Gender Recognition from Face Images Based on Convolutional Neural Network (CNN)

Harry Yuliansyah[1], Lutfi Arazi[2]

[1] Electrical Engineering Department, Institut Teknologi Sumatera, Lampung Selatan, INDONESIA

Abstract. Human face images store much information, where the face is a multidimensional visual model of humans that can show identity such as gender. This paper proposes gender recognition from face images based on a convolutional neural network (CNN). The dataset source was obtained from the Kaggle website, Github, and the acquisition itself using a mobile phone camera. The total number of datasets used in this project is 6856 images divided into three parts: training, validation, and test. There are two classes used, namely the female and male classes. The CNN model is designed using five convolution layers for the feature extraction part. Several optimizations were made to the parameters of the number of epochs, learning rate, and batch size to get the best performance. The CNN model achieved a performance rate of 97%.

Keywords: gender recognition, CNN, epochs, learning rate, batch size

1. Introduction

A lot of image recognition research based on face images has been done; this is due to the rapid development of the techniques used. Parts of the human body store a lot of information, one of which is the face, where the face is a multidimensional visual model of humans that can show identity such as gender. The ability of humans to recognize faces is excellent. Humans can recognize thousands of faces daily and identify them despite physical changes, aging, beard changes, mustaches, or hairstyle changes. This technology's applications can be used for security, biological data (biometrics), database search, computer vision, and others. [1]

There are various methods that have been developed by researchers to recognize facial patterns such as Support Vector Machine (SVM) [2], Naive Bayes, Principal Component Analysis (PCA) [3], Convolutional Neural Network (CNN) [4], and other methods. In this practical work topic, the method used for gender identification is CNN. The advantage of this method is a simple architecture designed with image detection accuracy. This method uses little pre-processing compared to other image classification algorithms relatively. [5]

CNN is a new form of multilayer perceptrons. Multilayer perceptrons usually refer to the network connected to each neuron in one layer to the next. This algorithm can be relied on for the current era in image classification because it has a reasonably high detection accuracy level. The CNN model is needed for the feature extraction process, where the process involves facial images as the main object on this topic. Then, from this extraction, the computer will classify the image into parts that can provide information such as the work system of the human brain in organizing an object related to images. Still, every model has advantages and disadvantages, so that researchers are expected to be able to develop more effective methods for the future. [6]

2. Literature review

Gil Levi and Tal Hassner [7], tested the CNN model for gender and age detection with a total data size of 26 000 images. In this study, CNN model architecture is more straightforward than the CNN model proposed in this study, where there is only three times the convolution process. The results of the accuracy in this study resulted in an accuracy rate of 86.8%. This research does not explain the optimization of parameters such as batch size, epoch, learning rate, and others. Therefore in this research, the CNN model design and optimization are carried out to produce a more optimal level of accuracy.

3. Proposed Model and Parameter Optimization

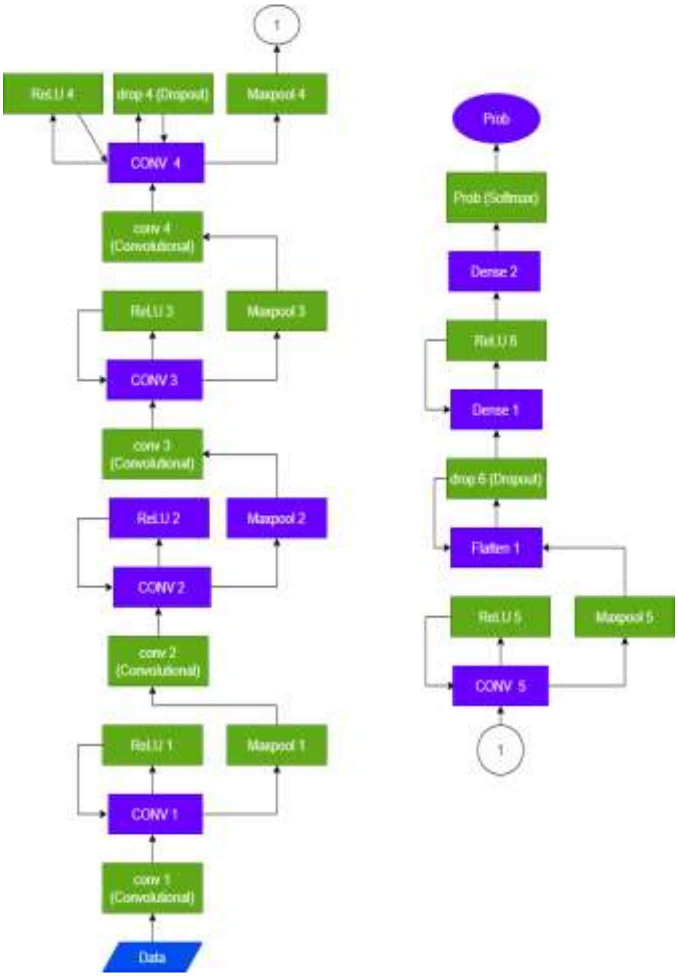


Figure 1 Complete schematic diagram of the CNN model

Figure 1 showed the architectural model of CNN proposed in this study. This architecture consists of five convolutional layers and three layers fully connected by some neurons. The model made is expected to reduce the risk of overfitting and underfitting. Overfitting occurs because the model only focuses on specific training datasets, so the model will be difficult to predict accurately if given another similar dataset. Overfitting will usually capture noise data that should not be needed in the prediction process.

Meanwhile, the underfitting condition occurs because the model cannot see the logic in the data's background. It cannot predict correctly where the dataset used is not relevant to the CNN model used in this condition. Another common cause of this problem is the amount and variation of data in the dataset that is used too little or can also be caused by the CNN model that is too complex.

Image resized to 150 x 150 pixels before processed, and then the next process can be defined as follows.

A. In the first layer, the filter size used is 32 with the filter dimensions 3 x 3, then the input form is 150 x 150 pixels of type RGB. The activation function used is ReLu. The next step is pooling to reduce the number of features with a size of 2 x 2.
B. In the second layer, the filter size used is 64, with the filter dimensions 3x3. The activation function used is ReLu with pooling in a 2 x 2 matrix.
C. In the third layer, the filter size used is 64, with the filter dimensions 3 x 3. The activation function used is ReLu with pooling in the form of a 2 x 2 matrix.
D. In the fourth layer, the filter size used is 128, with the filter dimensions 3 x 3. The activation function is ReLu with pooling in the form of a 2 x 2 matrix and a dropout layer.

E. In the fifth layer, the filter size used is 128, with the filter dimensions 3 x 3. The activation function used is ReLu with pooling in the form of a 2 x 2 matrix.
F. Data from the fifth process will enter the flattening stage. Pooling data in a 2-dimensional array is then converted into one-dimensional single vector data and a dropout layer. This data is connected to the Dense Layer, which has 512 nodes (or neurons) with the activation function used is ReLu.
G. The last layer has two softmax nodes; this process aims to return an array of two probability values when added together; the result is 1. Each node has a value indicating the probability that the image being processed is 1 of 2 label classes.

In this study, several parameter optimizations were carried out to increase the accuracy results. The optimized parameters include the number of epochs in the training process, the learning rate value, and the Batch Size value.

4. Result and Discussion

A. Dataset

The dataset contains various face images of men and women of different ages, ranging in age from 20 to 50. In this study, the total number of dataset images is 6856 images divided into three folders: training, validation, and test. Each face image has dimensions of 178 x 218 pixels with jpg and png extensions. These images are collected from Kaggle, GitHub, and some photos are taken manually using a cellphone camera.

B. Optimation of epoch parameter

Optimization will be carried out in referring to the number of epochs (iterations) based on the CNN model that has been made. The test is intended to determine the accuracy level of gender recognition using this method. The number of epochs used is 5, 10, 15, and 20. The following can be described in table 1.

Table 1 Accuracy and loss results

No	*Epoch*	Accuracy	*Loss*
1	5	0,93	0,1648
2	10	0,9433	0,188
3	15	0,9467	0,1387
4	20	0,9267	0,1988

The relationship between training loss and validation loss using epoch 5 and 10 indicates underfitting. In the training results with epoch 15 and 20, the results obtained have shown good accuracy results. However, on epoch 20, it can be seen that the level of accuracy has decreased compared to epoch 15. Based on the test results, the results obtained are the greatest accuracy on epoch 15, which is 0.9467, with the amount of training data used is 5402 and testing data 1309.

C. Optimation of leaning rate parameter

The use of the learning rate parameter affects the training time needed to achieve convergence. In general, the learning rate value is 0 <learning rate <1, the learning rate value is too small, causing the number of epochs to get bigger to reach the minimum global value. However, if the learning rate is too large, it can cause the training process never to reach the global minimum. It is essential to optimize the learning rate to maximize the accuracy of the CNN model that has been designed. The results of the learning rate parameter optimization can be seen in table 2.

Table 2 Learning rate value along with accuracy and loss results

Learning rate value	training time	Condition
0.01	150 s per-*epoch*	Underfitting
0.001	157 s per-*epoch*	Normal
0.002	164 s per-*epoch*	Normal
0.0001	167 s per-*epoch*	Underfitting

Underfitting occurs when the 0.01 and 0.0001 learning rates are used. However, normal conditions happen when the learning rate is 0.001 and 0.002. The training time required for each epoch is inversely proportional to the learning rate value. The smaller the learning rate value, the greater the training time per epoch.

D. Optimation of the batch size parameter

The batch size is the number of data samples distributed to the neural network into several subsets in the training process. In this study, a batch size optimization will be carried out on the dataset to determine the batch size effect. Table 3 showed the results of the batch size optimization.

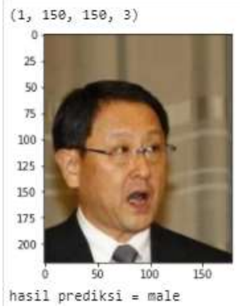
Tabel 3 Optimation of the batch size parameter

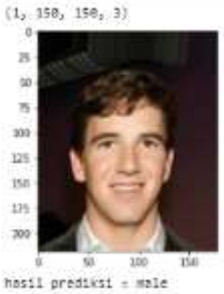
Sample	Training data	Test data	Validation data	Accuracy	Training time
A	120	120	3	0.93	126 s *per-epoch*
B	135	135	18	0.94	142 s *per-epoch*
C	150	150	33	0.92	157 s *per-epoch*
D	165	165	48	0.97	174 s *per-epoch*
E	180	180	63	0.94	185 s *per-epoch*

Based on table 3, the batch size value in sample D produces a reasonably high accuracy value, namely 97 %. However, when the batch size was added, the accuracy decreased to 94 %. A considerable batch size value does not always improve accuracy. It needs consideration in determining the batch size value, such as the hardware specifications used, the number of datasets, learning rate, epoch, and other factors.

Table 4 shows some of the results of gender detection based on facial images using the CNN model that has been made. The results of the gender prediction from the input image will be written under the image (male or female).

Table 4 Examples of gender detection results from face images





5. Conclusion

The gender detection based on face images using the CNN method showed success with reasonably high accuracy, namely 97%. Optimization is essential because each parameter influences the accuracy where the optimization must be adjusted to the CNN model, the number of datasets, classifier, activation used, image resolution, and others.

This study proves that CNN can be used for gender detection from face images because parts of the face have gender information/features for men or women. This feature extraction process is carried out by the CNN model that has been created. So it can be concluded that using facial images and the CNN method can be used for gender detection.

Acknowledgment

References

[1] B. Ammour, T. Bouden and L. Boubchir, "Face–iris multi-modal biometric system using multi-resolution Log-Gabor filter with spectral regression kernel discriminant analysis," in IET Biometrics, vol. 7, no. 5, pp. 482-489, 9 2018, doi: 10.1049/iet-bmt.2017.0251.

[2] J. G. da Silva Neto, J. L. M. Caldeira and D. D. Ferreira, "Face Recognition based on Higher-Order Statistics," in IEEE Latin America Transactions, vol. 16, no. 5, pp. 1508-1515, May 2018, doi: 10.1109/TLA.2018.8408448.

[3] B. Sugandi, I. Dewita and R. P. Hudjajanto, "Face Recognition Based on PCA and Neural Network," 2019 2nd International Conference on Applied Engineering (ICAE), Batam, Indonesia, 2019, pp. 1-5, doi: 10.1109/ICAE47758.2019.9221831.

[4] X. Yin and X. Liu, "Multi-Task Convolutional Neural Network for Pose-Invariant Face Recognition," in IEEE Transactions on Image Processing, vol. 27, no. 2, pp. 964-975, Feb. 2018, doi: 10.1109/TIP.2017.2765830.

[5] K. Bong, S. Choi, C. Kim, D. Han and H. Yoo, "A Low-Power Convolutional Neural Network Face Recognition Processor and a CIS Integrated With Always-on Face Detector," in IEEE Journal of Solid-State Circuits, vol. 53, no. 1, pp. 115-123, Jan. 2018, doi: 10.1109/JSSC.2017.2767705.

[6] A. Kishk and M. El-Hadidi, "AmpliconNet: Sequence Based Multi-layer Perceptron for Amplicon Read Classification Using Real-time Data Augmentation," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 2413-2418, doi: 10.1109/BIBM.2018.8621287.

[7] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks," 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, 2015, pp. 34-42, doi: 10.1109/CVPRW.2015.7301352.