基于单个卷积神经网络的面部多特征点定位

朱 虹 李千目 李德强

(南京理工大学计算机科学与工程学院 南京 210094)

摘 要 深度学习在面部特征点定位领域取得了比较显著的效果。然而,由于姿态、光照、表情和遮挡等因素引起的面部图像的复杂多样性,数目较多的面部特征点定位仍然是一个具有挑战性的问题。现有的用于面部特征点定位的深度学习方法是基于级联网络或基于任务约束的深度卷积网络,其不仅复杂,且训练非常困难。为了解决这些问题,提出了一种新的基于单个卷积神经网络的面部多特征点定位方法。与级联网络不同,该网络包含了3组堆叠层,每组由两个卷积层和最大池化层组成。这种网络结构可以提取更多的全局高级特征,能更精确地表达面部特征点。大量的实验表明,所提方法在姿态、光照、表情和遮挡等变化复杂的条件下优于现有的方法。

关键词 深度学习,卷积神经网络,面部特征点定位,数据扩增,无约束条件

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.04.046

Facial Multi-landmarks Localization Based on Single Convolution Neural Network

ZHU Hong LI Qian-mu LI De-qiang

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract Facial landmarks localization methods using deep learning network technology have achieved prominent effect. However, the localization of larger number of facial landmarks still has lots of challenges due to the complex diversities in face images caused by pose, expression, illumination and occlusion, etc. The existing deep learning methods for face and mark localization are based on cascaded networks or tasks-constrained deep convolutional network (TCDCN), which are complicated and difficult to train. To solve these problems, a new method of facial multi-landmarks location based on single convolution neural network was proposed. Unlike cascaded networks, the network consists of three stacks, and each group consists of two convolutional layers and a max-pooling layer. This network structure can extract more global high-level features, which express the facial landmarks more precisely. Extensive experiments show that the approach outperforms the existing methods in the complex conditions such as pose, illumination, expression and occlusion.

Keywords Deep learning, Convolution neural network, Facial landmarks localization, Data augmentation, Unconstrained condition

1 引言

面部特征点定位是计算机视觉中的重要问题,因为许多视觉任务依赖于准确的面部特征点定位结果,如面部识别、面部表情分析、面部动画等。虽然近几年面部特征点定位被广泛研究,并且得到了巨大的成功,但是,由于部分遮挡、光照、较大的头部旋转和夸张的表情变化等因素导致人脸图像具有复杂多样性(见图1),因此面部特征点定位仍然面临着巨大的挑战。卷积神经网络已经被证明在提取特征和分类方面具有有效性[1-3],同时它也被证明针对遮挡具有鲁棒性[4]。因此,本文提出一个新颖的卷积神经网络来直接预测面部多特征点的坐标,其在精度和速度上都取得了良好的效果。

图1 人脸图像的复杂表现

Fig. 1 Complex facial expression

虽然深度卷积网络具有很强的学习能力,但它需要从丰富的样本中进行训练。为了弥补训练图像特征点标注的不足,本文提供了数据扩充策略,包括缩放、旋转、平移和翻转4

到稿日期:2017-03-02 返修日期:2017-06-16 本文受江苏省重大研发计划社会发展项目:大数据驱动的隧道等城市快速路交通违章取证关键技术研究(SBE2017741114)资助。

朱 虹(1993-),女,硕士生,主要研究方向为计算机视觉、深度学习等,E-mail:2267907241@qq.com;李千目(1979-),男,教授,博士生导师,主要研究方向为机器学习、计算机视觉、信息安全等,E-mail:liqianmu@126.com(通信作者);李德强(1990-),男,博士,主要研究方向为深度学习与计算机视觉。

种操作。以这种方式学习得到的模型将对姿态旋转等变化更具鲁棒性。

2 相关工作

面部特征点定位方法大致分为两类:传统方法和基于深度学习的方法。典型的传统方法包括基于模型的方法和基于回归的方法。

2.1 基于模型的方法

基于模型的方法在给定平均初始形状的情况下学习形状增量。如主动形状模型(ASM)[5-6]和主动外观模型(AAM)[7-8]采用统计模型,如主成分分析(PCA),来分别捕获形状和外观变化。然而,它们并不能获得具有较大头部姿态变化和夸张面部表情等人脸图像的精确形状,因为单一的线性模型很难刻画现实场景数据中的复杂非线性变化。

2.2 基于回归的方法

基于回归的方法通过训练外观模型来预测关键点位置。Xiong 等人[9]通过在 SIFT 特征上应用线性回归来预测形状增量。Cao 等人[10]和 Burgos-Artizzu 等人[11]使用像素强度差异作为特征顺序学习了一系列随机蕨回归,并逐步退化了学习级联的形状。他们对所有参数同时进行回归,有效地利用形状约束。这些方法主要根据初始的估计迭代地修改预测的特征点位置,因此最终结果高度依赖于初始化。相比之下,本文使用卷积神经网络的方法将脸部图像作为输入,而不进行任何初始化,并在深层结构的较高层提取全局高级特征,可以有效地预测特征点。

2.3 基于深度学习的方法

到目前为止,只有几种基于深度学习的方法。Sun 等人[12]提出了采用 Cascade CNN 进行面部特征点定位的新方法。这种方法将脸划分为不同的部分,每个部分分别由卷积神经网络训练。最后,它实现了 5 个特征点的定位,即左眼睛、右眼睛、鼻尖、左嘴角、右嘴角。然而,由于级联网络的复杂性,其检测速度很慢;并且将人脸划分成了多个部分进行定位,忽略了人脸的整体性。Zhang 等人[13]训练了一个多任务学习(辅助属性)的深层卷积网络 TCDCN。每个任务对应人脸图像的一个属性,例如姿态、微笑、性别等,这使得特征点定位具有鲁棒性。结果表明,该方法的特征点定位的精度较高。然而,多任务学习对数据集的需求更高,并且不能重复进行复杂的训练。

显然,上述网络的结构和训练过程都非常复杂,而且现有的基于深度学习的方法大多是针对数目较少的特征点定位。当特征点的数目变多时,定位的准确性会降低。本文提出的单个卷积神经网络可以很好地解决这些问题。本文设计的网络从面部的整体性着手,既不需要级联网络也不需要多任务学习,从而显著降低了模型的复杂性,简化了训练过程,同时能够实现更好的性能。

3 卷积神经网络的面部多特征点定位

3.1 网络设计

本文设计的单个卷积神经网络包含 8 个卷积层,其后连接了 3 个全连接层,以学习全局高级特征。该网络基于 VGG 网[14],其堆叠的多个卷积层(之间没有最大池化层)共同形成复杂的特征。图 2 给出了该网络的详细结构。

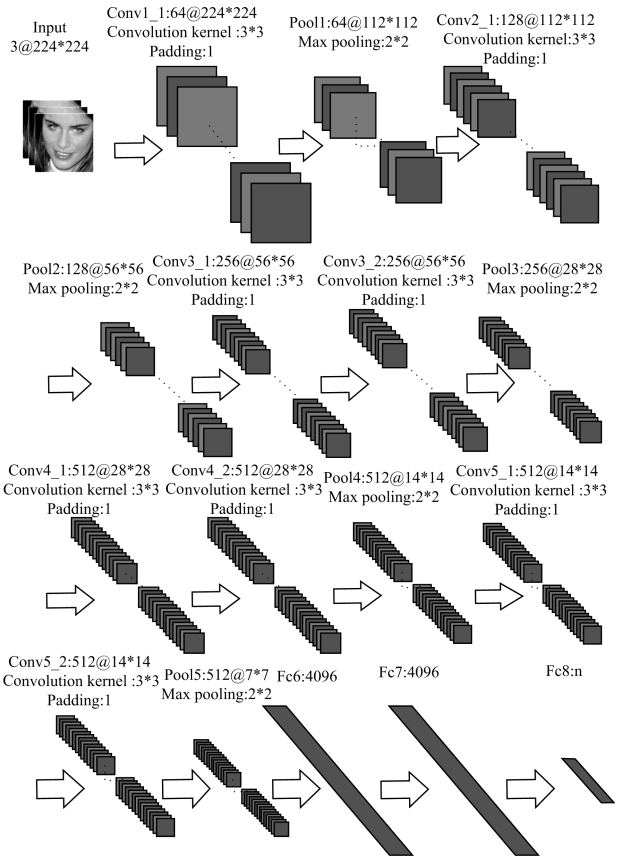


图 2 网络详细结构图

Fig. 2 Detailed architecture of the proposed network

网络的输入是 224×224×3 的彩色脸部图像及相应的面部特征点坐标 n。其中,n 是面部特征点总数的两倍。例如,对于 300-W[15]数据集,68×2=136。网络层数的确定参考经典网络 Alexnet 八层结构[16],前五层是卷积层,后三层是全连接层。本文将网络分为 8 组,第 1 组和第 2 组分别包括一个卷积层(Conv)和一个最大池化层(Pool);第 3 组、第 4 组和第 5 组都分别由堆叠的两个卷积层和一个最大池化层组成;第 1 个完全连接层(Fc6)连接 Pool5 的神经元,其输出数量为 4096;第 2 个完全连接层(Fc7)连接 Fc6 的神经元,其输出数量为 4096;第 3 个全连接层(Fc8)连接输出数为 $n(n/2$ 个面部特征点坐标)。

在卷积层中,卷积核的大小为 3×3,采用 3×3 大小的卷积[18]是能够捕获上下左右和中心概念的最小尺寸,此外两个卷积层 $n\times n$ 的堆叠具有 $(2n-1)\times(2n-1)$ 的有效接收域[17],即两个 3×3 卷积层串联能够拥有 5×5 大小的感受野,可以替代更大的卷积尺寸并保持较小卷积的优点,从而能够有效减少参数个数且拥有更多的非线性变换,使得提取的特征比单个更具区分性[18]。由于卷积操作会损失图像边缘,为了保证卷积后的图像大小与原图一致,设置相应的步长为 1,使得像素逐个滑动,并将边缘扩充设置为 1,即宽度和高度都扩充了 2 个像素。卷积运算表示为:

$$y^j=\sum_i k^{ij} * x^i + b^j \tag{1}$$

其中,x^i 和 y^j 分别是第 i 个输入图和第 j 个输出图,k^{ij} 表示第 i 个输入图和第 j 个输出图之间的卷积核,b^j 是第 j 个输

出图的偏差，* 表示卷积。

在池化层中，采用最大池化的方式，即使邻域内特征点取最大，因为该方式能更好地提取纹理。最大池化表示为：

$$y_{j,k}^{i}=\max_{0\leqslant m,n<h}\{x_{j,h+m,k,h+n}^{i}\} \tag{2}$$

其中，第 i 个输入映射 x^i 中的每个 $h\times h$ 局部区域被合并为第 i 个输出映射中的神经元。设置池化核的大小为 3×3，步长为 2，相邻池化窗口之间会有重叠区域，此时池化窗口的大小大于步长，即重叠池化，与传统池化（池化窗口的大小等于步长）相比，重叠池化在训练期间更不容易过拟合且还会降低错误率[19]。池化可以减少参数，降低特征维度，加快网络训练速度，这是因为池化使得特征图缩小，从而提升了计算速度，但这有可能会影响网络的准确度，因此通过增加特征图的数目来弥补。特征图是卷积核与图片进行卷积的结果，因此卷积核的数目与特征图的数目相等。鉴于在池化层中特征图的大小是翻倍减小的，为了更有效地保存图像信息，需成倍增加卷积核数目[20]，其中卷积核数的初始值参照 VGG 网络。

在每个卷积层之后添加非线性单元 ReLU（Rectified Linear Unit）（$y=\max(0,x)$）作为激活函数以加速网络收敛。本网络不对最后一个完全连接层（Fc8）进行 ReLU 操作，以保留重要的信息。为了防止过拟合，在完全连接层 Fc6 和 Fc7 增加 Dropout 操作，其表达式如下：

$$r=m.*a(Wv) \tag{3}$$

其中，v 是 $n*1$ 维列向量，W 是 $d*n$ 维的矩阵，m 是一个 $d*1$ 的列向量，$a(x)$是一个满足 $a(0)=0$ 的激发函数形式。这里，m 和 $a(Wv)$相乘是对应元素相乘。

3.2　实现细节

为了解决缺乏训练图片的问题，避免严重的过拟合，需要扩充训练样本。由于原始库中的图像包括各种各样的背景，因此先根据数据集所提供的与每个样本对应的面部特征点坐标来确定人脸边框，样例结果如图 3 所示。具体处理方式（伪代码）如算法 1 所示。

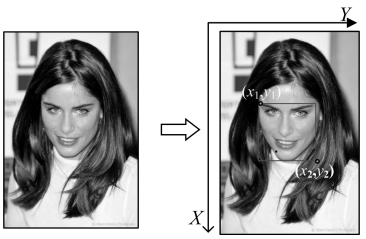


图 3　人脸边框样例图

Fig. 3　Sample of face frame

算法 1

输入：一张测试图片 IMG，与该测试图像对应的面部特征点坐标（x_i，y_i）$=\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$（$i\in\{1,\cdots,m\}$），其中 m 为特征点的个数

输出：该测试图片中的人脸边界框的坐标$\{(X_1,Y_1),(X_2,Y_2)\}$

1. $X_1\leftarrow\min(x_i)$
2. $Y_1\leftarrow\min(y_i)$
3. $X_2\leftarrow\max(x_i)$
4. $Y_2\leftarrow\max(y_i)$
5. End

然后再对数据进行扩充。采用缩放、平移和旋转操作来扩增数据。此外，通过将左眼的模型用于右眼、左眉毛用于右眉毛、左嘴角用于右嘴角来实现翻转图像。数据扩增的样例图如图 4 所示。

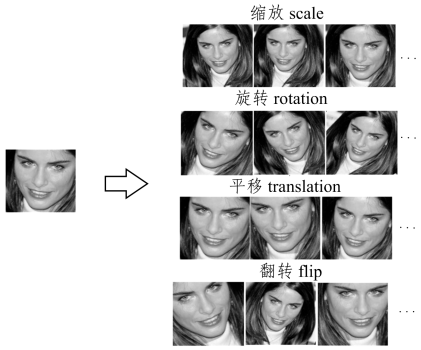


图 4　数据扩增样例图

Fig. 4　Examples of data augmentation

因为在测试时，不同的人脸检测器检测出的人脸边框稍有不同，通过缩放操作可以保留更多的上下文信息，使模型适应不同脸部边界框下的面部特征点定位；平移操作有助于提高特征点定位在微小的面部移动的条件下的鲁棒性；旋转操作使得模型可以学习适应复杂的姿态变化。值得一提的是，在缩放、平移、旋转和翻转的情况下，需要重新计算该人脸边框下与之对应的面部特征点的坐标。上述数据扩增方法能够有效防止深度网络模型训练的过拟合，增强自然环境下针对面部特征点定位各种不利变化的鲁棒性。

最后根据人脸边界框提取出人脸图像，将其归一化为 224×224 的像素大小。在归一化时，面部特征点的坐标位置通过原图与规范化后的比例关系调整。

4　实验

4.1　数据集

LFPW[21]是广泛使用的数据集之一，用于面部特征点定位。它包括 811 张训练图片和 220 张测试图片。该数据集包括互联网上拥有较大姿态、光照、遮挡和表情变化的非约束图片，由于数据集中提到的一些图片链接已经失效，本文使用 ibug 网站[15]上提供的 LFPW 图片，它已积累了所有有效图片以及对应的 68 个面部特征点标注。

Helen[22]数据集有 2000 张训练图片和 330 张测试图片，每张图片标注保持了 194 个面部特征点。为了与实验中的 68 点标注保持一致，本文同样使用 ibug 网站提供的 HELEN 数据集，其标注了 68 个面部特征点。

AFW 是由 Zhu 等人[23]创建的自然环境条件下的数据集，其标注了面部特征点。它拥有 337 张具有不同光照、姿势、属性和表情的面部图片。每张图片最初标注了 6 个面部特征点。然而，为了保持实验的一致性，本文使用 ibug 网站上提供的 AFW 数据集来进行实验，因为它包含了 68 个面部特征点的标注。

IBUG 是取自 300-W[15]数据集的 135 张图片的挑战子集。300-W 是由 IBUG，LFPW，Helen，AFW 和 XM2VTS[24]数据集组成，这些数据集都标注了 68 个面部特征点，包括眼睛、眉毛、鼻子、嘴巴和脸部外轮廓的位置，如图 5 所示。

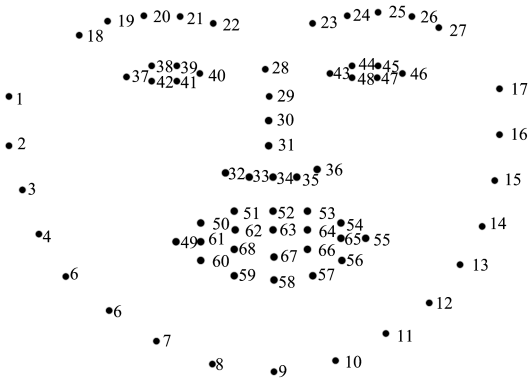


图 5 68 个面部特征点

Fig. 5 68 facial landmarks

本文使用的训练集包括 AFW,LFPW 和 HELEN 的训练集,共有 3148 张图像。验证集为 ibug 网站上提供的 300-W 测试集,其是在自然环境条件下新收集的 2×300 张图片(300 张室内和 300 张室外)。采用以下 3 种形式来执行测试:将来自 LFPW 和 Helen 的测试图片作为公共子集,将 IBUG 作为挑战子集,同时将公共子集和挑战子集联合作为具有 689 张图片的全集。其中,训练集、验证集、测试集均没有重叠。

4.2 评价指标

准确度:采用面部特征点的平均误差来度量所设计方法的性能。本文采用的平均误差定义如下(与文献[13]相同)。

$$RMSE=\frac{1}{N}\sum_{i=1}^{N}\frac{\frac{1}{M}\sum_{j=1}^{M}|p_{i,j}-g_{i,j}|_2}{|l_i-r_j|_2} \tag{4}$$

其中,M 是特征点的数目,p 是预测坐标,g 是真实坐标,l 和 r 分别是左眼角和右眼角的位置。

实时性:采用检测的平均耗时来度量。测试集上的平均耗时定义为:

$$t=t_{out}-t_{in} \tag{5}$$

其中,t_{in} 和 t_{out} 分别是检测的开始时间点和结束时间点。数据预处理的时间不包括在 t 中。

4.3 训练方法

采用深度学习框架 mxnet[25] 来训练网络,底层算法为 C++,外层为 python。虽然原始训练图片的数量只有 3283,本文采用缩放、平移和旋转操作来处理每张图片,使图片数量增加 10 倍,共训练 31480 张图片。当训练该网络时,将学习率设置为 $le-4$,每次处理数据的数量设置为 32。

图 6 是训练该网络时的学习曲线图,可以看出,训练和验证误差的收敛速度是快速且稳定的,并且当数据的训练次数 *epoch* 为 50 时,训练和验证的平均误差都开始趋于稳定。其中,1 次 *epoch* 相当于对训练集中的全部样本训练一次。

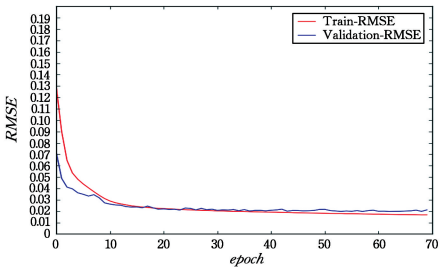


图 6 学习曲线图

Fig. 6 Learning curve

4.4 结果与分析

将本文所提方法与显式形状回归(ESR)[10]、鲁棒级联姿态回归(RCPR)[11]、监督下降法(SDM)[9]、基于局部二元特征(LBF)的回归[26]、由粗到细的自编码器网络(CFAN)[27]、基于回归树集合的回归(ERT)[28]、由粗到细的形状搜索(CFSS)[29]、CascadedCNN[12]和 TCDCN[13]等现有的主流方法进行比较,结果如表 1 所列。其中,TCDCN 在预训练阶段使用外部图像进行训练,显然这对于其他仅使用由 300-W 数据集提供的训练图像的方法是不公平的。与 CascadedCNN 和 TCDCN 相比,本文所提方法不需要级联网络和多任务学习。

表 1 300-W 数据集(68 个特征点)的平均误差

Table1 Mean errors on 300-W dataset(68 landmarks)

(单位:%)

方法	300-W		
	公共子集	挑战子集	全集
ESR	5.28	17.00	7.58
RCPR	6.18	17.26	8.35
SDM	5.57	15.40	7.50
LBF	4.95	11.98	6.32
CFAN	5.50	16.78	7.69
ERT	—	—	6.40
CFSS	4.73	9.98	5.76
TCDCN	4.80	8.60	5.54
CascadedCNN	5.12	10.92	6.24
Proposed	**4.74**	**6.01**	**4.99**

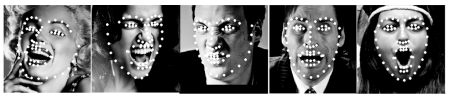
由表 1 中不难看出,本方法在公共子集上的平均误差是 4.74%,在挑战子集上的平均误差是 6.01%,在全集上的平均误差是 4.99%,显然,优于大多数现有方法。虽然在公共子集上其平均误差略高于 CFSS,但是在具有严重遮挡和较大头部旋转的挑战子集上其表现更好。因此所提方法在这些测试集上表现出的良好性能证明了该方法的优越性。从图 7 中可以看出,与传统的基于回归的方法 LBF 和 SDM 相比,本文提出的基于深度学习面部多特征点定位的方法对处理较大的头部旋转的图像表现出了卓越的能力。图 8 给出了该方法在不同条件下对面部多特征点进行定位的结果样例图。从图中观察到,在姿态、光照、表情和遮挡等变化复杂的情况下,本文方法表现出了优异的能力。因此,该方法对姿态、光照、表情和遮挡等复杂变化的人脸具有较强的鲁棒性。

图 7 本文方法与 LBF 方法、SDM 方法的定位特征点的对比结果图

Fig. 7 Comparision results of LBF, SDM and our method about landmarks

(a)姿态变化下的样例

(b)光照变化下的样例



(c)表情变化下的样例

(d)遮挡变化下的样例

图 8　结果样例图

Fig. 8　Results of samples

本文方法在 Intel Core i5 CPU 上处理单张图片需要 67 ms。这个速度比 TCDCN 的 17ms 慢，因为 TCDCN[13] 只预测了 5 个特征点坐标并且本文所提网络较为复杂，但与 Cascade CNN[12] 的 120ms 相比，本文所提方法具有一定的优势。

结束语　本文提出了一种用于面部多特征点定位的有效卷积网络，其使用只具有单个训练任务的单一深层卷积网络，精确地提取了全局高级特征，直接预测面部多特征点的坐标。实验结果表明，所提方法实现了非常高的精度并且具有比其他方法更好的性能。此外，该方法对姿态、光照、表情和严重遮挡(这在非受控场景中是常见的)具有鲁棒性。未来在进一步研究中将努力降低网络的复杂性，并且将人脸检测与面部特征点定位相结合以实现检测定位的一体化。

参 考 文 献

[1] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification [C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2014.

[2] SUN Y, CHEN Y, WANG X, et al. Deep learning face representation by joint identification-verification [C]// Neural Information Processing Systems. Canada: MIT Press, 2014.

[3] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2015.

[4] SUN Y, WANG X, TANG X. Deeply learned face representations are sparse, selective, and robust [C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2015.

[5] COOTES T F, TAYLOR C J, COOPER D H, et al. Active shape models-their training and application [J]. Computer Vision & Image Understanding, 1995, 61(1): 38-59.

[6] GU L, KANADE T. A Generative Shape Regularization Model for Robust Face Alignment [C]// European Conference on Computer Vision (ECCV). France: Springer-Verlag, 2008.

[7] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2001, 23(6): 681-685.

[8] MATTHEWS I, BAKER S. Active Appearance Models Revisited [J]. International Journal of Computer Vision, 2004, 60(2): 135-164.

[9] XIONG X, TORRE F D L. Supervised Descent Method and Its Applications to Face Alignment [C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2013.

[10] CAO X, WEI Y, WEN F, et al. Face Alignment by Explicit Shape Regression [J]. International Journal of Computer Vision, 2014, 107(2): 177-190.

[11] BURGOS-ARTIZZU X P, PERONA P, DOLLAR P. Robust Face Landmark Estimation under Occlusion [C]// IEEE International Conference on Computer Vision. New York: IEEE Press, 2013.

[12] SUN Y, WANG X, TANG X. Deep Convolutional Network Cascade for Facial Point Detection [C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2013.

[13] ZHANG Z, LUO P, CHEN C L, et al. Facial Landmark Detection by Deep Multi-task Learning [C]// European Conference on Computer Vision (ECCV). Zurich: Springer-Verlag, 2014.

[14] PARKHI O M, VEDALDI A, ZISSERMAN A. Deep Face Recognition [C]// British Machine Vision Conference (BMVC). UK: Springer-Verlag, 2015.

[15] SAGONAS C, TZIMIROPOULOS G, ZAFEIRIOU S, et al. A Semi-automatic Methodology for Facial Landmark Annotation [C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2013.

[16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc, 2012: 1097-1105.

[17] WU X, HE R, SUN Z. A Lightened CNN for Deep Face Representation [C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2015.

[18] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, 2015, 47(4): 1409-1556.

[19] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the Inception Architecture for Computer Vision [C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 2818-2826.

[20] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 770-778.

[21] BELHUMEUR P N, JACOBS D W, KRIEGMAN D J, et al. Localizing parts of faces using a consensus of exemplars [C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2011.

(下转第 284 页)

计意义。本文使用 KNN 和 SVM 两种分类器对腭裂高鼻音等级进行识别分类，从而形成对比。所提方法模型简单易懂，计算量小，识别率可以达到 91.67%。结果表明，使用小波分解系数倒谱特征的识别正确率优于其他经典特征，KNN 分类器的识别率优于 SVM 分类器，这充分说明了本文提出的小波分解系数倒谱特征和 KNN 分类器对腭裂高鼻音等级进行识别的方法的有效性和可行性。今后，随着研究的不断深入，可针对腭裂患者的其他病症特征，如共振异常、鼻漏气等，建立腭裂高鼻音等级识别系统，实现高鼻音等级识别方法的多样性。

致谢 感谢华西口腔医学院专家尹恒语音师提供了腭裂语音数据，并对腭裂语音数据进行 4 种等级（正常、轻度、中度、重度）的人工评估所做出的贡献。

参 考 文 献

[1] CHEN R J. The state and consider about speech therapy of cleft palate in China[J]. International Journal of Stomatology, 2012, 39(1): 1-5.

[2] ARIAS-LONDOÑO J D, GODINO-LLORENTE J I, SAeNz-LECHON N, et al. Automatic Detection of Pathological Voices Using Complexity Measures, Noise Parameters, and Mel-Cepstral Coefficients[J]. IEEE Transactions on Bio-medical Engineering, 2011, 58(2): 370-379.

[3] SMYTH A. Clinical grading system for submucous cleft palate [J]. British Journal of Oral & Maxillofacial Surgery, 2014, 52 (3): 275-276.

[4] VILLAFUERTE GONZALEZ R, et al. Acoustic analysis of voice in children with cleft palate and velopharyngeal insufficiency[J]. International Journal of Pediatric Otorhinolaryngology, 2015, 79(7): 1073-1076.

[5] MAIER A, HONIG F, HACKER C, et al. Automatic evaluation of characteristic speech disorders in children with cleft lip and palate[C]// Conference of the International Speech Communication Association. Brisbance, Australia, 2008: 270-278.

[6] ARROYAVE J R O, BONILLA J F V, et al. Automatic detection of hypernasality in children[C]// International Conference on Interplay Between Natural and Artificial Computation: New Challenges on Bioinspired Applications. Spain, 2011: 167-174.

[7] HE L, ZHANG J, LIU Q, et al. Automatic Evaluation of Hypernasality and Consonant Misarticulation in Cleft Palate Speech [J]. IEEE Signal Processing Letters, 2014, 21(10): 1298-1301.

[8] KUMARI V S R, DEVARAKONDA D K. A Wavelet Based Denoising of Speech Signal[J]. International Journal of Engineering Trends & Technology, 2013, 5(2): 107-115.

[9] CHEN Z, WANG S, YIN F. A Time Delay Estimation Method Based on Wavelet Transform and Speech Envelope for Distributed Microphone Arrays[J]. Advances in Electrical & Computer Engineering, 2013, 13(3): 39-44.

[10] MALLAT S G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1989, 11(7): 674-693.

[11] 成礼智，王红霞. 小波的理论与应用[M]. 北京：科学出版社，2004.

[12] 刘明才. 小波分析及其应用(第 2 版)[M]. 北京：清华大学出版社，2013.

[13] ZHAO L. Speech Signal Processing[M]. Beijing: China Machine Press, 2012.

[14] DAVE N. Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition[J]. Ijaret Org, 2013, 1(6): 1-5.

[15] ALI Z, ABBAS A W, THASLEEMA A W, et al. Database development and automatic speech recognition of isolated Pashto spoken digits using MFCC and K-NN[J]. International Journal of Speech Technology, 2015, 18(2): 271-275.

[16] AARON M, GANESH B, RATNADEEP R. Automatic Speech Recognition and Verification using LPC, MFCC and SVM[J]. International Journal of Computer Applications, 2015, 127(8): 47-52.

(上接第 277 页)

[22] LE V, BRANDT J, LIN Z, et al. Interactive facial feature localization[C]// European Conference on Computer Vision(ECCV). Italy: Springer-Verlag, 2012.

[23] RAMANAN D, ZHU X. Face detection, pose estimation, and landmark localization in the wild[C] // IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2012.

[24] MESSER K, MATAS J, KITTLER J, et al. Xm2vtsdb: the extended m2vts database[C]// Second International Conference on Audioand Video-based Biometric Person Authentication. Zurich: Springer-Verlag, 1999.

[25] CHEN T. MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems[J]. Statistics, 2015, 42(12): 87-129.

[26] REN S, CAO X, WEI Y, et al. Face Alignment at 3000 FPS via Regressing Local Binary Features[C] // IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2014.

[27] ZHANG J, SHAN S, KAN M, et al. Coarse-to-Fine Auto-Encoder Networks(CFAN) for Real-Time Face Alignment[C]// European Conference on Computer Vision (ECCV). Zurich: Springer-Verlag, 2014.

[28] KAZEMI V, SULLIVAN J. One millisecond face alignment with an ensemble of regression trees[C]// European Conference on Computer Vision(ECCV). Zurich: Springer-Verlag, 2014.

[29] ZHU S Z, LI C, CHEN C L, et al. Face alignment by coarse-to-fine shape searching[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2015.