Gender Classification Based on Face Recognition using Convolutional Neural Networks (CNNs)

R P Yuda\textsuperscript{1,ab}, C Aroef\textsuperscript{2b}, Z Rustam\textsuperscript{3c} and H Alatas\textsuperscript{4d}

\textsuperscript{1,2,3,4}Department of Mathematics, University of Indonesia, Depok, 16424, Indonesia, \textsuperscript{1}rivan.pratama@ui.ac.id, \textsuperscript{2}chelvian.aroef@ui.ac.id, \textsuperscript{3}rustam@ui.ac.id, \textsuperscript{4}hamimah.alatas@ui.ac.id

Abstract. Biometrics are physical or behavior characteristics of human that can be used to identify someone. One form of physical characteristics possessed by humans is fingerprints, retinal scanning, face and hand geometry, while one form of behavioral characteristics possessed by humans is handwriting, signatures, mouse usage analysis, walking patterns, etc. Basically, physical characteristics are more easily observed than behavioral characteristics. Therefore, physical characteristics are more often used in many aspects of security. One of the most common physical characteristics is face. By seeing the face, we can find out or predict how old they are, their gender and even their expression. However, there are still many mistakes in predicting a gender through person’s face. In fact, there are still many crimes in falsifying self-identity (such as gender). So, we need a method that is able to classify identity (gender) based on a person's face appropriately. One method that can be used is Convolutional Neural Networks (CNNs). Later, CNNs will classify a person's gender (male / female) based on a person's face image data. And based on.

1. Introduction

These Biometric is physical or behaviour characteristics of human which can be used to identify someone [1]. Fingerprint, retinal scans, face, hand geometry are the example of physical characteristics, while the example for behavioural characteristics are handwriting, signature, mouse use analysis, walking pattern, etc. [1]. Each individual has unique biometrics compare to others which make biometrics commonly used for security system. Physical characteristic is easier to be observed than behavioural characteristic. The most common physical characteristics is face.

Just by seeing someone face we can guess their origin, age, gender, and expression. Face recognition can be applied in many aspects. Determining the target of a product, we can advertise the product based on gender, the example is make up, women has higher chance for buying make up than men. In biometric sector, used to identify and grant access to someone. Give music, film, application on mobile phone also can be based on gender. If we do some searching for someone’s photo in a database, it would be much easier if we already know their gender, because it simplifies the searching process.

Everyone has the ability to recognize faces, but there could be some errors if there are a hundred or even a thousand faces. Therefore, an algorithm is needed to be able to do face recognition with large data. There were already some research trying to do face recognition with some methods, Support Vector Machine [3,4], Fuzzy Kernel C-Means [5], Fuzzy Kernel Learning Vector Quantization [6], Hidden
Markov Model [7], and so much more. Because in this study using classification, here are some methods already used in classification such as Support Vector Machines [8], Fuzzy Kernel C-Means [8], Fuzzy C-Means [9], Machine Learning [10], Normed Kernel Function-Based Fuzzy Possibilistic C-Means (NKFPCM) [11]. And this study will be using Convolutional Neural Network (CNN) to do face recognition. CNN already widely used for identification of shockable and non-shockable life-threatening ventricular arrhythmias [12], count palm trees [13], double JPEG compression forensics [14]. This study is expected to help to do face recognition which can be used in many aspects.

2. Research Methodology
The research method used in this experiment is Convolutional Neural Network (CNNs). Conventional Neural Network is one of a kind of Deep Neural Network that is the result of the development of the Multilayer Perceptron (MLP) [15,16]. As the name implies, Multilayer Perceptron has at least 3 or more layers used. Multilayer Perceptron can do classification with quite good results. Multilayer Perceptron can also be used to classify image data, but the drawback of this method is that the recognizable object must be in the middle. Therefore, the development of the Multilayer Perceptron is Convolutional Neural Networks (CNNs). Convolutional Neural Networks (CNNs) are a type of neural network commonly used in image data. CNNs can be used to detect and recognize objects in an image not just in the middle [18]. Broadly speaking, CNNs are not much different from neural networks usually because they consist of neurons that have a weight, bias and activation function. And then, CNNs have 2 major parts namely, Feature Extraction Layer and Fully-Connected Layer. Let see the following figure:

![Figure 1. Illustration of Convolutional Neural Network](image)

2.1. Feature Extraction Layer
Feature Extraction Layer "encodes" an image into features in the form of numbers that represent that image (Feature Extraction) [15]. So technically, CNNs is an architecture that consists of several stages. Each input (output) and output (output) of each stage consists of several arrays called feature maps. The Feature extraction layer consists of two important parts at each stage, namely, the Convolutional Layer and the Pooling Layer. Convolutional Layer is the main and most important layer to be used in the CNNs method. Followed by the Pooling Layer which is a layer of a combination that is used to extract the average value or maximum value on the pixel. Each input layer will have a different volume and is represented based on depth, height and width. The results that will be obtained are the amount that varies because it is influenced by the results of filtering the previous layers and also the number of filters used.

2.1.1. Convolutional Layer
Convolutional Layer consists of neurons arranged in such a way as to form a filter with length and height (pixels) [19]. The following is an illustration of the Convolutional Layer:
Figure 2. Illustration of Convolutional Layer

Based on Figure 2., all parts of the image will be shifted by the filters used. From each image position numbers are generated which are the results of the "dot product" between the image input and the values of the filters. The results of shifting these filters are called activation maps or feature maps. In the convolutional layer there are several components in it such as stride, zero padding and can find the dimensions of convolutional layer.

In accordance with the image channel below (in Figure 3.), where the first layer in the feature extraction layer is Convolutional Layers with a size of $5 \times 5 \times 3$. The purpose of the $5 \times 5 \times 3$ size is 5 pixels in length, 5 pixels in height and 3 in thickness. These filters will be shifted to all parts of the image. The output of the filter shift is called activation map or known feature map, by performing a "dot" operation between the input and the value of the filter. Consider the following illustration:

2.1.2. Stride

The next is stride which is a parameter in determining how many filter shifts in the convolutional layer size [19]. If the step value taken is 1, then the convolutional filter will shift horizontally first as much as 1 pixel, continue vertically as much as 1 pixel. However, the selection of steps or strides has no rules. But when using the smaller stride, more information obtained from an input. However, if the selected stride is smaller, the calculation of "dot product" between the input and the value of the filter is more
than choosing a larger stride. In addition, the small selection of strides does not necessarily get good performance.

2.1.3. Zero Padding
Padding or Zero Padding is a parameter that determines the number of pixels (if seen in figure 3 contains a zero number) that will be added to each input side [18]. This is used to manipulate the output dimensions of the convolutional layer (feature map). The purpose of using padding is that the dimensions of the output in the convolutional layer will always be smaller than the input (except, if we use of 1 × 1 filter with stride 1). The output will be reused to input for the convolutional layer, so that more information will be discarded. With padding, we can adjust the output dimension so that it can remain the same as the input dimension or so that it is not too reduced. So we can use convolutional layers with more successful extracted features. With padding too, the performance of the convolutional filter model will increase and focus on the actual information that is around the zero padding area.

2.1.4. The Dimensions of Feature Map
In Figure 3., the 5 × 5 input is the real dimension. If convolution is done with a 3 × 3 filter and stride 2 is chosen, a 3 × 3 feature map will be obtained (more information is built). And to calculate the dimensions of a feature map, we can use formulas [15]:

\[
\text{output} = \frac{W - N + 2P}{S} + 1
\]  

where,
- \( W \) = long or high of input
- \( N \) = long or high of filter
- \( P \) = as a padding
- \( S \) = as a stride

For example, based on Figure 3. will be obtained \( W = 5 \) \( N = 3 \) \( P = 1 \) \( S = 2 \) so that the output dimension of the feature map is 3. Overall, if the input of a convolutional layer is an image of size \( W_1 \times H_1 \times D_1 \), the output of that layer is a new "image" of size \( W_2 \times H_2 \times D_2 \), where [15]:

\[
W_2 = \frac{(W_1 - F + 2P)}{S} + 1 = H_2
\]

\[
D_2 = K
\]

Where:
- \( K \) = the number of filters used.
- \( F \) = the spatial size of the filter (width/height).
- \( S \) = the stride, or large filter shift in convolution.
- \( P \) = padding, the number of zeroes added to the image

2.1.5. Pooling Layer
After convolutional layer, pooling layer is usually done. In principle, the pooling layer consists of a filter with several sizes and stride that moves to follow all areas of the feature map [17]. Usually pooling is often used, namely max pooling and average pooling. For example, if the max pooling is 2 × 2 which will be used with stride 2, then for each filter move, the pixel area measuring 2 × 2 is chosen by its maximum value. If the value chosen is the average value, this is called average pooling. Consider the following illustration:
Figure 4. Illustration of Max Pooling

The purpose of using the pooling layer is to reduce the dimensions of the feature map (the other words down sampling), so that it accelerates the computational process because the updated parameters will decrease and overfitting.

2.2 Fully-Connected Layer

The feature map is created based on the feature extraction layer, which is still a multidimensional array, so it needs to be flattened or reshaped in the vector feature map which will be used as input from the fully-connected layer [20]. All layers that activate neurons from the previous layer are connected to neurons in the next layer (like a neural network) usually called the fully-connected layer. To be able to connect with all neurons in the layer, each activation of the previous layers needs to be converted first into one-dimensional data.

Fully-Connected layer usually uses a method called Multi-Perceptron Layer. This method has a target in the process of data so that it can carry out classification. The difference between fully-connected layers and convolution layers is the neurons from the convolution layer are connected to a specific input area, while the fully-connected layer has neuron connections to almost all parts [15, 20]. However, the two layers continue to do “dot product” operations, so the function is not much different.

3. Experimental Results and Discussions

3.1. Datasets

The images used in this dataset are face image data from each gender (male and female). The dataset used was taken from Google images and some of the results of taking photos of people's faces directly. The number of images used and taken as many as 500 sample images with the number of each gender as many as 250 samples of women's faces and 250 samples of men's faces again. The following are some examples of facial images obtained:

Figure 5. Example Training Set
Later the dataset that has been obtained will be used in testing the method by 70% of the images from the whole will be used as training data and 30% as testing data.

3.2. Style and spacing
In doing gender classification based on data on a person’s face image, the Convolutional Neural Networks (CNNs) method is used in Tensorflow. Tensorflow itself is a program framework or package that can help the CNNs system in processing image data. Later, on the selected CNNs model, 3 convolutional layers will be used, namely:
1. The first layer will use 32-3 x 3 filters
2. The second layer will use 64-3 x 3 filters
3. The third layer will use 128-3 x 3 filters

Each of these 3 layers will use a max-pooling of 2x2 with the number of strides 2 and zero padding added 1. The figure below is model of CNNs used for this research:

![Figure 7. Model of CNNs](image)

3.3. Evaluations
When the image data that has been obtained and also previously described in the dataset, the output results that will be obtained when applying the CNNs method in the dataset is the division of the dataset into 2 classes, namely classes for women and classes for men with accuracy and loss values. Following are the results of each accuracy value and loss value presented in graphical form:
Based on Figure 8 and Figure 9 above, it can be seen that both the accuracy value in the training accuracy and test accuracy in epoch 0-30 values increase quite high linearly, but does not increase much in subsequent epochs. Then, for accuracy values begin to stabilize or stagnate at epochs 80-100 with accuracy value training in the range of 0.94 and test accuracy value of 0.90-0.91.

In addition, overfitting is shown based on Figure 9. This is shown after the 13th epochs, the test loss value rises and is greater than the train loss value (although it does not increase too much).

4. Conclusions
Based on the results and discussion, the Convolutional Neural Network (CNNs) model that has been formed can classify which facial images constitute female gender and male gender. Thus, this method can be applied to individuals, especially an agency, in reducing acts of deception or falsification of self-identity (in gender). Based on evaluations obtained from epochs 1-100, the accuracy rate continues to increase and starts to be constant at 80-100 epochs with an accuracy level of 0.94. This shows that the CNNs method is able to be applied and applied to individuals or institutions in reducing lies or falsification of gender. Even though it has reached a good level of accuracy, it can be seen in epochs 0-30 that the accuracy is quite high because the CNN method is trying to memorize images.

5. Acknowledgments
This research supported financially by the Ministry of Research and Higher Education Republic of Indonesia (KEMENRISTEKDIKTI) with PDUPT 2019 research grant scheme, ID number 1621/UN2.R3.1/HKP.05.00/2019.
REFERENCES
[1] https://searchsecurity.techtarget.com/definition/biometrics accessed on 26th August 2019, 17.00 WIB.
[2] https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050/ accessed on 30th September 2019, 18.00 WIB.
[3] Rustam Z and Ruvita A 2018 Application Support Vector Machine on Face Recognition for Gender Classification Journal of Physics: Conference Series 1108 012067.
[4] Rustam Z and Faradina R 2018 Face Recognition to Identify Look-Alike Faces using Support Vector Machine Journal of Physics: Conference Series 1108 012071.
[5] Rustam Z and Faradina R 2019 Application of Fuzzy Kernel C-Means in face recognition to identify look-alike faces Journal of Physics: Conference Series 1218 012045.
[6] Rustam Z and Rika 2018 Face recognition using fuzzy kernel learning vector quantization Journal of Physics: Conference Series 1108 012068.
[7] Alhadi F H, & Fakhr M W and Seddik A F 2005 Hidden Markov Models for Face Recognition IASTED International Conference on Computational Intelligence 409-413.
[8] Rustam Z, Yaurita, F 2018 Insolvency Prediction in Insurance Companies Using Support Vector Machines and Fuzzy Kernel C-Means Journal of Physics: Conference Series 1028 012118.
[9] Rachman A A and Rustam Z 2016 Cancer classification using Fuzzy C-Means with feature selection Proceedings: International Conference on Mathematics, Statistics, and Their Applications (ICMSA) DOI: 10.1109/ICMSA.2016.7954302.
[10] Panca V and Rustam, Z 2017 Application of machine learning on brain cancer multiclass classification AIP Conference Proceedings 1862 DOI: 10.1063/1.4991237.
[11] Lestari A W and Rustam Z 2017 Normed kernel function-based fuzzy possibilistic C-means (NKFPCM) algorithm for high-dimensional breast cancer database classification with feature selection is based on Laplacian Score AIP Conference Proceedings 1862 DOI: 10.1063/1.4991247.
[12] Acharya U R et al 2017 Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network Future Generation Computer Systems 79 pages 952-959.
[13] Cheang E K, Cheang T K and Tay Y H 2017 Using Convolutional Neural Networks to Count Palm Trees in Satellite Images arXiv Preprint arxiv:1701.06462.
[14] Wang Q and Zhang R 2016 Double JPEG compression forensics based on a convolutional neural network EURASIP Journal on Information Security 23.
[15] O’Shea K and Nash R 2015 An Introduction to Convolutional Neural Networks arXiv Preprint arxiv:1511.08458v2.
[16] Krizhevsky A, Sutskever I and Hinton G E 2015 ImageNet Classification with Deep Convolutional Neural Networks Advances in Neural Information Processing Systems pp. 1907-1105.
[17] Li H, Lin Z, Shen X, Brandt J and Hua G A 2015 Convolutional Neural Network Cascade for Face Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 5325-5334.
[18] Turaga S C, Murray J F, Viren, Roth F, Helmstaedter M, Brigman K, Denk W and Seung H S 2010 Convolutional Networks can learn to generate affinity graphs for image segmentation. Neural Computation 22 pages 511–538.
[19] Ciresan D, Meier U, and Schmidhuber J 2012 Multi-column deep neural networks for image classification Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 3642–3649.
[20] Ciresan D C, Meier U, Masci J, Gambardella L M and Schmidhuber J 2011 Flexible, high performance convolutional neural networks for image classification IJCAI Proceedings-International Joint Conference on Artificial Intelligence 22 1237.