Applying a Deep Learning Convolutional Neural Network (CNN) Approach for Building a Face Recognition System: A Review

Mr. Binyam Tesfahun Liyew
M.Tech. Student
Department of Computer Science and Engineering
Parul Institute of Engineering and Technology, Vadodara, India

Abstract-- **Deep learning is a new advanced research area of machine learning. The Deep learning come up with a successful learning approach to make the machine learning closer to Artificial intelligence. In Deep learning, multiple processing layers are organized for simulating mathematical models that learn characteristics of data with various levels of abstraction. These techniques have greatly improved the state-of-the-art in pattern recognition. Deep learning finds out complex structure in large datasets by the help of backpropagation technique to show how the AI supposed to change its internal parameters that are used to calculate the characterization in each layer from the characterization in the previous layer. Basically, the Deep Convolutional nets have been used in the development to deal with images, speech, video, and audio. In contrast, recurrent nets give out a bright way on sequential data like text and speech. Recently, the deep learning CNN has reached propitious results in face recognition. In this survey paper, we are going to study about CNN-based face recognition systems. One of the objective of this study is to know more about the field of pattern recognition in our case face recognition, which is very interesting and have a tremendous benefits in real life applications that require peoples identity, such as Payments, Passport immigration, Access and security, Criminal identification, Advertising, Healthcare, Students attendance, and etc.**

Keywords-- **Biometrics, Pattern recognition, Face detection, Face recognition, Deep learning, Neural Network, Convolutional Neural Network (CNN).**

I. INTRODUCTION

In our daily life, the human face participates an important role to claim people's identity. A human face is one of a key feature for providing security. The biometric face recognition technology applied in various application areas that measure security in the first place, in both law enforcement and non-law enforcement. One of the basic reason that face recognition has a unique advantage than the other biometric systems such as fingerprint and iris is that of its non-contact process. The human face will capture from the distance without any interaction of the person being identified, and also during identification phase, there is no need for a person to interact. Furthermore, face recognition provides a great work in detecting criminals. The recorded human face will help to identify a person.

Deep learning is a sub-field of a broader family of machine learning which used to extract high-level abstractions or representations of data along multiple processing layers. The higher-level features are outlined from the lower-level concepts. An object, for example, an image can be represented in different ways might be in terms of a vector of pixels, e.g., a task to learn whether or the image is a human face or not?. So the thing is that a representation that makes the environment easy to learn tasks [1].

Deep learning is applied in different types of research areas such as neural network, artificial intelligence, graphical modeling, optimization, pattern recognition, signal processing, and natural language processing (NLP). Why Deep learning is so much popular in today's technology?. There are three basic reasons for popularity of deep learning: the

first one is the increased chip processing capacities i.e, GPGPUs, Second is tremendous amount of data that used for training, and the third one is the research going on development of advanced recent technologies in machine learning and signal or information processing or we can say on the development of new effective CNN architectures, like the VGG-16/19 networks, inception networks, and deep residual networks [1].

Identifying humans by using their facial feature is one of a biometrics approach that has gained popularity over the past few years and it is mostly used in various sectors such as, access controls, surveillance systems, and other security applications. It is the fact that, the facial features of each and every people are unique, that's why capturing this uniqueness is much easier than the other biometrics. In order to accomplish face recognition the first task to be performed is face detection. There are challenges during the face detection phase for example in Figure 1, off-angle faces, occlusions, low-resolution, and illuminations are the most important challenges in face detection that need to think about.

Figure 1. The results of face detection using MS-FRCNN technique [2].

There are four steps found in the conventional face recognition system: the first one is face detection, the second is face alignment, third is feature extraction or we call it face representation, and the last one is classification. From all of the entire stage, feature extraction is the most important one. Local Binary Patterns (LBP) and Local Phase Quantisation (LPQ) are a type of hand-crafted features that perform a good recognition rate in constrained environments. Unfortunately, these features will never be good or the performance will be reduced in unconstrained environments. Because in the unconstrained environment there might exist various challenges like the pose, occlusion, illumination, and expression. It is still a problem to create a robust face recognition system in unconstrained environments (FRUE). The deep learning Convolutional Neural Network (CNN) have done an impressive result on FRUE in the last three years. Instead of using the traditional way of hand-crafted features, the CNN learned features are more powerful to personal identification [3].

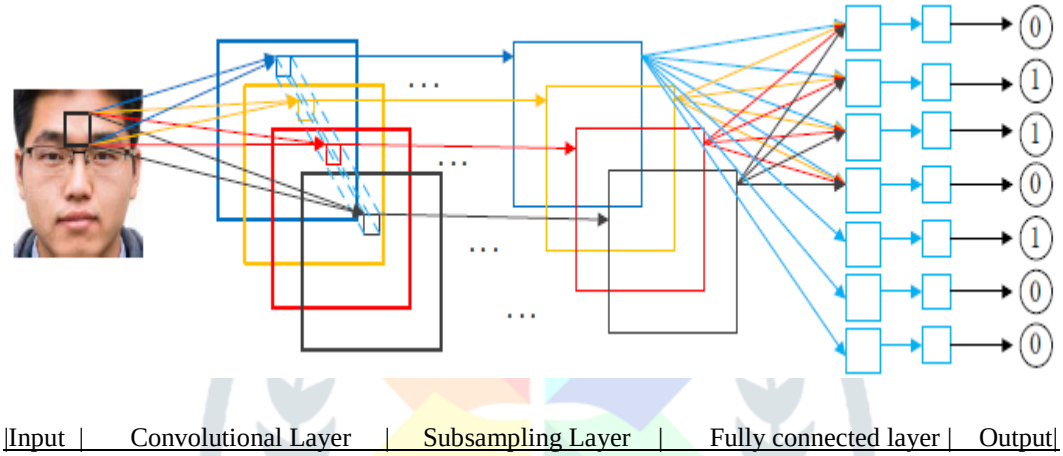
As discussed in the above section the traditional hand-crafted local non-linear features such as Local Binary Patterns, Local Phase Quantisation, and Fisher vectors are provide a good performance in constrained environments. Unfortunately, such features had also difficulties in uncontrolled environments, i.e, in the LFW standard. Metric learning can be used to improve the face recognition performance in such challenging situation. The learning mechanism of metric learning is by learning the transformation of features that pulls the objects which have the same label appearance while pushing the objects which have different label appearance [6].

One of the major benefits of using CNN is that each and every processing layer, beginning from the raw pixel-level input, have a configurable parameter which can be learned from data. For CNN model that learn a large number of parameter (might be in millions), it requires a very large training datasets. For example, in LFW standard the CNN model is trained using a dataset with millions of labeled faces. And fortunately it reached state-of-the-art performance, i.e., Facebook uses 4 million training samples to train the DeepFace model and Google use 200 million training samples to train the FaceNet model [6].

Convolutional Neural Networks are the deep learning framework, that is basically inspired by the biological network structure like the one exist in human brain visual cortex. For example, when we saw the LeNet-5 and HMAX, they

are both a ConvNet method and follow the model and feature of animal visual cortex structure that have a complex positioning of cells. The filtering mechanism of ConvNet model is examined just like the human visual cells to find out the relations in organic images [2].

Convolutional Neural Networks basically consists of three major layers, these are Convolutional Layer, Sub-Sampling or Pooling Layer, and Fully-Connected Layer. After merging all of these layers a CNN architecture has been set up. Let will try to see each layer shortly. Firstly, suppose there is a[32x32x3] inputted image in a raw pixel, this indicates an image with a width 32, height 32, and color RGB (Red, Green, Blue) format. By deciding the filter size, the Convolutional layer will calculate a dot product between the weights with each and every neuron. For example, the result of this CONV layer might be [32x32x12] if we used 12 filters. RELU layer will use an activation function called Rectified Linear Unit (RELU). Each neuron in the Convolutional layer uses these activation function to train the CNN several times faster. The RELU function is defined as a = max (z,o) where, a and z indicates the input, whereas o indicates the output of the activation function. The Sub-Sampling or POOL layer is used to reduce the dimensions (width, height), then the volume is like [16x16x12]. Finally, the Fully-Connected (FC) layer calculate the final values of each neuron and compare them, and resulting in a volume of with the size of [1x1x10] among the 10 categories. In FC layer, each neuron will be connected to all the numbers in the preceding volume [13].



|Input | Convolutional Layer | Subsampling Layer | Fully connected layer | Output|

Figure 2. The model of CNN in face recognition [9].

II. MOTIVATION

Nowadays, Image processing gives us a tremendous advantage to solve so many real-time problems. It is one type of recent technological advancements. So I am going to follow and work with such type of technological achievements. And also I am motivated on the success of deep learning in image processing and computer vision areas, and the artificial intelligence (AI) is doing great works in many applications and they make our life so much easy. Fortunately, this is all about my motivation to study in this area and to write this review paper.

III. EXISTING SYSTEMS

The existing system uses the traditional hand-crafted feature approach, that needs a lot of work from the human to determine and extract features and normally it is also has a less recognition rate [4].

Research discovered that there exist a challenging situations in the traditional hand-crafted features, such as pose, occlusion, facial expression, and illumination that affect the accurate identification and it has also a degraded performance, so that's why we are going to use the deep learning approach to enhance such type of issues [5].

IV. RELATED WORK

Towards a deep learning framework for unconstrained face detection [2] In this paper the authors describe about an approach called Multiple Scale Faster Region-based Convolutional Neural Network (MS-FRCNN) that used to strongly detect the landmark of peoples face from an image collected under various challenging circumstance or

conditions, i.e., occlusion, rotated face, illumination, low resolution, facial expression, and etc. There is two challenging face database used in this method, the Wider Face and the Face Detection and Benchmark (FDDB) database. As the authors describe these approaches achieve a state-of-the-art result same as with the other competitive methods.

When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition [3] In this paper the authors discuss CNN-based face recognition system (CNN-FRS). This CNN model is trained in one of the public datasets called Labeled Faces in the Wild (LFW). Experiments show that to have an efficient CNN-FRS and to improve the performance of face recognition, we have supposed to use the combination of multiple CNNs and metric learning. The multiple CNN is used to make the different networks capture the information from various section and plate to create a robust face representation.

Thermal face recognition using Convolutional neural network [4] In this paper authors present a Convolutional neural network (CNN) model for providing thermal face recognition. CNN is a type of neural network which is used to learn automatically the important features from the available raw data. The architecture is applied on RGB-D-T face dataset and according to the results, these new approaches give a better recognition rate than the traditional way of hand-crafted face feature recognition like Linear Binary Pattern (LBP), Histogram Oriented Gradients (HOG), and moments invariant. Three face detection challenges that reduce the recognition rate are considered, i.e., illumination, face rotation/pose, occlusion.

Local Binary Pattern Network: A Deep Learning Approach for Face recognition [5] In this paper the authors proposed a methodology called Local Binary Pattern Network (LBPNet), it is based on unsupervised deep learning mechanism and used to extract and compare high-level features in multilayer hierarchy. It's applied to the public standards, i.e., FERET and LFW. Two feature extraction techniques are used LBP and PCA. After completion of feature extraction, the LBPNet engage a simple network to compute the similarity of those extracted features. The LBPNet uses a deep network and regular network for feature extraction and classification, respectively.

Frankenstein: Learning Deep Face Representations using Small Data [6] In this paper the authors propose a technique to get a large number of training datasets of simulated images by combining real face images in a given dataset. This technique makes the model to learn from 10,000 training images. The authors propose a solution for training deep CNNs using datasets with small data. The method that the authors propose is a 2D data synthesis technique that used to expand the limited small datasets to larger ones. This face synthesis method will swap some facial parts of different face images to get/generate a new face. The authors apply this technique to LFW and CASIA NIR-VIS2.0 face databases.

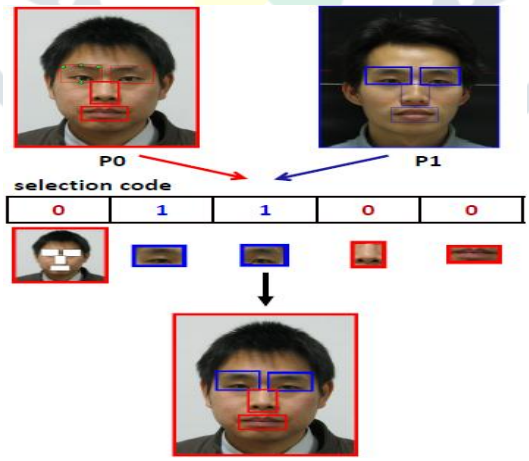


Figure 3. The synthesis process based on five parts: left-eye, right-eye, nose, mouth and the rest [6].

Face Recognition Using Deep Multi-Pose Representations [7] The authors introduce a multiple pose based deep learning model for face recognition. A face image is performed by various pose-distinct deep CNN models to produce multiple pose-distinct features. 3D rendering is applied to produce various face poses from the input image. Authors describe the experimental outcome on landmark detection, CNN layer selection and pose model selection.

Facial Landmark Detection Via Pose-induced Auto-encoder Networks [8] In this paper the authors propose a Pose-Induced Auto-encoder Networks (PIAN) method which use various pose-induced networks to handle landmark estimation in various pose conditions. By using this approach first the face images are classified into various pose conditions and the initial estimation will be learned, and then organize component-based local search and finally fine-tune each and every landmark point. The procedure that the authors proposed is shown in Figure 3.

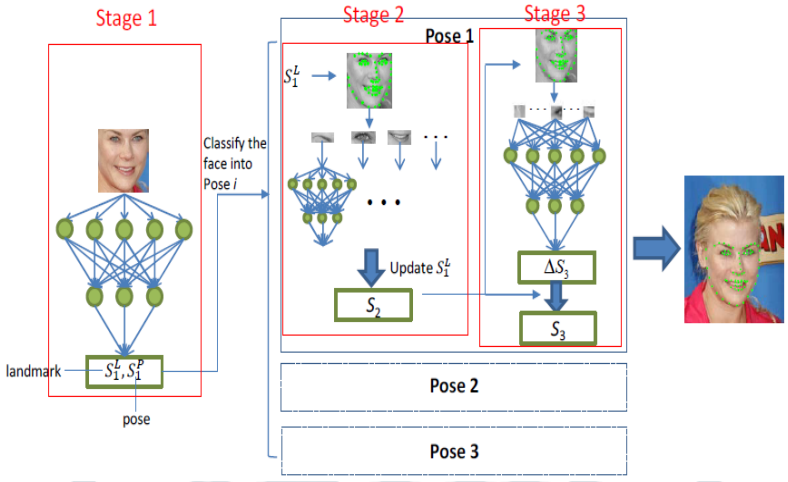
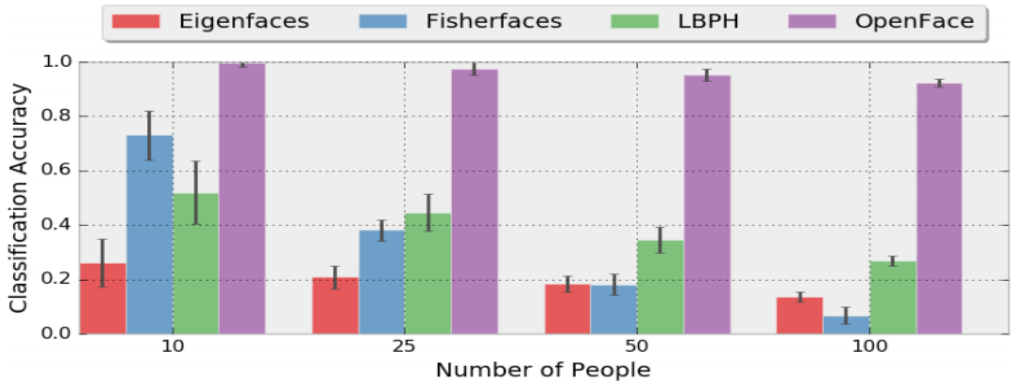


Figure 4. The process of facial landmark detection in the Pose-Induced Auto-encoder Networks (PIAN) [8].

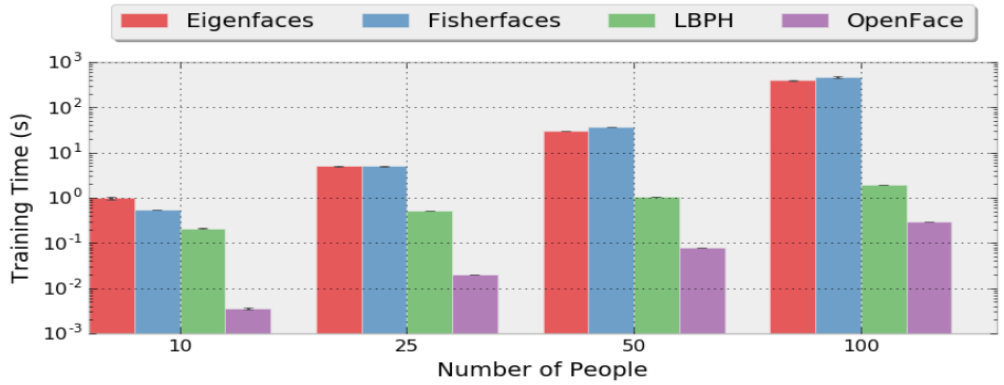
V. CONCLUSION

This survey paper focuses on the development of face recognition based on the deep learning approach called Convolutional Neural Network (CNN) by summarizing different literature studies. It is also discussed on the hand-crafted feature extraction and the deep learning CNN model. Deep learning is one branch of machine learning and an advanced way of learning mechanism that used in today's face recognition technology. It is also used to model and represent the data in a high-level abstraction. It provides multiple processing layers, and it is composed of various linear and non-linear transformation. These all make the face recognition system more accurate, even in a challenging situation, i.e., illumination, occlusion, rotated faces, etc.

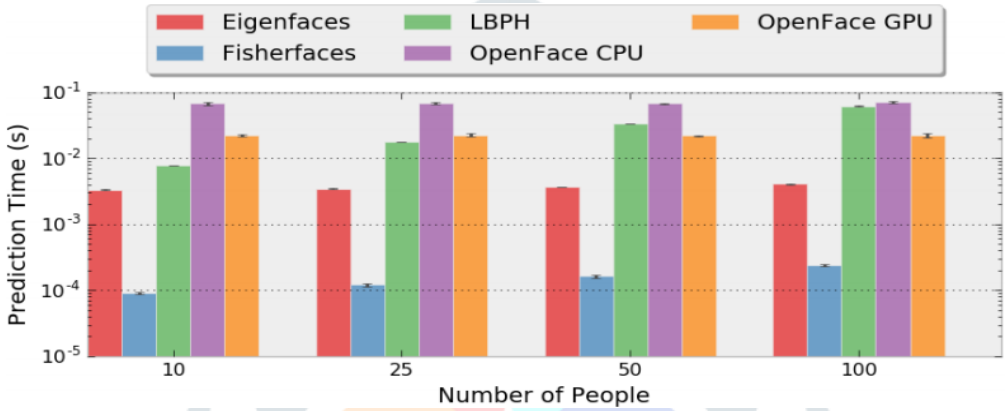
Now, we are going to compare the deep learning based CNN approach with the previous state-of-the-art face recognition methods. The result specifies that the deep learning based CNN approach is more advanced than the other hand-crafted feature extraction techniques in various circumstances.



a) Classification accuracy.



b) Training times.



c) Prediction time per image.

Figure 5. Comparing the deep learning approach with the other state-of-the-art face recognition methods according to accuracy, training time, and prediction time.

ACKNOWLEDEMENT

First of all, I want to thank God for giving me hope and the way to accomplish this paper. I would like to acknowledge the support that I had got from Dr. Vilin Parekh, Principal of Parul Institute of Engineering and Technology. I have a great thankful to Professor Harshal Shah, HOD of CSE. I want to express a warm gratitude to all staff members of Parul University Computer Science and Engineering Department.

REFERENCES

PAPERS

[1] Li Deng and Dong Yu, “Deep learning: methods and applications,” Foundations and Trends in Signal Processing, vol. 7, no. 3–4, pp. 197–387, 2014.

[2] Yutong Zheng, Chenchen Zhu, Khoa Luu, Chandrasekhar Bhagavatula, T. Hoang Ngan Le, Marios Savvides, "Towards a deep learning framework for unconstrained face detection", 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), 6-9 Sept. 2016, Niagara Falls, NY, USA, Pages: 1 - 8.

[3] Guosheng Hu, Yongxin Yang, Dong Yi, Josef Kittler, William Christmas, Stan Z. Li, Timothy Hospedales, "When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition", 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), 7-13 Dec. 2015, Santiago, Chile, Pages: 384 - 392.

[4] Zhan Wu, Min Peng, Tong Chen, "Thermal face recognition using Convolutional neural network", 2016

International Conference on Optoelectronics and Image Processing (ICOIP), 10-12 June 2016, Warsaw, Poland, Pages: 6 - 9.

[5] "Local Binary Pattern Network: A Deep Learning Approach for Face recognition", 2016 IEEE International Conference on Image Processing (ICIP), 25-28 Sept. 2016, Phoenix, AZ, USA, Pages: 3224 - 3228

[6] Guosheng Hu, Xiaojiang Peng, Yongxin Yang, Timothy M. Hospedales, Jakob Verbeek, "Frankenstein: Learning Deep Face Representations using Small Data", IEEE Transactions on Image Processing, 26 September 2017, Pages: 293 - 303, Volume: 27, Issue: 1.

[7] Wael AbdAlmageed, Yue Wu, Stephen Rawls, Shai Harel, Tal Hassner, Iacopo Masi, Jongmoo Choi, Jatuporn Lekust, Jungyeon Kim, Prem Natarajan, Ram Nevatia, Gerard Medioni, "Face Recognition Using Deep Multi-Pose Representations", 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 7-10 March 2016, Lake Placid, NY, USA, Pages: 1 - 9.

[8] Yu Chen, Wei Luo, Jian Yang, "Facial landmark detection via pose-induced auto-encoder networks", 2015 IEEE International Conference on Image Processing (ICIP), 27-30 Sept. 2015, Quebec City, QC, Canada, Pages: 2115 - 2119.

[9] Liping Yuan, Zhiyi Qu, Yufeng Zhao, Hongshuai Zhang, Qing Nian, "A Convolutional Neural Network based on Tensorflow for Face Recognition", 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 25-26 March 2017, Chongqing, China, Pages: 525 - 529.

[10] Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications", CMU School of Computer Science Carnegie Mellon University, June 2016, Pittsburgh.

WEBSITE

[11] Adam Geitgey, "Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning", Accessed on October 5, 2017.
https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78

[12] Cole Murray, "Building a Facial Recognition Pipeline with Deep Learning in Tensorflow", Accessed on October 5, 2017
https://hackernoon.com/building-a-facial-recognition-pipeline-with-deep-learning-in-tensorflow-66e7645015b8

[13] CS231n, "Convolutional Neural Networks for Visual Recognition", Accessed on December 28, 2017
http://cs231n.github.io/convolutional-networks/