International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:14, No:6, 2020

Two Concurrent Convolution Neural Networks TC*CNN Model for Face Recognition Using Edge

T. Alghamdi, G. Alaghband

***Abstract*—**In this paper we develop a model that couples Two Concurrent Convolution Neural Network with different filters (TC*CNN) for face recognition and compare its performance to an existing sequential CNN (base model). We also test and compare the quality and performance of the models on three datasets with various levels of complexity (easy, moderate, and difficult) and show that for the most complex datasets, edges will produce the most accurate and efficient results. We further show that in such cases while Support Vector Machine (SVM) models are fast, they do not produce accurate results.

***Keywords*—**Convolution neural network, edges, face recognition, support vector machine.

I. Introduction

FACE recognition refers to the technology capable of identifying or verifying the identity of subjects in images or videos [1]. Face recognition is used in many applications such as smart classes [2], human tracking [3], surveillance systems, access control, fraud detection, identity verification and social media [1], [4]. Face recognition is considered to be a more appealing technique than other biometric recognitions such as iris and fingerprint because it offers several advantages: good acceptability, low cost, and contactless acquisition [5]. There are many challenges associated with taking images in face recognition systems like occlusion, pose, illumination and expression variations [1], [5]. Many methods have been developed for face recognition such as Principal Component Analysis (PCA) [1], [6], Local Binary Patterns Histograms (LBPH) [1], [6] and K–Nearest Neighbor (KNN) [6], but CNN outperforms all the traditional methods [1], [6]. Different CNN structures have been developed, trained, and tested on different face datasets. These structures usually consist of multiple numbers of layers; convolution layers, pooling layers, fully connected layers and SoftMax regression layer [6], [7].

The first step in CNN based face recognition system is associated with input (image) preprocessing to enhance the performance of the model. The most popular methods for image preprocessing in CNN models are image resizing and normalization [6]. Another method is image dilation and erosion which focus on expanding and reducing the boundary of the connected components of the binary image [8]. Jun et al. developed CNN based face recognition model by feeding the CNN with the output of LBPH to extract features from the original image [9]. Another study handles the image quality before using in the CNN model [10] where this study converts the images to edges as a preprocessing method. The edge format preserves useful structural information about object boundaries and contains less data than the original image [11].

T. A. is PhD student and G. A. is professor and chair at the Computer Science and Engineering Department, University of Colorado Denver, Denver, CO 80217 USA (e-mail: thoria.alghamdi@ucdenver.edu, gita.alaghband@ucdenver.edu).

T. A. works at the Information Systems Department, King Abdulaziz University, Jeddah, KSA.

Many researchers working on face recognition have used a basic CNN model [2], [6], [7], [12]. In this study we develop a model that couples two concurrent CNNs with different filters (CC*CNN). The duplicated CNN model imitates the human brain visual perception [13]. Also, the duplicated CNN model is more flexible than the basic CNN model. It can use multiple basic CNN models with different characteristics and merge them in one or more intersections. The duplicated CNN model can gain better performance by using auxiliary inputs.

The input images in face recognition applications vary in complexity. The images taken under conditions dictate the degree of difficulty. For example, in authentication applications, the face usually aligns with small degree of poses [6], but in human tracking applications the face image may be captured under different illumination, with covered parts, or from the side [4], [14]-[16]. This study observed the impact of using edge input and duplicated models on the face recognition performance when it is applied to three face datasets, where each dataset represents different degree of challenge. The three datasets are: Extended Yale-B dataset [14], AR dataset [15], and LAFW dataset [16].

The goal of this paper is to show that our CC*CNN model produces highest accuracy for the complex datasets using image/edge pairs as input. This paper is organized as following: Section II explains the CNN based face recognition system. The description face datasets and the input format are illustrated in Section II. The obtained experimental results are analyzed in Section IV. Finally, the conclusion is in Section V.

II. The CNN Models

The basic CNN model consists of six layers as in Fig. 1. The first layer is convolution (Conv2D) with 16 neurons, (3x3) filter, and Relu activation function. The convolution layer extracts the features from the input. The second layer is pooling layer which used (2x2) filter and output was dropped out with probability 0.25. The pooling layer is used to reduce the computation work by reducing the dimension of the feature maps. The third layer is convolution (Conv2D) as the previous convolution layer but with 32 neurons. The fourth layer is pooling layer exactly like the previous pooling layer. The fifth layer is flatten layer. The sixth layer is dense layer with 3000 neurons and Relu activation function and used L2 regularization. Finally, full connected layer uses SoftMax regression. The full connected layer connects every filter to all neurons in the previous layer and classifies the data.

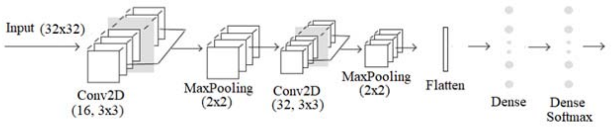

Fig. 1 The basic CNN model

A model is developed that couples two concurrent linear CNNs with different filters (and referred to as CC*CNN). In our CC*CNN implementations we experimented with different filter size combinations and report on two sets of configurations. And based on our experiments using two filter sizes (2x2) for image and (3x3) for edge in the linear CNNs will produce best results.

In configuration 1 depicted on Fig. 2, we used the CC*CNN model in several training combinations of input pairs to explore the impact of input format in producing more accurate results. The training and testing processes were carried out on two inputs of the same format:

- both images
- both edges

The first input is passed to filter (5x5) in the linear CNN and the second input is passed to filter (3x3) in the linear CNN.

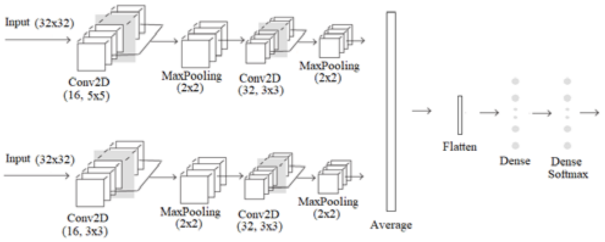

Fig. 2 The CC*CNN with filter 5x5 and the same inputs

In configuration 2 depicted on Fig. 3, we repeated the above experiments for smaller filter sizes: on two inputs of different format: image using (2x2) and edge of the same image using (3x3) filter.

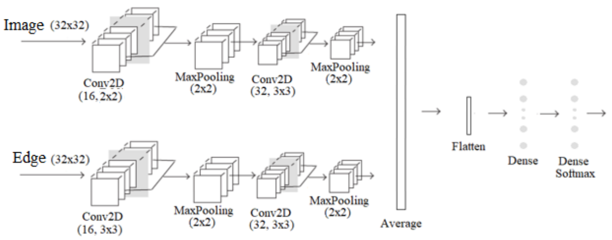

Fig. 3 The CC*CNN with filter 2x2 and different inputs

III. Experiments

A. Datasets Description

In this paper, we chose three datasets to represent degrees of difficulty in the face images for recognition: The extended Yale-B, AR, and LFW. The datasets are classified in three categories: easy, moderate, and difficult. The three factors we used to classify the datasets were: the number of labels, the size of the dataset, and the complexity of the images. A dataset is considered easy if there are few labels, large size dataset, and small poses, under same light, and the whole face appears in the image. More labels, small size dataset, many face expressions, different illumination, and hidden parts of the face in the image are considered as a difficult case. Table I shows the characteristics of the selected datasets.

TABLE I
DATASETS CLASSIFICATION

Dataset	Extend Yale-B	AR	LFW
Labels	28	126	102
Size	16128	1638	526
images/person	576	23	2-11
Complexity	Different poses and 64 illumination	Different expressions, from different side, and covering some part of the faces	Collected from the web
Difficulty	Easy	Moderate	Difficult

The extended Yale-B dataset contains 16128 color images of 28 persons; provides subjects under 9 poses and 64 illumination conditions. All images are 480 x 640 pixels [14]. The AR data set has over 4,000 color images of 126 persons at different facial expressions, illumination conditions and occlusions. The images for each person were collected in two sessions and each session was in a different day. All images are 768 x 576 pixels. In our experiments, a sub-database including 50 men and 50 women is selected, with 2600 color images, with different expressions (neutral, smile, anger, scream), from different side, and covering some part of the face by scarf or sunglasses [15]. The Labeled Faces in the Wild (LFW) dataset contains more than 13,000 images of faces collected from the web [16]. All images are 250 x 250 pixels. In the experiment, a subset of images includes people with names starting with A in 526 total images and with varying number of images per person in the range of [2-11] is selected. Fig. 4 shows a sample of faces from the Extended Yale-B dataset. Fig. 5 shows a sample of faces from AR dataset and Fig. 6 shows a sample of faces from LFW dataset.

B. Input Format

We used two types of input in our experiments: (a) the original images resized to 32x32, and (b) the edge generated from the resized image. Each input is represented by an array of three dimensions height, width, and channel. Fig. 7 shows the image and edge input pair before and after resizing. The detection algorithm was applied to LFW dataset before running the CNN. It is necessary to apply the detection algorithm to LFW dataset because it consists of images with different backgrounds. In this case, edge captures the features of both the face and the background and generates multiple edges for one face. Detection algorithm helps to extract the face and generate almost a unique edge for all faces belonging to one person regardless of the background.

Fig. 4 Samples of Extended Yale-B dataset

Fig. 5 Samples of AR dataset

Fig. 6 Sample of LFW dataset

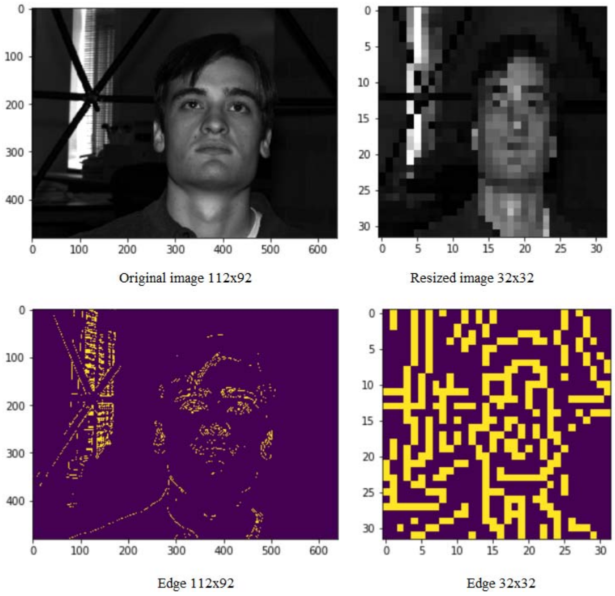

Fig. 7 Image and edge inputs

IV. Results

A. The Performance of SVM and Basic CNN

Fig. 8 shows and compares the behavior of SVM and the basic CNN. The performance of both techniques drops as the difficulty of datasets increase, but in contrast, SVM showed more sensitivity to dataset difficulty and edges input. In all datasets, basic CNN using edge showed better accuracy than using image. Fig. 9 shows the training computation time of SVM and basic CNN. For basic CNN the computation time increases when data size increases and decreases when using edge. For SVM however, we do not observe a direct relationship between data size and computation time, but the computation time drops significantly when using edges instead of images.

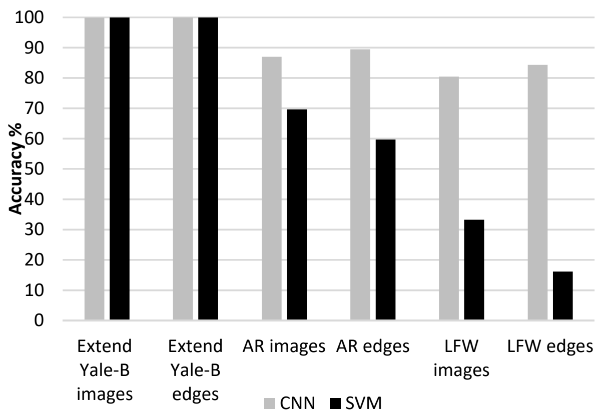

Fig. 8 The performance of SVM & basic CNN

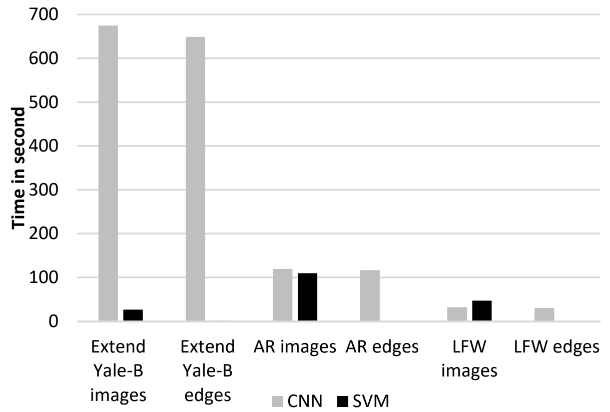

Fig. 9 The training computation time of SVM & basic CNN

B. The Respective Performance of Basic CNN and CC*CNN Using Edge vs. Image

To show the effectiveness of using edge vs. image as input to our face recognition models, we conducted the following experiment. Table II shows the accuracy of the basic CNN and the CC*CNN model using image and edge inputs. In this experiment, both inputs to CC*CNN are the same formats of image or edge for the purpose of comparison with the basic model. In the three datasets the accuracy of basic CNN model and CC*CNN model using edge outperformed the same model using the image. In AR dataset the accuracy increased slightly in CC*CNN model. Edge improves the accuracy by 0.03% and 0.16% in Extend Yale-B dataset, and by 2.87% and 2.54% in AR dataset, and by 4.77% and 5.56% in LFW dataset in each of the basic CNN and CC*CNN models. The gained performance when using the edges input instead of the image increased with the dataset difficulty in both models.

TABLE II
THE PERFORMANCE OF BASIC CNN AND CC*CNN WITH FILTER 5X5

	Extend Yale-B image	Extend Yale-B edge	AR image	AR edge	LFW image	LFW edge
Basic CNN	99.97	100	87.00	89.50	80.50	84.34
CC*CNN	99.84	100	87.89	90.12	79.20	83.60

Table III shows the computation time of training phase in minute when running the basic CNN and the CC*CNN models using images or edges inputs. The computation time increased with dataset size and decreased when using edges. Edge inputs decrease the computation time for training phase by 3.84% and 16.18% in Extended Yale-B dataset, and by 2.35% and 9.16% in AR dataset, and by 5.20% and 5.52% in LFW dataset in each of the basic CNN and CC*CNN models respectively. Edges inputs improve computation time for training phase as well as its accuracy especially when the model becomes more complex and the dataset has large size.

Table IV shows the accuracy when running the CC*CNN model with (2x2) filter using image/edge pairs together as input. The results show better accuracy for CC*CNN model than using basic CNN model in all the three face datasets. CC*CNN model allowed using two different inputs formats and each input was passed to its proper basic CNN model. Using image and edge together with small filter size improves the accuracy by 0.01% in Extended Yale-B dataset, 0.26% in AR dataset, and 5.45% in LFW dataset.

TABLE III
THE COMPUTATION TIME FOR BASIC AND CC*CNN WITH FILTER 5X5

	Extend Yale-B image	Extend Yale-B edge	AR image	AR edge	LFW image	LFW edge
Basic CNN	674.66	648.78	119.20	116.40	31.90	30.24
CC*CNN	856.54	717.98	132.10	120.00	40.94	38.68

TABLE IV
THE ACCURACY FOR THE CC*CNN MODEL USING IMAGE AND EDGE WITH FILTER 2X2

	Extend Yale-B image and edge	AR image and edge	LFW Image and edge
Complex CNN	99.98	87.23	84.89
(Improve/basic)%	0.01	0.26	5.45

V. CONCLUSION

In this study we developed two face recognition models based on CNN. The first model is basic CNN and consists of six layers. The second model couples two concurrent CNNs with different filters, CC*CNN. The experiments are conducted on several input format combinations. The first input is the resized original image from the face dataset. The second input is the edge extracted from the resized image. The models are trained and tested on three datasets. Each face dataset represents different levels of difficulty. These face datasets include Extended Yale-B, AR, and LFW. The results showed that feeding CNN models with edge can improve the model performance and accuracy. Edge input incur higher accuracy and less computational time in both of our CNN models than the traditional image input due to its ability to extract the main features in the image with less data. The study also conducted a comparison between SVM and basic CNN in face recognition system. The results showed that both SVM and basic CNN were impacted by the problem difficulty level but SVM was more negatively impacted, not just by the problem difficulty level but also to the edge input. Finally, we showed that our CC*CNN model produces the highest accuracy for image/edge pair input combinations among all test cases.

REFERENCES

[1] D. Trigueros, L. Meng, and M. Hartnett, “Face Recognition: From Traditional to Deep Learning Methods,” 2018.
[2] M. Z. Khan, S. Harous, S. U. Hassan, M. U. Ghani Khan, R. Iqbal, and S. Mumtaz, “Deep Unified Model for Face Recognition Based on Convolution Neural Network and Edge Computing,” *IEEE Access*, vol. 7, pp. 72622-72633, 2019.
[3] H. Manh and G. Alaghband, “Spatiotemporal KSVD Dictionary Learning for Online Multi-target Tracking,” *15th Conference on Computer and Robot Vision (CRV)*, Toronto, pp. 150-157. *2018.*
[4] Y. Pan and M. Jiang, “LRR-TTK DL for face recognition,” *IET Biometrics*, vol. 6, no. 3, pp. 165-172, May 2017.
[5] M. Taheri, “Robust face recognition via non-linear correlation filter bank, “ *IET Image Processing*, vol. 12, no. 3, pp. 408-415, March 2018.
[6] P. Kamencay, M. Benco, T. Mizdos, and R. Radill, “A New Method for Face Recognition Using Convolutional Neural Network,” Digital Image

Processing and Computer Graphics, vol. 15, No. 4, pp. 663-672. 2017.
[7] K. Yan, S. Huang, Y. Song, W. Liu, and N. Fan, "Face Recognition Based on Convolution Neural Network," Proceedings of the 36th Chinese Control Conference, Dalian, China. pp. 4077-4081 July 26-28, 2017.
[8] X. Li and D. Li, "Image preprocessing study on KPCA-based face recognition," Proc. SPIE 9813, MIPPR 2015: Pattern Recognition and Computer Vision, 981306, December 2015.
[9] Y. Jun, S. Kejia, G. Fei, and Z. Suguo. "Face biometric quality assessment via light CNN," Pattern Recognition Letters. 2017.
[10] H. Moon, C. H. Seo, and S. B. Pan. "A face recognition system based on convolution neural network using multiple distance face," Soft Computing, , vol. 21, no. 17, pp. 49-95, 2017.
[11] J. Canny, "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 8, pp. 679–714, 1986.
[12] S. Qiao and J. Ma, "A Face Recognition System Based on Convolution Neural Network," *2018 Chinese Automation Congress (CAC)*, Xi'an, China, pp. 1923-1927. 2018.
[13] Y. YC, H. KR, and D. YS. "A new image classification model based on brain parallel interaction mechanism," Neurocomputing. pp. 190–197, 2018.
[14] A. Georghiades, P. Belhumeur and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," PAMI, 2001.
[15] A.M. Martinez and R. Benavente, ``The AR face database," CVC Tech. Report #24, 1998.
[16] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," University of Massachusetts, Amherst, Technical Report 07-49, October, 2007.