Signal Processing: Image Communication ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Landmark perturbation-based data augmentation for unconstrained face recognition

Jiang-Jing Lv*, Cheng Cheng, Guo-Dong Tian, Xiang-Dong Zhou, Xi Zhou

Intelligent Multimedia Technique Research Center, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, PR China

ARTICLE INFO

Article history:
Received 8 September 2015
Received in revised form
31 March 2016
Accepted 31 March 2016

Keywords:
Feature representation
Face recognition
Landmark perturbation
Misalignment
Deep convolutional neural networks

ABSTRACT

Face alignment is a key component of face recognition system, and facial landmark points are widely used for face alignment by a number of face recognition systems. However, inaccurate locations of landmark points bring about spatial misalignment which degrades the performance of face recognition systems. In order to alleviate this problem, we propose a simple and efficient data augmentation approach, which uses artificial landmark perturbation to generate a huge number of misaligned face images, to train Deep Convolutional Neural Networks (DCNN) models robust to landmark misalignment. In our experiments, three types of facial landmark-based face alignment methods are applied to train DCNN models on CASIA-WebFace training database. Experimental results on Labeled Faces in the wild database (LFW) and YouTube Faces database (YTF) verify the effectiveness of our approach.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Automatic face recognition is an important vision task in many practical applications such as identity verification, intelligent visual surveillance and immigration automated clearance system. According to different application scenarios, it can be classified into two different tasks: face verification and face identification. The former aims to determine whether a given pair of face images is from the same person or not, while the latter is to recognize the person from a set of gallery face images and find the most similar one to the probe sample. Many approaches [1–4] have been proposed to improve the face verification performance in unconstrained environments and some of them have exhibited impressive results. For example, Schroff et al. [4] achieved 99.63% face verification accuracy on LFW database [5], which surpasses human accuracy of 97.53%. However, a good verification performance cannot guarantee a good identification performance. Hua et al. [6] concluded that some algorithms have already achieved impressive verification performance on LFW database but get poor performance in identification problem in real environment. In addition, there are still many factors which affect the face recognition performance, such as occlusions, poses, and expressions.

In face recognition systems, face alignment, which tends to warp face images into predefined canonical template, is very critical. Traditionally, facial landmark points are usually used for face alignment and accurate positions of facial landmark points are critical for good recognition performance [7–9]. According to the locations of landmark points, face images can be aligned to predefined canonical template. For instance, Sun et al. [10] aligned face images by using similarity transformation according to the several detected predefined landmark points. Berg et al. [7] incorporated piecewise transformation for face alignment. Taigman et al. [11] introduced a 3D face frontalization method according to the 67 landmark points of 2D face image. If landmark points are accurately located, faces can be well warped and each part of faces from different images will have a good correspondence between each other, which favors feature extracting and feature matching. However, facial landmark detection algorithms are seriously affected by several factors, such as blurring, pose, lighting, expression and occlusion. Fig. 1 shows some examples of misalignment. For a 128×128 face image, the alignment error of one landmark point may be up to more than 5 pixels.

Recently, DCNN based feature representation methods, such as FaceNet [4], DeepId [10] and DeepFace [11], have been widely used in face recognition tasks and have shown impressive results in unconstrained environment. Convolutional neural networks (CNN) was first proposed by LeCun et al. in [12] for handwritten code recognition. With large amounts of training data and computation resources such as GPU, since 2012, DCNN have become prevalent and variants of DCNN have been designed in image processing area. For example,

* Corresponding author.
E-mail addresses: lvjiangjing@cigit.ac.cn (J.-J. Lv), chengcheng@cigit.ac.cn (C. Cheng), tianguodong@cigit.ac.cn (G.-D. Tian), zhouxiangdong@cigit.ac.cn (X.-D. Zhou), zhouxi@cigit.ac.cn (X. Zhou).

http://dx.doi.org/10.1016/j.image.2016.03.011
0923-5965/© 2016 Elsevier B.V. All rights reserved.

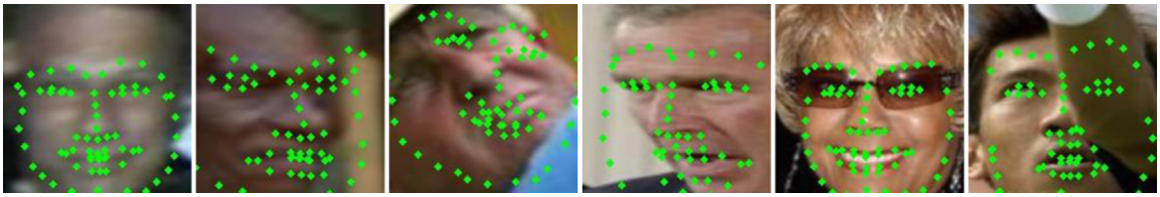

Fig. 1. Examples of misalignment face alignment.

Krizhevsky et al. [13] trained a large and deep convolutional network to classify images and achieved excellent recognition accuracy in ILSVRC-2012 competition [14]. Meanwhile, the architectures of DCNN , such as NIN [15], GoogLeNet [16] and VGG [17], tend to be much deeper and wider, leading to enormous parameters of the network. From Table 1, we can learn that GoogLeNet with 27 layers almost has 16,373K parameters. Thus, training a large DCNN is difficult because it is easy to be over-fitting or even divergency. As shown in Fig. 2, with large network and limited training data, even though the training error is continuously decreasing with increasement of epoches, the test error is increasing after several epoches. A large number of strategies have been proposed to address this problem. On the one hand, different regularization methods have been adopted to DCNN training, such as Dropout [18], Maxout [19] and DropConnect [20]. On the other hand, collecting more training data can essentially deal with this problem. Better performance can be achieved with more training data, however, it is difficult and expensive to collect a large number of labeled data. Therefore, data augmentation strategies, such as flipping [13], cropping [13,21], color casting [22] and blurring [23], have been proposed, which artificially generate large number of visual training data, and experimental results show that data augmentation can help the trained model get a strong generalization ability to unseen but similar patterns in the training data.

Inspired by data augmentation methods, we propose a simple and efficient landmark perturbation-based data augmentation method to alleviate the problem of misalignment. It automatically perturbs the landmark positions to generate a huge number of misaligned face images to train DCNN model, examples of landmark perturbation are shown in Fig. 3. There are some prior works similar to our work. In [24], the authors used several data augmentation methods to generate more training data, including flipping, shifting, rotation, scaling and cropping. In [25], data augmentation were used in augmentation of landmark points. Besides flipping and rotation, the authors also added Gaussian noise to the raw landmark points to generate more examples. Although related, our approach is different from previous ones in several ways: we can automatically generate different kinds of images (e.g., translation, rotation, scaling and shear) by using landmark perturbation-based data augmentation without complex composing of different data augmentation methods; different from [25] which just augments landmark points and extracts geometry feature of landmark points for facial expression recognition, we use the perturbed landmark points for face alignment and aim to generate more face images with misalignment for DCNN training; and experimental results on LFW [5] and YTF [26] show that the DCNN models trained by our method are robust to misalignment and significantly improve the face recognition rates.

The rest of this paper is organized as follows. Section 2 reviews some related previous works. Section 3 presents our data augmentation approach. Experimental setup and results are presented in Section 4. Section 5 offers our conclusions.

2. Related work

The curse of misalignment in face recognition was first proposed by Shan et al. in [27], which systematically evaluated Fisherface's sensitivity to misalignment problem by perturbing the eye coordinates and revealed that imprecise localization of the facial landmarks would abruptly degenerate the Fisherface system. Additionally, misalignment, which enlarges the within-class scatter and reduces the between-class scatter to some degree, increases the difficulties of face recognition. Many face recognition methods suffer from misalignment problem. Sparse representation-based classification (SRC) [28], which seeks a sparse linear representation of the probe images over the training images, is also sensitive to misalignment. Feature encoding methods, such as Fisher vector [29] and Hierarchical Gaussianization Vector [30], also need well-aligned images as the input.

In order to overcome the curse of misalignment problem, large number of methods have been proposed. Shan et al. [27] proposed

Table 1
The architecture of the GoogLeNet.

Name	Type	Filter size/ stride	Output size	Depth	#Params
Conv11	Convolution	7 × 7/2	64 × 64 × 64	1	2.7K
Pool1	Max pooling	3 × 3/2	32 × 32 × 64	0	
Conv21	Convolution	3 × 3/1	32 × 32 × 192	2	112K
Pool2	Max pooling	3 × 3/2	16 × 16 × 192	0	
Inception3a	Inception		16 × 16 × 256	2	159K
Inception3b	Inception		16 × 16 × 480	2	480K
Pool3	Max pooling	3 × 3/2	8 × 8 × 480	0	
Inception4a	Inception		8 × 8 × 512	2	364K
Inception4b	Inception		8 × 8 × 512	2	437K
Inception4c	Inception		8 × 8 × 512	2	463K
Inception4d	Inception		8 × 8 × 528	2	580K
Inception4e	Inception		8 × 8 × 832	2	840K
Pool4	max pooling	3 × 3/2	4 × 4 × 832	0	
Inception5a	Inception		4 × 4 × 832	2	1072K
Inception5b	Inception		4 × 4 × 1024	2	1388K
Pool5	Avg pooling	5 × 5/1	1 × 1 × 1024	0	
Linear1	Fully connection		1 × 1 × 10,575	1	10,575K
Cost	Softmax		1 × 1 × 10,575	0	
Total				22	16,373K

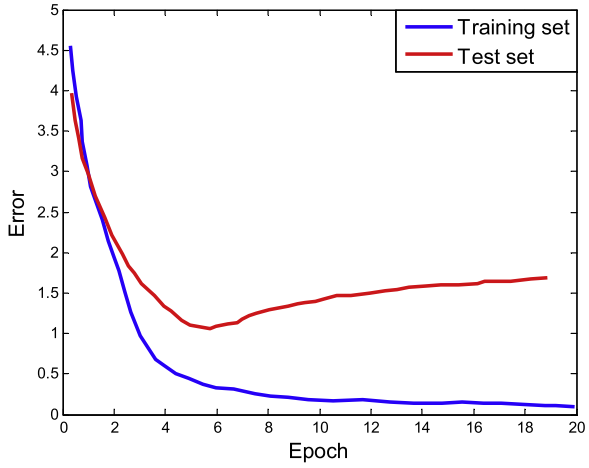

Fig. 2. The error curves of training set and test set with increasing of epoches.

Please cite this article as: J.-J. Lv, et al., Landmark perturbation-based data augmentation for unconstrained face recognition, Signal Processing-Image Communication (2016), http://dx.doi.org/10.1016/j.image.2016.03.011

Download English Version:

https://daneshyari.com/en/article/6941788

Download Persian Version:

https://daneshyari.com/article/6941788

Daneshyari.com