Very Deep Learning for Gender Prediction with Modified VGG16

By

Berlian Al Kindhi
Department of Electrical Automation Engineering Instititut Teknologi Sepuluh Nopember Surabaya, Indonesia

Maurdhi Hery Purnomo
Department of Computer Engineering Institut Teknologi Sepuluh Nopember Surabaya, Indonesia

Abstract

The application of deep learning to detect an object has been widely applied in various fields, one of which is detecting the gender of the object being tested. Gender detection is to determine the image of a man or woman's face. In this case, the Convolutional Neural Network (CNN) method has been able to recognize the existence of this gender difference. However, in some cases, objects detected in conditions wearing accessories; such as hats, bandanas, scarves, hair ribbon, hijab (headscarves for Muslim women), and so on; to cover their heads so that some of their faces are covered too. The partial closure of the face on the object is one of the obstacles in determining the symmetrical shape between men and women on the object being detected. Gender detection on objects wearing accessories is a challenge for research in the field of deep learning. In this study, we propose a very deep learning method using VGG16 to detect gender in the case of objects wearing accessories. We change the Fully Connected Layer (FC Layer) on VGG16 with the number of layers we propose. VGG-16 has three of fully connected layers with a large number of layers, that is 4096 layers, while the fully connected layer we propose has 9 layers with the largest number of layers 128 and the smallest 16. In addition, until now there are no data sets for women using scarves, we have built our own datasets and we use transfer learning added with data augmentation techniques. CNN can predict the genders only get 77% testing accuracy. But if the CNN combined with our proposed modified VGG-16 as a feature extraction layer, the test accuracy increased significantly by 13.56%, that becomes 90.56%.

Keywords: Very deep learning, VGG-16, Gender Prediction, Image Augmentation

Deep structured learning has been a lot of success lately [1] [2] [3]. Several architectural models have been proposed and have high levels of performance, like R-CNN [4], until Faster R-CNN [5], and VGG-16 [6]. Datasets with various models have also been available as learning objects. Deep learning is able to recognize Cancer in CT Scan images [7] or MRI [8], classify various objects in one image [9] [10], or social analysis [11]. One sub-research on social analysis that can be done in deep learning is gender recognition [12].

Deep structured learning consists of high level abstraction modeling algorithms in the dataset [13]. Deep learning uses a set of non-linear transformation functions in hidden neurons in layers and depths [14]. Deep learning techniques can be used both for supervised learning, unsupervised learning, and semi-supervised learning in various applications of image recognition, speech recognition, and text recognition [15]. Deep learning is called deep because the structure and number of neural networks on the algorithm is very large and reaches hundreds of layers [16].

The architectural difference between machine learning and deep learning is from the feature extraction side [17]. Before it is processed into machine learning, there is a preprocessing stage which functions to extract the dataset [18]. Whereas in deep learning, the dataset extraction process is a unit of deep learning itself, also called the feature extraction layer [19]. So that in machine learning requires two learning processes and in deep learning there is a unified learning process. Along with the development of technology every year, the number of layers used also increased with the results is equivalent to the resulting accuracy [20].

Besides supervised, unsupervised, and semi-supervised deep learning, there is also hybrid deep learning [20]. This model is a combination of various Convolutional Neural Network architectures that have been tested on various data sets available and have good performance [21]. In this study, we adopted one very deep learning architecture, namely VGG-16 and adapted it to our experiments.

Besides VGG-16, there are several methods that are able to recognize the gender of an object, one of which is the muti-stage learning method [22]. The muti-stage method performs an introduction by dividing the learning process into several stages and dividing the region object before doing the encoder and decoder process [23]. Pre-training observation can improve the results of guided learning [24]. Another approach in predicting gender is by random forest method, this method is able to combine mathematical analysis with psychological sensors from a person's face [25].

The benefits of gender detection itself can be implemented in various fields [26]. Gender detection can analyze factors that affect perfectionism and academic difficulties in graduate students based on their gender [27]. Gender prediction can also be used to analyze the level of suicide in epilepsy sufferers [28], recognize how to market a product by gender [29], and gender recognition based on analysis of one's handwriting patterns [30].

In this study we propose a very deep learning method with VGG-16 architecture modified in its fully connected layer. VGG-16 architecture has 1000 output neurons because VGG-16 is trained on image-net that has 1000 classes, we modify the output neurons into one output neuron because our case study is binary classification. We also build datasets in the form of women using scarves in female classes. The test results show that the very deep learning architecture that we propose is able to recognize gender differences in the tested datasets.

Very Deep Gender Prediction

Data Augmentation

To obtain optimal deep learning processing performance data, the data needed for deep learning must be in large quantities. Because the more image data is used, the learning process is getting better. From the training dataset we collected, there were only 400 male photos and 400 female photos. We obtain this data from the UI Faces API 300 images, the Random User Generator 300 images and our own datasets 200 images. For women's data, we take pictures on objects of women using accesories. In total there are 800 photos for training and 240 photos for testing which are divided into two classes, male and female.

Data from random user generators has a size of 128 x 128 pixels, but the image data originating from UI Faces has different sizes. So pre-processing must be done first to get the same photo size. With the same image size will facilitate the process of training and learning. But to produce optimal performance, the data is still insufficient. Therefore we use data augmentation techniques. Data augmentation is a technique for manipulating data without losing the core information from the data image.

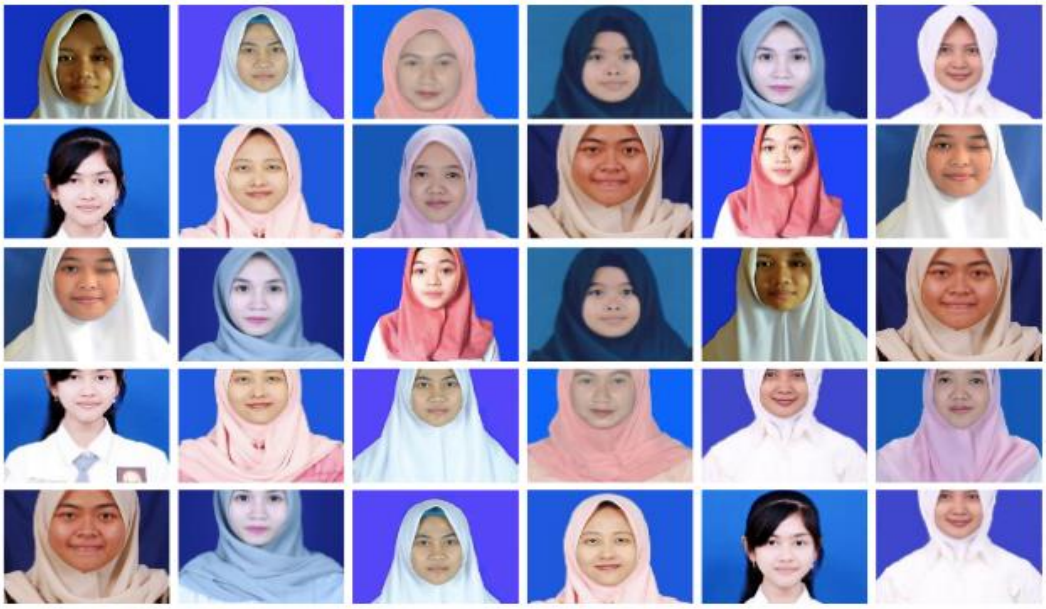


Fig. 1. *Example of our build datasets for women with accesories*

Fig 1. is an example of a dataset that we built using image objects from native Indonesian women. Most of the sample data were women using scarves and women with bangs. This amount of data is 100 images out of the total 200 images that we have collected and 25% of the total number of women images. Then from the total of all the data images, we do the augmentation process to increase the amount of training data with better results.

In the image data, augmentation can be rotated, flip, and crop. In this study, the augmentation techniques we use are zoom, flip, shear, and beyond whereas for the rescale parameter we use is dividing RGB values to simplify classification. Our RGB values from 0-255 divide by 255 so that the RGB range is 0 to 1. As for data testing, the data processing techniques we use are only rescale image.

After doing the augmentation technique, we do the process of converting data in the form of raw image into a dataset image for training and testing. We use the flow_from_directory method from Image Data Generator to process the dataset change. Then the sample data are ready to input to the deep learning

Very Deep Learning with VGG 16

As discussed in the previous sub-chapter to make a deep learning model that can recognize gender well, a good learning process is needed with a sufficient dataset. Besides using data augmentation techniques, we also use transfer learning techniques. Transfer learning is a method that utilizes models that have been trained in a dataset to solve other similar problems by transferring knowledge from the learning outcomes as the starting point of the learning process that is designed. The package transfer learning is so flexible that it can be modified and update its parameters to match our data set.

Our training process generally following VGG-16 arsitektur, the training process is carried out by optimizing multinomial logistic regression using the gradient descendant approach. Whereas in the evaluation process on deep learning architecture using VGG-16 it can be in the form of single-scale evaluation, multi-scale evaluation, and multi-crop evaluation.

The first evaluation conducted was to use a single scale model that is analyzing the performance of each ConvNet models on a single scale, where the layer configuration of a single scale can be observed in Equation 1-3. with Tr is Data Training, and Ts is Data Testing. In previous studies, it was concluded that training by designing image augmentation processes using a scale jittering approach is indeed better for multi-scale image statistics [6].

$Tr = Ts\,; for\ fixed\ Tr$ (1)
$Ts = 0.5(Tr_{min} + Tr_{max})\ ;\ for\ jittered$ (2)
where
$Tr \in =\ [Tr_{min}, Tr_{max}]$ (3)

The approach we take for the transfer learning process is the Pre-trained Model, the approach has 3 steps namely select model, reuse model, and tune model. Many research institutes release models on large datasets that we can use for selected candidate models. In this study, 75% of the model data used were obtained from the two research institutions described in the previous sub-chapter and the remaining 25% were data that we collected ourselves. The next step is to reuse the model, which is where the pre-trained model can be used as a starting point for the model in the second interesting task. This process can involve the use of all or part of the model, at this stage we use augmented modeling techniques. The final step of the transfer learning process is to tune the model, the model that has been processed needs to be adjusted or refined in the input-output pair data available for the desired task. In this study, we made adjustments to the number of input and output neurons that can be observed in Fig. 6.

The transfer learning model that we used in this study was VGG16. All weights on the VGG16 model have been trained in ImageNet data so that they have knowledge about the introduction of colors, textures, objects, and so on. So this transfer learning model can be used to extract features from all photos in the dataset we use.

Our Purpose Fully Connected Layer

Fully connected layers on the VGG16 architecture consist of 4096-4096-4096 neurons in the hidden layer and 1000 neurons in the output layer. This is because VGG-16 is trained on ImageNet which has 1000 classes. ImageNet is a dataset consisting of 1,200,000 image training and 100,000 for testing. The Image-Net dataset consists of 1000 classes so that each class on Image-Net consists of 1,200 images. Whereas the dataset in this study has only 2 classes, male and female, so we propose Fully Connected Layer with two output neurons. Transfer learning is a technique or method that utilizes a model that has been trained on a dataset to solve other similar problems by using it as a starting point, modifying and updating its parameters so that it matches the new dataset.

We use the VGG16 package so that we can arrange the learning layer by layer transfers and download the Weights. The include_top = False argument indicates that we do not use Fully Connected Layer on VGG 16, so that during the prediction process, the dataset will flow from the Feature Extraction Layer from VGG-16 to the layer we are proposing. The result is a feature map that we can save in the train_features.npy and val_features.npy file. The results in the file are presented and used to conduct training on the Fully Connected Layer that we propose. We use the feature map as training data and data testing for the proposed Fully connected layer.

The very deep learning architecture that we propose consists of 87.969 parameters that will continue to be updated during the training process. Whereas in the feature extraction layer, there are 4 Convolutional Layer, Zero Padding Layer, and Max Pooling Layer. While in the fully connected layer, there are two layers with 32 and 1. 32 are hidden layers in the fully connected layer and 1 is the output neuron. We do binary classification so that only requires one output neuron, the results of the training will produce binary 0 (male) or 1 (female).

The activation function that we use is the output layer is sigmoid activation with binary crossentropy as a loss function. Another way to look for two different results (output neurons) is to use the softmax activation function with categorical crossentropy as its loss function. For all hidden layers we use ReLU as the activation function and for the layer output we use sigmoid to activate the function. To get better training results, we also use ADAM as an optimizer.

ADAM is one of the optimization algorithms that has been specifically designed to train deep neural networks adaptively. Adam's optimization algorithm is an extension of the decrease in a stochastic gradient that is usually used in training data on deep learning. ADAM is developing broader optimization for deep learning applications, image processing, and natural language processing. ADAM combines the benefits of AdaGrad and RMSProp. ADAM's advantage was to adapt the level of learning parameters based on the average of the first moment (average) as in the RMSProp, as well as utilizing the average of the second moment of the gradient (not centralized variant).

Analysis

We tested the Fully Connected Layer that we proposed before combining it with the VGG-16 architecture. The purpose of this test is to determine the performance of the architecture. One way to find out the performance of deep learning architecture is to test it. Architecture with the number of neurons and layers of hundreds does not necessarily produce good performance and vice versa. But the increasing number of datasets will facilitate the learning process in predicting an object.

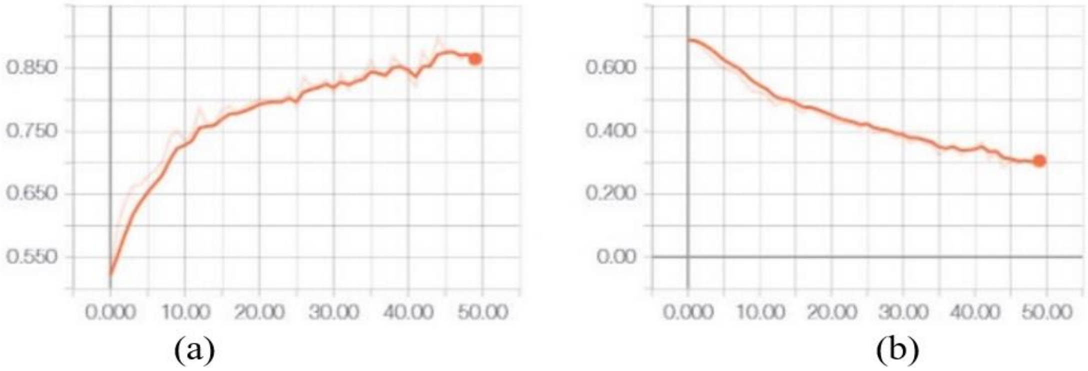


Fig.2. *Analysis of the results of training data testing: (a) accuracy and (b) loss error*

At Fig. 2. it can be observed that the accuracy value of the training data after the 50th epoch shows good results, that is, the accuracy value is above 85%, the performance value has dropped at one 40th epoch but the accuracy trend has increased again. Along with the increasing value of accuracy, the amount of training data that experiences a loss error is also less, so the value of the loss error at the 50th epoch reaches 0.3.

After carrying out the learning process using as many as 800 training data we carried out the testing process using 240 test data images. The test results can be observed in Fig. 3. In contrast to the training process, trends from the test results appear to be unstable and experience fluctuations in each of the epochs, although there is still an increase in accuracy in each epoch, the trend of accuracy testing can be observed in Fig. 3. (a). While for the loss error value, the testing process has a significant change in value, marked by the high up and down line in Fig. 3. (b).

The test results using the two class architecture for output neurons only reach an accuracy value of 77% and crossentropy loss value of 0.4982. So it can be concluded that there are

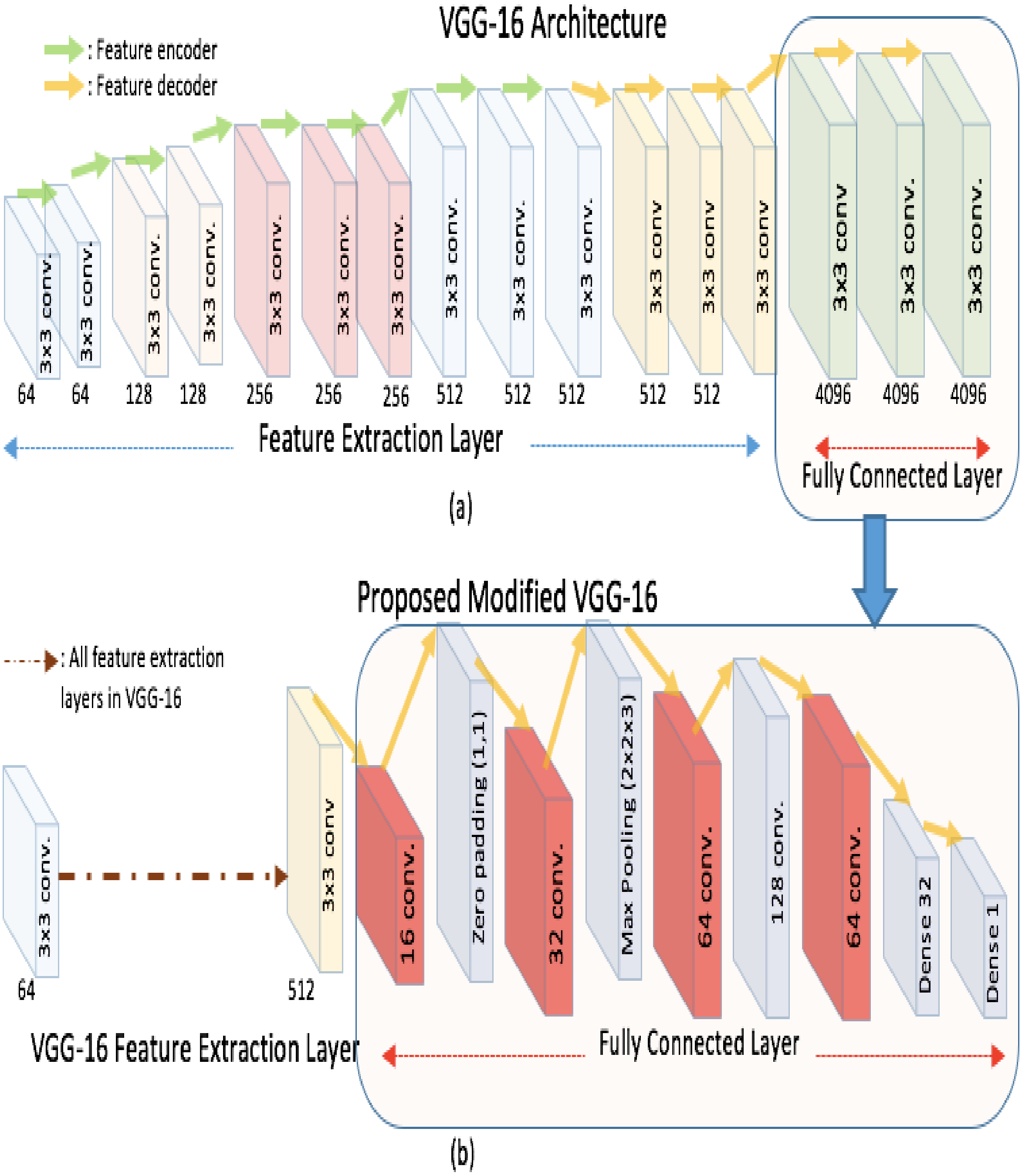


Fig. 6. *The difference between the VGG-16 architecture (a) and the modified VGG-16 architecture that we proposed (b)*

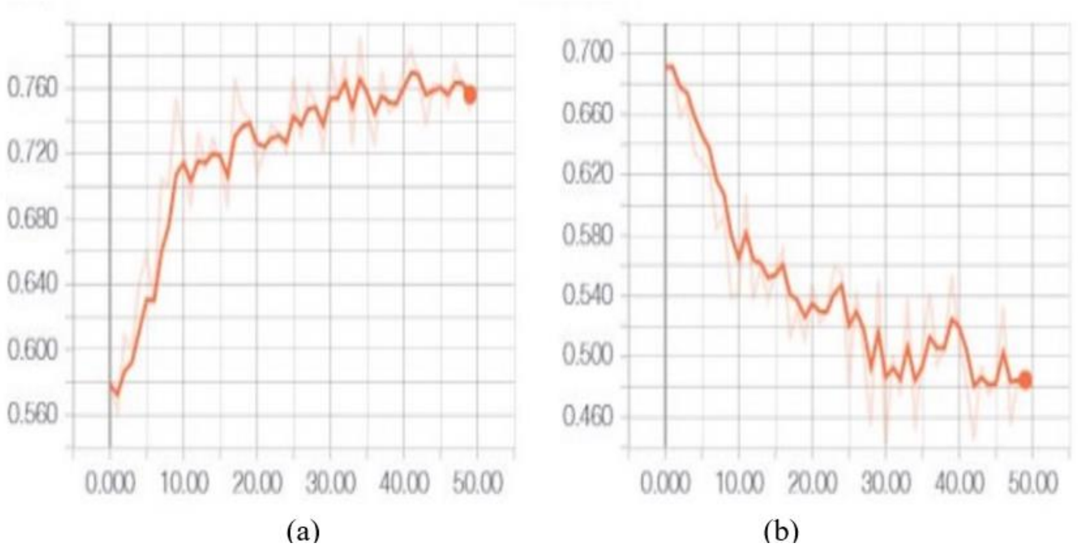


Fig.3. *Analysis of test results using 240 datasets: (a) accuracy and (b) loss error*

indications of overfitting which causes low loss and accuracy. Therefore we need a deep learning model that is able to transfer the learning outcomes by combining the VGG-16 model with the the model we tested earlier. We use the VGG-16 model for the Feature Extraction Layer and the model we propose for Fully Connected Layer.

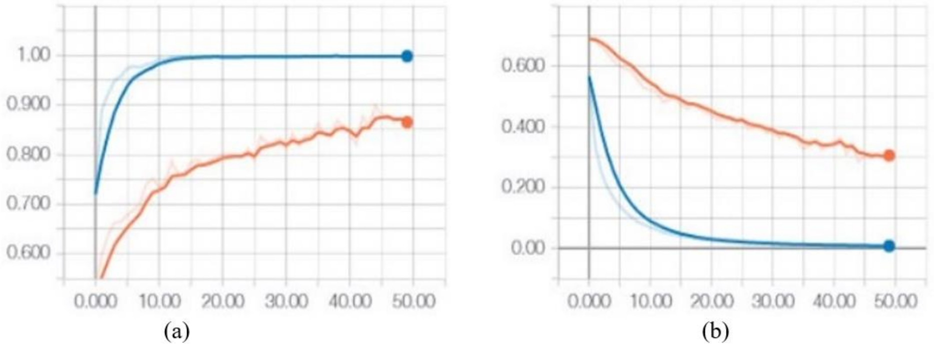
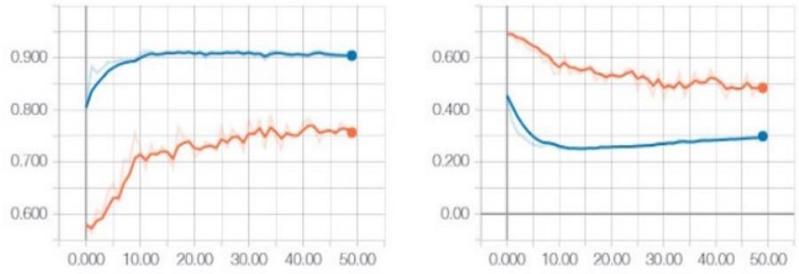


Fig. 4. *The results of purposed modified VGG16 learning: (a) accuracy and (b) loss error*

At Fig. 2-5., In the graph (a), the x axis shows the epoch iteration and the y axis shows the level of accuracy. Whereas in figure (b), the y axis is the level of loss error for each epoch (y axis). After combining the VGG-16 model as a feature extraction layer with the fully connected layer that we propose, the learning results show a high accuracy value of 1, or it can also be said that there is no prediction error, if the accuracy of predictions reaches one, the loss error 0. The results of the proposed modified VGG-16 can be observed in Fig. 4. For the results of the test, as we can observe in Fig. 5., the value of accuracy reaches 90, 56%, although not yet 100%, but the value is far better than the results of deep learning without using VGG-16 as its Feature Extraction Layer.

In this study, the process of augmented data and transfer learning is very helpful in increasing the level of prediction accuracy. Augmented data significantly adds to the number of learning models and the transfer learning process makes it easy for us to use deep learning architectures that have been well tested to be changed and adapted to our needs.



(a) (b)

Fig. 5. The test results used CNN (red graph) and purposed modified VGG-16 (blue graph) with (a) accuracy and (b) loss error

Discussion

The application of very deep learning in detecting faces on the image net has been done a lot, one of which is the VGG-16 architecture. The VGG-16 architecture has high accuracy to test the net image, in this study we simplified the VGG-16 architecture for gender recognition, which changed it from 1000 classes to 2 classes. As shown in Fig. 2 and Fig. 3., the results of learning and testing on architectures with two classes for neuron output are not very good. In the training test data only reached 86% and for 240 data sets only reached 74%. Therefore architecture is needed, the architecture that has been trained in the image database, one of which is VGG-16.

Combining VGG-16 as a Feature Extraction layer shows a significant improvement in performance. Initially, when testing only used CNN with two neuron outputs it only produced 77%, but after adding VGG-16, the value of accuracy rose by 13.56% which was 90.56%. In Table 1. it can be observed that the results of training and testing data processing using data testing produce an architecture that we propose to be superior, namely CNN by using transfer learning with VGG 16 modified as the model.

Table 1. *comparison CNN accuracy and our propose architecture*

No.	Architecture	Data set	Accuracy	Error
1	CNN	Training	0.85	0.3
2	CNN + modified VGG-16	Training	1	0
3	CNN	Testing	0.77	0.6
4	CNN + modified VGG-16	Testing	0.91	0.3

From our experiments, several learning methods produced predictions of two different faces on one object who was using accessories and not. The system should be able to detect that it is the same person. In addition, the difficulty of deep learning in recognizing women with hijabs is because some of their faces are covered by scarves and sometimes the scarf model used can vary so that it can cover different parts of the face. That will be the focus of our next research.

Conclusion

The use of deep learning for gender prediction is usually done on a normal image dataset on female and male. In this study, we propose a case study with an object who was using accessories. However, the current dataset does not include the group of women with hijab (head scraf for women). So when we encounter these cases, deep learning systems will be difficult to detect because they have never been trained in the previous dataset. In this study we propose a dataset in which there are women using a scarf in the female class.

In the trial process, we tested the CNN architecture with less than maximal results of 76% accuracy, then we applied the transfer learning method at the fully connected layer for better results. The data set we use is only 800 images for the total number of male and female subjects, therefore in the process of data normalization, we use the augmented method.

In this research, we proposed a modified VGG-16 method that we adjust Fully Connected layer with the number of classes that we predict. The results of the training showed that the accuracy value of the modified VGG-16 that we proposed reached 100% with loss error 0. As for the test results, there were 90.56% of the data that was successfully predicted correctly.

References

A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, A. Maida, "Deep learning in spiking neural networks," *Neural Networks,* 2018.

M. Yan, M. Li, H. He, J. Peng, "Deep Learning for Vehicle Speed Prediction," *Energy Procedia,* vol. 152, pp. 618-623, 2018.

Y. Yuan, G. Xun, Q. Suo, K. Jia, A. Zhang, "Wave2Vec: Deep representation learning for clinical temporal data," *Neurocomputing,* vol. 324, pp. 31-42, 2019.

R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *arXiv:1311.2524v5,* 2014.

S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object," *arXiv:1506.01497v3,* 2016.

K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *International Conference on Learning Representations*, San Diego, 2015.

M. P. McBee, O. A. Awan, A. T. Colucci, C. W. Ghobadi, N. Kadom, A. P. Kansagra, S. Tridandapani, W. F. Auffermann, "Deep Learning in Radiology," *Academic Radiology,* vol. 25, pp. 1472-1480, 2018.

A. S. Lundervold, A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik,* 2018.

L. Zhong, L. Hu, H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sensing of Environment,* vol. 221, pp. 430-443, 2019.

C. Augusta, R. Deardon, G. Taylor, "Deep learning for supervised classification of spatial epidemics," *Spatial and Spatio-temporal Epidemiology,* 2018.

A. Chatterjee, U. Gupta, M. K. Chinnakotla. R. Srikanth, M. Galley, P. Agrawal, "Understanding Emotions in Text Using Deep Learning and Big Data," *Computers in Human Behavior,* vol. 93, pp. 309-317, 2019.

A. Dhomne, R. Kumar, V. Bhan, "Gender Recognition Through Face Using Deep Learning," *Procedia Computer Science,* vol. 132, pp. 2-10, 2018.

A. Sboev, I. Moloshnikov, D. Gudovskikh, A. Selivanov, R. Rybka, T. Litvinova, "Deep Learning neural nets versus traditional machine learning in gender identification of authors of RusProfiling texts," *Procedia Computer Science,* vol. 123, pp. 424-431, 2018.

D. Triantafyllidou, P. Nousi, A. Tefas, "Fast Deep Convolutional Face Detection in the Wild

Exploiting Hard Sample Mining," *Big Data Research*, vol. 11, pp. 65-76, 2018.

R. Garcia, A. C. Telea, B. C. da Silva, J. Tørresen, J. L. D. Comba, "A task-and-technique centered survey on visual analytics for deep learning model engineering," *Computers & Graphics*, vol. 77, pp. 30-49, 2018.

A. Sboev, I. Moloshnikov, D. Gudovskikh, A. Selivanov, R. Rybka, T. Litvinova, "Automatic gender identification of author of Russian text by machine learning and neural net algorithms in case of gender deception," *Procedia Computer Science*, vol. 123, pp. 417-423, 2018.

Z. Md. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, K. Mizutani, "State-of-the-Art Deep Learning: Evolving Machine Intelligence Toward Tomorrow's Intelligent Network Traffic Control Systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, p. 69, 2017.

D. Zhang, X. Han, C. Deng, "Review on the research and practice of deep learning and reinforcement learning in smart grids," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 3, pp. 362-370, 2018 , Volume: 4 , Issue: 3.

T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, K. Yan, "A Deep Neural Network-Driven Feature Learning Method for Multi-view Facial Expression Recognition," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2528 - 2536, 2016.

S. Shen, M. Sadoughi, M. Li, Z. Wang and C. Hu, "Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries," *Applied Energy,*, vol. 260, p. 114, 2 2020.

H. Goh, N. Thome, M. Cord, J. Lim, "Learning Deep Hierarchical Visual Feature Coding," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 12, pp. 2212-2225, 2014 , Volume: 25 , Issue: 12.

M. Hasan, A. K. Roy-Chowdhury, "A Continuous Learning Framework for Activity Recognition Using Deep Hybrid Feature Models," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 1909-1922, 2015.

J. Fang, Y. Yuan, X. Lu, Y. Feng, "Muti-Stage Learning for Gender and Age Prediction," *Neurocomputing*, 2019.

Y. Choi, Y. Kim, S. Kim, K. Park, J. Park, "An on-device gender prediction method for mobile users using representative wordsets," *Expert System with Applications*, vol. 64, pp. 423-433, 2016.

G. Antipov, M. Baccouche, S. Berrania, J. Dugelay, "Effective training of convolutional neural networks for face-based gender and age prediction," *Pattern Recognition*, vol. 72, pp. 15-26, 2017.

T. Chaudhuri, D. Zhai, Y. C. Soh, H. Li, L. Xie, "Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology," *Energy and Buildings*, vol. 166, pp. 391-406, 2018.

M. Duan, K. Li, C. Yang, K. Li, "A hybrid deep learning CNN–ELM for age and gender classification," *Neurocomputing*, vol. 275, pp. 448-461, 2018.

M. E. Cowie, L. J. Nealis, S. B. Sherry, P. L. Hewitt, G. L. Flett, "Perfectionism and academic difficulties in graduate students: Testing incremental prediction and gender moderation," *Personality and Individual Differences*, vol. 123, pp. 223-228, 2018.

V. V. Kalinin, D. A. Polyanskiy, "Gender and suicidality prediction in epilepsy," *Epilepsy & Behavior*, vol. 7, no. 4, pp. 657-663, 2005.

D. Cooke, A. P. Fernandes, P. Ferreira, "Product market competition and gender discrimination," *Journal of Economic Behavior & Organization*, 2018.

N. Bi, C. Y. Suen, N. Nobile, J. Tan, "A multi-feature selection approach for gender identification of handwriting based on kernel mutual information," *Pattern Recognition Letters*, 2018.