

基于并行卷积神经网络的人脸关键点定位方法研究^{*}

杨海燕^{1,2}, 蒋新华², 聂作先²

(1. 福建工程学院 信息科学与工程学院, 福州 350118; 2. 中南大学 信息科学与工程学院, 长沙 410083)

摘要: 为解决人脸关键点定位受到姿态、光线、表情以及遮盖问题的影响而使得定位效果可靠性不佳的问题, 提出了一种基于并行卷积神经网络的人脸关键点定位方法。利用卷积神经网络对局部细节特征提取以及深度学习的特点, 设计实现了一种并行卷积神经网络, 该网络把人脸图像、上半人脸以及下半人脸分别送入到相同结构的卷积网络进行训练学习, 通过对图像进行局部卷积以及下采样, 提取人脸关键点附近的细节特征, 并对三级并行网络定位结果进行加权合成, 实现人脸特征点定位。在 LWF 人脸库上定位实验结果表明, 该方法在准确性以及可靠性方面都得到很大程度提升, 能实现人脸关键点的鲁棒准确估计。

关键词: 人脸特征点定位; 卷积神经网络; 图像卷积; 下图像采样

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1001-3695(2015)08-2517-03

doi:10.3969/j.issn.1001-3695.2015.08.067

Facial key points location based on parallel convolutional neural network

Yang Haiyan^{1,2}, Jiang Xinhua², Nie Zuoxian²

(1. School of Information Science & Engineering, Fujian University of Technology, Fuzhou 350118, China; 2. School of Information Science & Engineering, Central South University, Changsha 410083, China)

Abstract: Facial key points localization suffered from face pose diversification, illumination, expressions and collusion, the location result is unreliable. To solve these problem, this paper proposed a facial key points location method based on parallel convolution neural network. Convolution neural network could preserve the structural feature of image texture, which had widely used in face detection and OCR recognition successfully. Thus, the method used the merit of CNN, designed and fulfilled a parallel CNN. Firstly, it used the entire face, up half face and down half face to train the CNN. The CNN adopted local convolution and down sampling to extract the features around the face key points, and then located the key points. At last layer, it used the average result to fined the final result. The experiment on LWF face database show that this method outperforms other methods both in detection accuracy and reliability, which can locate the key points efficiently.

Key words: facial features localization; convolution neural network (CNN); image convolution; image down sampling

0 引言

人脸关键特征点定位是人脸识别中核心问题之一,也是基于面部特征的疲劳检测等应用中的必要前提,错误的关键点定位会导致提取的人脸特征与事实不符,即使是不精确的对齐也会导致识别性能快速下降。眼睛、鼻子和嘴巴是最为关键的面部特征,尤其是眼睛,现有的多数系统均采用眼睛进行对齐。但是由于人脸样本图像受到光线、姿态等影响,导致定位结果不具有鲁棒性,容易受到姿态光线等影响,因此有必要对关键点定位进行进一步研究。

目前人脸关键点定位方法可以分成两类。一是基于搜索空间的人脸成分定位方法,该方法是判定局部区域是否为关键点。例如文献[1]基于条件回归树的人脸关键点定位方法,学习人脸图像块中的特征点与整个人脸图像条件关系,实现特征点的定位;文献[2~4]基于 AdaBoost 人眼定位方法,针对同一个训练集训练不同的弱分类器,然后把把这些弱分类器集合起来构成强分类器,实现眼睛定位;文献[5]提出基于成分的人脸对齐系统,通过把人脸的各个形状部分构造独立高斯模型以保

持尽可能多的局部形状描述信息,使对齐具有鲁棒性。这类方法是对每个关键点所在的位置或区间进行检测,是一种判定人脸局部区域是否为关键区域的方法。由于局部特征可能存在模糊,以致可能将人脸区间中多个子区域判断为人脸关键区域,也可能在整个人脸区域中不能找到合适的候选区域。另一类是预测关键点位置方法,该类方法首先从人脸样本集合中学习出关键点模型,利用训练学习得出的模型对测试样本进行预测。例如主动形状模型(ASM)^[6~8],该类方法的思想是将已经手工标注好的人脸形状的人脸关键点信息形成学习集,测试时,对任意给定的新的人脸图像形状,估计该图像与学习集中训练图像之间的相似度,并将同样的相似度用于人脸形状的重构。目前这两类人脸关键点定位方法已经取得一定的效果,但是在图像样本具有光线、姿态等挑战时,性能下降快。

卷积神经网络(CNN)最初是应用在手写字符识别^[9,10]中,近年来获得极大发展并逐步应用到各个领域,如文献[11]应用到行人跟踪上,文献[12]应用到语音识别上,文献[13]通过对手写体字符识别算法进行改进用到车牌识别中。卷积神经网络能够从输入图像中提取细节结构信息,同时又能使得这

收稿日期: 2014-07-07; **修回日期:** 2014-08-25 **基金项目:** 福建省教育厅 A 类基金资助项目(JA13219,JA13223);福建省自然科学基金资助项目(2015J01652,2012J01245)

作者简介: 杨海燕(1980-),女,湖南衡山人,副教授,博士研究生,主要研究方向为图像处理、模式识别(yanghaiyantb@126.com);蒋新华(1956-),男,教授,博导,主要研究方向为交通信息、智能控制;聂作先(1972-),男,副教授,博士,主要研究方向为智能控制、工作流。

些结构信息具有特征旋转等空间不变性,非常适合用于图像中的检测识别问题。基于此,文中提出一种新的基于并行卷积神经网络的人脸特征点定位方法。

在卷积神经网络的研究中,网络结构的设计是最主要问题之一,不同的网络结构对识别结果影响非常大。在传统卷积与下采样不断重复的网络结构中,网络层数过多,结构复杂,因此本文设计并实现了一种并行卷积神经网络拓扑结构。

1 卷积神经网络

卷积神经网络(CNN)^[9,13,14]是在传统神经网络的基础上发展起来,并引起广泛重视的一种高效识别方法。卷积神经网络已经成为众多科学领域的研究热点之一,特别是在人脸检测、字符识别等模式分类领域。由于该网络避免了对图像的复杂前期预处理,可以直接输入原始图像,所以得到了广泛的应用。

卷积神经网络的核心思想是通过局域感受野卷积、权共享和下采样对神经网络结构进行优化,减少网络中神经元个数以及权值,同时,采用池化技术来保持特征,使得特征具有位移、缩放和扭曲不变性。区域感受野卷积是指网络中下一层神经元只与上一层的小邻域内的神经元连接,大大减少了神经元个数。通过设计卷积滤波模板对该小邻域进行卷积,提取视觉特征,在同一个卷积中,这些权值相同,也就是权值是共享的,因此大大减少了需要学习的权参数,使得卷积神经网络具有更少的参数,需要相对少的训练数据。下采样减小图像分辨率,同时又能保持图像全局结构特性。典型的 CNN 中,开始几层都是卷积求取特征映射与下采样的交替,也就是说先通过局部卷积得出特征映射图,进行下采样减少数据维数,再次进行新的卷积得出新的特征映射图,在最后得出特征映射图的规模以及特征数量达到要求之后,在靠近输出层时,把二维的特征映射图按照某种方式重新排列成一维特征作为最后一层的输入向量,即将所有二维特征映射图转换为全连接的一维网络的输入,采用全连接的一维 BP 神经网络学习训练作为最后一层的输出结果。卷积神经网络能够从输入图像中提取细节结构信息,同时又能使得这些结构信息具有特征旋转等空间不变性,非常适合用于图像中的检测识别问题。

2 基于并行卷积神经网络的人脸关键点定位算法

文中的人脸关键点定位,是指对左右眼睛中心、鼻尖、左右嘴角这五个人脸关键点的定位,在实际应用中,由于人脸图像受到姿态、光线、旋转角度、表情以及模糊等方面的影响,使得定位结果很容易受到干扰。为了使检测结果更具有稳定性,文中设计了并行卷积神经网络。该网络是由三个独立但结构相同的网络分别对关键点进行定位,只是各个网络定位的关键点位置不完全相同。考虑到人脸关键点中,鼻尖是在人脸中处于人脸中间位置,也是相对稳定的人脸关键点,该点不会像嘴角、眼睛一样随着表情的变化而有很大变化,因此在进行卷积网络对关键点检测时,除了对全部人脸图像输入卷积网络进行五个关键点检测,同时把人脸图像分成包括鼻子、左右眼睛这三个关键点在内的上半边脸,以及鼻子、左右嘴角三个关键点的下半边脸分别训练。在训练之前,首先对所有训练集中的关键点位置进行手动标注,将该标注的位置作为网络输出指导卷积神经网络进行有监督学习。

2.1 并行卷积神经网络结构图

根据卷积神经网络特点以及如上所述的人脸特征点定位问

题,本文设计如下并行独立网络以及网络的级联结构。文中要检测五个人脸关键点:左眼(LE)、右眼(RE)、鼻子(N)、左嘴角(LM)、右嘴角(RM)。采用三个独立卷积神经网络预测特征点位置,三个卷积神经网络分别为 F、EN、NM。其中 F 的输入是整张人脸;EN 的输入是左眼、右眼、鼻子三个特征点的上半部分人脸;NM 的输入是鼻子、左下嘴角、右下嘴角三个特征点的下半部分人脸。每个网络独立预测人脸特征点,最后的输出是将三个独立卷积神经网络预测点进行加权,得到五个特征点的预测位置值。图 1 是提出的并行卷积神经网络的结构图。

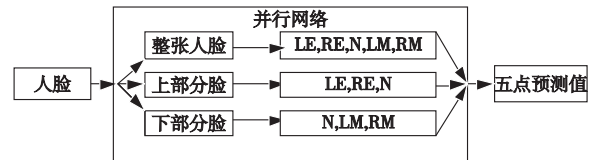


图1 人脸关键点定位的并行网络结构图

并行卷积神经网络的结构为三级并行的独立卷积网络,其中每级卷积网络设计为五层结构。其中,第一层是输入层,第二层为卷积层,第三层是下采样层,最后两层采用全连接层,其中第四层为普通 BP 网络层,最后一层为输出的关键点位置。上半脸卷积神经网络 EN、下半脸卷积神经网络 NM 和人脸卷积神经网络结构完全一样,这三个网络并联预测五个特征点位置。

2.2 各级独立卷积神经网络结构

与文献[9]中采用复杂的 LeNet-5 的结构不同,本文采用一种更简单的卷积网络模型,结构如图 2 所示。

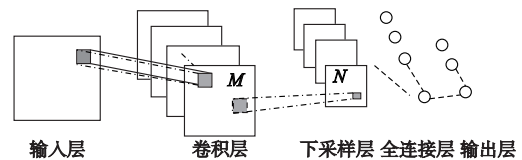


图2 卷积神经网络的框架

第一层是输入层,输入层是直接输入一幅图像,在 FN 中第一层中图像大小为 39×39 像素,卷积层是由输入层的人脸图像经过四个能提取不同纹理细节特征的卷积核对输入图像进行卷积得到四个特征映射,因此也称为特征映射层。该层把一个单一的输入图像通过多个 5×5 的卷积核得出多个卷积图像,卷积核能保持特定局部纹理信息。该层利用图像局部相关性的原理对图像进行子抽样,可以减少数据处理量,同时保留图像的整体结构信息。最后靠近输出的层是全连接层,功能与 BP 神经网络一样。对一维特征进行最后的分类识别。

2.3 关键点位置修正

卷积神经网络在训练的时候,通过有监督的训练方式,训练出并行卷积网络中各神经元结构以及权值。由于全脸、上半人脸以及下半人脸的网络同时对关键点进行预测分别得出的关键点位置略有不同,最后根据各级网络输出结果进行均值加权,最终的特征点定位表达式如下:

$$x = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (1)$$

其中: n 为检测到的关键点数目,取值为 2 或 3; x_n 为其坐标位置。

3 实验结果

本文实验采用人脸样本库的训练集来自 LWF 人脸库^[15],这是在各种光线姿态下从新闻中采集选择的人脸图像库,文中从中选择 10 000 张样本图片作为训练集。测试集有 3 466 张,其中人脸均具有不同角度、侧面、戴眼镜等不同姿态。由于原

始库中包括各种各样背景图像,所以,首先根据人脸位置从原始 LWF 人脸图像中标记出人脸区间,并保存左上角以及右下角坐标位置,同时将每个人脸图像中的五个关键点的坐标位置保存下来,将人脸区间归一化为 39×39 的像素大小。在归一化时,关键点的坐标位置通过原图与规范化后的比例关系调整坐标位置。实验中采用的软件环境为 VS2012 和 OpenCV 2.4.4,在 Windows7 平台下进行测试。在训练时,把所有的训练以及测试人脸集合归一为 39×39 像素大小,由于在并行网络中各级网络结构完全一样,所以先从规范化后的人脸图像中分离出上半人脸和下半人脸图像,并通过高度的尺度变化成与人脸图像相同的大小,然后把 10 000 组训练样本成批输入到并行卷积神经网络中进行训练,网络的最后输出将三级网络的预测结果进行平均加权,得到最终的预测结果;测试时,本文对 3 466 张人脸进行测试,测试样本也如前作规范化和分割处理,对人脸左眼、右眼、鼻子、左嘴角、右嘴角进行预测,预测得出的结果与标定位置进行对比,计算得出平均误差在 1.6 个像素左右。图 3 是对各种表情姿态和光线下的定位结果,从中可以看出对于带眼睛以及姿态偏转严重的样本也能有较小的误差。



图3 基于卷积神经网络人脸特征点定位效果

为了说明本文并行结构对定位性能的影响,文中对单人脸卷积网络和并行卷积网络进行实验,结果如表 1 所示。从中可以看出,设计成单个人脸的卷积神经网络时,测试样本的总体平均误差和每个特征点的误差都比设计成单人脸和上下半人脸的三级并行卷积神经网络的大,采用并行卷积神经网络训练模型可以更精确地定位人脸特征点位置。因此,采用 F、EN、NM 三级并行卷积网络的设计模型,提高了特征点定位的精确度。

表1 训练模型及测试结果

训练模型	左眼	右眼	鼻子	左嘴角	右嘴角
单人脸 CNN	1.690	1.346	2.520	1.926	1.798
并行 CNN	1.482	1.241	2.083	1.751	1.592

为了进一步验证本文提出的并行卷积网络模型,文中与其他人脸特征点定位算法进行比较,训练和测试都采用相同的样本集。测试时,选择的样本数为 200 张,实验结果如表 2 所示,其中的误差数用像素表示。

表2 不同方法及测试结果

方法	左眼	右眼	鼻子	左嘴角	右嘴角
ASEF 算法 ^[16]	2.390	3.195	3.114	2.92	1.902
本文算法	1.312	1.220	1.901	1.401	1.601

从表 2 可以看出,本文算法与其他算法在对同一样本的训练和测试上,定位准确性要好些。

4 结束语

基于并行卷积神经网络的人脸特征点定位算法,充分利用

了卷积神经网络的深度学习能力。由于卷积是对图像局部区域进行处理,能保持图像局部区域性,这样网络可以有效地提取人脸特征点附近区域的特征,在 LWF 数据库上的实验结果表明,提出的并行模型结构有很低的像素误差率,并且模型简单、检测速度快,通过三级并行卷积网络分别对人脸、人脸上半部分,以及人脸下半部分同时进行训练学习,进一步提高检测准确率。由此可见文中所使用的卷积神经网络用于人脸特征点定位有较强的鲁棒性,能实现对多姿态和多种光线条件下的人脸图像进行定位,具有一定的实用价值。

参考文献:

- [1] Dantone M, Gall J, Fanelli G, *et al.* Real-time facial feature detection using conditional regression forests [C]//Proc of International Conference on Computer Vision and Pattern Recognition. 2012:2578-2585.
- [2] Zhang Wencong, Chen Hong, Yao Peng, *et al.* Precise eye localization with AdaBoost and fast radial symmetry [C]//Proc of International Conference on Computational Intelligence and Security. 2006:725-730.
- [3] Liu Xiaoming. Generic face alignment using boosted appearance model [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2007:1-8.
- [4] Lin Liang, Rong Xiao, Fang Wen, *et al.* Face alignment via component-based discriminative search [C]//Proc of the 10th European Conference on Computer Vision. 2008:72-85.
- [5] Huang Yuchi, Liu Qingshan, Metaxas D N, *et al.* A component-based framework for generalized face alignment systems [J]. *IEEE Trans on Man, and Cybernetics; Part B*, 2011, 41 (1): 287-298.
- [6] Hu Jiani, Li Yu, Deng Weihong, *et al.* Locating facial features by robust active shape model [C]//Proc of the 2nd IEEE International Conference on Network Infrastructure and Digital Content. 2010:196-200.
- [7] Xiong Pengfei, Huang Lei, Liu Changping. Initialization and pose alignment in active shape model [C]//Proc of the 20th International Conference on Pattern Recognition. 2010:3971-3974.
- [8] Lu Huchuan, Shi Wengang. Skin-active shape model for face alignment [C]//Proc of Computer Graphics, Imaging and Vision Conference. 2005:187-190.
- [9] Lauer F, Suen C Y, Bloch G. A trainable feature extractor for handwritten digit recognition [J]. *Pattern Recognition*, 2007, 40 (6): 1816-1824.
- [10] Ahranjany S S, Razzazi F, Ghassemian M H. A very high accuracy handwritten character recognition system for Farsi/Arabic digits using convolutional neural networks [C]//Proc of the 5th IEEE International Conference on Bio-Inspired Computing: Theories and Applications. 2010:1585-1592.
- [11] Fan Jialue, Xu Wei, Wu Ying, *et al.* Human tracking using convolutional neural networks [J]. *IEEE Trans on Neural Networks*, 2010, 20 (10): 1610-1623.
- [12] Sainath T N, Mohamed A R, Kingsbury B, *et al.* Deep convolutional neural networks for LVCSR [C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 2013:8614-8618.
- [13] 赵志宏, 杨绍普, 马增强. 基于卷积神经网络 LeNet-5 的车牌字符识别研究 [J]. *系统仿真学报*, 2010, 22 (3): 638-641.
- [14] Neubauer C. Evaluation of convolutional neural networks for visual recognition [J]. *IEEE Trans on Neural Network*, 1998, 9 (4): 685-696.
- [15] Berg T L, Berg A C, Edwards J, *et al.* Names and faces in the news [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004:848-854.
- [16] Bolme D S, Draper B A, Beveridge J R. Average of synthetic exact filters [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2009:2105-2112.