Face recognition system Using Deep Neural Network with Convolutional Neural Networks

Erick Fernando¹, Denny Andwiyan², Dina Fitria Murad³, Derist Touriano⁴, Muhamad Irsan⁵

¹Information System Department, School of Information Systems, Bina Nusantara University, Indonesia
²,⁵Universitas Islam Syekh Yusuf (UNIS), School of Information Technology, Indonesia
³Information Technology Department, Faculty of computer science and engineering, University of Adi Wangsa, Indonesia
⁴Information System Department, Binus Online Learning, Bina Nusantara University, Indonesia

Email: dinafitriamurad@gmail.com

Abstract. Face recognition has long been a hot topic and challenging research point in areas such as image processing, pattern recognition, and machine vision. The face is a biometric feature with the intrinsic nature of a human. So that the face has self-stability, deep individual differences and can be an ideal basis for verification of an identity. In this research use, Deep Learning Network method uses to perform detection or face recognition. In this study, we present a framework that can be used to detect faces. This research is also able to present a DNN model that is used to study data sources from the data stream in sequence. The most important part of this study is able to adjust the capacity of the model from the simple one. This research uses experimental design method. The first step is a collection of face image data. Then the architecture design starts from the determination of the depth of the network, layout layers, and the selection of layer types that will be used to get the model based on input dataset and label name index.

1. Introduction

Technological developments are now growing steadily. This development also occurs in face detection. A highly accurate human identity authentication system is indispensable today due to the increasing number of crimes and losses through identity fraud. Token-based systems and knowledge-based systems have a high risk of being stolen or forgotten passwords, so the current system widely uses biometric systems such as access control, criminal identification, autonomous vending, and automated banking because of the uniqueness of biometric features and non-transferable characteristic[1][2]. The human face is one of the most common patterns facing the human visual system. The face is one part of humanity that can be said to have different characteristics [3].

Face recognition systems have been widely used in various fields and various systems within an organization or a company such as used in security systems, use of verification of a credit card, identification system of a crime, use in video telephony, and teleconference. Conventional face
recognition systems are able to provide very useful features. Thus recognizing facial images using a visual technique or reading with a computer and general knowledge of facial elements, this process has been carried out but does not produce a satisfactory process in recognizing images. In the face recognition process occurs a lot of difficulty in detection due to the position of the face (looks face), the presence of hair, or the expression of a face. Other factors that may affect accuracy include lighting, occlusion, and facial expression. Illumination is a change in the distribution of light because of the nature of the skin's reflectance and the control of the internal camera that can cast shadows on some parts of the face. Occlusion is the result of an object that includes the face, such as a scarf, turban, etc. Examples of smiling face expressions, laughing, angry, sad, shocked, and scared.

One of the most well-known techniques in overcoming this problem is the technique of artificial neural network classification (ANN). Artificial neural networks are the network of a group of small processing units that are modeled on the basis of human neural networks. ANN is an adaptive system that can change its structure to solve problems based on external and internal information flowing through the network in other words this technique has the ability to learn from experience.

In this research use Deep Learning Network method used to perform detection or face recognition. In this work, we present a novel framework for detection, which is able to learn DNN models from data streams sequentially, and more importantly, is able to adapt its model capacity from simple.

2. Theoretical Foundation

2.1. Deep Learning

A large number of emerging works adopt the principle of "shallow to deep". This approach exploits something that a superficial model meets faster than a deeper model, and this idea has been executed in several ways. Another approach does so implicitly by modifying the network architecture and objective functions for the network to allow input and flow through the network and gradually adapt to deep representation learning, such as Highway Nets.

However, everything that has been designed to optimize the loss function is based on the output obtained from the innermost layer. When batch convergence increases, the inferences made by the innermost layer take considerable time for convergence. Over the years, it has witnessed an effort in studying the architecture of neural networks, which incorporate hyper-parameters architecture into an optimization goal. Starting from overcomplete networks, they use regularizer that helps in removing neurons from the network.

2.2. DNN architecture

DNN consists of convolution, convection, and full connection (shown in Figure 1). This DNN consists of convection layer resolution, max-pooling and fully connected. A typical hierarchical characteristic extractor that maps the raw image intensity of the input image into the feature vector to be classified by multiple connected layers. DNN also has more maps per layer, and thus more connections and weights. After performing the 1st to the 4th stage on the convolutional layer and the union of multiple full layers connected further incorporates the output into the 1D vector feature.

The output layer is always a fully connected layer with one neuron per class. The softmax activation function used for the last layer ensures that each activation of the neuron output can be interpreted as a certain probability the class image input is a general hierarchical characteristic extractor that maps the pixel default intensity of the input image to the feature vector to be classified. by several fully-connected layers. All of these parameters can be adjusted optimized together with minimizing error classification error in the training set. Each convolutional layer performs a 2D convolution from its input map with a square filter. The output activation is obtained by summing the existing convolution response through the nonlinear activation function.
2.3. Face Recognition

Face recognition is a procedure for detecting and performing facial recognition of objects such as human faces, animals, or others that are done by an automated process using a computer [4], [5], [12]. Facial recognition has long been a hotspot point and a challenging research point in areas such as image processing, pattern recognition and machine vision [4]. Because the face is a biometric feature of the intrinsic nature of man, which has self-stability and deep individual differences and can be an ideal basis for identity verification [12].

In contrast to other biometric features such as iris, fingerprints. Thus there is no doubt that the facial nature is the most natural way, making it much more convenient and available for the user to be employed directly as a verification or detection of a security system process or the other. Face recognition has a complicated object recognition problem caused by many variations of expressions face, face position change and lighting [3]. The proposed facial recognition technique consists of three parts: preprocessing, feature extraction, and introductory steps.

2.4. Convolutional Neural Networks (CNN)

CNN is a variation of Multilayer Perceptron that is inspired by human neural networks. Research conducted by Hubel and Wiesel [13], which became the basis of this discovery, has carried out a study of visual cortex in cat vision. Visual cortex in animals is very strong in the visual processing system that ever existed. Convolutional Neural Networks is a layer that has a 3D neuron arrangement (width, height, depth) [13]. The width and height are the sizes of the layer while the depth refers to the number of layers. In general, type of layer on CNN can be divided into two namely: The image feature extraction layer, located at the beginning of the architecture is composed of several layers and each layer is composed of neurons connected to the local region of the previous layer [14]. The first type of layer is the convolution layer and the second layer is the layer pooling. Each layer is enforced by an activation function. Its position is intermittent between the first type and the second type. This layer accepts image inputs directly and processing [15].

The process to produce vector output to be processed at the next layer. Layer classification, arranged on several layers and each layer is composed of neurons that are fully connected with other layers [14]. This layer accepts input from the output layer of the feature extraction feature in the form of vector images and then transformed like Multi Neural Networks with the addition of several hidden layers. The output is a class scoring for classification.

Thus CNN is a method for transforming the original image layer per layer from the image pixel value into the class scoring value for classification. And every layer has a hyperparameter and some have no parameters (weight and bias on neurons). The Convolutional Layer first receives direct image input on the architecture. Operation at this layer is the same as convolution operation that is doing
combination operation of the linear filter to the local area. The filter is a representation of the receptive plane of neurons connected into the local connectivity of the image input.

3. Method
This research uses experimental design method. The steps taken are, the first step is a collection of face image data. Then the architecture design starts from the determination of the depth of the network, layout layers, and the selection of layer types that will be used to get the model based on input dataset and label name index. The steps of CNN algorithm to get model with network training as indicated in Figure 2.

![Figure 2. CNN Process Flowchart](image.png)

4. Experimental Result
In this research, experimenting by executing real-time face recognition implementation. The recognition process is executed by capturing the detected face image, converted into 48x48 grayscale. Grayscale images will be done by pre-processing ELBP into image feature extraction. Extraction results will enter the model and classification of data. Output classification model in the form of a label name that will be displayed on the monitor screen exactly above the left bounding box location detected a face (shown in Figure 3).
Testing Data with same faces position as shown on Table 1 to Table 6:

**Table 1.** Face at straight position

| Range | Time Detection | Motion | Skin detection status |
|-------|----------------|--------|-----------------------|
| 1000  | 47ms/21.28fps  | 0.15   | 0.12 Detects          |
| 2000  | 38ms/26.32fps  | 0.19   | 0.12 Detects          |
| 3000  | 34ms/29.41fps  | 0.20   | 0.12 Detects          |
| 4000  | 40ms/25.00fps  | 0.21   | 0.12 Detects          |
| 5000  | 25ms/40fps     | 0.21   | 0.12 Detects          |

**Table 2.** Face at left position 45°

| Range  | Time Detection | Motion | Skin detection status |
|-------|----------------|--------|-----------------------|
| 1000  | 47ms/21.28fps  | 0.15   | 0.12 Detects          |
| 2000  | 38ms/26.32fps  | 0.19   | 0.16 Detects          |
| 3000  | 34ms/29.41fps  | 0.20   | 0.12 Detects          |
| 4000  | 40ms/25.00fps  | 0.21   | 0.12 Detects          |
| 5000  | 25ms/40fps     | 0.21   | 0.17 Detects          |

**Table 3.** Face at right position 45°

| Range  | Time Detection | Motion | Skin detection status |
|-------|----------------|--------|-----------------------|
| 1000  | 45ms/21.28fps  | 0.15   | 0.12 Detects          |
| 2000  | 34ms/26.32fps  | 0.19   | 0.14 Detects          |
| 3000  | 33ms/29.41fps  | 0.20   | 0.12 Detects          |
| 4000  | 38ms/25.00fps  | 0.23   | 0.17 Detects          |
| 5000  | 23ms/40fps     | 0.21   | 0.17 Detects          |

**Table 4.** Face at straight position with accessories

| Range  | Time Detection | Motion | Skin detection status |
|-------|----------------|--------|-----------------------|
| 1000  | 46ms/21.28fps  | 0.15   | 0.12 Detects          |
| 2000  | 35ms/26.32fps  | 0.17   | 0.12 Detects          |
6

| Range | Time Detection | Motion | Skin detection | status   |
|-------|----------------|--------|----------------|----------|
| 1000  | 33ms/29.41fps  | 0.20   | 0.12           | Detects  |
| 2000  | 38ms/25.00fps  | 0.23   | 0.12           | Detects  |
| 3000  | 25ms/40fps     | 0.25   | 0.12           | Detects  |

Table 5. Face at left position

| Range | Time Detection | Motion | Skin detection | status   |
|-------|----------------|--------|----------------|----------|
| 1000  | 26ms/38.46fps  | 0.18   | 0.09           | Not Detects  |
| 2000  | 21ms/47.62fps  | 0.26   | 0.12           | Not Detects  |
| 3000  | 19ms/29.41fps  | 0.20   | 0.12           | Not Detects  |
| 4000  | 15ms/25.00fps  | 0.21   | 0.12           | Not Detects  |
| 5000  | 11ms/40fps     | 0.21   | 0.12           | Not Detects  |

Table 6. Face at right position

5. Conclusion and Future Research
The results of this study conducted through experiments, CNN is done with the camera in real time to detect faces. Face detection is done with some face position, among others, from straight face position, facing left, facing right, and down position. From the experimental process that the detection results with the face straight and down that detected and detection time with fast time, while the face left and right are not detected. From the results of this research in the future must provide the facial dataset that in real time can be stored well and become new data that can be used for the detection process so as to further accelerate the detection process.

References
[1] Howard W R 2008 Biometric System and Data Analysis: Design, Evaluation, and Data Mining20092Ted Dunstone and Neil Yager. Biometric System and Data Analysis: Design, Evaluation, and Data Mining . New York, NY: Springer 2008. 268 pp., ISBN: 978‐0‐387‐77625‐5 (contact www.springer.com for current price),” Kybernetes, 38 9, pp. 1637–1637, Oct. 2009.
[2] Siagian P and Fernando E 2014 Authentication Login E‐Library With Multimodal Biometrics System in International Conference on Electrical Engineering, Computer Science and Informatics, 2014 August pp 20–21.
[3] Charifa M S, Suliman, and Bikdash M 2007 Face Recognition Using a Hybrid General Backpropagation Neural Network IEEE Int. Conf. Granul. Comput. (GRC 2007), pp. 510–515.
[4] Rizk M R M and Taha A 2002 Analysis of neural networks for face recognition systems with feature extraction to develop an eye localization based method in 9th International Conference on Electronics, Circuits and Systems 2002 3 pp 847–850.
[5] Sekhar G R 2013 Face Recognition Using Neural Networks 3 3 pp. 1–5.
[6] Hecht-Nielsen R 1989 Theory of the Backpropagation Neural Network Proc. Int. Jr. Conf. Neural Networks 1 pp. 593–605.
[7] Ciresan D, Meier U and Schmidhuber J 2012 Multi‐column Deep Neural Networks for Image Classification Cvpr, pp. 3642–3649.
[8] Varshney S, Srivastava L, and Pandit M 2012 NN based integrated security assessment of power system using parallel computing *Int. J. Electr. Power Energy Syst.* **42**, pp. 49–59.

[9] Alvarez J M and Salzmann M 2016 Learning the Number of Neurons in Deep Networks Nips.

[10] Riesenhuber M and Poggio T Computational Models of Object Recognition in Cortex: A Review,” MIT, 2000.

[11] Ciresan D, Meier U, Masci J, and Schmidhuber J 2012 Multi-column deep neural network for traffic sign classification *Neural Networks* **32**, January 23, 2012, pp. 333–338.

[12] Yang G and Xu L 2012 Face recognition algorithm combined with DCT, PCA and BPNN *IEEE 5th Int. Conf. Adv. Comput. Intell. ICACI* 2012 pp. 682–685.

[13] Fukushima K 1980 Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position *Biol. Cybern.* **36** 4 pp. 193–202.

[14] Giusti A, Ciresan D C, Masci J, and Gambardella L M and Schmidhuber J 2013 Fast Image Scanning with Deep Max-Pooling Convolutional Neural Networks *Int.Conf.on Image Processing (ICIP)*

[15] Collobert R and Weston J 2008 A unified architecture for natural language processing: Deep neural networks with multitask learning *Proc. 25th Int. Conf. MachLearn* pp. 160–167.