JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering

Volume 9, Issue 2, October 2022, pp. 149-158
ISSN 2355-5068 ; e-ISSN 2622-4852
**DOI:** 10.33019/jurnalecotipe.v9i2.3044

# Design of Real-Time Face Recognition and Emotion Recognition on Humanoid Robot Using Deep Learning

Muhammad Iqbal[1], Bhakti Yudho Suprapto[2], Hera Hikmarika[3], Hermawati[4], Suci Dwijayanti[5]
[1,2,3,4,5]Department of Electrical Engineering, Faculty of Engineering, Universitas Sriwijaya,
Jl. Raya Palembang-Prabumulih Km. 32, Indralaya 30662, Indonesia

## ARTICLE INFO

**Article historys:**

Received: 12/06/2022
Revised: 25/08/2022
Accepted: 09/09/2022

**Keywords:**

Accuracy**,** Convolution Neural Network**,** Emotion Recognition, Face Recognition, Humanoid Robot

## ABSTRACT

A robot is capable of mimicking human beings, including recognizing their faces and emotions. However, current studies of the humanoid robot have treated face recognition and emotion recognition as separate problems. Thus, this study proposed a combination of face and emotion recognition implemented in the real-time system. Face and emotion recognition systems were developed concurrently in this study using convolutional neural network architectures. The proposed architecture was compared to the well-known architecture, AlexNet, to determine which architecture would be better suited for implementation on a humanoid robot. Primary data from 30 respondents was used for face recognition. Meanwhile, emotional data were collected from the same respondents and combined with secondary data from a 2500-person dataset. Surprise, anger, neutral, smile, and sadness were among the emotions. The experiment was carried out in real-time on a humanoid robot using the two architectures. The accuracy of face and emotion recognition using the AlexNet model was 88 % and 56 %, respectively. Meanwhile, the proposed architecture achieved accuracy rates of 96 % for face recognition and 68 % for emotion recognition, respectively. Thus, the proposed method performs better in terms of recognizing faces and emotions, and it can be implemented on a humanoid robot.

**Corresponding Author**:

Suci Dwijayanti
Department of Electrical Engineering, Faculty of Engineering, Universitas Sriwijaya,
Jl. Raya Palembang Prabumulih Km. 32, Indralaya 30662, Indonesia
Email: sucidwijayanti@ft.unsri.ac.id.

## 1. INTRODUCTION

The rapid advancement of technology has accelerated robotics research. A robot is an automatic device that can be interpreted as a piece of equipment that can be operated with or without human assistance. Many robots, such as humanoid robots, are currently in use to assist people in their daily lives. A humanoid robot is a human-shaped robot with a body, hands, and head, and a motion system that is designed automatically using various sensors and other supporting components. Humanoid robots have a variety of abilities, one of which is the ability to recognize people around them using their faces. This is accomplished by capturing facial images with a camera embedded as a robot’s eye. Humanoid robots can use their ability to recognize faces to interact with humans. The face is a multidimensional visual stimulus that provides a variety of individual information as a person’s identity, such as gender, age, race, mood, and emotions. The face has been widely used as a biometric in everyday life. Emotion is another type of information that can be obtained from the face. Emotions are feelings that can motivate people to act or respond to a stimulus [1]. Emotions or facial expressions at the time and condition experienced by a human being can be used to assess a psychological condition. Because facial and emotional recognition systems utilize still images and video sequences on still images, they can be used as an accurate medium for recognition. The face can be identified by its eyes, mouth, nose, and other features, whereas emotions take many forms, such as smiles, anger, and happiness, which are not the same from one person to the next. Thus, the differences in these features can be extracted into features for facial recognition in order to recognize a person's identity.

Several studies have been conducted using various algorithms to recognize faces and facial expressions. In their research, S. Madanny, Samsuryadi and N. Yusliani discussed facial recognition using the hypersausage neural network method [2]. H. Zhi and S. Liu used the Principal Component Analysis (PCA) method to extract features of the gray-scale face, followed by the genetic algorithm to optimize the extracted features and support vector machine as the classifier [3]. Then, for face recognition, [4] compared two methods: multi-layer perceptron and radial basis function regardless the image quality and illumination. Zhang et al. developed a method for recognizing facial emotions using convolutional neural network and image edge computing [5]. AK. Bahreini, W. Van der Vegt, and W. Westera discussed the use of fuzzy systems for facial emotion recognition [6]. Adeyanju et al. evaluated the performance of various support vector machine kernels for face emotion recognition [7].

The methods mentioned above are quite good at facial recognition and facial expressions, but they have shortcomings such as low accuracy [2-7] and have not been implemented in real-time. In addition, face recognition and expression recognition were not combined into a single recognition system.

Based on the problems described above, this study aims to use a deep learning convolutional neural network (CNN) method to increase the accuracy during testing. CNN is one of the neural networks that has been used in image classification research. CNN is known to be superior to other deep learning in image classifications because of its high level of accuracy. In addition, the facial and emotion recognition system developed in this study is then implemented on a humanoid robot. As a result, the humanoid robot can recognize a person and his emotions in real-time.

The paper is structured as follows: Section 2 describes the methodology used in this study. Section 3 contains the findings and discussions. Finally, Section 4 concludes the paper.

## 2. RESEARCH METHOD

### 2.1. Hardware Design

In this study, several hardware accessories were used to support the implementation of humanoid robot, such as:

1. Webcam Camera
2. JX Servo 60KG
3. Arduino
4. Raspberry Pi
5. DOT Matrix

The positioning of the components used in this study was planned as shown in Figure 1 (A-D).

Volume 9, Issue 2, October 2022, pp. 149-158
ISSN 2355-5068 ; e-ISSN 2622-4852
**DOI:** 10.33019/jurnalecotipe.v9i2.3044

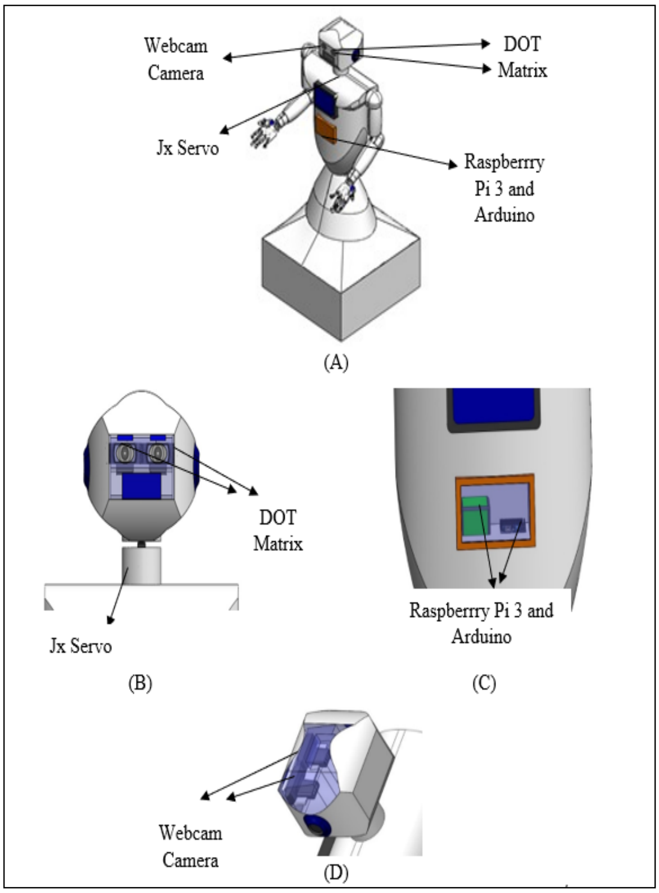

**Figure 1.** Component design on humanoid robot (A-D)

## 2.2. Data Collection

The face dataset for this study was obtained from 30 Universitas Sriwijaya students. Figure 2 depicts samples of face data. The data were obtained using a webcam with a resolution of 640 × 480 pixels. The collected data was then processed and extracted in order to perform face recognition and emotion recognition in the image.

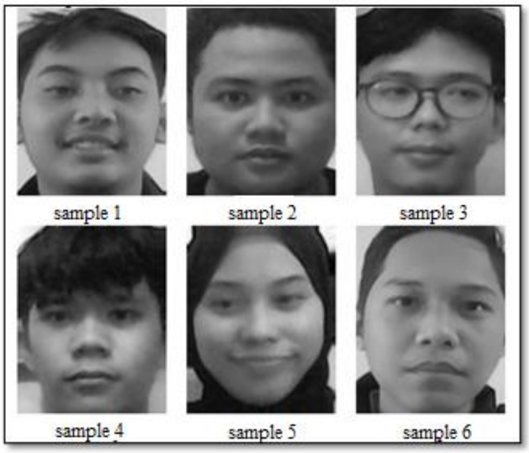

**Figure 2.** Samples of faces

Meanwhile, the same respondents provided the emotion data. The dataset from Kaggle [8] was used in this study to add variation. Figure 3 depicts a sample of emotions from Kaggle. In this study, the five facial emotions used were: smile, angry, surprise, neutral, and sad.

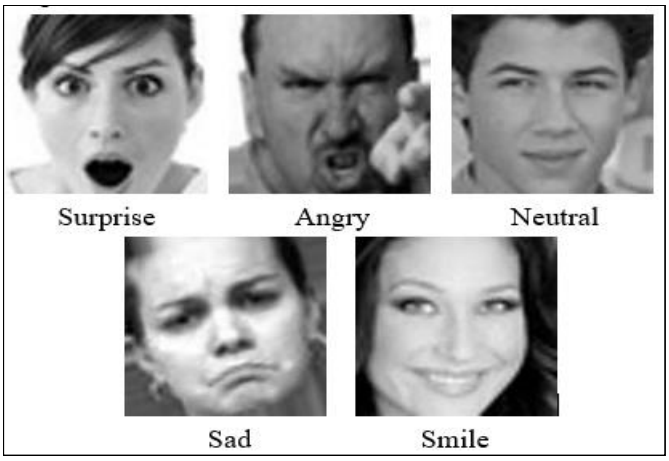

**Figure 3.** Samples of emotion (surprise, angry, neutral, sad, smile)

### 2.3. Face and Emotion Recognition Algorithm

Figure 4 depicts the design of the recognition system embedded in the robot to recognize a person's face.

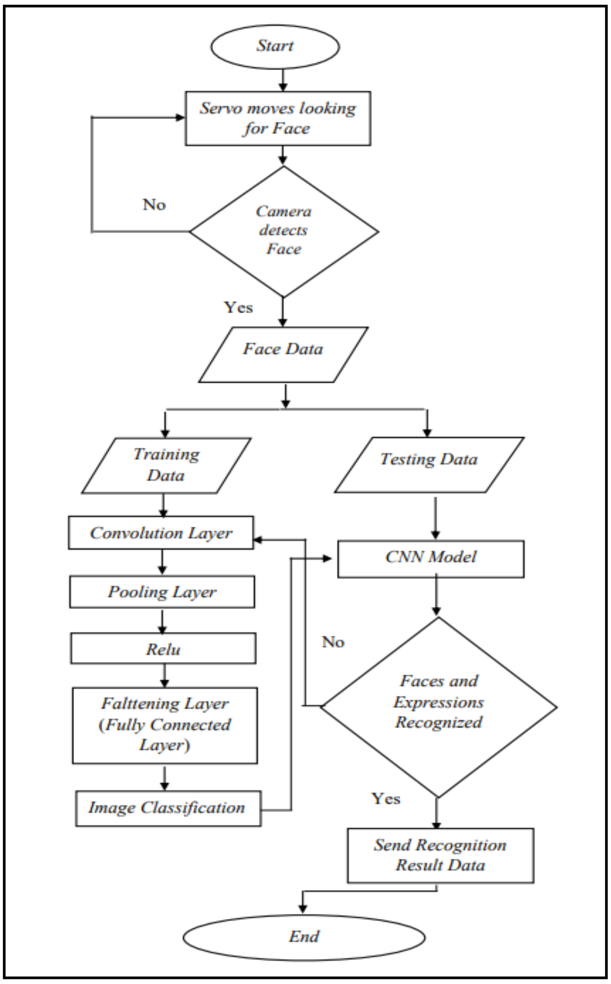

**Figure 4.** Flowchart of face recognition and emotion recognition

In this study, a convolutional neural network was used. A convolution neural network (CNN) is a variation of the multilayer perceptron that contains or modifies the neural network which is inspired by the visual perception of living creatures [9].

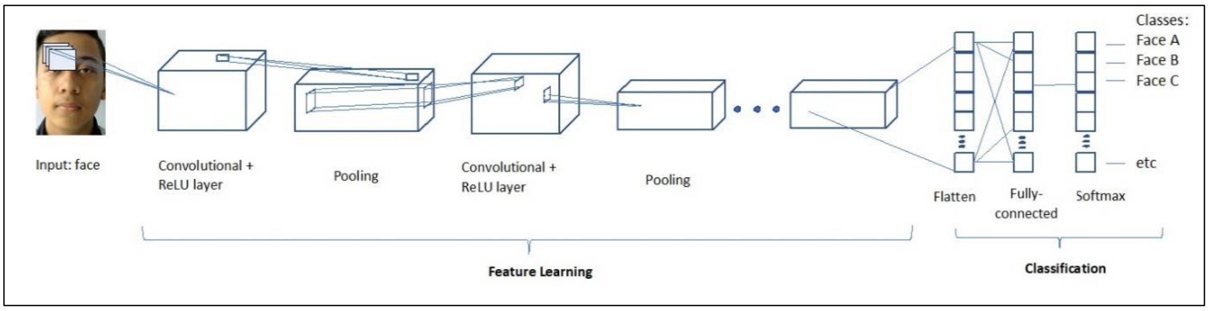

**Figure 5.** Convolutional neural network [10]

As shown in Figure 5, a layer in CNN has a 3-dimensional arrangement of neurons (width, height, depth). The width and height of the layer are measured, while the depth refers to the number of layers. In general, layers in CNN can be classified into 2 types:

1. The image feature extraction layer, which is at the top of the architecture. This layer is made up of several layers, each of which contains neurons connected to the previous layer's local region.
2. The classification layer is made up of several layers, each with its own set of neurons that are fully connected to the others.

AlexNet is one of several CNN architectures that can be used to carry out the training process. AlexNet contains eight layers, the first five of which are convolutional layers, some of which were followed by max-pooling layers, and the final three of which were fully connected layers [11]. It used the non-saturating ReLU activation function, which showed improved training performance over tanh and sigmoid [11]. Figure 6 depicts the AlexNet architecture.

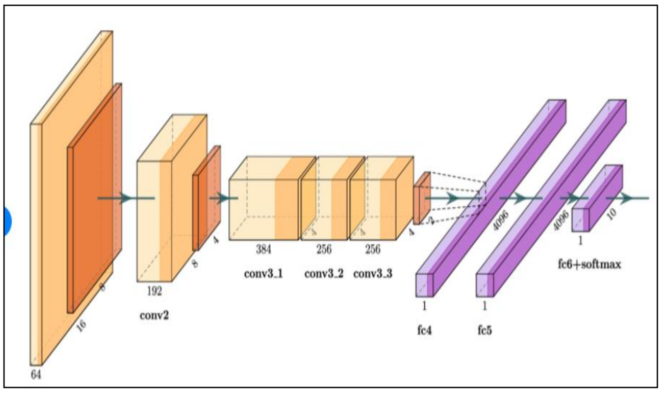

**Figure 6.** AlexNet architecture

### 2.4. System Evaluation

The performance of the facial recognition and expression recognition system that has been created is evaluated. It includes accuracy:

$$\text{accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}}, \qquad (1)$$

Where TP is true positive, TN is true negative, FP is false positive, and TN is true negative. This formula is used to get the accuracy of face recognition or emotion recognition. A calculation of the accuracy value shows the level of effectiveness of a classification per class.

## 3. RESULTS AND DISCUSSION

This section discusses the results obtained for real-time face and emotion recognition on a humanoid robot. A new CNN architecture is proposed in this study. The proposed architecture has the same numbers of layers as the AlexNet but with some layer and parameter changes such as changing the stride and size. Table I shows the parameters used in the proposed architecture. This proposed architecture was also compared to the AlexNet.

**Table 1.** Parameter of Proposed Architecture

| Parameter | Value |
|---|---|
| Optimizer | Adam, SgD |
| Dropout After Pooling Layer | 0.05, 0.1 |
| Dropout Fully Connected Layer | 0.25, 0.1 |
| Dense Layer | 64, 128 |
| Learning Rate | 0.0001 |
| Batch size | 16 |

### 3.1. Face Recognition

In this study, the proposed architecture was compared to AlexNet. The training loss of each class can be seen in Figure 7. As shown in the figure, the proposed architecture has lower training loss compared to the AlexNet. The training loss of the proposed architecture and the AlexNet were 0.05 and 0.13, respectively. These results showed that the proposed architecture may give better performance than the AlexNet for recognizing the face.

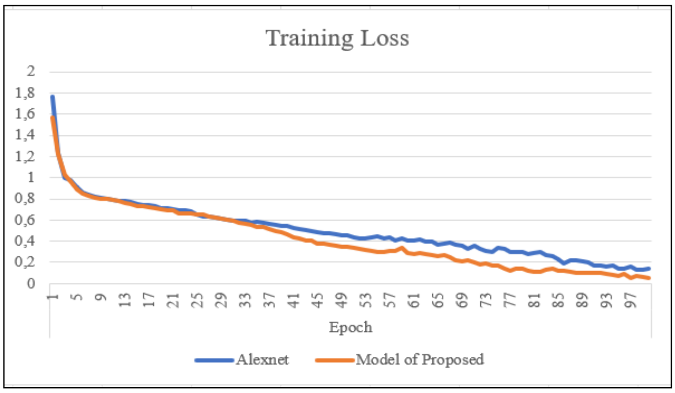

**Figure 7.** Training loss for face recognition

The accuracy obtained from the testing data is shown in Table 2. As shown in the table, the model based on the proposed architecture outperforms AlexNet. The proposed architecture recognized all of the testing data samples. Meanwhile, AlexNet recognized 26 of testing data. It encountered an error while recognizing samples 2, 8, 10, and 30. The accuracy of the AlexNet and the proposed architecture was 88% and 96%, respectively These results demonstrated that the model of the proposed architecture was capable of accurately recognizing faces.

### 3.2. Emotion Recognition

Figure 8 depicts the results of training loss for emotion recognition. Based on the graph, AlexNet's and the proposed architecture's training losses were 0.44 and 0.33, respectively. The results once again showed that the proposed architecture outperformed the AlexNet. It could be due to the parameters in the proposed architecture.

Volume 9, Issue 2, October 2022, pp. 149-158
ISSN 2355-5068 ; e-ISSN 2622-4852
**DOI:** 10.33019/jurnalecotipe.v9i2.3044

**Table 2.** Face recognition results

| **Face Recognition** | | | | | |
|---|---|---|---|---|---|
| **No** | **Name** | ***Accuracy*** | | ***Recognition / Unrecognition*** | |
| | | **AlexNet** | **Model of proposed architecture** | **AlexNet** | **Model of proposed architecture** |
| 1 | Sample 1 | 96 % | 98 % | √ | √ |
| 2 | Sample 2 | 65 % | 82 % | x | √ |
| 3 | Sample 3 | 100 % | 100 % | √ | √ |
| 4 | Sample 4 | 92 % | 97 % | √ | √ |
| 5 | Sample 5 | 91 % | 97 % | √ | √ |
| 6 | Sample 6 | 85 % | 98 % | √ | √ |
| 7 | Sample 7 | 90 % | 99 % | √ | √ |
| 8 | Sample 8 | 70 % | 87 % | x | √ |
| 9 | Sample 9 | 84 % | 92 % | √ | √ |
| 10 | Sample 10 | 71 % | 89 % | x | √ |
| 11 | Sample 11 | 93 % | 98 % | √ | √ |
| 12 | Sample 12 | 96 % | 97 % | √ | √ |
| 13 | Sample 13 | 89 % | 96 % | √ | √ |
| 14 | Sample 14 | 92 % | 100 % | √ | √ |
| 15 | Sample 15 | 90 % | 98 % | √ | √ |
| 16 | Sample 16 | 84 % | 95 % | √ | √ |
| 17 | Sample 17 | 94 % | 100 % | √ | √ |
| 18 | Sample 18 | 92 % | 92 % | √ | √ |
| 19 | Sample 19 | 97 % | 96 % | √ | √ |
| 20 | Sample 20 | 83 % | 96 % | √ | √ |
| 21 | Sample 21 | 95 % | 98 % | √ | √ |
| 22 | Sample 22 | 93 % | 98 % | √ | √ |
| 23 | Sample 23 | 91 % | 96 % | √ | √ |
| 24 | Sample 24 | 100 % | 98 % | √ | √ |
| 25 | Sample 25 | 90 % | 98 % | √ | √ |
| 26 | Sample 26 | 70 % | 91 % | √ | √ |
| 27 | Sample 27 | 92 % | 99 % | √ | √ |
| 28 | Sample 28 | 94 % | 100 % | √ | √ |
| 29 | Sample 29 | 81 % | 93 % | √ | √ |
| 30 | Sample 30 | 77 % | 89 % | x | √ |

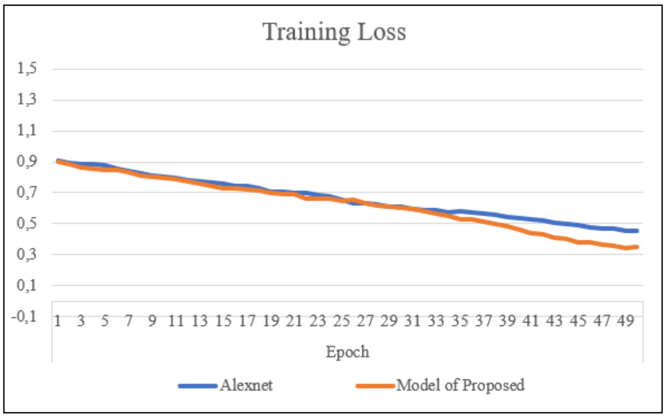

**Figure 8.** Training loss for emotion recognition

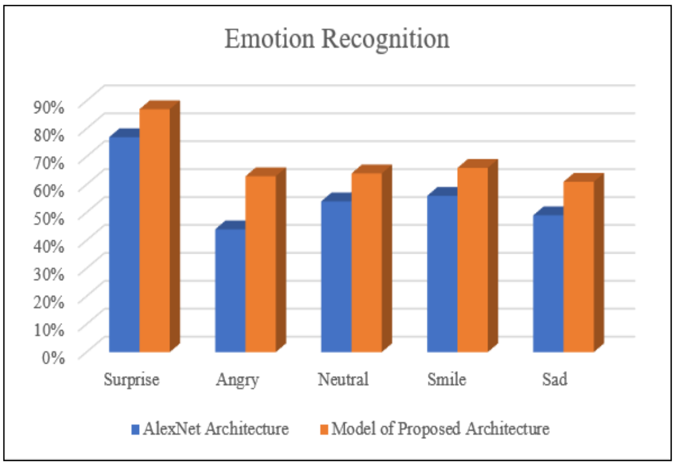

**Figure 9.** Results of emotion recognition

The accuracy of the proposed architecture and the AlexNet are shown in Figure 9. As shown in the figures, the proposed architecture is more accurate at recognizing emotions. For both models, the emotion of surprise provided the highest accuracy. This could be due to the difference in expression where the person opened his mouth. Sad expression yielded the lowest accuracy; for this expression, the AlexNet provided an accuracy of 56% whereas the proposed model's accuracy was 68%.

### 3.3. Face Recognition and Emotion Recognition in Real-time on Humanoid Robot Using Webcam Module

The experiment was carried out in real-time with the help of a camera mounted on the head of a humanoid robot. The test was carried out by combining face recognition and emotion recognition together into one frame and then obtaining the results. This system can identify not one but two people. Figure 10 shows the result of face recognition, emotion recognition, and coordinate point recognition.

Volume 9, Issue 2, October 2022, pp. 149-158
ISSN 2355-5068 ; e-ISSN 2622-4852
**DOI:** 10.33019/jurnalecotipe.v9i2.3044

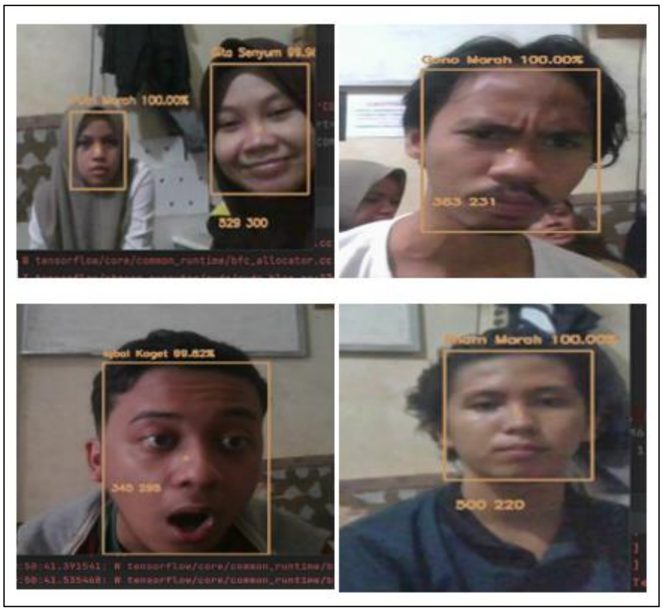

**Figure 9.** Face and emotion recognition in real-time

As shown in Figure 10, the accuracy of the proposed architecture was high enough to recognize both faces and emotions at the same time. The proposed recognition system has the advantage of being able to recognize two people at the same time. In addition, the proposed recognition system can determine the coordinate point, which is useful for determining the robot's position.

## 4. CONCLUSION

Based on the simulation and testing of face recognition and emotion recognition, it is possible to conclude that the proposed architecture model has better performance than the AlexNet. The AlexNet's accuracy in face recognition and emotion recognition was 88 % and 56 %, respectively. Meanwhile, the proposed architecture model has an accuracy of 68 % for emotion detection and 96 % for face recognition.

This study also showed that the proposed recognitions for humanoid robots can be implemented in real-time. However, the proposed system must be improved to send the coordinate position of the recognized object for future work so that the humanoid robot can work like a human.

### Acknowledgments

The research/publication of this article was funded by DIPA of Public Service Agency of Universitas Sriwijaya 2022. SP DIPA-023.17.2.677515/2022, on December 13, 2021. In accordance with the Rector's Decree Number: 0109/UN9.3.1/SK/2022, on April 28, 2022.

## REFERENCES

[1] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, "Human emotion recognition: Review of sensors and methods," *Sensors*, vol. 20, no. 3, p.592., 2020.

[2] S. Madanny, Samsuryadi,N. Yusliani, "Face Recognition Using Hyper Sausage Neuron Networks," in *Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*, pp. 480-484, 2020.

[3] H. Zhi, and S. Liu, "Face recognition based on genetic algorithm," *Journal of Visual Communication and Image Representation*, vol. *58*, pp.495-502, 2019

[4] T.A. Mohammed, A. Alazzawi, O.N. Uçan, and O. Bayat, "Neural network behavior analysis based on transfer functions MLP & RB in face recognition," In *Proceedings of the First International Conference on Data Science, E-learning and Information Systems*, pp. 1-6, 2018.

[5] H. Zhang, A. Jolfaei, and M. Alazab, "A face emotion recognition method using convolutional neural network and image edge computing," *IEEE Access*, vol. 7, pp. 159081–159089, 2019.

[6] K. Bahreini, W. Van der Vegt, and W. Westera, "A fuzzy logic approach to reliable real-time recognition of facial emotions," *Multimedia Tools and Applications*, vol. 78, no. 14, pp.18943-18966, 2019

[7] A. Adeyanju, E. O. Omidiora, and O. F. Oyedokun, "Performance evaluation of different support vector machine kernels for face emotion recognition," *IntelliSys 2015 - Proc. 2015 SAI Intell. Syst. Conf.*, pp. 804–806, 2015. doi: 10.1109/IntelliSys.2015.7361233.

[8] G. Sharma, "CK+48 5 Emotions," 2019. [Online]. Available: https://www.kaggle.com/datasets/gauravsharma99/ck48-5-emotions.

[9] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp.354-377, 2018.

[10] S. Dwijayanti, R.R. Abdillah, H. Hikmarika, Z. Husin, and B.Y. Suprapto, "Facial Expression Recognition and Face Recognition Using a Convolutional Neural Network," In *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 621-626, 2020.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks, "*Advances in neural information processing systems*, vol. 25, 2012.