基于跨连接 LeNet-5 网络的面部表情识别

李 勇 [1, 2] 林小竹 [1] 蒋梦莹 [1, 2]

摘 要 为避免人为因素对表情特征提取产生的影响, 本文选择卷积神经网络进行人脸表情识别的研究. 相较于传统的表情识别方法需要进行复杂的人工特征提取, 卷积神经网络可以省略人为提取特征的过程. 经典的 LeNet-5 卷积神经网络在手写数字库上取得了很好的识别效果, 但在表情识别中识别率不高. 本文提出了一种改进的 LeNet-5 卷积神经网络来进行面部表情识别, 将网络结构中提取的低层次特征与高层次特征相结合构造分类器, 该方法在 JAFFE 表情公开库和 CK+ 数据库上取得了较好的结果.

关键词 卷积神经网络, 面部表情识别, 特征提取, 跨连接

引用格式 李勇, 林小竹, 蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别. 自动化学报, 2018, **44**(1): 176−182

DOI 10.16383/j.aas.2018.c160835

Facial Expression Recognition with Cross-connect LeNet-5 Network

LI Yong[1, 2] LIN Xiao-Zhu[1] JIANG Meng-Ying[1, 2]

Abstract In order to avoid the influence of human factors on facial expression feature extraction, convolution neural network is adopted for facial expression recognition in this paper. Compared with the traditional method of facial expression recognition which requires complicated manual feature extraction, convolutional neural network can omit the process of feature extraction. The classical LeNet-5 convolutional neural network has a good recognition rate in handwritten digital dataset, but a low recognition rate in facial expression recognition. An improved LeNet-5 convolution neural network is proposed for facial expression recognition, which combines low-level features with high-level features extracted from the network structure to construct the classifier. The method achieves good results in JAFFE expression dataset and the CK+ dataset.

Key words Convolutional neural network, facial expression recognition, feature extraction, cross-connect

Citation Li Yong, Lin Xiao-Zhu, Jiang Meng-Ying. Facial expression recognition with cross-connect LeNet-5 network. *Acta Automatica Sinica*, 2018, **44**(1): 176−182

随着计算机的快速发展, 人机交互越来越多地出现在日常生活中, 如何让计算机更好地理解人类的心理是人机交互必须要解决的问题. 人的面部表情中包含丰富的信息, 研究指出, 面部表情可以比动作和语言更好地表达人类的心理活动[1], 面部表情识别也因此成为了人机交互中不可或缺的部分. 一个普通人可以很好地读取别人的面部表情并做出相应的判别, 但对于计算机来说这是一项十分困难的任务, 为此大量的专家学者投入到该领域的研究中来. 面部表情是一个十分复杂的系统, 各国研究者们构建了不同的模型来实现表情分类, 其中最具代表性的就是 Ekman 等[2] 在 1978 年提出了面部动作编码系统 (Facial action coding system, FACS), 随后在 1984 年根据不同动作单元的组合定义了 6 种基本的表情: 生气、厌恶、开心、悲伤、惊讶、恐惧. 研究者们据此构建了不同的表情库, 其中日本的 JAFFE 公开库就是采用这六种基本表情构建的表情库, 并且在此基础上增加了第 7 类表情: 无表情. 卡耐基梅隆大学的 Lucey 等[3] 于 2010 年在 Cohn-Kanade dataset 的基础上发布了 The extended Cohn-Kanade dataset (CK+), 这个库中包括了 123 个人的 593 个视频序列, 其中有 327 个序列是包含表情标签的序列, 该库中的表情除了 6 种基本的表情以外增加了蔑视和无表情两种. 进入上世纪 90 年代以后, 面部表情识别与分析迅速发展起来, 研究者们提出了不同的算法来提高识别的准确率, 主要的方法有两种: 1) 是基于几何的方法, 例如 Lanitis 等[4] 采用几何特征方法进行识别, 该方法是通过标记人眼、口、鼻等特征点, 计算其相对位置来识别表情, 虽然这种方法大大减少了输入的数据, 但是仅用有限的点来表示复杂的人脸表情显然会丢

收稿日期 2016-12-23 录用日期 2017-05-04
Manuscript received December 23, 2016; accepted May 4, 2017
国家自然科学基金 (60772168) 资助
Supported by National Natural Science Foundation of China (60772168)
本文责任编委 胡清华
Recommended by Associate Editor HU Qing-Hua
1. 北京石油化工学院信息工程学院 北京 102617 2. 北京化工大学信息科学与技术学院 北京 100029
1. School of Information Engineering, Beijing Institute of Petrochemical Technology, Beijing 102617 2. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029

失很多重要的信息, 因而整体的识别率并不太高; 2) 是基于整体的识别方法, 例如 Praseeda 等[5] 使用 Gabor 小波和 SVM 相结合的方式进行面部表情识别, 首先用 Gabor 滤波器对表情图像滤波、提取特征, 将提取到的特征用于 SVM 训练, 训练分类器进行表情的分类识别, 这种方法同样依赖于前期人工提取特征的优劣, 人为干扰因素较大. 近年来, 随着计算机运行速度的提高, 处理大数据成为可能, 同时互联网的快速发展, 研究者采集大量的数据变得相对容易, 在此基础上, 深度卷积神经网络被证实了在图像识别领域有巨大的优势. Krizhevsky 等[6] 于 2012 年在 ImageNet 图像数据集上使用 AlexNet 卷积神经网络结构取得惊人的成绩, 其识别率远超传统的识别方法. 这个数据集包含约 120 万张训练图像、5 万张验证图像和 10 万张测试图像, 分为 1 000 个不同的类别, 传统的特征提取方法被网络结构取代, 网络可以自行提取特征并分类而不需要人工干预.

1　深度学习与卷积神经网络

2006 年, 机器学习领域泰斗 Hinton 与他的学生在 *Science* 上发表的文章[7] 掀起了深度学习研究的浪潮, 多隐层的神经网络再次回到人们视野之中. 在那以后, 斯坦福大学、纽约大学、蒙特利尔大学等名校迅速成为深度学习研究的重要场所, 甚至美国国防部 DARPA 计划也首次资助了深度学习项目[8]. 卷积神经网络作为深度学习的一支, 也迅速受到了广泛的关注. 现今, 深度学习广泛地应用于监控视频事件检测[9]、自然语言处理[10]、语音信号的基音检测[11]、图像分类与识别等领域[12−15].

卷积神经网络虽然是在近年来才受到广泛的关注和应用, 但早在 1962 年 Hubel 等[16] 就通过对猫视觉皮层细胞的研究, 提出了感受野 (Receptive field) 的概念. 1984 年日本学者 Fukushima 等[17] 基于感受野的概念而提出的神经认知机 (Neocognitron) 可以看作是第一个实现了的卷积神经网络, 这也是感受野概念在人工神经网络领域的首次应用. Le Cun 等[18] 提出的深度卷积神经网络, 就是以神经认知机为基础, 并使用了反向传播算法来识别手写数字, 后来在 1998 年正式确定的 LeNet-5 模型, 在文档识别中取得了很好的效果[19], 该模型当年成功用于美国大多数银行支票的手写数字识别, 是卷积神经网络在工业界最早的应用. Le Cun 设计的 LeNet-5 卷积网络结构图如图 1 所示.

LeNet-5 卷积神经网络可以看成是一个多隐层的人工神经网络, 其基本结构主要包括输入层、卷积层、池化层、全连接层和输出层. 其中卷积层和池化层会交替出现, 构成了特殊的隐层. 如图 1 所示, Input 是输入层, 输入大小为 32 像素 × 32 像素的图片, MNIST 手写数字库中图片大小为 28 像素 × 28 像素, 所以实际使用时将其扩展为大小为 32 像素 × 32 像素的图片使用. Layer 1 层是卷积层, 共有 6 个特征图, 每一副输入为 32 像素 × 32 像素的图片都与 6 个不同的大小为 5 像素 × 5 像素卷积核卷积, 得到大小为 28 像素 × 28 像素的特征图. Layer 2 层是池化层, Layer 1 中 6 个大小为 28 像素 × 28 像素的特征图经过池化后得到 6 个大小为 14 像素 × 14 像素的特征图. Layer 3 层是卷积层, 共有 16 个大小为 10 像素 × 10 像素的特征图. 每个 10 像素 × 10 像素的特征图是由前一层的某几个或全部特征图与 5 像素 × 5 像素的卷积核卷积得到, 具体连接方式如表 1 所示. Layer 4 层是池化层, 共 16 个 5 像素 × 5 像素的特征图, 由 Layer 3 层 16 个特征图经过池化得到. Layer 5 层是卷积层, 共有 120 个 1 像素 × 1 像素的特征图, 每一个特征图都是由 Layer 4 层所有的特征图与 5 像素 × 5 像素卷积核卷积得到. Full 层是全连接层, 共有 84 个单元, Layer 6 层与 Layer 5 层为全连接关系. Output 层是输出层, 输出分类结果.

不同于传统的神经网络, 卷积神经网络采取的是局部连接 (Locally-connection) 的方式, 不仅有效地减少了与神经元连接的参数个数, 而且在误差反向传播过程中, 让梯度在一个较小的广度范围内传播[20], 使得训练变得更加容易. 权值共享也是卷积神经网络的一个特点, 对于输入图像的每一个小块, 用相同的卷积核进行卷积操作, 这种方法来源于局部感受野的概念, 可以使得图像具有平移不变性. 最后, 卷积神经网络中的池化操作, 就是一次下采样操

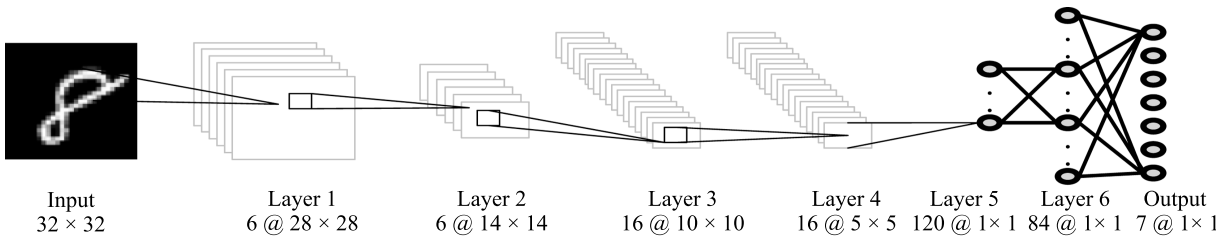


图 1　LeNet-5 结构图

Fig. 1　The LeNet-5 convolutional neural network

表 1 LeNet-5 网络 Layer 2 与 Layer 3 之间的连接方式

Table 1 Connection between LeNet-5 network′s Layer 2 and Layer 3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	√				√	√	√			√	√	√	√		√	√
2	√	√				√	√	√			√	√	√	√		√
3	√	√	√				√	√	√			√		√	√	√
4		√	√	√			√	√	√	√			√		√	√
5			√	√	√			√	√	√	√		√	√		√
6				√	√	√			√	√	√	√		√	√	√

作, 将相邻的几个像素点用一个像素点代替, 根据采样方法的不同分为 Max-pooling 和 Avg-pooling, 该操作可以使图片具有一定的缩放不变性.

2 改进的 LeNet-5 网络

LeNet-5 卷积神经网络是通过不同的卷积核自行提取特征, 将原始数据经过一些简单的非线性的模型转变为更高层次的、更加抽象的表达, 最终使用高层次的特征进行分类识别. 然而这种分类方法没有考虑到低层次的细节特征, 而且随着网络深度加深, 网络训练的困难程度在增加, 尤其是梯度消失或爆炸问题[21], 为了能够解决深层网络的训练问题, 研究者提出了跨层的连接方式. 早期训练的多层感知机通常将输入作线性变换后加到输出上[22], 近年来 Srivastava 等[23] 提出了一种新的网络连接结构 Highway networks, 该结构主要特点是提供了一种门限机制, 一部分的特征不需要经过处理直接通过某些网络层, 该结构更加容易优化, 并且在 CIFAR-10 数据集上表现优于 Romero 等[24] 提出的 FitNets. 而 He 等[25] 提出的深度残差网络同样在标准的前馈卷积网络上, 增加了一些跨层的连接, 目的也是为了降低训练的难度, 该结构不仅在层数上刷新了记录, 而且在 ImageNet 和 COCO 几个主要的任务中都取得了优异的成绩. Sun 等[26] 提出的 DeepId 网络中也有设计将最后的池化层和卷积层与全连接层相连, 张婷等[27] 提出的跨连的卷积神经网络 (Cross-connected convolutional neural net-work, CCNN) 可以有效地将低层次特征与高层次特征结合起来, 构造出更好地分类器, 在性别分类中取得了不错的结果. 本文在 LeNet-5 的结构基础上引入跨连的思想, 将 LeNet-5 网络的两个池化层与全连接层相结合用于最后的分类器构造中.

虽然 LeNet-5 在手写数字集上取得了巨大的成功, 但是将该结构用于表情识别时却难以得到理想的结果, 本文提出了改进的 LeNet-5 结构如图 2 所示, 包括一个输入层、3 个卷积层、2 个池化层、一个全连接层和一个输出层. 网络输入是 32 像素 × 32 像素的图片, 经过卷积池化操作后将前两个池化层与全连接层结合起来作为 softmax 分类器的输入, 最终获得 7 种表情的分类输出. 表 2 为各层的网络参数.

整个网络的训练过程分为正向传播和反向传播, 其中正向传播过程就是隐层提取特征的过程, 主要是卷积和池化操作. 反向传播采用 BP 反向传播算法传递误差, 使用随机梯度下降算法, 更新权值参数. 给定输入, 整个网络的计算过程如下:

1) 对于卷积层输入 X, 卷积过程

$$con = f\left(\sum_{i,j\in\ M} x_{i,j} * w_{m-i,n-j} + b\right) \qquad (1)$$

其中, x 表示输入 X 中卷积区域 M 中的元素, w 表示卷积核中的元素, m, n 表示卷积核的大小, b 表示偏置, $f(\cdot)$ 表示 ReLU 激活函数. 卷积核大小及卷积步长如表 2 所示.

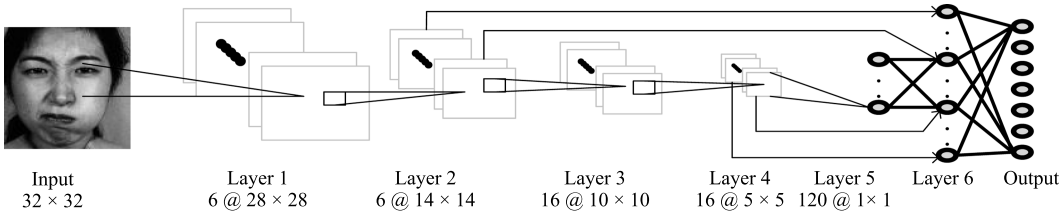


图 2 改进的 LeNet-5 卷积神经网络

Fig. 2 Improved LeNet-5 convolutional neural network

表 2　卷积网络参数

Table 2　Convolutional network parameters

输入	输入尺寸	卷积核大小	池化区域	步长	输出尺寸
Input	32 × 32	5 × 5		1	28 × 28
Layer 1	6 @ 28 × 28		2 × 2	2	6@14 × 14
Layer 2	6 @ 14 × 14	5 × 5		1	10 × 10
Layer 3	16 @ 10 × 10		2 × 2	2	16 @ 5 × 5
Layer 4	16 @ 5 × 5	5 × 5		1	120@1 × 1
Layer 5	120 @ 1 × 1				1 × 84
Layer 6	1 × 1 660				1 × 7
Output	1 × 7				

2) 对于池化层输入 Y, 池化过程

$$pool = down(\max(y_{i,j})), \quad i,j \in p \tag{2}$$

其中, y 表示池化层输入 Y 中池化区域 p 中的元素, $down(\cdot)$ 是下采样过程, 保留池化区域中的最大值. 池化区域大小及步长如表 2 所示.

3) 对于全连接层输入 Z

$$full = f(w \times z + b) \tag{3}$$

其中, z 表示输入 Z 中的元素, w 表示权值, b 表示偏置, $f(\cdot)$ 表示 ReLU 激活函数.

4) 对于输出层输入 X

$$p(y=j|x;w) = \frac{\mathrm{e}^{w_j \times x}}{\sum\limits_{l=1}^{k} \mathrm{e}^{w_l \times x}} \tag{4}$$

式 (4) 为 Softmax 分类器的假设函数, 计算的是输入分类为类别 j 时的概率, w 是权值参数, k 为总的类别数. 其损失函数为

$$J(w) = -\left[\sum_{l=1}^{k} 1\{j=1\} \frac{\log(\mathrm{e}^{w_j \times x})}{\sum\limits_{l=1}^{k} \mathrm{e}^{w_l \times x}}\right] \tag{5}$$

其中, $1\{\cdot\}$ 是示性函数, $1\{$值为真的表达式$\} = 1$, $1\{$值为假的表达式$\} = 0$.

反向传播过程如下: 输入样本得到实际输出之后首先需要计算每一层的反馈传递误差

$$\begin{cases} \delta^7 = out - y \\ \delta^6 = \left[\delta^7 \times (w^7)^{\mathrm{T}}\right] \circ f' \\ \delta^5 = \left[\delta_j^6 \times (w^6)^{\mathrm{T}}\right] \circ f', & 1\,177 \le j \le 1\,260 \\ \delta_i^4 = (\delta_j^5 \oplus w^5), & 1 \le i \le 16,\ 1 \le j \le 120 \\ \delta^3 = up(\delta^4) \circ f' \\ \delta_i^2 = (\delta_j^3 \oplus w^3), & 1 \le i \le 6,\ 1 \le j \le 16 \\ \delta^1 = up(\delta^2) \circ f' \end{cases} \tag{6}$$

式 (6) 为网络各层的反馈传递误差, 式中符号 $\circ$ 表示矩阵或向量中对应的元素相乘. 参考图 2 可知, δ^7 是输出层 (Output) 的反馈传递误差, out 表示网络的实际输出, y 表示网络的目标输出. δ^6 是 Layer 6 层的反馈传递误差, w^7 是 Layer 6 与输出层之间的权值. 本算法采用的是 ReLU 激活函数, f' 表示 ReLU 激活函数的导数. 全连接层是由跨连接组合而成, Layer 5 层只与该层中 ($1\,177 \le i \le 1\,260$) 部分连接, 故误差传递时只需使用 δ^6 ($1\,177 \le j \le 1\,260$) 参与计算, 式中 δ^5 是 Layer 5 层的反馈传递误差, w^6 是 Layer 5 与 Layer 6 之间的权值. δ_i^4 是 Layer 4 层第 i 个特征图对应的反馈传递误差, w^5 是 Layer 4 与 Layer 5 之间的卷积核, 对于每一个 δ_i^4, 都是将 δ_j^5 ($1 \le j \le 120$) 与 w^5 进行外卷积得到, $\oplus$ 表示的是外卷积操作, 参考文献 [27] 定义外卷积与内卷积如下: 假设有 A 和 B 两个矩阵, 大小分别为 $M \times N$, $m \times n$, 其中 M, $N \ge m$, n. 内卷积 $C = A \odot B$, C 中所有元素

$$C_{i,j} = \sum_{p=1}^{m}\sum_{q=1}^{n} a_{i+M-p,j+N-q} b_{p,q} \tag{7}$$

其中, $1 \le i \le M-m+1$, $1 \le j \le N-n+1$. 外卷积定义为

$$A \oplus B = \hat{A} \odot B \tag{8}$$

其中, $\hat{A}$ 是对 A 进行补零扩展, 使其矩阵大小为 (M

$+ 2m - 2) \times (N + 2n - 2)$.

δ^3 是 Layer 3 层的反馈传递误差, $up(\cdot)$ 是一个上采样操作, 同时需要乘以激活函数的导数. δ_i^2 是 Layer 2 层的反馈传递误差, 它的计算过程与 Layer 4 层相似, 每个 δ^2 是由 δ^3 $(1 \le j \le 16)$ 与 w^3 进行外卷积得到, 不过这里要注意 Layer 2 与 Layer 3 之间的连接方式, 具体可参见表 2. δ^1 是 Layer 1 层的反馈传递误差, 与 Layer 3 层类似.

计算权值和偏置的偏导数:

$$\begin{cases} \frac{\partial J}{\partial w^7} = (f6)^{\mathrm{T}} \times \delta^7 \\ \frac{\partial J}{\partial w^6} = (f5)^{\mathrm{T}} \times \delta^6 \\ \frac{\partial J}{\partial b^6} = \delta^6 \\ \frac{\partial J}{\partial w_i^5} = \sum\limits_N [(f4) \odot (\delta^5)] \\ \frac{\partial J}{\partial b_i^5} = \sum\limits_N \delta^5, & 1 \le i \le 120 \\ \frac{\partial J}{\partial w_i^3} = \sum\limits_N [(f2) \odot (\delta^3)] \\ \frac{\partial J}{\partial b_i^3} = \sum\limits_N \delta^3, & 1 \le i \le 16 \\ \frac{\partial J}{\partial w_i^1} = \sum\limits_N [(input) \odot (\delta^1)] \\ \frac{\partial J}{\partial b_i^1} = \sum\limits_N \delta^1, & 1 \le i \le 6 \end{cases} \tag{9}$$

式 (9) 即为各层权值和偏置的偏导数, 其中 $f6$ 为 Layer 6 的特征图, 以此类推, $f5$、$f4$、$f2$ 分别为对应层的特征图, $input$ 是输入图像. 式中 $\odot$ 表示内卷积操作, 定义见式 (7).

对于训练集 $S = \{x^l, y^l\}$ 算法流程如下:

1) 确定迭代次数、网络结构、学习步长, 随机初始化卷积核以及偏置. 其中初始学习步长选择为 0.005, 随着训练次数的增加, 测试误差变化较小时, 将学习步长除以 10, 直至学习步长降至 0.00005, 停止训练.

2) 输入样本, 正向传播, 计算实际输出.

3) 优化目标函数, 采用反向传播算法计算反馈传递误差.

4) 计算参数的修正量.

5) 通过梯度下降法更新参数值.

3 实验

本文所有实验均在 Matlab7.0 上实现, 硬件平台为 Lenovo Tian-Yi 100: Intel(R) Core(TM) i5-5200u CPU, 主频为 2.20 GHz, 内存为 4.00 GB.

3.1 数据集

本实验分别采用 JAFFE 表情数据库和 CK+ 数据库进行实验. JAFFE 表情数据库包含 7 种表情, 分别属于 10 名女性, 每个人每种表情有 2～4 张, 共 213 张图片. 图 3 为 7 种表情的示例图像.

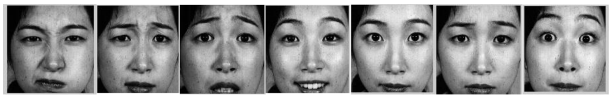


图 3 JAFFE 表情库 7 种表情示例图像

Fig. 3 7 kinds of facial expression image in JAFFE expression dataset

CK+ 数据库中有 123 个人的不同表情序列, 为保持一致性, 在 CK+ 数据库中也只考虑七种表情, 从库中取出七类表情共 990 幅图片, 图 4 为 7 种表情示例图像.



图 4 CK+ 表情库 7 种表情示例图像

Fig. 4 7 kinds of facial expression image in the CK+ expression dataset

3.2 实验结果及分析

将数据集中图片统一裁剪采样至大小为 32 像素 × 32 像素, 采用交叉验证的方法, 将 JAFFE 数据集中图片分为 3 份, 每次取其中两份为训练数据, 另一份为测试数据; 将 CK+ 中图片分为 5 份, 每次取其中 4 份作为训练数据, 1 份作为测试数据.

表 3 为本文算法在 JAFFE 表情库中不同表情的分类结果, 表 4 为本文算法在 CK+ 数据库中不同表情的分类结果. 由表 3 和表 4 可以看到, 在一些测试集上, 整个网络表现较好, 正确率高, 但是在另一些上表现相对较差, 其原因可能是训练样本数据中能提取出的表情特征不足, 无法获得足够的特征进行训练, 导致分类器分类效果较差, 这个现象也反映出了样本数据对于卷积神经网络的重要.

表 5 为不加跨连方式的网络与加了跨连接之后的网络在测试集上的正确率对比. 实验过程中, 由于样本较少, 不加跨连方式的网络训练难度大, 参数调整困难, 而跨连接网络收敛速度快, 训练更容易. 由表 5 可以发现, 直接采用 LeNet-5 网络结构识别正确率很低, 不能很好地分类, 主要原因是 LeNet-5 是设计用于手写数字识别, 相对于数字而言, 面部表情特征更复杂, 然而样本数量却更少, 仅用高层次特征不足以训练得到好的分类器, 本文加入了跨连接的方法后, 低层次特征参与最后的分类器的构造后, 识别效果显著提高, 即使在小样本中也有不错的正确率. 由表 5 还可以发现, 同样的结构在 JAFFE 表情库中得到的正确率高于 CK+ 库中的正确率, 原因是

表 3 JAFFE 表情库不同表情的分类正确率 (%)

Table 3 Classification accuracy of different expressions in JAFFE expression dataset (%)

	生气	厌恶	害怕	高兴	中性	悲伤	惊讶	整体
测试集 1	100	80	100	100	100	90.91	88.89	94.37
测试集 2	100	90	90	81.82	100	100	100	92.96
测试集 3	100	100	81.82	90.91	100	100	100	95.77
整体	100	89.66	90.63	90.63	100	96.77	96.55	94.37

表 4 CK+ 数据库不同表情的分类正确率 (%)

Table 4 Classification accuracy of different expressions in CK+ dataset (%)

	生气	厌恶	害怕	高兴	中性	悲伤	惊讶	整体
测试集 1	88.89	94.44	80	92.86	70.83	96	93.94	88.89
测试集 2	70.37	77.78	80	96.30	68	84	96.97	82.32
测试集 3	77.78	85.71	84.62	100	64	72	93.94	83.33
测试集 4	62.96	94.29	88	89.29	60	80	87.88	80.81
测试集 5	81.48	85.71	72	92.86	64	79.17	100	83.33
整体	76.30	87.59	80.92	94.26	65.37	82.23	94.55	83.74

表 5 网络是否跨连接正确率对比 (%)

Table 5 Classification accuracy of the network whether cross connection or not (%)

方法	参数量	JAFFE 表情库中平均正确率	CK+ 数据库中平均正确率
LeNet-5	14 444	62.44	32.32
本文方法	25 476	94.37	83.74

JAFFE 中只有 10 名亚洲女性的表情图像, 而在 CK+ 中包含了 123 个不同性别不同肤色的人的表情图像, 后者更为复杂, 因而需要的样本数量更多, 否则难以学习到足够多和足够好的特征进行分类.

表 6 为本文方法与传统非深度学习方法的比较, 可以看出, 相对于 SVM 等浅层学习方法, 本文算法在 JAFFE 表情数据上表现较为优异.

表 6 不同方法在 JAFFE 上的对比 (%)

Table 6 The comparison of different methods on JAFFE (%)

来源	方法	正确率
Kumbhar 等[28]*	Image feature	60 ∼ 70
Praseeda 等[5]*	SVM	86.9
本文算法	跨连的 LeNet-5	94.37

* 数据来源于文献 [15]

4 结论

卷积神经网络的特点是自动地、隐式地学习特征, 不需要人为地定义特征, 如果有足够多的样本用于训练, 网络可以学习到很好的特征进行分类. 相反如果没有足够多的样本进行训练, 那么卷积神经网络就不如人为地定义特征能更快地找到样本之间的联系, 从而达到好的分类效果. 本文在 LeNet-5 的网络基础上, 引入跨连接的方法, 设计出新的卷积神经网络结构, 将其应用于面部表情识别. 实验结果表明, 低层次特征的应用可以一定程度上弥补样本数量的不足, 获得不错的分类效果. 另外, 由本次实验可知, 卷积神经网络现在没有一种通用的结构可以很好地解决多种问题, 在手写数字库上表现非常好的 LeNet-5 结构在表情识别中表现较差, 所以不同的问题需要设计不同的结构来解决问题, 这给卷积神经网络的普及带来了一定的困难.

下一步研究计划是寻找各层特征之间的关系, 运用反卷积等方法实现卷积神经网络各层特征的可视化, 更好地理解各层特征, 进而找到更加通用的卷积神经网络结构设计方法.

References

1 Pantic M, Rothkrantz L J M. Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, 2000, **18**(11): 881−905

2 Ekman P, Friesen W V. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press, 1978.

3 Lucey P, Cohn J F, Kanade T, Saragih J, Ambadar Z, Matthews I. The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). San Francisco, CA, USA: IEEE, 2010. 94−101

4 Lanitis A, Taylor C J, Cootes T F. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 743−756

5 Praseeda Lekshmi V, Sasikumar M. Analysis of facial expression using Gabor and SVM. *International Journal of Recent Trends in Engineering*, 2009, **1**(2): 47−50

6 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, USA: NIPS, 2012. 1097−1105

7 Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, **313**(5786): 504−507

8 Yu Kai, Jia Lei, Chen Yu-Qiang, Xu Wei. Deep learning: yesterday, today, and tomorrow. *Journal of Computer Research and Development*, 2013, **50**(9): 1799−1804
(余凯, 贾磊, 陈雨强, 徐伟. 深度学习的昨天、今天和明天. 计算机研究与发展, 2013, **50**(9): 1799−1804)

9 Wang Meng-Lai, Li Xiang, Chen Qi, Li Lan-Bo, Zhao Yan-Yun. Surveillance event detection based on CNN. *Acta Automatica Sinica*, 2016, **42**(6): 892−903
(王梦来, 李想, 陈奇, 李澜博, 赵衍运. 基于 CNN 的监控视频事件检测. 自动化学报, 2016, **42**(6): 892−903)

10 Xi Xue-Feng, Zhou Guo-Dong. A survey on deep learning for natural language processing. *Acta Automatica Sinica*, 2016, **42**(10): 1445−1465
(奚雪峰, 周国栋. 面向自然语言处理的深度学习研究. 自动化学报, 2016, **42**(10): 1445−1465)

11 Zhang Hui, Su Hong, Zhang Xue-Liang, Gao Guang-Lai. Convolutional neural network for robust pitch determination. *Acta Automatica Sinica*, 2016, **42**(6): 959−964
(张晖, 苏红, 张学良, 高光来. 基于卷积神经网络的鲁棒性基音检测方法. 自动化学报, 2016, **42**(6): 959−964)

12 Sui Ting-Ting, Wang Xiao-Feng. Convolutional neural networks with candidate location and multi-feature fusion. *Acta Automatica Sinica*, 2016, **42**(6): 875−882
(随婷婷, 王晓峰. 一种基于 CLMF 的深度卷积神经网络模型. 自动化学报, 2016, **42**(6): 875−882)

13 Wang Wei-Ning, Wang Li, Zhao Ming-Quan, Cai Cheng-Jia, Shi Ting-Ting, Xu Xiang-Min. Image aesthetic classification using parallel deep convolutional neural networks. *Acta Automatica Sinica*, 2016, **42**(6): 904−914
(王伟凝, 王励, 赵明权, 蔡成加, 师婷婷, 徐向民. 基于并行深度卷积神经网络的图像美感分类. 自动化学报, 2016, **42**(6): 904−914)

14 Chang Liang, Deng Xiao-Ming, Zhou Ming-Quan, Wu Zhong-Ke, Yuan Ye, Yang Shuo, Wang Hong-An. Convolutional neural networks in image understanding. *Acta Automatica Sinica*, 2016, **42**(9): 1300−1312
(常亮, 邓小明, 周明全, 武仲科, 袁野, 杨硕, 王宏安. 图像理解中的卷积神经网络. 自动化学报, 2016, **42**(9): 1300−1312)

15 Sun Xiao, Pan Ting, Ren Fu-Ji. Facial expression recognition using ROI-KNN deep convolutional neural networks. *Acta Automatica Sinica*, 2016, **42**(6): 883−891
(孙晓, 潘汀, 任福继. 基于 ROI-KNN 卷积神经网络的面部表情识别. 自动化学报, 2016, **42**(6): 883−891)

16 Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat′s visual cortex. *The Journal of Physiology*, 1962, **160**(1): 106−154

17 Fukushima K, Miyake S, Ito T. Neocognitron: a neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 1983, **SMC-13**(5): 826−834

18 Le Cun Y, Boser B, Denker J S, Howard R E, Habbard W, Jackel L D, Henderson D. Handwritten digit recognition with a back-propagation network. *Advances in Neural Information Processing Systems 2*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1989. 396−404

19 Le Cun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278−2324

20 Bengio Y. Learning deep architectures for AI. *Foundations and Trends® in Machine Learning*, 2009, **2**(1): 1−127

21 Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS) 2010. Sardinia, Italy: Chia Laguna Resort, 2010. 249−256

22 Ziegel R. Modern Applied Statistics with S-plus (3rd edition), by Venables W N and Ripley B D, New York: Springer-Verlag, 1999, *Technometrics*, 2001, **43**(2): 249

23 Srivastava R K, Greff K, Schmidhuber J. Highway networks. *Computer Science*, arXiv: 1505.00387, 2015.

24 Romero A, Ballas N, Kahou S E, Chassang A, Gatta C, Bengio Y. FitNets: hints for thin deep nets. *Computer Science*, arXiv: 1412.6550, 2014.

25 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. arXiv: 1512.03385, 2016. 770−778

26 Sun Y, Wang X G, Tang X O. Deep learning face representation from predicting 10,000 classes. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, 2014. 1891−1898

27 Zhang Ting, Li Yu-Jian, Hu Hai-He, Zhang Ya-Hong. A gender classification model based on cross-connected convolutional neural networks. *Acta Automatica Sinica*, 2016, **42**(6): 858−865
(张婷, 李玉鑑, 胡海鹤, 张亚红. 基于跨连卷积神经网络的性别分类模型. 自动化学报, 2016, **42**(6): 858−865)

28 Kumbhar M, Jadhav A, Patil M. Facial expression recognition based on image feature. *International Journal of Computer and Communication Engineering*, 2012, **1**(2): 117−119



李 勇 北京化工大学硕士研究生. 主要研究方向为图像处理与模式识别, 深度学习. E-mail: 15117965051@163.com
(**LI Yong** Master student at the College of Information Science and Technology, Beijing University of Chemical Technology. His research interest covers image processing, pattern recognition, and deep learning.)



林小竹 北京石油化工学院教授. 主要研究方向为图像处理与模式识别, 深度学习, 信号与系统. 本文通信作者.
E-mail: linzhu1964@163.com
(**LIN Xiao-Zhu** Professor at the School of Information Engineering, Beijing Institute of Petrochemical Technology. His research interest covers image processing and pattern recognition, deep learning, and signals and systems. Corresponding author of this paper.)



蒋梦莹 北京化工大学硕士研究生. 主要研究方向为图像处理与模式识别, 深度学习. E-mail: 18810493772@163.com
(**JIANG Meng-Ying** Master student at the College of Information Science and Technology, Beijing University of Chemical Technology. Her research interest covers image processing, pattern recognition, and deep learning.)