Computer Science and Application 计算机科学与应用, 2020, 10(10), 1843-1852
Published Online October 2020 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2020.1010194

Hans汉斯

用于人脸表情识别的卷积神经网络研究

孙丽萍[1]，陈红倩[1]，李　慧[2*]

[1]北京工商大学计算机学院，北京
[2]北京联合大学管理学院，北京
Email: *3056145528@qq.com

收稿日期：2020年10月6日；录用日期：2020年10月21日；发布日期：2020年10月28日

摘　要

为了研究卷积神经网络在人脸表情识别中的应用，设计了一种10层的卷积神经网络模型识别人脸表情，最后一层用Softmax函数将表情的分类结果输出。首先，研究了卷积神经网络的卷积和池化算法并设计了模型的结构。其次，为了更形象地展示卷积层提取的特征，把提取的特征做了可视化处理并以特征图的形式展示。本文的卷积神经网络模型在Fer-2013数据集上进行了实验，实验结果展示了识别率的优越性。为了验证模型识别的泛化能力，最后自制了一个自然状态下的人脸表情数据集，并对人脸图片做了裁剪，灰度化以及像素调整等一系列的预处理。用本文模型识别该数据集中的人脸表情图片，识别的准确率达85.1010%。

关键词

表情识别，卷积神经网络，深度学习，特征提取，图像分类

Research on Facial Expression Recognition Based on Convolutional Neural Network

Liping Sun[1], Hongqian Chen[1], Hui Li[2*]

[1]School of Computer and Information Engineering, Beijing Technology and Business University, Beijing
[2]School of Management, Beijing Union University, Beijing
Email: *3056145528@qq.com

Received: Oct. 6th, 2020; accepted: Oct. 21st, 2020; published: Oct. 28th, 2020

Abstract

In order to study the application of CNN in the field of facial expression recognition, the 10-layer

*通讯作者。

文章引用: 孙丽萍, 陈红倩, 李慧. 用于人脸表情识别的卷积神经网络研究[J]. 计算机科学与应用, 2020, 10(10): 1843-1852. DOI: 10.12677/csa.2020.1010194

CNN model is designed. The last layer of said model employs Softmax function to output the expression classification results. Firstly, this study concentrates on the convolution and pooling algorithm as well as the design structure of the model. In addition, the study visualized the extracted features and displayed them in the form of feature maps to show the features extracted by every convolutional layer. The study conducted experiments on the Fer-2013 dataset, and the result demonstrated the efficacy of the model. It is known that the Fer-2013 dataset contains data collected in an experimental environment. Therefore, to prove the effectiveness of the model, the study created a facial expression dataset by collecting facial expression images in a natural, spontaneous setting. The trained model, which was previously applied to the Fer-2013 dataset, was tested out on the new dataset. The experiment yielded promising results, one of which in the form of a recognition accuracy rate as high as 85.1010%.

Keywords

Expression Recognition, CNN, Deep Learning, Feature Extraction, Image Classification

1. 引言

随着人工智能的发展，人机交互被广泛研究。根据人类的表情让机器去学习人类的情感是人机交互的一个重要研究部分。人脸表情识别从广义上来讲是一个交叉学科，它的研究涉及计算机视觉，图形图像处理以及心理学等。研究人脸表情识别能够促进更好的人机交互，也是人机交互中一个不可或缺的部分。

研究人脸表情识别的目的：1) 在人机交互中更好地理解人类的情感，从而提升人机交互的体验；2) 在视频片段中可以对人脸表情进行跟踪和识别；3) 研究人脸表情的编码模式，从而更有利于传输以及存储有人脸表情的图片。人脸表情识别在安防，心理学，医疗，客户满意度分析以及网络教学等领域有着广阔的应用前景。

从上世纪 70 年代开始研究人类面部表情，并对人类的表情进行了分类。在传统的表情识别系统中，把这一过程分为了特征提取和表情分类。提取脸部特征的方法有 Gabor 滤波器[1]；方向直方图 HOG；离散余弦变换(DCT)和尺度不变特征变换(SIFT)等，然后利用 SVM [2]或者 PCA [3]对表情进行分类。随着深度学习的发展，深度学习被应用在了表情识别中。深度神经网络模型能够同时提取图像特征和图像分类，因此也为表情识别带来了极大的便利。

在计算机视觉中，卷积神经网络由于自身的卷积与池化操作，因此在处理图形图像中比其他的神经网络有更好的性能。本文设计了一种新颖的卷积神经网络结构进行表情特征的提取与分类。本文的模型借鉴了 VGG 网络的思想，设计了一个卷积网络结构，并对网络结构的参数进行了调整。我们受到 GoogLeNet [4]网络的启发在卷积神经网络的第一层添加了一个 1 * 1 的卷积核来增加输入的非线性表示。最后在全连接层我们通过丢弃一部分神经元，来简化了模型的复杂性。

本文的贡献：1) 借鉴 VGG 网络思想设计了一种新的卷积神经网络结构。2) 通过 Fer-2013 数据集训练模型并验证模型识别的准确率。3) 自制了一个在自然状态下人脸表情数据集，并根据自制的数据集验证模型的识别泛化能力。

2. 相关工作

卷积神经网络分为正向传播和反向传播两个过程，正向传播进行卷积和池化操作，这两个操作是为了提取图像特征和处理图像特征。反向传播是采用 BP 算法传递误差，从而使用优化算法进行模型参数的更新。

2012 年在 ILSVRC 挑战赛中，Krizhevsky 等人[5]将深度卷积神经网络用于图像分类中，并在挑战赛取得了很好的效果，自此卷积神经网络被广泛应用在图像识别中。陈等人[6]研究了卷积和池化算法识别人脸表情，指出固定池化的一些局限性，并提出动态自适应的池化算法。卢等人[7]和 Jeon 等人[8]分别通过设计一种卷积神经网络模型进行表情识别，表情识别的准确率不太理想。Arriaga [9]等人分别设计了同时识别性别和表情的卷积神经网络。徐等人[10]设计了并行的卷积神经网络识别表情，在模型训练过程中减少了训练时间。为提高识别的准确率，卷积神经网络通常还会融合另一种模型进行表情识别[11] [12]。例如王等人[11]融合了卷积神经网络和支持向量机，卷积神经网络只提取特征，用支持向量机代替全连接层进行分类。黄等人[13]和李等人[14]分别提出了跨连接的卷积神经网络，不同卷积层提取的特征不同，利用跨连接保留不同层的特征从而提高识别率。钱等人[15]使用卷积神经网络提取了不同视角下的人脸表情特征，从而提取的特征比较准确详细更有利于分类表情。

3. CNN 结构设计

卷积神经网络是一种在提取图像特征方面有独特优势的神经网络。表情识别是属于分类监督性学习，利用含有标签的表情图片训练一个卷积神经网络的分类模型。卷积神经网络模型的正向传播是卷积和池化操作，利用反向传播算法传递误差，使用随机梯度下降(stochastic gradient descent, SGD)优化算法对模型的参数进行训练和优化。本文设计的用于表情识别的卷积神经网络由输入层，4 个卷积层，3 个池化层，2 个全连接层和 SoftMax 层组成，其结构如图 1。

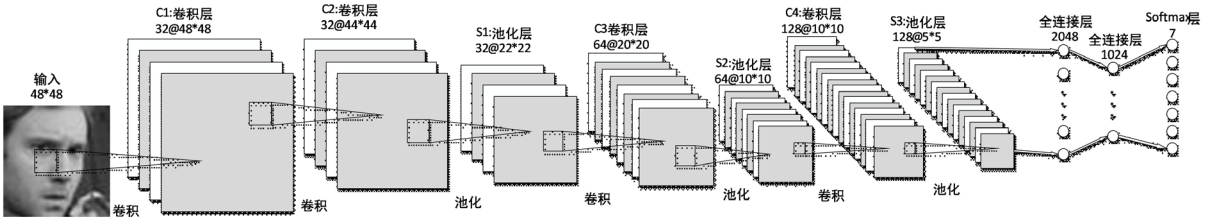

Figure 1. The structure of our CNN
图 1. 本文的卷积神经网络结构

3.1. 卷积层

卷积神经网络的卷积层在人脸表情图片上进行卷积操作，提取人脸表情特征。输入层直接用图片像素作为输入值，然后对输入值进行卷积操作。本文采用了不同大小的卷积核进行特征提取，不同大小的卷积核代表了感受野的不同，因而使用了不同的卷积核提取不同感受野的表情特征，卷积层的表达式如公式 1。

$$\boldsymbol{C}_i = f\left(\boldsymbol{x} * \boldsymbol{w}_i + \mathrm{b}_i\right) \tag{1}$$

其中 $\boldsymbol{C}_i$ 表示第i个卷积得到的输出结果，f(·)表示激活函数，激活函数选择了修正线性单元函数(Rectified Linear Units, ReLU)，$\boldsymbol{x}$ 表示输入的图像值，*表示卷积操作，$\boldsymbol{w}_i$ 表示第i个卷积核，b_i 表示第i个卷积核的偏置。ReLU 函数的表达式如公式 2。

$$\mathrm{ReLU}(y)=\begin{cases} y, y \geq 0 \\ 0, y < 0 \end{cases} \tag{2}$$

其中 $y = \boldsymbol{x} * \boldsymbol{w}_i + \boldsymbol{b}_i$。本文总共使用了 4 个卷积层，卷积核大小依次为：1 * 1，5 * 5，3 * 3，3 * 3，卷积核个数依次为 32-32-64-128。卷积层之后经过激励层输出，我们在第二层卷积之前使用了 1 * 1 的卷积核是为了增加输入的非线性表示，同时加深了模型的网络结构，提升模型的表达能力。图片的输入为 48 * 48 的矩阵，使用 32 个 1 * 1 的卷积核卷积后，输出 32 个 48 * 48 的特征图。第二层卷积使用 5 * 5 的卷积核是为了首先在大的感受野内提取特征，之后缩小卷积核的尺寸，在更小的区域内提取特征。在 48 * 48 的特征图上使用 5 * 5 的卷积核进行卷积，得到 32 个(48 – 5 + 1) * (48 – 5 + 1)的特征图，使用 32 个卷积核是提取了 32 个不同的局部表情特征，第三和第四个卷积层分别使用了 3 * 3 的卷积核，每一层网络具体的参数值见表 1。

卷积神经网络的每一层卷积操作都进行了特征提取，本文对每一层不同卷积核提取的特征进行了融合，并对提取的特征图进行了可视化展示。使用了 Fer2013 数据集中的一张人脸表情图片做了演示，每一层卷积操作提取特征后的特征图如图 2 所示。

Table 1. The parameters of CNN
表 1. 卷积神经网络结构的参数

网络层	输入尺寸	卷积核大小	池化区域	步长	有无填充	输出尺寸
Layer 1 (卷积)	48*48	32@1*1		1	无	32@48*48
Layer 2 (卷积)	32@48*48	32@5*5		1	无	32@44*44
Layer 3 (池化)	32@44*44		2*2	2	无	32@22*22
Layer 4 (卷积)	32@22*22	64@3*3		1	无	64@20*20
Layer 5 (池化)	64@20*20		2*2	2	无	64@10*10
Layer 6 (卷积)	64@10*10	128@3*3		1	有	128@10*10
Layer 7 (池化)	128@10*10		2*2	2	无	128@5*5
Layer 8	1*3200		全连接层 Dropout(0.6)			1*2048
Layer 9	1*2048		全连接层 Dropout(0.4)			1*1024
Layer 10	1*1024		SoftMax 层			1*7

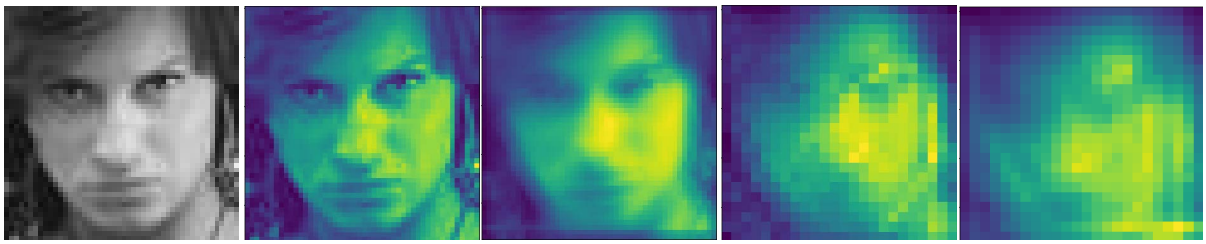

Figure 2. Feature Map after convolution
图 2. 卷积后的特征

3.2. 池化层

卷积神经网络的池化层通常设计在卷积层之后，特征图的数量会随着卷积层数的增加而增加，然而特征维数的增加会造成维度灾难，因此在卷积层后面通常连接池化层进行降维。本文利用了最大池化操

作保持一个池化区域中最显著的特征。池化层可表示为如公式 3。

$$S_i = down\left(\max\left(y_{a,b}\right)\right) \quad a,b \in p_i \tag{3}$$

其中 S_i 表示第 i 个池化区域的最大池化结果，$down(\cdot)$表示下采样过程(保留池化区域的最大值)，$y_{a,b}$ 表示池化区域中的值，p_i 表示第 i 个池化区域。本文网络结构的第三层是池化层，输入第三层的特征图是 44 * 44，池化区域为 2 * 2，所以在特征图中，2 * 2 代表一个池化窗口，每个池化窗口得出一个最大池化结果，因而特征图最终池化的结果为(44/2) * (44/2)。每一层池化层的参数详见表 1。

3.3. 全连接层

全连接层的神经元和上一层的神经元两两相连接，从而把特征维度转化为一维数据。本文最后一层池化层连接全连接层，最后一层池化层输出 128 个卷积的 5 * 5 的特征图，转化为一维数据为：128 × 5 × 5 = 3200，然后输入 1 × 3200 的数据到全连接层，全连接层的表示如公式 4。

$$\text{Full} = f(w \times z + b) \tag{4}$$

其中 Full 表示全连接层输出结果，f(·)是 ReLU 激活函数，$\boldsymbol{w}$ 代表连接的权重值，$\boldsymbol{z}$ 是输入全连接层的值，b 是偏置，本文的网络结构设计了两个全连接层，为了降低网络结构的复杂度和防止过拟合，采用了神经元的随机失活(Dropout)。

3.4. Softmax

本文网络结构的最后一层是 SoftMax 函数把人脸的 7 种表情进行分类。本层有 7 个神经元，每个神经元代表一个表情类别，针对每个输入的人脸图片，SoftMax 层的 7 个神经元分别输入 0 到 1 之间的概率，输入概率值最大的神经元，则代表此神经元对应的表情概率最大。SoftMax 分类的表示如公式 5 所示。

$$p\left(y = c \middle| m; w\right) = \frac{e^{w_c \times m}}{\sum_{i=1}^{k} e^{w_i \times m}} \tag{5}$$

其中 $p\left(y = c \middle| m; w\right)$ 表示输入的图片 m 是表情种类c的概率，w 为权重参数值(待拟合)，k 为类别总数 7。表情种类 c 的取值为{0, 1, 2, 3, 4, 5, 6}。

4. 实验

本文的实验使用 python 实现，是基于 Keras 的深度学习平台上进行，另外我们把其中对比的两篇论文中的模型也使用了 Python 进行了复现，为了对实验结果公平比较，我们使用了统一数据集对不同模型进行了训练。

4.1. 数据集

本文使用了两个数据集，一个是 Fer2013 [16]人脸表情数据集，另一个是我们自制的数据集。Fer2013 表情数据库有 35,886 张人脸表情图片，其中训练集有 28,708 张，验证集和测试集各有 3589 张。每张灰度图片的大小都是 48 * 48，数据集共有 7 种表情如：生气，厌恶，恐惧，开心，伤心，惊讶，中性。

Fer2013 数据集是在实验室环境下进行的采取，从而不能很好的验证模型在自然状态下的人类表情的识别情况，因此我们从网络上搜索了人类在自然状态下一些表情图片，然后对图片的尺寸，像素，背景等进行了预处理，把图片统一转换为了灰度图。最终形成一个小型的数据集。自制数据集上人脸表情分为了 7 种表情，共 396 张图片。本文使用了以上两个数据集共同验证提出的卷积神经网络模型的性能。

4.2. 模型的训练

为了训练更准确的模型，更高效率的利用表情图片，对表情图库经过一系列的随机变换进行了数据扩增，如图 3。

Figure 3. Data enhancement
图 3. 数据增强

本文利用的损失函数是多分类的交叉熵损失函数，如公式 6：

$$\text{loss} = -\sum_{i=1}^{n} y_{i1} \log a_{i1} + y_{i2} \log a_{i2} + \cdots + y_{i7} \log a_{i7} \tag{6}$$

其中 a 是神经元实际的输出值，y 是期望输出值。

训练目标就是最小化损失值，用反向传播算法传播误差值，采用 SGD 优化算法沿着梯度下降的方向更新参数值。SGD 算法如公式 7：

$$\frac{\partial loss}{\partial \theta_{i1}} = -\sum_{i=1}^{n} \frac{\theta_{i1}}{a_{i1}} \tag{7}$$

故而参数更新如公式 8：

$$\theta_j = \theta_j - a\frac{\partial loss}{\partial \theta_j} \tag{8}$$

其中 θ_j 是待更新参数，a 为学习率，$\frac{\partial loss}{\partial \theta_j}$ 在梯度下降方向减少的值。本模型中的学习率为 0.01，为了让训练收敛到最佳结果，本文设置随着训练次数的增加学习率逐渐衰减，故而学习步长逐渐减小。

本文首先使用了 Fer2013 数据集中的训练集训练模型，然后用验证集验证识别的准确率，当验证集的准确率下降，损失值上升的时候，停止训练。

4.3. 实验结果与分析

本文把卢[7]等人提出的卷积神经网络模型，李[14]等人提出的 LeNet-5 模型用 python 进行了复现。

并用 Fer2013 数据集训练并在测试集上计算了准确率。本文的模型，卢[7]等人的模型以及李[14]等人的模型的训练结果分别如图 4~6。

对于训练结果，我们分别选取在验证集上准确率最高的训练模型，各种模型在测试集中的准确率总结如表 2。从表 2 中可以看出我们提出的模型在测试集上的准确率相对较高，准确率为：72.92%。

Table 2. The accuracy of all model
表 2. 各模型的准确率

	迭代次数	验证集准确率	测试集准确率
文献[7]	76	0.5811	0.6455
文献[8]			0.7074
LetNet 模型[14]	116	0.5646	0.7142
本文模型	76	0.6400	0.7292

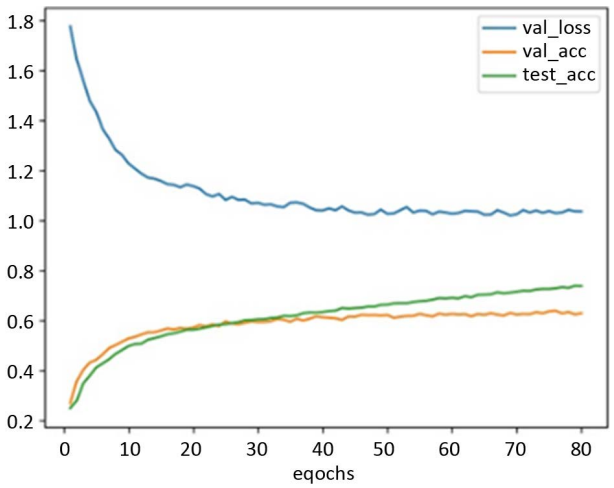

Figure 4. The result of our model
图 4. 本文模型的训练结果

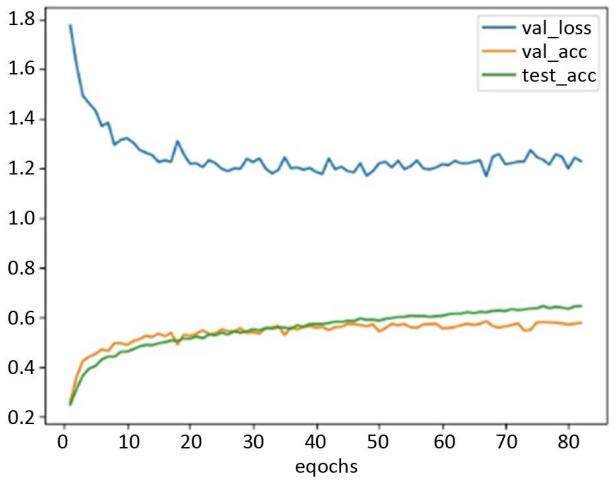

Figure 5. The result of Lu *et al.*
图 5. 卢等人模型的训练结果

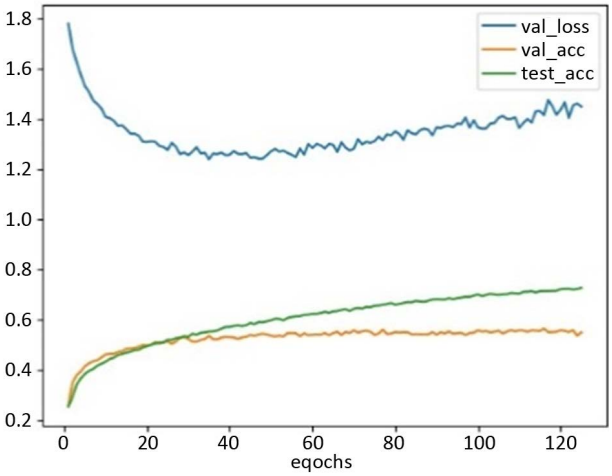

Figure 6. The result of Li *et al.*
图 6. 李等人模型的训练结果

本文的模型用 Fer2013 数据集训练后并保存训练模型，然后用训练的模型识别我们自制数据集中的图片，例如单个图片识别结果如图 7。我们对整个我们自制的数据集进行识别，识别结果的混淆矩阵如表 3。

Table 3. Our dataset confusion matrix
表 3. 识别自制数据集的混淆矩阵

		预测							
		生气	厌恶	恐惧	开心	伤心	惊讶	中性	识别率
实际	生气	33	0	3	0	0	1	0	89.1891%
	厌恶	5	10	0	2	3	0	2	45.4545%
	恐惧	0	0	22	0	1	13	2	57.8947%
	开心	1	0	1	134	0	0	5	95.0354%
	伤心	0	0	2	1	20	0	7	66.6667%
	惊讶	0	0	1	2	0	39	1	90.6977%
	中性	1	0	2	2	1	0	79	92.9412%

Figure 7. The result of expression recognition
图 7. 表情识别结果

从混淆矩阵中，我们可以看出开心，中性，惊讶表情的识别的准确率比较高，但是在识别厌恶和恐惧表情时识别效果比较差。对于厌恶表情，每个人的表达方式不同，面部表情也有很大差异，所以在识别厌恶表情时，识别结果比较零散，可能会识别成各类表情。识别恐惧表情时，易识别于惊讶，主要是提取眼睛部位的特征时，恐惧和惊讶都容易让人的瞳孔放大，因此恐惧会识别为惊讶。在自制的自然状态下的表情数据集上总体的准确率为 337/396 = 85.1010%。

在表情识别中我们分析了人脸表情和表情识别中存在的几个难点，人类是一种复杂的动物，内心世界也是很丰富的，人脸上的表情有时是多种情感交织在一起，例如人脸表情可能同时存在惊讶，生气，无奈等多种表情，这对于识别就有一定难度。人类有时不同的表情可能表达相同的情感，相同的表情针对不同的人可能情感不同，这就要求严格的提取人脸的细微特征。最后，人类的五官特征都有各自的特色，不能一概而论，例如在表情识别过程中大眼睛的人的表情更容易识别成惊讶或者恐惧。

5. 结论

卷积神经网络能够自动隐式地学习人脸的表情特征，无需人为提取，可以使用图像的像素点作为输入进行训练模型。本文利用了卷积神经网络处理图片的优点，设计了一种卷积神经网络结构模型进行人脸表情的识别。使用表情数据集 Fer2013 训练模型，在该数据集上的实验结果表明了所提出方法的优越性。另外我们在自制的自然状态下的数据集上测试该模型的识别泛化能力，泛化能力相对较好。卷积神经网络需要大量的数据集训练，训练的模型才能学到很好的分类效果，因此收集更多自然状态下的人脸表情的图片训练模型，表情识别的模型将更具有泛化能力。

基金项目

国家自然科学基金(31701517)；北京市社会科学基金(17GLC060)；“十三五”时期北京市属高校高水平教师队伍建设支持计划－青年拔尖人才培育计划项目(CIT&TCD201704039)。

参考文献

[1] Kumbhar, M., Jadhav, A. and Patil, M. (2012) Facial Expression Recognition Based on Image Feature. *International Journal of Computer and Communication Engineering*, **1**, 117-119. https://doi.org/10.7763/IJCCE.2012.V1.33

[2] Reddy, C., Reddy, U. and Kishore, K. (2019) Facial Emotion Recognition Using NLPCA and SVM. *Traitement Du Signal*, **36**, 13-22. https://doi.org/10.18280/ts.360102

[3] 邓洪波, 金连文. 一种基于局部 Gabor 滤波器组及 PCA+LDA 的人脸表情识别方法[J]. 中国图象图形学报, 2007, 12(2): 322-329.

[4] Szegedy, C., Liu, W., Jia, Y.Q., *et al*. (2014) Going Deeper with Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 1-9. https://doi.org/10.1109/CVPR.2015.7298594

[5] Krizhevsky, A., Sutskever, I. and Hinton, G. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, **25**, 1106-1114.

[6] 陈航, 邱晓晖. 基于卷积神经网络和池化算法的表情识别研究[J]. 计算机技术与发展, 2019, 29(1): 61-65.

[7] 卢官明, 何嘉利, 闫静杰, 等. 一种用于人脸表情识别的卷积神经网络[J]. 南京邮电大学学报: 自然科学版, 2016, 36(1): 16-22.

[8] Jeon, J., *et al*. (2016) A Real-Time Facial Expression Recognizer Using Deep Neural Network. *Proceedings of the* 10*th International Conference on Ubiquitous Information Management and Communication*, 1-4. https://doi.org/10.1145/2857546.2857642

[9] Arriaga, O., Valdenegro-Toro, M. and Plöger, P. (2017) Real-Time Convolutional Neural Networks for Emotion and Gender Classification.

[10] 徐琳琳, 张树美, 赵俊莉. 构建并行卷积神经网络的表情识别算法[J]. 中国图象图形学报, 2019, 24(2): 227-236.

[11] 王忠民, 李和娜, 张荣. 融合卷积神经网络与支持向量机的表情识别[J]. 计算机工程与设计, 2019, 40(12):

3594-3600.

[12] 孙晓, 潘汀, 任福继. 基于 ROI-KNN 卷积神经网络的面部表情识别[J]. 自动化学报, 2016, 42(6): 883-891.

[13] 黄倩露, 王强. 基于跨连特征融合网络的面部表情识别[J]. 计算机工程与设计, 2019, 40(10): 2969-2973.

[14] 李勇, 林小竹, 蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别[J]. 自动化学报, 2018, 44(1): 176-182.

[15] 钱勇生, 邵洁, 季欣欣, 等. 基于改进卷积神经网络的多视角人脸表情识别[J]. 计算机工程与应用, 2018, 54(24): 12-19.

[16] Goodfellow, I.J., Erhan, D., Carrier, P.L., *et al*. (2013) Challenges in Representation Learning: A Report on Three Machine Learning Contests. *Neural Networks*, **64**, 59-63.