Facial Recognition using Convolutional Neural Networks and Supervised Few-Shot Learning

Gallardo-García, R., Beltran, B., Vilariño, D, Martínez, R.

Language & Knowledge Engineering Lab
Faculty of Computer Science
Benemérita Universidad Autónoma de Puebla
Puebla, Puebla, México
rafael.gallardo@alumno.buap.mx, bbeltran,darnes,beetho(@cs.buap.mx)

Abstract

This paper presents a feature-based face recognition method. The method can be explained in two separated processes: A pretrained CNN-Based face detector looks for faces in images and return the locations and features of the found faces, this face detector will be used to train the models for the classifiers and then will be used to find unknown faces in new images. The used classifiers are: K-Nearest Neighbors, Gaussian Naive Bayes and Support Vector Machines. Each model will be trained with a different quantity of training examples in order to obtain the best version of the method. When the models are ready, each classifier will try to classify the faces with the previously trained models. The accuracy of each classifier in few-shot face recognition tasks will be measured in Recognition Rate and F1 Score, a comparative table of the results is presented. This paper has the goal to show the high accuracy achieved by this method in datasets with several individuals but few examples of training.

1. Introduction

Artificial vision is one of the artificial intelligence disciplines which tries to develop, improve and research new methods trough which computers are able to acquire, process, analyze and understand real-world images in order to get numerical or symbolic information that can be more easily processed by computers. The face recognition is one of the most common applications of the Biometric Artificial Intelligence, a facial recognition system is capable of identify or verify persons in digital images or videos. Face recognition technology can be used in wide range of applications such as identity authentication, access control, and surveillance. A face recognition system should be able to deal with various changes in face images (Guodong, Stan & Kapluk, 1970). A typical facial recognition system uses special features of faces in order to recognize people.

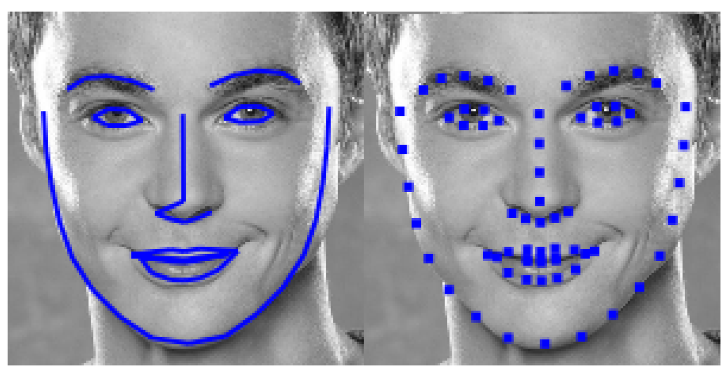

Figure 1: *Facial features.*

2. Theoretical Approach

The theoretical base of this research consists in a Convolutional Neural Networks that detects faces in images and then extract its landmarks, when this information is available, one of the three classifiers will try to find if the detected face belongs to a known people. This system use k-Nearest Neighbors, Support Vector Machines and Gaussian Naive Bayes classifiers. A complete mathematical foundation is available in the full paper. For further information consult bibliography and SOTA papers.

3. Facial Recognition System Structure

In general, the facial recognition method consists in a face detector, and a clustering algorithm, for experimental purposes, three different classifiers were tested, technically the *Table 1* presents a comparison between the methods, which have the same CNN-based face detector, but its accuracy will depend on the effectiveness of each classifier when clustering and recognizing faces. Details are specified below.

3.1 CNN-based Face Detector

This method used a *Maximum-Margin Object Detector(MMOD)* (King, 2015) with CNN-based features. The face detector were trained with various datasets like ImageNet, PASCAL VOC, VGG, WIDER and Face Scrub, the training dataset for the face detector contains 7220 images even so, a good CNN-based face detector can be obtained with a training dataset with just 4 examples. The CNN version of MMOD tested with the 10-fold cross-validation version of the FDDB challenge(Jain & Learned-Miller, 2010), gives the following results:

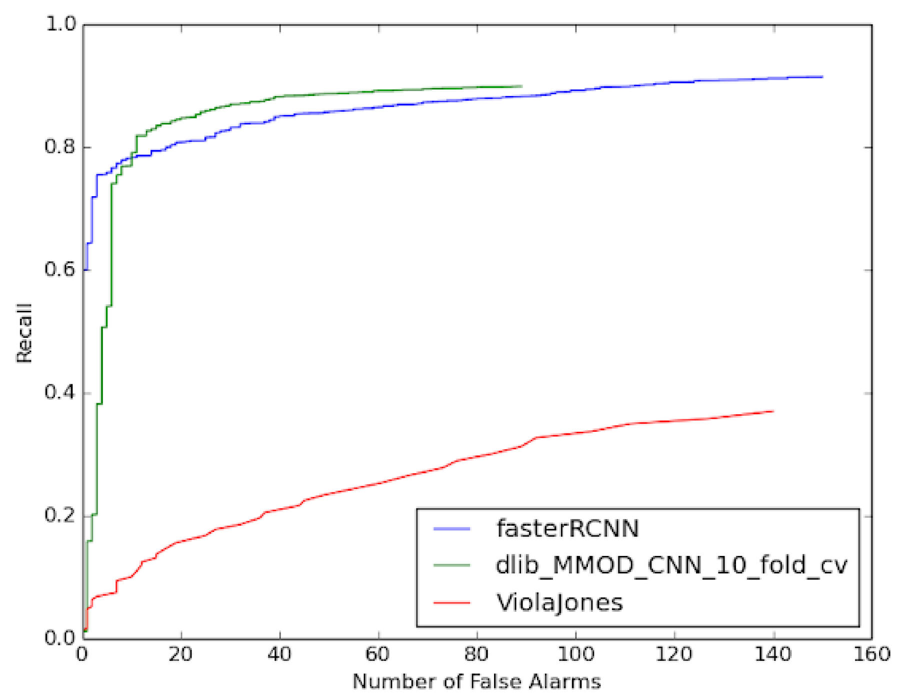

Figure 2: *CNN-MMOD on FDDB challenge.*

The results presented on the chapter 5 were realized by running the CNN face detector over the *NVIDIA CUDA Deep Neural Network (cuDNN)* library, with a NVIDIA 960M GPU with 4GB of VRAM with Maxwell architecture and 5.x compute capability.

3.2 Structure of the experiments

In order to test the facial recognition method and the accuracy of the classifiers as few-shot learning, 10 models were trained per each classifier, each model has a different quantity of training examples per subject, from 1 training example to 10 training examples per subject per classifier, 30 models at the end of training. The experiments were performed over the *MIT-CBCL Face Recognition Database*(Weyrauch, Huang, Heisele & Blanz, 2004), this database contains face images of 10 subjects in high resolution, including frontal, half-profile and profile views. Test images have a size of 115x115 pixeles.

Figure 3: *Expected recognition behavior in the system.*

4. Experimental Results

The following table presents the obtained scores with the three different classifiers, when using the CNN-MMOD as face detector and the classifiers as face recognizer, this experiments were performed over the *MIT-CBCL Face Recognition Database* (Weyrauch, Huang, Heisele & Blanz, 2004), the results on the *table 1* belongs to an experiment performed over a subset of 119 known faces with 11 unknown faces to be recognized, *RR* means Recognition Rate measure.

Table 1: *Scores of the facial recognition method.*

# Shots	Classifier	RR	Precision	Recall	F1
	k-NN	99.15	0.991	0.991	0.991
1-Shot	SVM	N/A	N/A	N/A	N/A
	GNB	99.09	0.921	1	0.959
	k-NN	99.159	0.991	0.991	0.991
2-Shots	SVM	100	0.915	1	0.955
	GNB	99.159	0.921	1	0.959
	k-NN	99.15	0.991	0.991	0.991
3-Shots	SVM	100	0.915	1	0.955
	GNB	99.159	0.921	1	0.959
	k-NN	100	0.991	1	0.995
4-Shots	SVM	100	0.915	1	0.955
	GNB	100	0.922	1	0.959
	k-NN	100	0.991	1	0.995
5-Shots	SVM	100	0.915	1	0.955
	GNB	100	0.922	1	0.959
	k-NN	100	0.991	1	0.995
6-Shots	SVM	100	0.915	1	0.955
	GNB	100	0.922	1	0.959
	k-NN	100	0.991	1	0.995
7-Shots	SVM	100	0.915	1	0.955
	GNB	100	0.922	1	0.959
	k-NN	100	0.991	1	0.995
8-Shots	SVM	100	0.922	1	
	GNB	100	0.922	1	0.959
	k-NN	100	0.991	1	0.995
9-Shots	SVM	100	0.922	1	0.959
	GNB	100	0.922	1	0.959
	k-NN	100	0.991	1	0.995
10-Shots	SVM	100	0.929	1	0.963
	GNB	100	0.937	1	0.967

5. Conclusions

The experiments are clear, the presented facial recognition method can be considered as a success, the three used classifiers gave a $> 99\%$ of Recognition Rate and a F1 score > 0.9, mixing the accuracy of a CNN-based face detector with face clustering techniques gives very interesting results. It is clear that the k-Nearest Neighbors classifier has a recognition rate similar to the other classifiers but it's noticeably better in Recall and Precision scores, in consequence it is better in the F1 score. Surprisingly, the quantity of training examples per subject did not greatly affect the accuracy of the facial recognition method. k-Nearest Neighbors was notably better while recognizing known people and was better while labeling faces as "unknown", Support Vector Machines and Gaussian Naive Bayes versions of the facial recognition method gave excellent results while recognizing known faces but gave disappointing results while trying to identify unknown faces, labeling faces as "known" and giving false positives. Execution times of the CNN-MMOD face detector is very slow when it runs on a Intel Core i7-6700HQ CPU, with a 4.5 FPS average, but this FPS improve if the face detector runs over a CUDA GPU, reaching up to 200 FPS, this FPS makes CNN-MMOD+k-NN a feasible method for real-time facial recognition. k-NN reaches high accuracy scores, both in F1 and in Recognition Rate, with feasible times by using parallel programming, this method could be improved by optimizing the algorithms and the CNN achieve better execution times and trying to obtain similar scores.