基于深度学习的人脸姿态分类方法

邓宗平,赵启军,陈 虎
(四川大学 计算机学院 视觉合成图形图像技术国防重点学科实验室,四川 成都 610065)

摘 要:人脸姿态通常表达着有用的信息,准确地把握人脸的姿态,往往在人脸对齐、人类行为分析以及司机疲劳驾驶监控等方面有着重要的作用。文中方法与以往姿态估计方法不一样,是一种基于卷积神经网络,应用深度学习做人脸姿态分类的方法。首先,第一次网络对姿态在 yaw 方向上进行 5 分类,同时在 roll 方向具有鲁棒性。之后,将第一次输出正脸的结果进入第二次网络,对姿态在 pitch 方向进行 3 分类。所有的输出结果对光照都具有鲁棒性。文中采用级联的方法在公开库上做测试,准确率高达 95% 以上。在实际监控视频中,姿态估计不仅有较高的准确率,而且有惊人的速度。由于本身实验的设计独特性,只做了自身实验对比。结果充分展示了用合理的神经网络与网络级联的方法在姿态估计上面的发展潜力。

关键词:姿态分类;级联;深度学习;卷积神经网络
中图分类号:TP301 **文献标识码**:A **文章编号**:1673-629X(2016)07-0011-03
doi:10.3969/j.issn.1673-629X.2016.07.003

Face Pose Classification Method Based on Deep Learning

DEND Zong-ping,ZHAO Qi-jun,CHEN Hu
(National Key Laboratory of Fundamental Science on Synthetic Vision,School of Computer Science,
Sichuan University,Chengdu 610065,China)

Abstract:Face pose usually contains useful information,so detecting it accurately plays an important role in face alignment,human behavior analysis and drivers' fatigue driving monitoring. A novel method is proposed in this paper which applies deep learning to human face pose classification based on convolutional neural networks. It can be divided into two steps mainly. First,layer one classifies pose into 5 categories at direction yaw,and it's of robustness at direction roll. Then layer two takes the result of step one as input to classify pose into 3 categories at direction pitch. All outputs are robust to illumination. The cascade connection is used to test on public benchmark,and the result shows that its accuracy is 95%. In real surveillance video,it has both high accuracy and fast estimating speed. Due to the particularity of experiment,it only contrasts the result to itself. Experimental results show that well-designed cascade connection of neural network can estimate pose well.

Key words:pose classification;cascade;deep learning;convolutional neural network

0 引 言

人脸姿态估计是模式识别与计算机视觉领域重要的研究课题。人脸姿态估计在人脸对齐、人脸识别方面有着重要的影响。比如,在已知不同的姿态前提下可以有更精确的算法来进行对齐或匹对工作。此外,人脸姿态自身也具有重要的实际应用价值,比如疲劳驾驶监控[1]、消费者购物行为分析等。

研究人员提出了很多姿态估计的方法,比如:基于模型的方法[2]、基于人脸外观的方法[3]、基于 3 维人脸的方法[4]。同时在文献[5]中,还提出了一种基于椭圆模板和眼睛嘴巴位置的人脸姿态估计方法。常用算法包括支持向量机、基于特征空间的方法[6],以及使用神经网络[7]的姿态估计,等等。

文中运用深度学习的方法对人脸姿态进行大角度分类。首先用第一级网络进行 5 种姿态分类(左+,左,正,右,右+),在第一级分类为正脸的前提下进行第二层网络,再区分 3 种姿态(俯,正,仰),然后进行相关参数对比实验。实验结果表明,用深度学习的方

收稿日期:2015-11-12 **修回日期**:2016-03-09 **网络出版时间**:2016-06-22
基金项目:国家自然科学基金资助项目(61202160,61202161);科技部重大仪器专项(2013YQ49087904)
作者简介:邓宗平(1990-),男,硕士研究生,研究方向为模式识别、计算机视觉;赵启军,副教授,硕士生导师,研究方向为深度学习、模式识别、机器学习、计算机视觉等;陈 虎,讲师,硕士生导师,研究方向为模式识别、图像处理等。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20160622.0845.060.html

法进行姿态分类是可行的,而且可以得到令人满意的效果。实验的关键在于合理的卷积神经网络设计,有效的数据处理与级联方式,以及在训练网络过程中选择合适的参数。

1 深度学习与卷积神经网络

深度学习的概念源自人工神经网络,其是包含多隐藏层的多层感知器(MLP)。它通过对输入数据的低层特征学习来形成更加抽象的高层特征表示,从而学习到数据的分布规律[8]。文献[9]揭示了深度学习的发展潜力。此后,文献[10-12]也展示了深度学习在图片分类、人脸识别、行人检测、信号处理等领域所取得的成果。深度学习之所以称之为"深度"是相较于传统的机器学习方法而言的,比如支持向量机(SVM)[13]、提升方法(boosting)、最大熵方法等。

深度学习网络的非线性操作的层数比较多,属于非监督的学习。深度学习常用的算法包括:自动编码器、稀疏自动编码器、受限玻尔兹曼机、卷积神经网络、深度信念网络[14]等。实验选用卷积神经网络。

卷积神经网络是人工神经网络的一种,是计算机视觉与模式识别、语音分析等领域研究的热点。卷积神经网络是由多层的神经网络构成,每一层又可以有多个二维平面,每一个平面有多个独立的神经元。典型的卷积神经网络如图 1 所示。

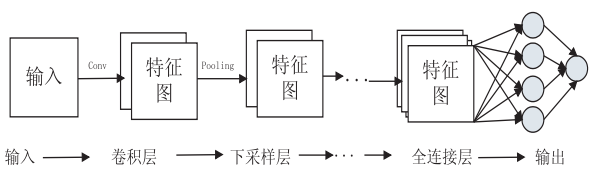

图 1 卷积神经网络

文中实验也按照这样的"卷积-pooling-卷积-pooling"设计模式进行网络设置,并且采用两层级联的方法进行相应的姿态分类。

2 实验评估参数

测试实验评估包括两个方面:一方面是训练模型实时输出性能指标,包括损失函数(loss)、训练中测试准确率(accuracy);另一方面是模型测试性能指标,包括分类准确率、姿态分类时间。

2.1 训练实时输出性能指标分析

假设网络最后的全连接层输出 5 维数据是 $\theta_i(i=0,1,\cdots,4)$,则输出概率函数可表示为:

$$\sigma_i(\theta)=e^{\theta_i}/\sum_{i=1}^{n}e^{\theta_i},i=0,1,\cdots,4 \tag{1}$$

损失函数定义如下:

$$\text{loss}=-\log(\sigma_i(\theta)) \tag{2}$$

其中,$\sigma_i(\theta)$ 表示属于第 i 类的概率大小,即预测出来的属于每一类的概率大小。由于 $\sigma_i(\theta)$ 在[0,1]之间,则对应的 loss 值就应该取值在(0, + ∞)。因此,loss 值越小,表明学习到的特征越好。

网络最后接 accuracy 层输出,其为评判分类准确度,accuracy 越接近 1,说明训练分类效果越好。

对比实验主要是调节三个参数,它们分别是学习率(base_lr)、学习步长(stepsize)、最大迭代次数(max_iter)。学习率是指神经网络学习数据特征的能力,学习步长是每学习一定次数之后相应改变学习率的次数大小,最大迭代次数是学习总的次数。

2.2 测试结果性能指标

对训练好的模型进行性能评估,包含了两个重要指标:一是测试已知类型的姿态图片获取的准确率;二是进行姿态分类估计所需要耗费的时间(ms)。

3 实验与测试

3.1 数据集

训练数据分别来源于 FERET,Multi-pie,Cas-peal 和 point-04 公开数据库。四个数据库的共同特点在于,它们的数据是按照不同角度进行的分类,因此便于对数据集进行加标签分类。实验第一个网络把姿态角分为 5 类,第二个网络姿态分为 3 类。

3.2 数据处理方式

由于原始数据的数量有限性,实验需要对原始数据进行扩充。实验之前,对相应的原始数据按 1:5 的比例划分测试集与训练集,然后对训练数据进行扩充。数据处理方法如下:

(1)人脸检测;

(2)获取人脸框之后,用旋转、平移、缩放的方式对训练数据进行扩充;

(3)同时为了让姿态分类模型适应不同的环境条件,实验对部分图片进行加光照、加噪声的处理。

在各种处理方式下,最终数据分布均匀,训练图片全部归一化为 32 * 32 大小的灰度图,如图 2 所示。

图 2 实验灰度图(旋转角(上),俯仰角(下))

3.3 网络设置

实验网络使用 3 层卷积层,3 层下采样,网络层之间加适当的变换层,最后的网络进行两次全连接。第

一次网络输出一组 5 维向量,第二次网络在第一次预测结果为正脸的情况下进行再分类,精细姿态角是俯仰还是平视角度,最后的输出都是 3 维向量。最终经过两层网络,实验得到的结果是 7 分类问题。级联方式如图 3 所示。

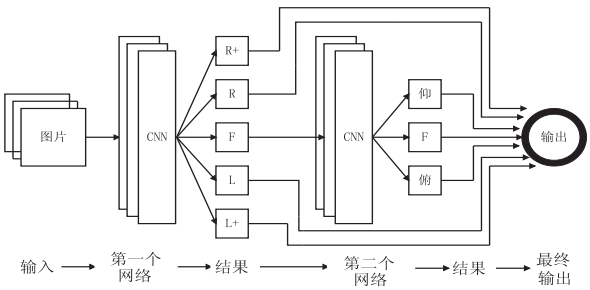

图 3　级联示意图

3.4　对比训练方法与评估

表 1 是实验对部分参数的尝试并且修改部分参数后获取的结果。进行了多组实验,结果表明,当初始 base_lr 设置过大(大于 0.01),网络很快出现饱和状态,反之,学习能力很弱,获取的有效信息很少。同时,对学习步长和学习迭代次数的设置也会直接影响学习的效果。因此,设计了 5 组实验进行对比,对比结果如表 1 所示,两层网络的训练方法是一致的。

3.5　测试方法与结果

根据上面的实验比较,把实验 4 训练出的模型作为测试模型,测试方法是根据已知角度的图片进行定量分析,测试集不仅包括训练的各种角度,而且包含了光照、旋转的变化。所有的测试图片预处理方法与训练数据一致。同时,实验对人脸由输入网络到输出结果的所需时间进行计算,准确地记录姿态估计所需要的时间。实验取得了较高的准确率,在测试速度上也同样令人满意。测试结果与测试环境如表 2 所示。

表 1　参数对比实验及结果

实验	base_lr	stepsize	Mean-loss	Max-accuracy	Last-accuracy
实验 1	0.001	5 000	0.146 9	0.978	0.970
实验 2	0.000 1	5 000	0.485 9	0.877	0.810
实验 3	0.001	10 000	0.019 7	0.998	0.996
实验 4	0.001	20 000	0.017 4	1.000	0.998
实验 5	0.01	10 000	0.724 0	0.818	0.690

注:Mean-loss 表示平均误差,参数越小越好;Max-accuracy 表示训练中最大准确率;Last-accuracy 表示最终收敛的准确率。

表 2　网络实验结果与时间统计

标签	第一层网络					第二层网络		
	Right+	Right	Front	Left	Left+	俯角	正脸	仰角
角度范围	[-60,-90]	[-30,-60)	(-30,+30)	[30,60)	[60,90]	[-60,-30]	(-30,30)	[30,60]
测试数据	257	5 801	9 321	5 820	249	3 539	5 123	2 788
正确分类	245	5 633	9 082	5 756	239	3 336	4 892	2 593
准确率/%	0.953	0.971	0.975	0.989	0.959	0.943	0.954	0.930
耗时/ms	1.74	1.38	1.32	1.61	0.81	1.24	1.21	0.91

注:测试电脑配置,4 核 cpu 处理器:intel i5-4690 3.50 GHz,显卡:NVIDIA GeForce GTX750。

3.6　实验总结

由实验结果可知,使用深度学习方法进行姿态估计,不管在数据库还是在实际应用场景都是可行的。

4　结束语

文中应用深度学习在人脸姿态问题上进行了尝试,结果表明,深度学习是一种高效、高准确度的新方法。但是,实验还存在不足之处,有待进一步研究。总的说来,深度学习下进行人脸姿态估计,是一个可行且高效的方法。为了突破已有成果与应用的限制,更大的数据、更深的网络将会是一个契机,在大数据的支持下,把更多更好的人脸姿态估计应用到实际生活中,比如:场景监控、司机驾驶监控,甚至是物体形态分析,都将成为可能,相信在深度学习的浪潮中人脸姿态估计会有更好的发展。

参考文献:

[1] Ghaffari A,Rezvan M,Khodayari A,et al. A robust head pose tracking and estimating approach for driver assistant system [C]//Proc of IEEE international conference on vehicular electronics and safety. [s. l.]:IEEE,2011:180-186.

[2] Yang R,Zhang Z. Model-based head pose tracking with stereovision[C]//Proceedings of fifth IEEE international conference on automatic face and gesture recognition. [s. l.]:

(下转第 18 页)

来识别动作,目前已经取得了一些研究进展和成果。

参考文献:

[1] 宋鸣侨. 浅析人机交互技术的发展趋势[J]. 现代装饰:理论,2012(2):148-148.

[2] Carroll J M. Human-computer interaction: psychology as a science of design[J]. Annual Review of Psychology,1997,48(1):61-83.

[3] Iwai Y,Watanabe K,Yagi Y,et al. Gesture recognition using colored glove[C]//Proceedings of the 13th international conference on pattern recognition. [s. l.]:[s. n.],1996:662-666.

[4] Weissmann J,Salomon R. Gesture recognition for virtual reality applications using data gloves and neural networks[C]//Proceedings of international joint conference on neural networks. [s. l.]:IEEE,1999:2043-2046.

[5] Viola P,Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of accepted conference on computer vision and pattern recognition. [s. l.]:[s. n.],2001:511-518.

[6] Barrho J,Adam M,Kiencke U. Finger localization and classification in images based on generalized Hough transform and probabilistic models[C]//Proceedings of 9th international conference on control, automation, robotics and vision. [s. l.]:[s. n.],2007:1-6.

[7] Lee D,Lee S G. Vision-based finger action recognition by angle detection and contour analysis[J]. ETRI Journal,2011,33(3):415-422.

[8] Guo K,Zhang M,Sun C,et al. 3D fingertip tracking algorithm based on computer vision[J]. Journal of Computer Research and Development,2010,47(6):1013-1019.

[9] 李博男,林 凡. 基于曲率的指尖检测方法[J]. 南京航空航天大学学报,2012,44(4):587-591.

[10] 梅萍华,李 斌,朱中的,等. 基于径向对称变换的实时指尖检测算法[J]. 中国科学技术大学学报,2011,41(2):101-107.

[11] 刘 佳,郑 勇,张小瑞,等. 基于 Kinect 的手势跟踪概述[J]. 计算机应用研究,2015,32(7):1921-1925.

[12] 宋海声,刘平和,王全州,等. 基于人体骨骼和深度图像信息的指尖检测方法[J]. 计算机工程与科学,2014,36(9):1788-1794.

[13] Wikipedia. Kinect[EB/OL]. 2011-01-13. http://en.wikipedia.org/wiki/Kinect.

[14] Clark R A,Pua Yong-Hao,Karine F,et al. Validity of the Microsoft Kinect for assessment of postural control[J]. Gait & Posture,2012,36(3):372-377.

[15] Dawod A Y,Abdullah J,Alam M J. Adaptive skin color model for hand segmentation[C]//Proceedings of international conference on computer applications and industrial electronics. [s. l.]:[s. n.],2010:486-489.

[16] Homma K,Takenaka E. An image processing method for feature extraction of space-occupying lesions[J]. J Nucl Med,1985,26(12):1472-1477.

(上接第页 13)

IEEE,2002:255-260.

[3] Ng J,Gong S. Composite support vector machines for detection of faces across views and pose estimation[J]. Image & Vision Computing,2002,20(5):359-368.

[4] Baggio D L,Emami S,Escriva D M,et al. Mastering OpenCV with practical computer vision projects[M]. Birminghan:Packt Publishing Ltd,2012:208-254.

[5] 施 华. 头部姿态估计与跟踪系统的研究与实现[D]. 上海:华东师范大学,2015.

[6] Darrell T,Moghaddam B,Pentland A P. Active face tracking and pose estimation in an interactive room[C]//Proc of 2013 IEEE conference on computer vision and pattern recognition. [s. l.]:IEEE Computer Society,1996:67-67.

[7] Hogg T,Rees D,Talhami H. Three-dimensional pose from two-dimensional images: a novel approach using synergetic networks[C]//Proceedings of IEEE international conference on neural networks. [s. l.]:IEEE,1995:1140-1144.

[8] 余 凯,贾 磊,陈雨强,等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展,2013,50(9):1799-1804.

[9] Krizhevsky A,Sutskever I,Hinton G E. Imagenet classification with deep convolutional neural networks[M]//Advances in neural information processing systems. [s. l.]:[s. n.],2012:1097-1105.

[10] Luo P. Hierarchical face parsing via deep learning[C]//Proc of IEEE conference on computer vision and pattern recognition. [s. l.]:IEEE,2012:2480-2487.

[11] Sun Y,Wang X,Tang X. Deep convolutional network cascade for facial point detection[C]//Proceedings of IEEE computer society conference on computer vision and pattern recognition. [s. l.]:IEEE,2013:3476-3483.

[12] Zhu Z,Luo P,Wang X,et al. Deep learning identity-preserving face space[C]//Proceedings of 2013 IEEE international conference on computer vision. [s. l.]:IEEE Computer Society,2013:113-120.

[13] 王 辉. 主成分分析及支持向量机在人脸识别中的应用[J]. 计算机技术与发展,2006,16(8):24-26.

[14] Huang G B. Learning hierarchical representations for face verification with convolutional deep belief networks[C]//Proc of IEEE conference on computer vision and pattern recognition. [s. l.]:IEEE,2012:2518-2525.