学习一组固定的滤波器。这组过滤器是根据从一组训练图像中获取的图像块的统计信息来学习的。

分别提取这三类高维视觉特征并将其映射到不同的特征空间。每个特征空间都用来描述图像的某种视觉属性。由于这些特征是基于直方图的特征, 因此用  $\chi^2$  核函数作为直方图的基核。给定图像  $I$  和  $J$  的两个特征向量  $u$  和  $v$ , 定义  $\chi^2$  核:

$$\chi^2 = \frac{1}{2} \sum_{i=1}^D \frac{u_i - v_i}{u_i + v_i} \quad (1)$$

其中:  $D$  是特征向量的维度。  $u_i$  和  $v_i$  是特征向量  $u$  和  $v$  的第  $i$  维。这样, 第  $c$  个特征直方图的核函数  $K_c(I, J)$  定义为:

$$K_c(I, J) = e^{-x^2(u, v) / \sigma_c} \quad (2)$$

其中:  $\sigma_c$  是所有图像对之间的  $x^2$  距离的平均值。

利用这些不同权重的基核的线性组合来逼近视觉特征的核, 并通过交叉验证来确定这些权值。图像之间不同的视觉相似性更精确地表示为:

$$K_h(I, J) = \sum_{l=1}^{\tau} \beta_l K_l(I, J), \sum_{l=1}^{\tau} \beta_l = 1 \quad (3)$$

其中:  $\tau$  是图像核的数量。  $\beta_l \geq 0$  是第  $l$  个图像核  $K_l(I, J)$  的权重, 可通过交叉验证来估计。在本文中,  $\tau$  设置为 3, 对应三个基本核函数。

### 1.3 自动图像清理

为了进一步对原始数据集的弱相关图像进行交互过滤, 需要在上述初始数据集中显示图像, 然后直接删除弱相关图像。在自动图像清理中有三个步骤:

1) 采用近邻传播 (AP)<sup>[24]</sup> 算法将图像聚类为多个类

别。考虑到 2000 个左右的图像很难清晰显示，而且很多图像都是相似的，最好将几个有代表性的图像显示给人们，以确定弱相关的图像。所以所有的图像都用 AP 方法聚类成一些子集，然后选择几个图像作为每个类别的代表性图像。

2) 采用双曲可视化方法，也就是说根据用户的非线性视觉相似性背景，使用双曲线可视化技术显示大量返回的图像，获得初始数据集<sup>[25]</sup>的全局视图，以帮助用户交互地评估用户的查询意图与初始图像之间的相关性。

3) 不相关的图像作为垃圾图像被过滤掉。首先，通过在双曲线可视化空间上点击并拖动鼠标来选择被视为无关的图像；然后，将与每个所选图像高度相似的一些图像作为垃圾图像滤除。

如果认为仍然存在许多不相关的图像，则将 AP 聚类再次用于剩余图像，并且重复上述交互式过滤过程。

### 1.4 手动图像清理

通过聚类自动清理后，只有少量图像被错误标记。由于这些图像的数量相对较少，我们选择手工消除这些错误标记的图像。

## 2 多任务深度卷积神经网络表情识别方法

参考文献 [12]，得到如图 2 所示分层多任务学习框架。使用两层树分类器来代替传统深层网络中的平面 softmax 分类器，利用人脸表情标签和人脸标签来帮助深度网络学习更具辨别力的深度特征。

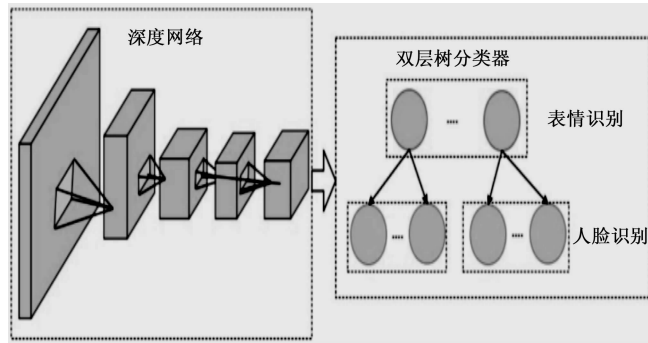


图 2 分层多任务学习框架

对于输入图像  $I$ ，属于第  $m$  类人脸表情的对象图像  $I$  的预测概率可以用式 (4) 计算：

$$P(I \in e_m) = \left\{ \frac{\exp(W_m^T x_e + b_e)}{\sum_{i=1}^M \exp(W_i^T x_e + b_e)} \right\} \quad (4)$$

其中： $e_m$  表示第  $m$  类人脸表情； $x_e$  表示人脸表情的深度特征向量； $W_m$  表示人脸表情识别分类器对应第  $m$  类表情的权重参数； $T$  表示表示转置； $M$  表示人脸表情类别数； $b_e$  为偏置参数。

人脸表情识别的预测概率可用于选择特定的人脸识别分类器。对于对应于第  $l$  类表情的低层人脸识别分类器，输入图像  $I$  属于第  $n$  个人脸类别的预测概率可以表示为式 (5)：

$$P(I \in f_n) = \left\{ \frac{\exp(V_n^T x_f + b_f)}{\sum_{j=1}^N \exp(V_j^T x_f + b_f)} \right\} \quad (5)$$

其中： $f_i$  表示第  $j$  类人脸； $x_f$  表示人脸的深度特征向量； $V_n$  表示人脸识别分类器对应第  $n$  类人脸的权重参数； $T$  表示表示转置； $N$  表示人脸类别数； $b_f$  为偏置参数。

由于最终预测概率可计算为  $P(I \in e_m) P(I \in f_n)$ ，因此深层网络的损失函数可以表示为式 (6)：

$$(W, V, x_e, x_f) = - \sum_{n=1}^N \log \left\{ \frac{\exp(V_n^T x_f + b_f)}{\sum_{j=1}^N \exp(V_j^T x_f + b_f)} \right\} - \sum_{m=1}^M \log \left\{ \frac{\exp(W_m^T x_e + b_e)}{\sum_{i=1}^M \exp(W_i^T x_e + b_e)} \right\} \quad (6)$$

该目标函数旨在最大化人脸表情识别和人脸识别的正确预测概率，因此在每次迭代期间可以联合优化多级分类器中的权重参数。

## 3 实验结果

首先，采用第二节描述的方法对每类表情收集到 3000 幅图像，得到共 21000 幅表情图像数据集。然后，将提出的方法与两个最流行的人脸表情数据集 (CK+ 和 Oulu-Casias) 上的现有类似方法进行比较。每个数据集的图像序列数量如表 1 所示。

表 1 每个数据集的图像序列分布

	愤怒	蔑视	厌恶	恐惧	幸福	悲伤	惊讶	合计
CK+	45	18	59	25	69	28	83	327
Oulu-Casias	80	-	80	80	80	80	80	480

所有实验均在 HP 工作站上进行，该工作站配备 Intel i5-7400 CPU，带有 8G 内存的 Nvidia GTX1080；16G 内存，512G SSD。所有的实验都在 PyTorch 平台上进行。

### 3.1 CK+数据集实验结果

CK+数据集<sup>[25]</sup>：扩展 CohnKanade \* (CK+) 数据集是用于评估人脸表情识别方法的最常用数据集之一。CK+ 包含来自 123 名受试者的 593 个视频序列，其中有 327 个被标记。CK+数据集未为每幅图像提供精确的标签，只图像序列有标签，序列中包含了从平静到表情表现峰值的图像。因此通过对应表情序列取表情峰值附近的 3 帧。之后，九个子集用于训练，另一个子集用于验证。在本文中，所有 118 个具有精确标签的受试者被分配到 7 个人脸表情类别中。

在 CK+数据集上的识别准确率如表 2 所示，可以很容易地发现，本文所提出的方法在准确率方面取得非常有竞争力的表现。不仅可以在依赖于人的实验中获得最佳性能。更重要的是，提出的方法可以在不依赖于人的实验上得到更多的改善。在识别快乐、愤怒、厌恶和惊讶方面表现得非常好，并且在其他表情识别方面表现也不错。

### 3.2 Oulu-Casias 的实验结果

Oulu-Casias 数据集<sup>[31]</sup> 包含 2880 个图像序列和 80 个标

记有 6 种基本表情的主题。与 CK + 类似, 提取最后三帧 (仅考虑强照明条件) 作为对应标签的峰值帧。所有图像的九个子集用于训练, 另一组用于验证。在该实验中, 构建

表 2 CK + 数据集上不同表情识别方法的准确率

方法	准确率(%)
VGG face (依赖于人)	96.42
文献[12]方法(依赖于人)	97.53
本文方法(依赖于人)	98.24
VGG face (不依赖于人)	92.70
文献[12]方法(不依赖于人)	95.64
本文方法(不依赖于人)	96.32
Inception V3(依赖于人)	100.0
文献[12]方法(依赖于人)	100.0
本文方法(依赖于人)	100.0
Inception V3(不依赖于人)	93.35
文献[12]方法(不依赖于人)	96.02
本文方法(不依赖于人)	96.83
Resnet 34(依赖于人)	100.0
文献[12]方法(依赖于人)	100.0
本文方法(依赖于人)	100.0
Resnet 34(不依赖于人)	94.21
文献[12]方法(不依赖于人)	96.62
本文方法(不依赖于人)	97.16

的两层分层结构包括 6 个粗粒度节点 (人脸表情) 和 480 个细粒度节点 (每个人脸表情下 80 个人)。

如表 3 所示, 提出的方法仍然在 Oulu-Casia 数据集上实现了最佳性能。实验中发现提出的方法可以在愤怒, 恐惧, 快乐和惊讶方面达到很高的准确率, 但在识别厌恶和悲伤方面表现相对差一些。

表 3 Oulu-Casia 数据集的准确率

方法	准确率(%)
VGG face (依赖于人)	88.70
文献[12]方法(依赖于人)	92.75
本文方法(依赖于人)	93.14
VGG face (不依赖于人)	76.42
文献[12]方法(不依赖于人)	77.04
本文方法(不依赖于人)	77.97
Inception V3(依赖于人)	88.55
文献[12]方法(依赖于人)	93.04
本文方法(依赖于人)	93.85
Inception V3(不依赖于人)	77.05
文献[12]方法(不依赖于人)	77.82
本文方法(不依赖于人)	78.43
Resnet 34(依赖于人)	89.16
文献[12]方法(依赖于人)	93.33
本文方法(依赖于人)	94.16
Resnet 34(不依赖于人)	77.78
文献[12]方法(不依赖于人)	78.64
本文方法(不依赖于人)	79.32

## 4 结论

利用社交图像构建了一个新的自发表情数据集, 社会标签图像是通过特定的关键词从图像搜索引擎中获取的, 利用垃圾图像清理算法去除标记错误的图像; 然后, 采用双层树分类器代替传统深层 CNN 中的平面 softmax 分类器, 构成在表情识别同时考虑人脸识别的多任务学习框架, 利用自然表情图像数据集对网络进行预训练, 有效提高了表情识别准确率。在未来的研究中可以从以下几方面进行更深入研究: 一是进一步扩大人脸表情数据集; 二是对容易混淆的表情采用特殊处理方法; 三是将更多人脸属性考虑进去构成更完善的多任务深度学习框架。

### 参考文献:

- [1] Li S Z, Jain A K. Handbook of Face Recognition. Springer Publishing Company [M]. London: Springer-Verlag, 2011.
- [2] Zeng Z H, Maja P, Roisman G, et al. A survey of affect recognition methods: audio, visual, and spontaneous expressions [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2008, 31 (1): 39-58.
- [3] Andrew J, Calder A, Bunon M, et al. A Principal Component Analysis of Facial Expressions [J]. Vision Research, 2001, 41 (9): 1179-1208.
- [4] Mahmud F, Khatun M T, Zuhori S T. Face recognition using Principle Component Analysis and Linear Discriminant Analysis [C]. Dhaka: International Conference on Electrical Engineering and Information Communication Technology, 2015. 1-4.
- [5] Cheon Y, Kim D. Natural facial expression recognition using differential-AAM and manifold learning [J]. Pattern Recognition, 2009, 42 (7): 1340-1350.
- [6] Kirtac K, Dolu O, Gokmen M. Face recognition by combining Gabor wavelets and nearest neighbor discriminant analysis [C]. Krakow: International Symposium on Computer and Information Sciences, 2016. 1-5.
- [7] 侯小红, 郭敏. 一种基于 Harris-SIFT 特征点检测的 LBP 人脸表情识别算法 [J]. 西北大学学报 (自然科学版), 2017, 47 (2): 209-214.
- [8] XIA Z, LIN J, FENG X. Trademark image retrieval via transformation-invariant deep hashing [J]. Journal of Visual Communication and Image Representation, 2019, (59): 59108-116.
- [9] Parkhi O. M, Vedaldi A, Zisserman A. Deep face recognition [C]. Swansea: 26th British Machine Vision Conference, 2015, 6.
- [10] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision [C]. Seattle: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. 2818-2826.
- [11] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [C]. Seattle: IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), 2016. 770–778.

- [12] 彭先霖, 张海曦, 胡琦瑶. 基于多任务深度卷积神经网络的人脸/面瘫表情识别方法 [J]. 西大学学报 (自然科学版), 2019, 49 (2): 187–192.
- [13] Pantic M, Valstar M, Rademaker R, et al. Webbased database for facial expression analysis [C]. Amsterdam; In IEEE International Conference on Multimedia and Expo (ICME), 2005. 1–5.
- [14] Kamachi M, Lyons M, Gyoba J. The Japanese female facial expression (jaffe) database [ED/OL]. <http://www.kasrl.org/jaffe.html>, 1998.
- [15] Kanade T, Cohn J F, Tian Y. Comprehensive database for facial expression analysis [C]. Grenoble; In Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000. 46–53.
- [16] Belhumeur P N, João P Hespanha, Kriegman D J. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19 (7): 711–720.
- [17] Lyons M J, Akamatsu S, Kamachi M G, et al. Coding facial expressions with gabor wavelets [C]. Nara; In IEEE International Conference on Automatic Face and Gesture Recognition, 1998. 200–205.
- [18] Sim T, Baker S, Bsat M. The cmu pose, illumination, and expression (pie) database [C]. Washington, DC; In IEEE International Conference on Automatic Face and Gesture Recognition, 2002. 46–51.
- [19] Peng X L, Xia Z Q, Li L, et al. Towards Facial Expression Recognition in the Wild: A New Database and Deep Recognition System [C]. Las Vegas; 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016. 1544–1550.
- [20] Viola P, Jones M J. Robust Real-Time Face Detection [J]. International Journal of Computer Vision, 2004, 57 (2): 137–154.
- [21] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24 (7): 971–987.
- [22] Ojansivu V, Janne Heikkil?. Blur insensitive texture classification using local phase quantization [C]. Cherbourg; International Conference on Image & Signal Processing, 2008. 236–243.
- [23] Kannala J, Rahtu E. BSIF: Binarized statistical image features [C]. Tsukuba; In International Conference on Pattern Recognition (ICPR), 2012. 1363–1366.
- [24] Frey B J, Dueck D. Clustering by Passing Messages Between Data Points [J]. Science, 2007, 315 (5814): 972–976.
- [25] Wang X B, Feng X Y, Peng J Y. A novel facial expression database construction method based on web images [C]. Chengdu; In Proceedings of the Third International Conference on Internet Multimedia Computing and Service, 2011. 124–127.
- [26] Lucey P, Cohn J F, Kanade T, et al. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression [C]. San Francisco; IEEE Computer Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010. 94–101.
- [27] Klaser A, Marszalek M, Schmid C. A spatio-temporal descriptor based on 3d-gradients [C]. Leeds; 2008 19th British Machine Vision Conference, 2008. 1–10.
- [28] Scovanner P, Ali S, Shah M. A 3-dimensional sift descriptor and its application to action recognition [C]. Augsburg; 15th ACM International Conference on Multimedia, 2007. 357–360.
- [29] Liu M, Li S, Shan S, et al. Deeply learning deformable facial action parts model for dynamic expression analysis [C]. Singapore; Asian Conference on Computer Vision, 2014. 1749–1756.
- [30] Jung H, Lee S, Yim J, et al. Joint fine-tuning in deep neural networks for facial expression recognition [C]. Santiago; 2015 IEEE International Conference on Computer Vision (ICCV), 2015; 2983–2991.
- [31] Zhang K, Huang Y, Du Y, et al. Facial expression recognition based on deep evolutionary spatial-temporal networks [J]. IEEE Transactions on Image Processing, 2017, 26 (9): 4193–4203.
- [32] Zhao G Y, Huang X H, Taini M, et al. Facial expression recognition from near-infrared videos [J]. Image and Vision Computing, 2011, 29 (9): 607–619.
- [33] Ekman P. Universal and cultural differences in facial expression of emotion [J]. Nebraska Symposium on Motivation, 1972, 19: 207–284.
- [34] 彭进业, 杨瑞靖, 冯晓毅, 等. 人脸疼痛表情识别综述 [J]. 数据采集与处理, 2016, 31 (1): 43–55.
- [35] Craig K D, Hyde S A, Patrick C J. Genuine, suppressed and faked facial behavior during exacerbation of chronic low back pain [J]. Pain, 1991, 46 (2): 161–171.
- [36] Prkachin K M. The consistency of facial expressions of pain: A comparison across modalities [J]. Pain, 1992, 51 (3): 297–306.
- [37] Prkachin K M, Solomon P E. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain [J]. Pain, 2008, 139 (2): 267–274.
- [38] Brahmam S, Chuang C-F, Sexton R S, et al. Machine assessment of neonatal facial expressions of acute pain [J]. Decision Support Systems, 2007, 43 (4): 1242–1254.
- [39] Monwar M, Rezaei S, Prkachin K. Eigenimage based pain expression recognition [J]. IAENG International Journal of Applied Mathematics, 2007, 36 (2): 1–6.