

Comparative Study of VGG16 and MobileNetV2 for Masked Face Recognition

Faisal Dharma Adhinata¹, Nia Annisa Ferani Tanjung², Widi Widayat³, Gracia Rizka Pasfica⁴, Fadlan Raka Satara⁵

^{1,2,4,5}Department of Software Engineering, Faculty of Informatics, Institut Teknologi Telkom Purwokerto, Indonesia

³Department of Informatics Engineering, Faculty of Informatics, Institut Teknologi Telkom Purwokerto, Indonesia

ARTICLE INFO

Article history:

Received May 28, 2021

Revised June 22, 2021

Accepted July 09, 2021

Keywords:

Coronavirus;
Face Recognition;
MobileNetV2;
Transfer Learning;
VGG16

ABSTRACT

Indonesia is one of the countries affected by the coronavirus pandemic, which has taken too many lives. The coronavirus pandemic forces us to continue to wear masks daily, especially when working to break the chain of the spread of the coronavirus. Before the pandemic, face recognition for attendance used the entire face as input data, so the results were accurate. However, during this pandemic, all employees use masks, including attendance, which can reduce the level of accuracy when using masks. In this research, we use a deep learning technique to recognize masked faces. We propose using transfer learning pre-trained models to perform feature extraction and classification of masked face image data. The use of transfer learning techniques is due to the small amount of data used. We analyzed two transfer learning models, namely VGG16 and MobileNetV2. The parameters of batch size and number of epochs were used to evaluate each model. The best model is obtained with a batch size value of 32 and the number of epochs 50 in each model. The results showed that using the MobileNetV2 model was more accurate than VGG16, with an accuracy value of 95.42%. The results of this study can provide an overview of the use of transfer learning techniques for masked face recognition.

This work is licensed under a [Creative Commons Attribution-Share Alike 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



Faisal Dharma Adhinata,

Department of Software Engineering, Faculty of Informatics, Institut Teknologi Telkom Purwokerto, Indonesia

Email: faisal@ittelkom-pwt.ac.id

1. INTRODUCTION

More than 1,390,000 people have recently been infected with the coronavirus, which comes from Wuhan, China. The month of March 2021 has caused more than 37,000 deaths in Indonesia [1]. The Indonesian government has taken numerous measures to limit the spread of the coronavirus, including mask-wearing, hand washing, and social distancing. The use of masks is essential in preventing the coronavirus's transmission, protecting ourselves and others [2]. In offices, schools, or universities, employees who are about to start work always check their body temperature and wear masks correctly. A suitable mask is worn to protect oneself from other people's coughing and sneezing, as well as from the surrounding air, which may contain viruses in the form of aerosols that mix in the air for an extended period [3].

Air contaminated with the coronavirus is dangerous for people who move around it, especially for workers. It is a particular concern when working in offices, one of the activities when starting work is doing attendance. Some agencies or offices have used attendance using facial biometric data [4][5]. Biometric data is biological data that humans have since birth. Humans can use biometric data to provide information about an individual's identity based on physical characteristics that distinguish individuals from one another [6][7]. The use of facial data for identity recognition is popular because, visually, it can be easily recognized through the sense of sight. Facial characteristics often used for identity recognition are face shape, skin color, facial components such as nose, mouth, and eyes [8]. However, the coronavirus pandemic requires employees to carry out activities using masks that occlusion some facial features when identity recognition is carried out. This mask makes the identity recognition system using facial biometric data become less accurate. Almost half

of the faces are occluded by masks. The occlusion mask on the face recognition requires application in Artificial Intelligence [9][10][11], especially the use of deep learning for training masked face data.

Previous research related to masked face recognition using deep learning resulted in an accuracy value of 94.5% [12]. This result is not quite optimal for accuracy. Therefore, we propose another deep learning method to increase the accuracy value. Computer vision is a science field that extracts image patterns to obtain information or image descriptions [13]. One of the techniques often used in computer vision is using deep learning that processes image data by convolution. This masked facial data is not too much for one person's identity. A deep learning technique for processing data that is not too much is using transfer learning, which means that the model has been previously trained with other data [14][15]. One of the transfer learning techniques from the Keras library is MobileNetV2 [16] and VGG16 [17].

Researchers Hu and Ge [18] used the MobileNetV2 transfer learning model to recognize facial expressions. This research results in an accuracy of 97.98%. The result is pretty accurate in the case of facial expression recognition. Other studies have also used pre-trained MobileNetV2 to classify interest rates accurately above 96% [19]. Then, the research of Ruvunga et al. [17] used VGG16 to recognize human concentration levels based on their faces. The system created is good enough that it results in an accuracy of 93.47%. These studies provide an overview of transfer learning techniques that can be used to classify facial data with optimal accuracy and speed. Therefore, we will analyze the use of the two models, namely MobileNetV2 and VGG16, to recognize human faces wearing masks. We hope to find the best accuracy values for masked facial recognition systems.

2. RESEARCH METHOD

Fig. 1 shows the masked face recognition system architecture. The masked face recognition system begins with the training dataset stage. This study uses training and validation data as input in people's faces and the identities of people's names. The identity of the person's name will be as the label used for the testing phase on other image data with the same identity. Before the dataset is processed using transfer learning, the data is resized to adjust the size of the transfer learning model that has previously been trained using ImageNet data. A small amount of data input requires data augmentation to increase the amount of data. The results of augmentation data are trained using transfer learning VGG16 or MobileNetV2. The use of two transfer learning models because it has been proven in other cases previous studies have resulted in accuracy above 93% for VGG16 [17] and above 96% for MobileNetV2 [18][19]. The training result is an h5 model, which will be evaluated to find the best model based on loss and accuracy values.

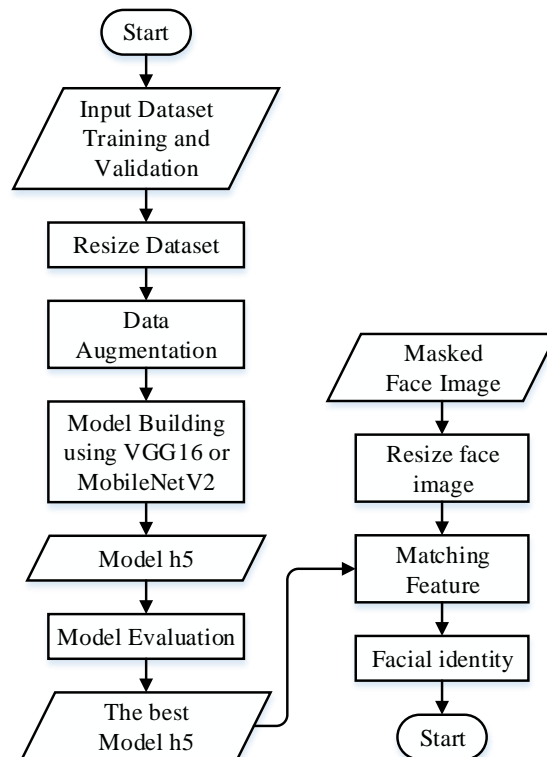


Fig. 1. The proposed system of masked face recognition

The testing phase in this study uses other face image data with the same person to evaluate system accuracy. The face images are resized so that they are the same size as the trained dataset. The next stage is a matching feature that generates facial identity. The results of facial identity recognition will be evaluated to test the performance of the system being built.

2.1. Data Acquisition

We used primary data in the form of the facial image representing 35 distinct identities in our experiment. Each identity has ten facial images, so that the total amount of data used in this study is 350. From each identity data, 80% of which are used for training and 20% for validation. Each image in this study is composed of three colors channels: red, green, and blue (RGB). Fig. 2 shows an example of the data and labels used in this study. Before performing transfer learning on the original face dataset, it is scaled to 224×224 pixels. The purpose of this resized is to adjust the size of the resolution of the VGG16 and MobileNetV2 pre-trained models.



Fig. 2. Samples of data and label

2.2. Data Augmentation

This study uses augmentation data to increase the number of datasets. This stage was carried out because the number of original training datasets only amounted to 8 images. Data augmentation in this study uses a library from Keras, namely ImageDataGenerator, often used as geometric augmentation [20]. In the training data using augmentation with rescaling parameters of $1./255$. Then, the rotation_range parameter is used to rotate the image by 45 degrees. The width_shift_range and height_shift_range parameters allow controlling the amount of horizontal and vertical shift by 0.3, respectively. The horizontal_flip parameter is used for flipping inputs horizontally. Next, the fill_mode parameter uses nearest. The fill_mode parameter is used to resample the image if the target size is different from that of the loaded image. Data validation only uses augmentation data with rescale parameters of $1./255$.

2.3. Deep Learning

The recognition of objects in computer vision aims to make decisions related to actual physical objects based on images obtained from sensors, such as cameras. Object recognition often uses machine learning technology. Machine learning is a technique to imitate how humans learn and generalize through the training process of data. The result is drawing conclusions based on training data [21]. As technology develops, machine learning has evolved into deep learning. Deep learning is an artificial neural network method that processes input data and processes it using many hidden layers. Then the result is a non-linear transformation of the input data to calculate the output value [22]. Deep learning that is often used to process image data is often known as a Convolutional Neural Network. Deep learning is usually used on large amounts of data, but the data is not too much like masked facial recognition in this research. One technique for processing less data is transfer learning [23], which means that the model has been previously trained with other data [14]. Examples of transfer learning models are VGG16 and MobileNetV2.

2.4. VGG16

K. Simonyan and A. Zisserman of the University of Oxford proposed the VGG16 convolutional neural network model [24]. The VGG16 architecture is a Convolutional Neural Network (CNN) utilized to win the 2014 ILSVR (ImageNet) competition. It is widely regarded as one of the most exemplary vision model architectures to date. The most distinguishing feature of VGG16 is that, rather than having a vast number of hyper-parameters, they concentrated on 3×3 convolution layers with stride one and always employed the same padding and max-pooling layers 2×2 filter with stride 2. It maintains this order of convolution and max-pooling layers throughout the architecture. Finally, it includes two FC (Fully Connected layers) and a softmax for

output. The 16 in VGG16 alludes to the fact that it has sixteen weighted layers. It is a vast network with approximately 138 million (estimated) parameters [25]. The VGG16 architecture is shown in Fig. 3.

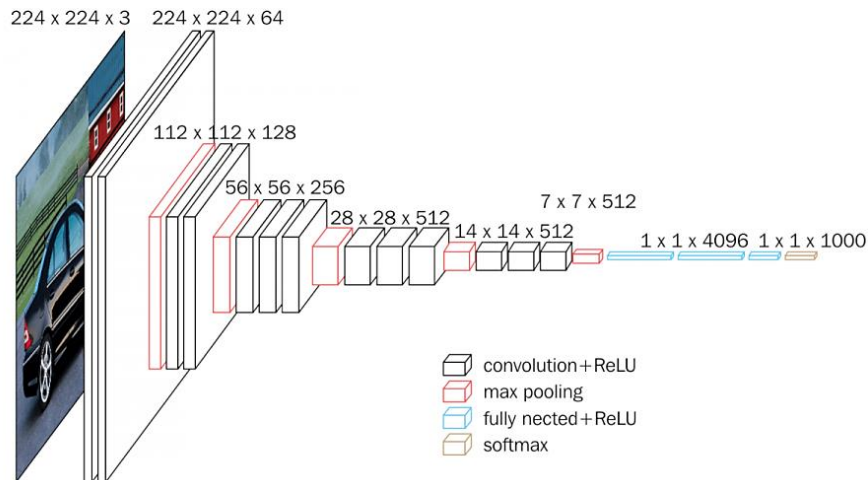


Fig. 3. VGG16 architecture [26]

2.5. MobileNetV2

In April 2017, MobileNet debuted its second version. MobileNetV2, like MobileNetV1, employs depthwise and pointwise convolution. MobileNetV2 adds two new CNN layers: an inverted residual layer and a linear bottleneck layer. MobileNetV2 introduces two new capabilities: 1) bottlenecks that are linear and 2) shortcut links between bottlenecks [27]. There is input and output between models in the bottleneck region, whereas the inner layer or layer encapsulates the model's ability to convert lower-level notions (i.e., pixels) to higher-level descriptors (i.e., image categories).

Finally, like with residual connections in classical CNNs, bypassing bottlenecks enables faster training and increased accuracy. As illustrated in Table 1, the MobileNetV2 network is mainly constructed using the inverted residual layer presented in the research [28]. Additionally, the model may modify object detection and semantic segmentation by utilizing the network as a feature extractor.

Table 1. MobileNetV2 architecture [28]

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2 1x1	-	k	-	-

2.6. System Evaluation

Research on masked face recognition is discussed in terms of the importance of training and validation accuracy. All training and validation results will be saved for each epoch and batch size combination. After the model is trained, we will evaluate the system's accuracy. Classification accuracy is a simple-to-understand concept that quantifies the proportion of instances in which a classifier correctly identifies the class of testing examples. Calculate the accuracy of a classifier using the formula shown in (1).

$$CR = \frac{C}{A} \quad (1)$$

where, CR is the correct rate, C is the number of image data testing recognized correctly and A is the number of all image data testing.

3. RESULTS AND DISCUSSION

This study uses a 3-channel color dataset, namely Red, Green, and Blue, with the size of each data being 224×224. This size adjusts the size of the pre-trained model from ImageNet used in the VGG16 and MobileNetV2 models. The pre-trained model will classify the identities of 35 people who wear masks. Before testing using other image data, we tuned the batch size and epoch parameters to see the accuracy values for training and validation. We experimented with batch size values 2, 4, 8, 16, and 32. Then the best results from the batch size values were used to experiment with variations in the number of epochs, namely 10, 20, 30, 40, and 50. In this study, we used Adam as optimizer and binary_crossentropy as a loss. Table 2 shows the experimental results on the variation in the value of the batch size. Then, Table 3 shows the experimental results on the variation in the value of the number of epochs in the training data process.

3.1. Training Result Analysis

In this study, several metrics related to batch size variations and the number of epochs were recorded, as shown in Table 2 and Table 3. The measurement metrics seen in the training process were the value of training accuracy and validation. We used multiple variations in the batch size variation experiment, namely 2, 4, 8, 16, and 32. A batch size of 8 indicates that the dataset will be divided into batches for Neural Network training, with each batch containing eight data.

Based on Table 2, the greater the number of batches, the better the accuracy value for training and validation. Even on MobileNetV2 in batch four, the training accuracy and validation values are 100%. Using the VGG16 model, resulting a maximum accuracy of 98% is obtained using batch size 32. In all variations of the experiment, the best batch size is 32 in both transfer learning models. Therefore, in the number of epoch experiments, we used batch size 32.

Table 2. Experiments in the training data stage using batch size variations

Pre-trained Model	2 batch		4 batch		8 batch		16 batch		32 batch	
	Train Acc.	Val Acc.	Train Acc.	Val Acc.	Train Acc.	Val Acc.	Train Acc.	Val Acc.	Train Acc.	Val Acc.
VGG16	0.9276	0.9571	0.9529	0.9706	0.9841	0.9844	0.9751	0.9844	0.9857	0.9844
MobileNetV2	0.9558	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 3. Experiments in the training data stage using the number of epochs variations

Pre-trained Model	10 Epoch		20 Epoch		30 Epoch		40 Epoch		50 Epoch	
	Train Acc.	Val Acc.	Train Acc.	Val Acc.	Train Acc.	Val Acc.	Train Acc.	Val Acc.	Train Acc.	Val Acc.
VGG16	0.1293	0.2188	0.5707	0.4688	0.8628	0.8906	0.9569	0.9219	0.9857	0.9844
MobileNetV2	0.9723	0.9688	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

An epoch is a point at which the complete dataset is trained on a Neural Network until it is reset to the beginning for one round. Based on Table 3, the value of training and validation accuracy using the MobileNetV2 model is 100% at epoch 20, in contrast to epoch on the VGG16 model, which has not reached 100% on epoch 50. These results show that the best model is obtained using the pre-trained MobileNetV2 model.

Fig. 4 shows that the use of the VGG16 model is slightly underfitting because, in certain epochs, the training accuracy is greater than the validation accuracy. For example, suppose that around the epoch of 20 to 30, it appears that the validation accuracy value is smaller than the training accuracy value. In the loss chart on training and validation, it is not so visible that overfitting or underfitting is because the difference in the value of training loss and validation is not too significant. Meanwhile, using the MobileNetV2 model, the training accuracy and validation graph do not show overfitting or underfitting, like loss training and validation results. Based on these results, to perform testing using other image data on the same identity, we use a batch size of 32 and a maximum number of epochs of 50. In this study, we used the early stop function to prevent continuous overfitting or underfitting. If the seven epochs do not decrease the loss validation value, the data training process will be stopped.

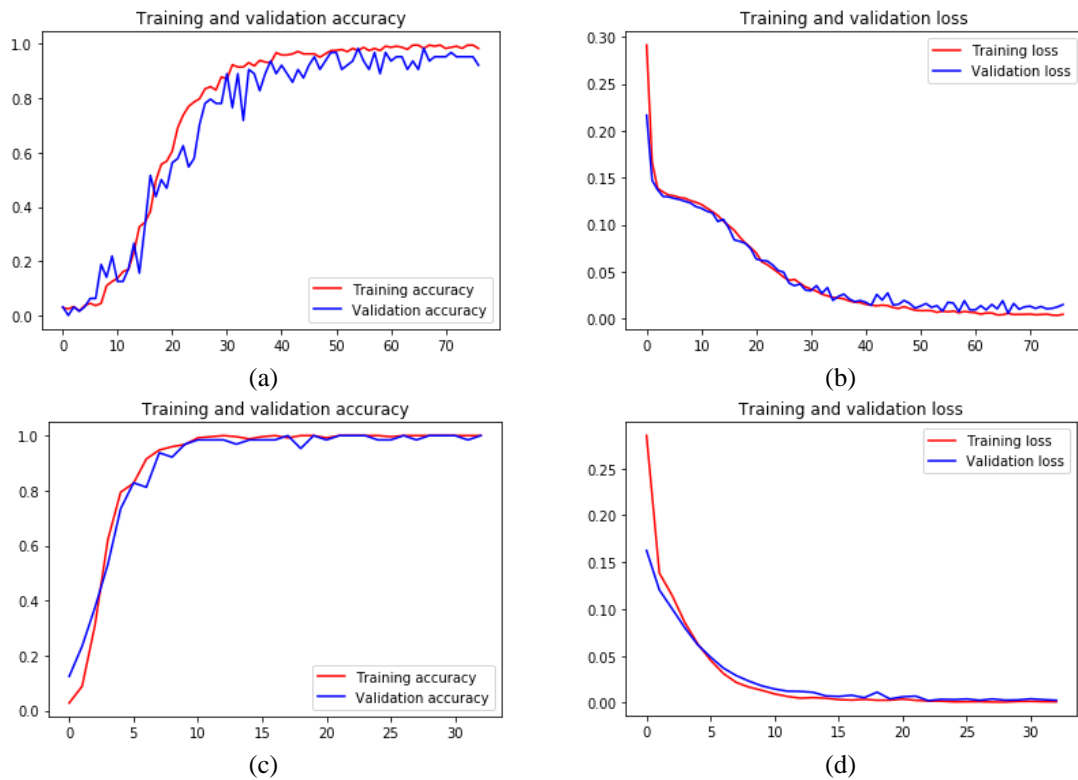


Fig. 4. The result of training data, (a), (b) VGG16, (c), (d) MobileNetV2

3.2. Testing Result Analysis

In this research, the proposed system needs to be tested to see the accuracy value generated using different data with the same identity. Table 4 shows the accuracy results for the proposed system. We used 175 testing data with 35 identities in the same training data. The VGG16 model contained 13 false identities, while using the MobileNetV2 model only had eight false identities. Then in Fig. 5 shows an example of the results of matching data with the model that has been made.

Table 4. The accuracy results matching data testing with VGG16 and MobileNetV2

Pre-trained Model	Accuracy
VGG16	92.57%
MobileNetV2	95.42%



Fig. 5. The result of matching data testing with model

Fig. 5 shows the results of matching the image with the model that has been previously trained. The matching results show the face image and its identity. However, as a result of this matching, not all face images match with the correct identity. This incorrect face identification because the CNN model cannot classify it correctly. Based on Table 4, the accuracy value in using MobileNetV2 is higher than using the VGG16 model. The difference in accuracy is because, at the time of training model building, the value of training accuracy

and validation was also enormous for MobileNetV2, which reached 100%. These results prove that transfer learning techniques can be used to recognize masked faces. Future research can continue to calculate processing time when using video data as testing data. The results of this study were better than previous studies [12], with an accuracy of 94.5%. In other words, the accuracy increased by 0.92% to 95.42%. Previous research also used a combination of two deep learning techniques, namely RetinaFace and VGGFace2. Meanwhile, the proposed research only uses one deep learning based on MobileNetV2 transfer learning.

4. CONCLUSION

This research focuses on the face recognition of a person wearing a mask. The coronavirus pandemic forces us to continue to carry out health protocols, namely wearing masks when on the move, especially working. We propose the use of transfer learning techniques for facial feature extraction and classification according to identity. The results showed that the MobileNetV2 transfer learning model was better than VGG16, with an accuracy of 95.42%. These results indicate that using the MobileNetV2 model is better in classifying 35 different people's identities. This study also uses only one deep learning model, which is undoubtedly more efficient than previous studies using two deep learning models. The accuracy value also increased by 0.92% compared to previous studies. Future research can be developed on real-time video data so that the data processing speed can be known.

Acknowledgments

Thanks to LPPM Institut Teknologi Telkom Purwokerto through the Internal Research Grant for funding in the publication process of this journal.

REFERENCES

- [1] WHO, "World Health Organization. Coronavirus disease 2019 (COVID-19)," 2021.
- [2] N. P. E. D. Yanti, I. M. A. D. P. Nugraha, G. A. Wisnawa, N. P. D. Agustina, and N. P. A. Diantari, "Public Knowledge about Covid-19 and Public Behavior During the Covid-19 Pandemic," *J. Keperawatan Jiwa*, vol. 8, no. 4, p. 491, 2020. <https://doi.org/10.26714/jkj.8.4.2020.491-504>
- [3] S. E. Hwang, J. H. Chang, O. Bumjo, and J. Heo, "Possible Aerosol Transmission of COVID-19 Associated with an Outbreak in an Apartment in Seoul, South Korea, 2020," *Int. J. Infect. Dis.*, no. xxxx, pp. 0–3, 2020. <https://doi.org/10.1016/j.ijid.2020.12.035>
- [4] A. Patil, P. K. P. P. More, A. Joshi, and A. R. Kamble, "Attendance Monitoring using Face Recognition and Machine Learning," *Int. J. Futur. Gener. Commun. Netw.*, vol. 13, no. 3, pp. 94–102, 2020. <https://doi.org/10.2139/ssrn.3529319>
- [5] M. Surve, P. Joshi, S. Jamadar, and M. Vharkate, "Automatic Attendance System Using Face Recognition Technique," *Int. J. Recent Technol. Eng.*, vol. 9, no. 1, pp. 2134–2138, 2020. <https://doi.org/10.35940/ijrte.A2644.059120>
- [6] J. Mason, R. Dave, P. Chatterjee, I. Graham-Allen, A. Esterline, and K. Roy, "An Investigation of Biometric Authentication in the Healthcare Environment," *Array*, vol. 8, no. August, p. 100042, 2020. <https://doi.org/10.1016/j.array.2020.100042>
- [7] A. Casanova, L. Cascone, A. Castiglione, W. Meng, and C. Pero, "User recognition based on periocular biometrics and touch dynamics," *Pattern Recognit. Lett.*, vol. 148, pp. 114–120, 2021. <https://doi.org/10.1016/j.patrec.2021.05.006>
- [8] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, present, and future of face recognition: A review," *Electron.*, vol. 9, no. 8, pp. 1–53, 2020. <https://doi.org/10.3390/electronics9081188>
- [9] A. Elmahmudi and H. Ugail, "Deep face recognition using imperfect facial data," *Futur. Gener. Comput. Syst.*, vol. 99, pp. 213–225, 2019. <https://doi.org/10.1016/j.future.2019.04.025>
- [10] L. Yang, J. Ma, J. Lian, Y. Zhang, and H. Liu, "Deep representation for partially occluded face verification," *Eurasip J. Image Video Process.*, vol. 2018, no. 1, 2018. <https://doi.org/10.1186/s13640-018-0379-2>
- [11] F. K. Zaman, A. A. Shafie, and Y. M. Mustafah, "Robust face recognition against expressions and partial occlusions," *Int. J. Autom. Comput.*, vol. 13, no. 4, pp. 319–337, 2016. <https://doi.org/10.1007/s11633-016-0974-6>
- [12] V. Aswal, O. Tupe, S. Shaikh, and N. N. Charniya, "Single Camera Masked Face Identification," *Proc. - 19th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2020*, pp. 57–60, 2020. <https://doi.org/10.1109/ICMLA51294.2020.00018>
- [13] V. Wiley and T. Lucas, "Computer Vision and Image Processing: A Paper Review," *Int. J. Artif. Intell. Res.*, vol. 2, no. 1, p. 22, 2018. <https://doi.org/10.29099/ijair.v2i1.42>
- [14] Y. Gultom, A. M. Arymurthy, and R. J. Masikome, "Batik Classification using Deep Convolutional Network Transfer Learning," *J. Ilmu Komput. dan Inf.*, vol. 11, no. 2, p. 59, 2018. <https://doi.org/10.21609/jiki.v11i2.507>
- [15] A. Brodzicki, M. Piekarski, D. Kucharski, J. Jaworek-Korjakowska, and M. Gorgon, "Transfer Learning Methods as a New Approach in Computer Vision Tasks with Small Datasets," *Found. Comput. Decis. Sci.*, vol. 45, no. 3, pp. 179–193, 2020. <https://doi.org/10.2478/fcds-2020-0010>
- [16] K. Y. Lum, Y. H. Goh, and Y. Bin Lee, "American sign language recognition based on MobileNetV2," *Adv. Sci. Technol. Eng. Syst.*, vol. 5, no. 6, pp. 481–488, 2020. <https://doi.org/10.25046/aj050657>
- [17] C. Ruvinga, D. Malathi, and J. D. Dorathi Jayaseeli, "Human concentration level recognition based on vgg16 cnn architecture," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 6 Special Issue, pp. 1364–1373, 2020.

- <http://sercs.org/journals/index.php/IJAST/article/view/9271>
- [18] L. Hu and Q. Ge, "Automatic facial expression recognition based on MobileNetV2 in Real-time," *J. Phys. Conf. Ser.*, vol. 1549, no. 2, 2020. <https://doi.org/10.1088/1742-6596/1549/2/022136>
- [19] W. Dai, Y. Dai, K. Hirota, and Z. Jia, "A Flower Classification Approach with MobileNetV2 and Transfer Learning," *9th Int. Symp. Comput. Intell. Ind. Appl.*, pp. 1–5, 2020. <https://isciiia2020.bit.edu.cn/docs/20201114083020836285.pdf>
- [20] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019. <https://doi.org/10.1186/s40537-019-0197-0>
- [21] A. Ahmad, "Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning," *J. Teknol. Indones.*, no. October, 2017. https://www.academia.edu/download/54674088/Perbedaan_Deep_learn.pdf
- [22] R. Primartha, "Belajar Machine Learning; Teori dan Praktik," 2018. http://repo.unikadelasalle.ac.id/index.php?p=show_detail&id=12893&keywords=
- [23] F. D. Adhinata, D. P. Rakhmadani, M. Wibowo, and A. Jayadi, "A Deep Learning Using DenseNet201 to Detect Masked or Non-masked Face," vol. 9, no. 1, pp. 115–121, 2021. <https://doi.org/10.30595/juita.v9i1.9624>
- [24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015. <https://arxiv.org/abs/1409.1556v6>
- [25] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection," *Constr. Build. Mater.*, vol. 157, no. September, pp. 322–330, 2017. <https://doi.org/10.1016/j.conbuildmat.2017.09.110>
- [26] W. Nash, T. Drummond, and N. Birbilis, "A review of deep learning in the study of materials degradation," *npj Mater. Degrad.*, vol. 2, no. 1, pp. 1–12, 2018. <https://doi.org/10.1038/s41529-018-0058-x>
- [27] D. Yu, Q. Xu, H. Guo, C. Zhao, Y. Lin, and D. Li, "An efficient and lightweight convolutional neural network for remote sensing image scene classification," *Sensors (Switzerland)*, vol. 20, no. 7, 2020. <https://doi.org/10.3390/s20071999>
- [28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018. <https://doi.org/10.1109/CVPR.2018.00474>

BIOGRAPHY OF AUTHORS



Faisal Dharma Adhinata obtained a Master of Computer Science (M.Cs) degree in Computer Science from Universitas Gadjah Mada, Indonesia, in 2020. Currently, he is a lecturer at the Department of Software Engineering, Institut Teknologi Telkom Purwokerto. His research interests are Artificial Intelligence, machine learning techniques, image processing, and computer vision. Email: faisal@ittelkom-pwt.ac.id



Nia Annisa Ferani Tanjung was accepted in the Department of Physics at the State University of Medan in 2010 and completed her master's degree in 2018 in Physics, Universitas Gadjah Mada. Her research field is subsurface modeling in geophysics and has joined the Physical Society of Indonesia since 2018 until now. Starting in 2018, she has been a Lecturer at the Institut Teknologi Telkom Purwokerto until now. Email: nia@ittelkom-pwt.ac.id



Widi Widayat received his Bachelor of Computer Science (S.Kom) from the Universitas Islam Indonesia in 2014. After working in the software development field for about 2 years, then take his master degree at UGM. In 2019 completed his Master of Engineering with research interest in Machine learning, NLP and Sentiment Analysis. Starting in 2020, he has been a Lecturer at the Institut Teknologi Telkom Purwokerto until now. Email: widi@ittelkom-pwt.ac.id



Gracia Rizka Pasfica is an active student majoring in Software Engineering at the Institut Teknologi Telkom Purwokerto since 2019 until now. She has a strong interest in research and organization. Her research interests include Artificial Intelligence, particularly Computer vision and Machine Learning. Email: 19104064@ittelkom-pwt.ac.id



Fadlan Raka Satura Graduated in 2019 at SMK SMTI Yogyakarta with a major in Industrial Chemistry, then continued studying at Institut Teknologi Telkom Purwokerto with a major in software engineering until now. Email: 19104072@ittelkom-pwt.ac.id