Scientific Journal of Informatics
Vol. 11, No. 1, Feb 2024
p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

YOLO vs. CNN Algorithms: A Comparative Study in Masked Face Recognition

Muhammad Ridho Dewanto[1*], Mifta Nur Farid[2], Muhammad Abby Rafdi Syah[3], Aji Akbar Firdaus[4], Hamzah Arof [5]

[1,2,3]Electrical Engineering, Department of Electrical Engineering, Institut Teknologi Kalimantan, Balikpapan, Indonesia
[4]Department of Engineering, Universitas Airlangga, Surabaya, Indonesia
[5]Department of Electrical Engineering, Usniversity of Malaya, Kuala Lumpur, Malaysia

Abstract.
Purpose: This research investigates the effectiveness of YOLO (You Only Look Once) and Convolutional Neural Network (CNN) in real-time face mask recognition, addressing the challenges posed by mask-wearing in infectious disease prevention.
Method: Utilizing a diverse dataset and employing YOLO's object detection and a combined Haar Cascade Algorithm with CNN, the study evaluated key performance indicators, including accuracy, framerate, and F1 Score.
Results: Results indicated that CNN outperformed YOLO in accuracy (99.3% vs. 79.3%) but operated at a slightly lower framerate. YOLO excelled in recall and precision, presenting a compelling choice for specific application needs. The research underscores the importance of considering factors beyond accuracy for informed decision-making in the realm of face mask recognition.
Novelty: This research evaluates the real-time performance of YOLO and CNN algorithms in masked face recognition, highlighting the crucial balance between framerate efficiency and detection accuracy.

Keywords: Convolutional neural network (CNN); Face mask recognition; Infectious disease prevention; Real-time object detection; YOLO (You only look once)
Received November 2023 / **Revised** February 2024 / **Accepted** February 2024

INTRODUCTION

The field of Artificial Intelligence (AI) is rapidly growing within computer science, with computer vision emerging as an extensively researched subfield [1]–[3]. This subfield has a specific case study, face recognition, which has attracted significant attention from researchers due to its potential applications [4], [5]. The ability of face recognition to identify individuals based on their facial features provides numerous advantages across various domains [6]. For instance, organizations such as schools or companies can enhance efficiency in attendance systems by incorporating face recognition technology [7]. In addition, face verification, a specialized application within face recognition, plays a vital role in ensuring secure access to functions like smartphone locks and payment systems [8].

Several studies have demonstrated the impressive performance of face recognition systems. A study proposed a method for constructing a face recognition system, attaining an accuracy of 95.97% on the AR Face dataset with 120 individuals and 97.20% on the VTU-BEC-DB multimodal database [9]. Similarly, another study developed a face recognition system for school attendance, achieving an accuracy of 97.29% in their proposed attendance system [10]. Notably, surveys employing various methodologies conducted by several researchers in recent years have consistently indicated satisfactory performance of observed face recognition systems [11]–[13]. Although face recognition seems to be a well-addressed topic in AI development, this does not diminish the need for ongoing research by scholars exploring different cases and conditions in face recognition systems.

One of the very new cases is regarding the use of face masks that might prevent face recognition from functioning. This is particularly true during the outbreak, in which a highly concerning infectious disease

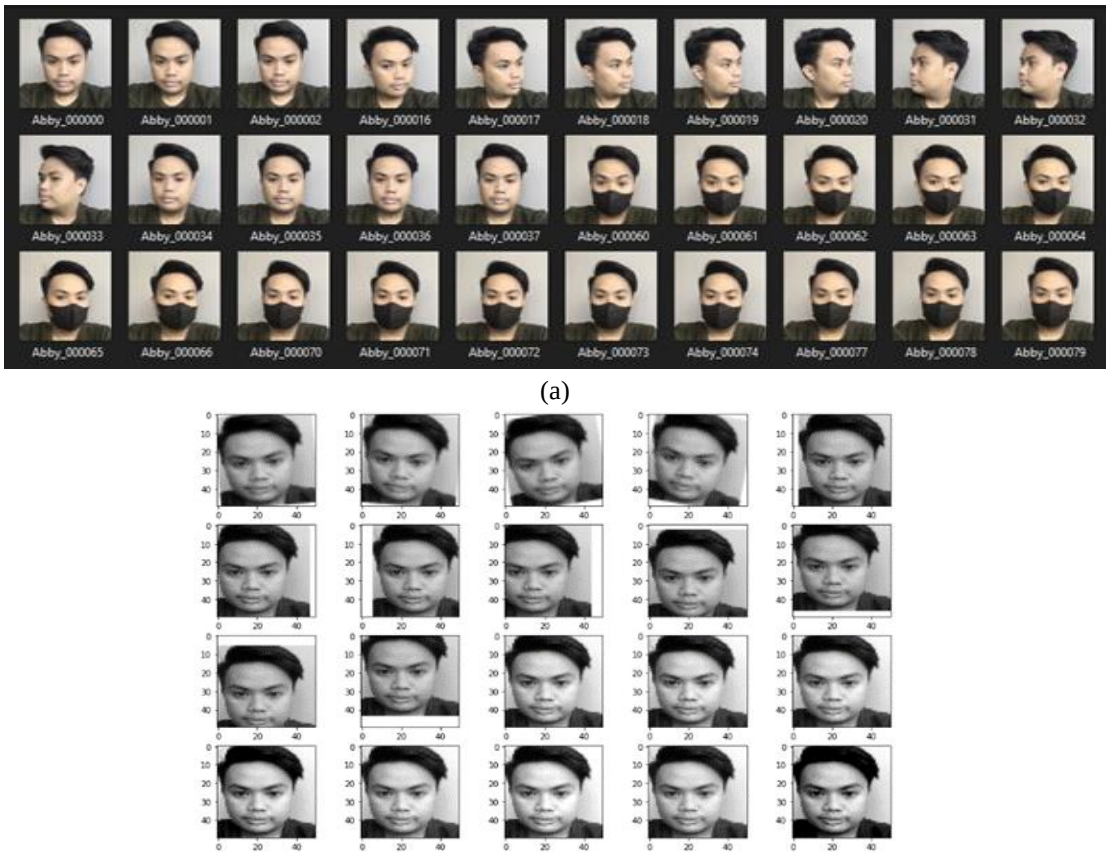
*Corresponding author.
Email addresses: ridho.dewanto@lecturer.itk.ac.id (Dewanto)
DOI: 10.15294/sji.v11i1.48723

that spreads through touches and droplets requires people to wear masks, both indoors and outdoors. This precautionary measure obstructs facial visibility, making it difficult for individuals to recognize each other. For this reason, a study developed masked face recognition systems, achieving higher accuracy (97% and improved true positive rates, respectively) by including masked face images in their training data [14]. Another study simulated face dataset augmentation with nonphysical masks [15]. Other studies have found that YOLO-V3 and YOLO V3-tiny achieved higher accuracy in detecting face masks compared to CNN. A study reported an average accuracy of 91.28 for YOLO-V3, an average precision value of 86.65 for CNN, and an accuracy of 95% for YOLO V3-tiny and 84% for CNN [16]. Meanwhile, another study proposed a novel dataset and methods for real-time detection of masked and unmasked faces, achieving an accuracy of 99.5% using YOLO and a CNN architecture [17].

However, until now, there has been no comprehensive comparison between the YOLO algorithm and the Haar-Cascade CNN algorithm. Thus, this research focuses on face mask recognition using deep learning models, namely YOLO and CNN. Both methods are employed for face mask recognition due to Yolo's known speed, as its frame detection architecture utilizes a regression model and does not require a complex pipeline. On the other hand, the use of dimensions greater than 1 in CNN will affect the overall scale of an object. In this research, deep learning experimentation will be conducted for face recognition, in which two deep learning models will be used for comparison to determine which of these algorithms can, at the very least, be identified as superior when applied to a camera for mask-wearing identification.

METHODS

The dataset was carried out by three individuals, each assigned the task of providing 15 images with masks and 15 images without masks, resulting in a total of 90 images. Each photo had a resolution of 286x286 pixels and was captured using a Logitech C920 camera. Various perspectives were captured in each photo, creating a diverse dataset for testing with the intended implementation of detection methods. Figure 1 serves as an exemplar of the dataset employed for the implementation database.



(a)

(b)

Figure 1. (a) Database photos, (b) Data sensing photo

In this study, the decision was made to compare two different detection methods: utilizing YOLO (You Only Look Once) [18], [19] and a combined method of the Haar Cascade Algorithm with Convolutional Neural Network (CNN) [20]. YOLO is a real-time object detection algorithm that divides the image into a grid and predicts bounding boxes and object classes in a single processing step [21]. Meanwhile, the Haar Cascade Algorithm is a classical method that detects objects based on visual features, and CNN is a type of artificial neural network architecture effective in understanding spatial data structures [22]. The CNN layers used in data processing include normalization to adjust the pixel value range [23], filtering layer to extract crucial features, ReLU (Rectified Linear Unit) to introduce non-linearity, Max pooling to reduce spatial dimensions, flattening to transform data into a vector, fully connected layer for feature combination, and softmax as the activation function for final classification [24].

In this research, the Keras library was used to create a face detector, and detect_face can be declared as a cropping function. Following this, the augmentation stage was reached to enhance data diversity. Image augmentation techniques were used to create new variations of existing faces [25], including rotation, flipping, or cropping to generate similar but different images [26]. This contributes to an increased diversity of available data for training the face recognition model, allowing the model to better recognize various facial characteristics. As a result, a total of 1890 augmentations and originals with three labels each were generated. Expanding the dataset of facial images with more variations and labels is aimed at improving the training of a more effective face recognition model. In this research, 20% of the data were used as test data, while 80% served as training data. Each variable checked the number of samples and dimensions of the training and testing data to understand the data structure used in model training and testing.

In terms of research indicators, the study carefully selected three key variables: accuracy, framerate, and F1 Score. Accuracy is a holistic measure of the model's correctness, providing an overall assessment of its performance. Framerate, measured in frames per second, offers insights into the real-time applicability of the detection methods. The inclusion of the F1 Score, a metric combining precision and recall, ensures a balanced evaluation that considers both false positives and false negatives [27].

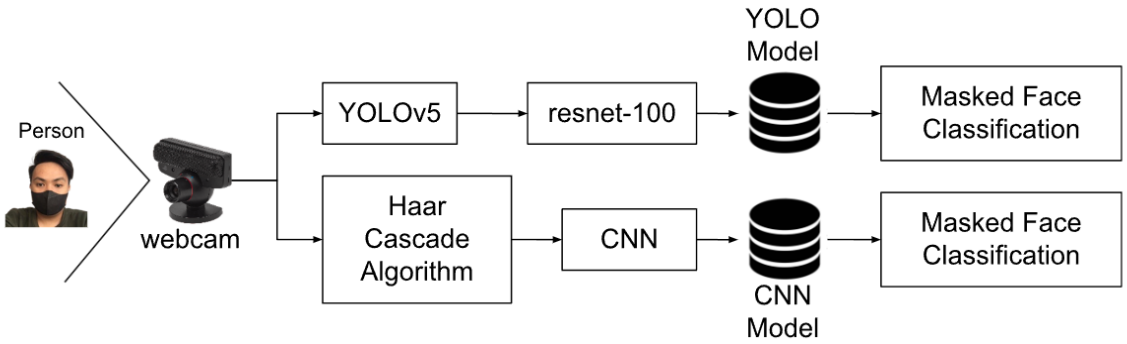


Figure 2. The workflow of the system execution

Figure 2 depicts the flow diagram of the system utilized in this research. The figure illustrates that an individual's webcam feed was analyzed through the YOLO and CNN algorithmic systems [28], [29]. Different layers were employed based on the system's operational principles [30]–[32].

RESULTS AND DISCUSSIONS

This research began the analysis by collecting data images from three individuals during the research experiment. Subsequently, the system was tested using two algorithmic comparisons, namely YOLO and CNN. Three instances of person were employed to test both YOLO and CNN in each respective trial.

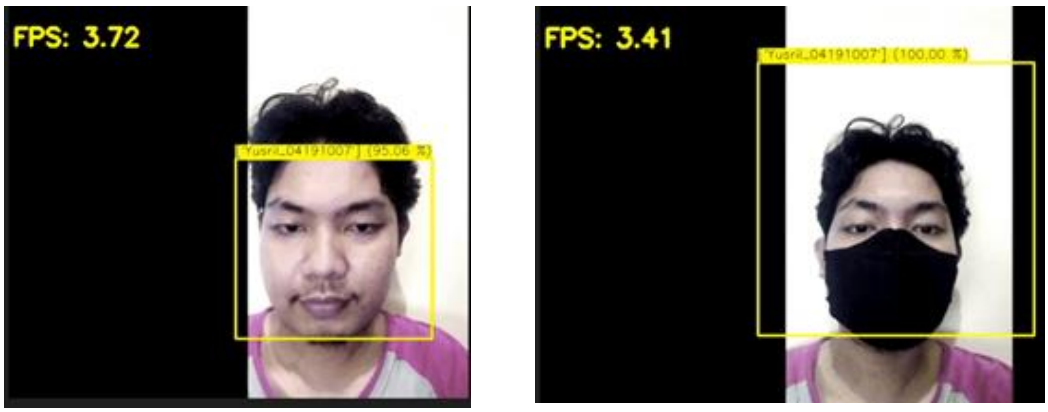


Figure 3. Realtime test CNN object person 1

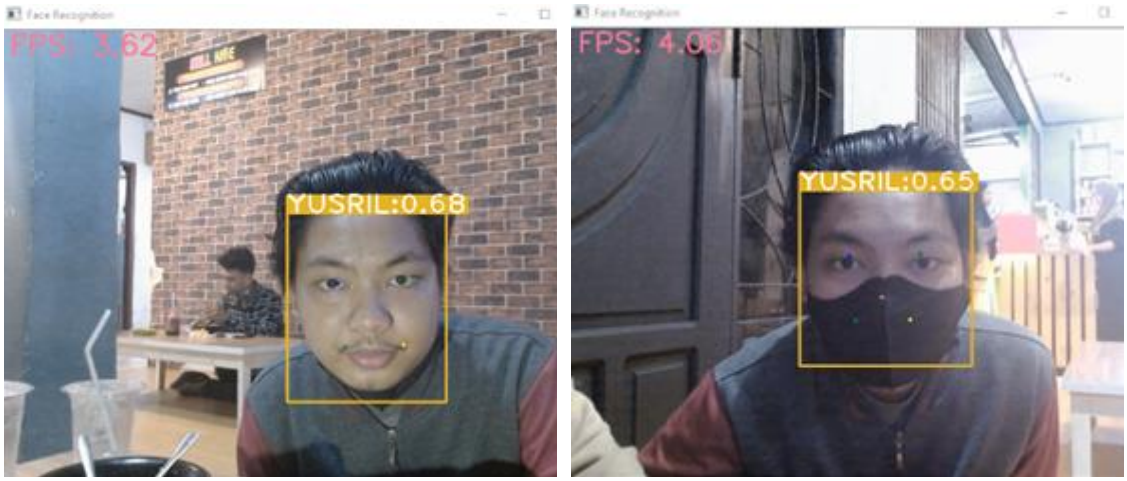
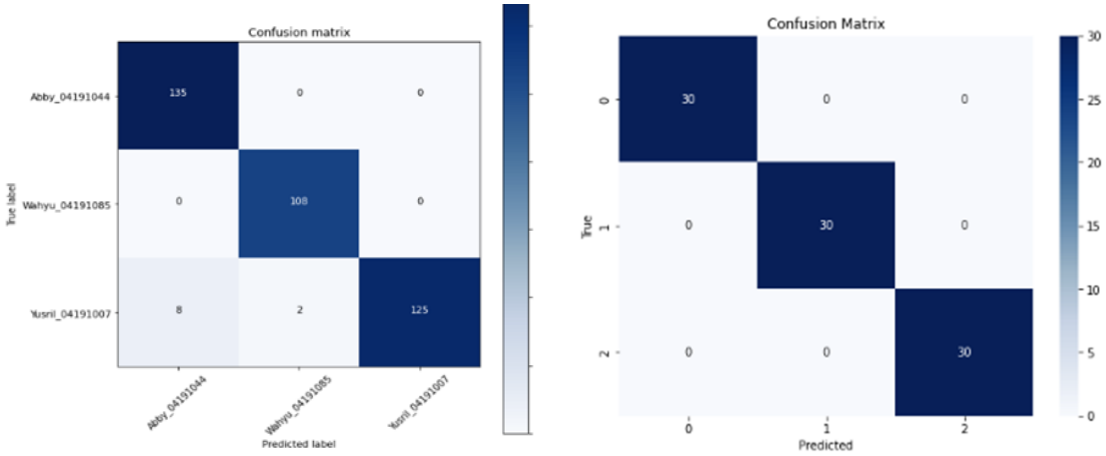


Figure 4. Realtime test YOLO object person 1

Table 1 provides a comprehensive comparison of statistical analyses for mean rainfall observations using two different methods: CNN and YOLO. In terms of frame rate, CNN operated at 3.57 FPS with a slight advantage over YOLO at 3.67 FPS. Figures 3 and 4 illustrate the success of both systems in facial detection and their capability to distinguish individuals effectively. However, when considering accuracy, CNN outperformed YOLO with a rate of 99.3% compared to 79.3%. The precision of both methods was high, with CNN at 97.3% and YOLO at 81.3%. Interestingly, both methods achieved perfect recall, indicating their ability to capture all relevant instances. The F1-Score, which balances precision and recall, favored CNN with a score of 0.97, while YOLO scored a perfect 1.00. These findings suggest that CNN excels in accuracy and F1-Score, while YOLO demonstrates competitive performance, especially in terms of recall and precision. The choice between the two methods may depend on specific priorities, such as real-time processing (where CNN has a slight edge in frame rate) or a balance between precision and recall (where YOLO excels).



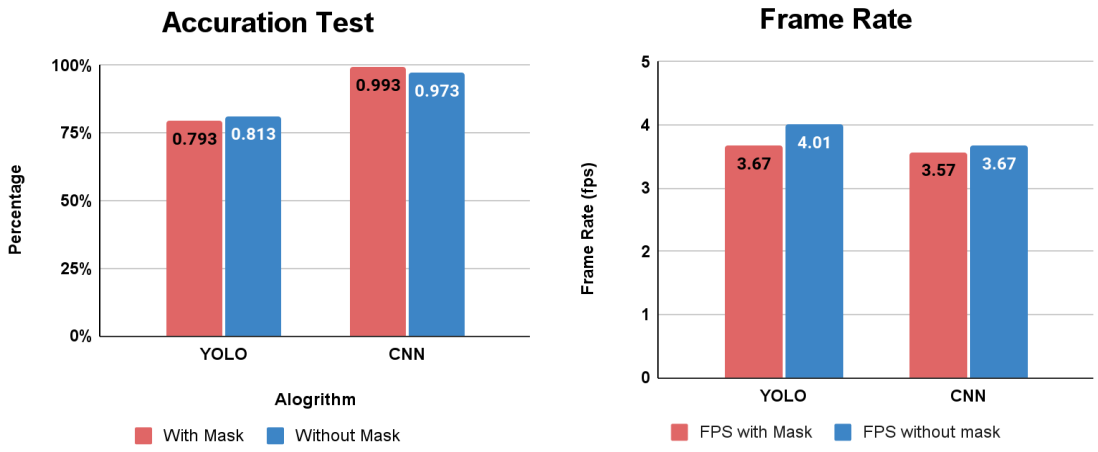
(a) (b)

Figure 5 (a) Classification of CNN matrix confusion, (b) Classification of YOLO matrix confusion

The comparison in Figure 5 indicates that the YOLO algorithm can achieve higher detection accuracy than CNN images that exhibit detection errors. When considering these results, it becomes evident that both CNN and YOLO possess their own strengths and excel in specific areas. CNN's superior accuracy and high F1-Score, for example, make it a robust option for applications where precision and overall model performance are paramount. On the contrary, YOLO's impeccable precision and recall imply an impressive capability to accurately detect and classify instances. The choice between these methods may be subjective and contingent upon the specific requirements of the application. If prioritizing real-time processing is crucial, the slight advantage in frame rate for CNN might be considered a decisive factor. However, if the emphasis lies on a well-balanced model with high precision and recall, YOLO could be presented as a compelling choice. Figure 6 shows the test accuracy of YOLO and CNN, as well as their frame rate rests. The classification results can be seen in Figure 7.

Table 1. Comparison of statistical analysis of the mean rainfall observations

Method	Frame Rate (FPS)		Accuracy (%)				
	With Mask	Without Mask	With Mask	Without Mask	Precision	Recall	F1-Score
CNN	3.57 Fps	3.67 Fps	99.3 %	97.3 %	0.94	1.00	0.97
YOLO	3.67 Fps	4.00 Fps	79.3 %	81.3 %	1.00	1.00	1.00



(a) (b)

Figure 6. (a)Accuration test, (b) Frame rate test

The table, overall, provides a valuable foundation for decision-making, but further considerations, such as computational efficiency and specific application requirements, should be taken into account to make an informed choice between CNN and YOLO for mean rainfall observations. The generated data can be compared to the previous study, which researched CNN and YOLO solely on full-face images without masks. Additionally, this study involved a real-time application where detection was performed directly, providing accuracy comparisons, wherein the CNN algorithm demonstrated a detection accuracy of 99.3%, while YOLO showed 79.3% accuracy in the case of mask usage. Furthermore, CNN achieved a higher Frames Per Second (FPS) at 3.57 FPS compared to YOLO, which recorded 3.67 FPS.

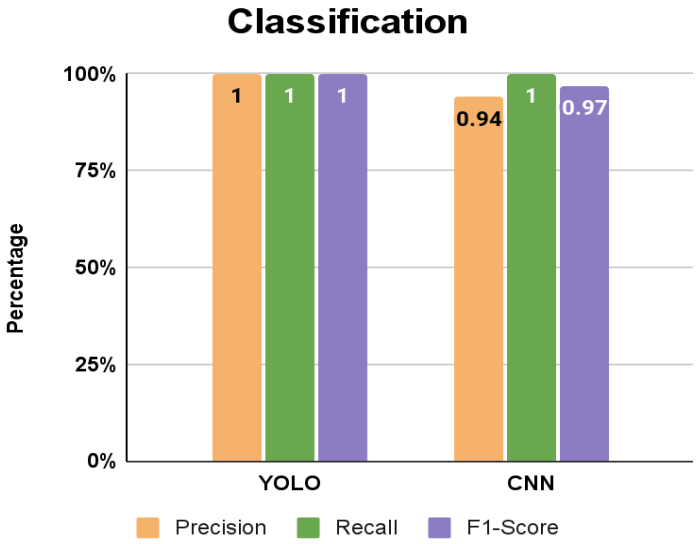


Figure 7. Classification results.

CONCLUSION

In conclusion, the CNN method demonstrated high accuracy in recognizing both masked and unmasked faces, with an accuracy rate of 99.3% and 97.3%, respectively. However, it operated at a slightly slower processing speed, achieving an FPS of 3.27, and exhibited some prediction errors as indicated by imperfections in precision, recall, and F1-score. Despite these limitations, the CNN method remains viable for effective face mask recognition applications. On the other hand, the YOLO algorithm offered a comparable average processing speed of 3.8 FPS, but its accuracy was slightly lower, at 79.3% for masked faces and 81.3% for unmasked faces. Therefore, the choice between these methods depends on the specific priorities regarding accuracy and processing speed in the context of face mask recognition applications. Both methods have their strengths and limitations, and selecting the most suitable one should be based on the specific requirements of the application at hand.

REFERENCES

[1] J. Harika, P. Baleeshwar, K. Navya, and H. Shanmugasundaram, "A Review on Artificial Intelligence with Deep Human Reasoning," *Int. Conf. Appl. Artif. Intell. Comput.*, 2022, doi: 10.1109/ICAAIC53929.2022.9793310.

[2] R. Rofik, R. Aulia, K. Musaadah, S. S. F. Ardyani, and A. A. Hakim, "Optimization of Credit Scoring Model Using Stacking Ensemble Learning and Oversampling Techniques," *J. Inf. Syst. Explor. Res.*, vol. 2, no. 1, pp. 11–20, 2023, doi: 10.52465/joiser.v2i1.203.

[3] H. Hadiq, S. Solehatin, D. Djuniharto, M. A. Muslim, and S. N. Salahudin, "Comparison of the suitability of the otsu method thresholding and multilevel thresholding for flower image segmentation," *J. Soft Comput. Explor.*, vol. 4, no. 4, pp. 242–249, 2023, doi: 10.52465/joscex.v4i4.266.

[4] G. G. Dordinejad and H. Çevikalp, "Face Frontalization for Image Set Based Face Recognition," *Signal Process. Commun. Appl. Conf.*, 2022, doi: 10.1109/SIU55565.2022.9864911.

[5] A. Adimas and S. Y. Irianto, "Image Sketch Based Criminal Face Recognition Using Content Based Image Retrieval," *Sci. J. Informatics*, vol. 8, no. 2, pp. 176–182, 2021, doi:

10.15294/sji.v8i2.27865.
[6] Y. Lin and H. Xie, “Face Gender Recognition based on Face Recognition Feature Vectors,” *IEEE 3rd Int. Conf. Inf. Syst. Comput. Aided Educ.*, 2020, doi: 10.1109/ICISCAE51034.2020.9236905.
[7] A. K. Sirivarshitha, K. Sravani, K. S. Priya, and V. Bhavani, “An approach for Face Detection and Face Recognition using OpenCV and Face Recognition Libraries in Python,” *9th Int. Conf. Adv. Comput. Commun. Syst.*, 2023, doi: 10.1109/ICACCS57279.2023.10113066.
[8] F. Gu, J. Lu, G. Xia, and Z. Feng, “Face Verification Technology Based on FaceNet Similarity Recognition Network,” *IEEE 10th Data Driven Control Learn. Syst. Conf.*, 2021, doi: 10.1109/DDCLS52934.2021.9455715.
[9] S. A. A. and S. M. Hatture, “Face Recognition Through Symbolic Modeling of Face Graphs and Texture,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 33, no. 12, 2021, doi: https://doi.org/10.1142/S0218001419560081.
[10] B. Contoh *et al.*, “No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 1, no. 1, p. 2019, 2019, [Online]. Available: http://www.ghbook.ir/index.php?name=فرهنگ و رسانه های نوین&option=com_dbook&task=readonline&book_id=13650&page=73&chkhashk=ED9C9491B4&Itemid=218&lang=fa&tmpl=component%0Ahttp://www.albayan.ae%0Ahttps://scholar.google.co.id/scholar?hl=en&q=APLIKASI+PENGENA
[11] S. P. Samadhi and E. Izquierdo, *Deep-learned faces: a survey*, vol. 2020, no. 1. EURASIP Journal on Image and Video Processing, 2020. doi: 10.1186/s13640-020-00510-w.
[12] M. Wang and W. Deng, “Deep face recognition: A survey,” *Neurocomputing*, vol. 429, pp. 215–244, 2021, doi: 10.1016/j.neucom.2020.10.081.
[13] M. Norouzi, “A Survey on Face Recognition Based on Deep Neural Networks,” pp. 1–15, 2022, [Online]. Available: https://doi.org/10.21203/rs.3.rs-1367031/v1
[14] I. Q. Mundial, M. S. U. Hassan, M. I. Tiwana, W. S. Qureshi, and E. Alanazi, “Towards Facial Recognition Problem in COVID-19 Pandemic,” *4rd Int. Conf. Electr. Telecommun. Comput. Eng.*, 2020, doi: 10.1109/ELTICOM50775.2020.9230504.
[15] Y. Li, K. Guo, Y. Lu, and L. Liu, “Cropping and attention based approach for masked face recognition,” *Appl. Intell.*, vol. 51, no. 5, pp. 3012–3025, 2021, doi: 10.1007/s10489-020-02100-9.
[16] A. A. Shaik, R. T. Prabu, and S. Radhika, “Detection of Face Mask using Convolutional Neural Network (CNN) based Real-Time Object Detection Algorithm You Only Look Once-V3 (YOLO-V3) Compared with Single-Stage Detector (SSD) Algorithm to Improve Precision,” *Int. Conf. Adv. Comput. Commun. Appl. Informatics*, 2023, doi: 10.1109/ACCAI58221.2023.10200890.
[17] S. Abbasi, H. Abdi, and A. Ahmadi, “A Face-Mask Detection Approach based on YOLO Applied for a New Collected Dataset,” *26th Int. Comput. Conf. Comput. Soc. Iran*, 2021, doi: 10.1109/CSICC52343.2021.9420599.
[18] Y. Ma, J. Yang, Z. Li, and Z. Ma, “YOLO-Cigarette: An effective YOLO Network for outdoor smoking Real-time Object Detection,” *Ninth Int. Conf. Adv. Cloud Big Data*, 2021, doi: 10.1109/CBD54617.2021.00029.
[19] M. Zou *et al.*, “Feature Compression Applications of Genetic Algorithm,” *Front. Genet.*, vol. 13, no. March, pp. 1–13, 2022, doi: 10.3389/fgene.2022.757524.
[20] J. Shah and A. K. Pandey, “Estimation of Face Attributes Using Standard CNN Features,” *3rd Int. Conf. Adv. Comput. Innov. Technol. Eng.*, 2023, doi: 10.1109/ICACITE57410.2023.10183194.
[21] W. Yang, D. BO, and L. S. Tong, “TS-YOLO:An efficient YOLO Network for Multi-scale Object Detection,” *IEEE 6th Inf. Technol. Mechatronics Eng. Conf.*, 2022, doi: 10.1109/ITOEC53115.2022.9734458.
[22] H. Wang and J. Han, “Research on military target detection method based on YOLO method,” *IEEE 3rd Int. Conf. Inf. Technol. Big Data Artif. Intell.*, 2023, doi: 10.1109/ICIBA56860.2023.10165623.
[23] S. F. Kak, F. M. Mustafa, and A. Varol, “Design and Enhancement of a CNN Model to Augment the Face Recognition Accuracy,” *3rd Int. Informatics Softw. Eng. Conf.*, 2022, doi: 10.1109/IISEC56263.2022.9998236.
[24] O. P. Yakubu, A. Y. A. . M. Ismail, M. L. Abdulrahman, I. Z. Yahkubu, and L. Step, “A Deep Learning Approach for Detecting Face Mask Using an Improved Yolo-V2 With Squeezenet,” *IEEE 6th Conf. Inf. Commun. Technol.*, 2022, doi: 10.1109/CICT56698.2022.9997956.
[25] C. Liu and J. Liu, “Application Analysis of Face Recognition Technology Based on Computer Vision,” *3rd Int. Acad. Exch. Conf. Sci. Technol. Innov.*, 2021, doi:

10.1109/IAECST54258.2021.9695689.
[26] S. Watcharabutsarakham, S. Suntiwichaya, C. Junlouchai, and A. Kitvimorat, "Comparison of Face Classification with Single and Multi-model base on CNN," *15th Int. Jt. Symp. Artif. Intell. Nat. Lang. Process.*, 2020, doi: 10.1109/iSAI-NLP51646.2020.9376825.
[27] V. Mudeng, M. N. Farid, G. Ayana, and S.-W. Choe, "Domain and Histopathology Adaptations-Based Classification for Malignancy Grading System," *SSRN Electron. J.*, 2022, doi: 10.1016/j.ajpath.2023.07.007.
[28] K. E. Ewald, Z. L. Zeng, C. B. Mawuli, H. S. Abubakar, and A. Victor, "Applying CNN with Extracted Facial Patches using 3 Modalities to Detect 3D Face Spoof," *17th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process.*, 2020, doi: https://doi.org/10.1109/iccwamtip51612.2020.9317329.
[29] F. Firdaus and R. Munir, "Masked Face Recognition using Deep Learning based on Unmasked Area," *2022 Second Int. Conf. Power, Control Comput. Technol.*, 2022, doi: 10.1109/ICPC2T53885.2022.9776651.
[30] A. E. B. Alawi and A. M. Qasem, "Lightweight CNN-based Models for Masked Face Recognition," *2021 Int. Congr. Adv. Technol. Eng.*, 2021, doi: 10.1109/ICOTEN52080.2021.9493424.
[31] N. Ragesh, R. Ranjith, and P. Sivraj, "Fast R-CNN based Masked Face Recognition for Access Control System," *2022 4th Int. Conf. Inven. Res. Comput. Appl.*, 2022, doi: 10.1109/ICIRCA54612.2022.9985509.
[32] M. Mobaraki *et al.*, "Masked Face Recognition Using Convolutional Neural Networks and Similarity Analysis," *2023 24th Int. Conf. Digit. Signal Process.*, 2023, doi: 10.1109/DSP58604.2023.10167977.