p–ISSN: 2723 - 6609 e-ISSN: 2745-5254
Vol. 5, No. 10, October 2024 http://jist.publikasiindonesia.id/

Implementation of the YOLO Method for Detection of Human Emotions Based on Facial Mimics

Thomas Felison[1*], Erwin conery firtan[2], Steven[3], Willyam Chandra[4], Saut Dohot Siregar[5]
Universitas Prima Indonesia, Indonesia
Email: Thomas.tba2016@gmail.com[1*], Winsmithryneco123@gmail.com[2], Swangs370@gmail.com[3], williamchandra969@gmail.com[4], Sautdohotsiregar@gmail.com[5]

*Correspondence

Keywords: emotion detection, facial expressions, digital imagery, yolo method.

ABSTRACT

The specific purpose of this study is to test the accuracy of the YOLO method in recognizing human facial expressions through tests that involve several types of expressions such as angry, surprised, happy, neutral, and afraid. Emotion detection through facial expression recognition plays an important role in everyday life, such as how to respond correctly to emotional expressions in social interactions so that you can establish and build verbal or nonverbal communication with other people and so on. Facial expressions are facial changes in response to a person's emotional state, intentions, or social communication. Face detection is the first step that must be taken in facial analysis, including facial expression recognition. Many methods can be used to carry out the face detection process, such as the YOLO method. This YOLO method reframes object detection as a single regression problem, directly from image pixels to bounding box coordinates and class probabilities. By using the YOLO method, the process only needs to look once at the input image, to predict what objects are in the image and where those objects are. Based on the results of the tests carried out, the YOLO method can be used to detect human facial expressions with a success rate of 80%, with neutral, surprised, and disgusted facial expressions having a good level of accuracy and fearful facial expressions having a poor accuracy level. The YOLO method can detect the facial expressions of humans who wear accessories such as glasses.

Introduction

The use of digital images has been widely used to identify a shape object, one of which is to recognize human facial expressions. Research on facial expression analysis has many uses, such as detecting facial expressions in the presence system, using facial

feature points to control the movement of 3D Augmented Reality objects, creating a synthesis of human faces in a human-computer dialogue system, or detecting the presence of micro-expressions of subtle movements in the face (Rosiani et al., 2018). One of the real applications of human facial expression analysis is to recognize the facial expressions of e-learning users, where learner interaction is a weakness that must be considered in e-learning learning (Husdi, 2016). The process of recognizing human facial expressions is also applied to the MOODSIC music player application, which will play music according to the user's emotions obtained by detecting the user's facial expressions (Wijaya et al., 2018). Emotion detection through facial expression recognition plays an important role in daily life, such as how to properly respond to emotional expressions in social interactions so that it can establish and build verbal or nonverbal communication with others and so on. Another advantage is being able to see and understand the intention of the interlocutor so that it will minimize deception and falsehood. The inability to recognize facial emotional expressions can lead to inaccuracies in interpreting other people's emotions/feelings, which will automatically lead to ambiguity and inaccurate decision responses (Hartanto, 2019).

Facial expressions are facial changes in response to a person's emotional state, intentions, or social communication. Face detection is the first step that must be done in facial analysis, including facial expression recognition. Face detection aims to determine whether or not there is a face in the image, and if there is a location of the face and the size of each face in the image. (Budiyanta et al., 2021). In face detection, there are several challenges such as the position of the face not facing directly to the camera, face scale, facial expressions, face obstructed by other objects, and lighting conditions. (Prasetyawan, 2020). Many methods can be used to carry out the face detection process, such as Jatmoko has implemented the Viola-Jones algorithm in facial recognition, with the results of research from all experiments obtaining an average accuracy score of 65% (Jatmoko et al., 2020). Another research conducted by Putra and Krishna using the eigenface and haar cascade classifier methods to carry out the facial recognition process, with the results of the study obtaining facial recognition accuracy with a maximum distance of 3 meters from the camera is 63% (Putra et al., 2023). To solve this problem, the You Only Look Once (YOLO) method can be used. This YOLO method reframes object detection as a single regression problem, directly from the image pixel to the bounding box coordinates and class probability. Using this YOLO method, the process only needs to look once at the input image, to predict what object is contained in the image and where it is located. (Redmon, 2016).

Some similar studies that have been conducted before can be detailed as follows:

Table 1
Previous Research

Author's Name (Year)	Heading	Recency

Rizki Rafiif Amaanullah, Gracia Rizka Pasfica, Satria Adi Nugraha, Mohammad Rifqi Zein, Faisal Dharma Subordination (2022)	Implementation of convolutional neural network for emotion detection through faces. The results of the study obtained an accuracy of 81.92% for training and 81.69% for testing.	Maximum accuracy only as large as 81.92%, so it is expected that the YOLO method can have more accuracy good.
Evan Tanuwijaya, Timotius, David Christian Kartamihardja, Timothy Leonardo Lianoto (2021)	Detection of Human Facial Expressions Using Convolution Neural Network on Online Learning Images. An accuracy of 0.94 was obtained during *training*.	Method CNN has a long execution process, so it is expected that the YOLO method can have a faster execution time.

The purpose of the research is to implement the YOLO method in carrying out the process of recognizing human facial expressions. After that, tests will be carried out on the YOLO method so that the accuracy of the YOLO method in carrying out the process of recognizing human facial expressions can be known. (Bamba et al., 2022).

The benefit of this study is to enrich the literature related to the detection of human emotions through facial expression analysis using the YOLO method, as well as to provide a reference for further research in the development of more efficient and accurate methods of object detection and facial expressions. Practically, the results of this research have the potential to be applied in various fields, such as human-computer interaction (HCI), supervision and monitoring systems in work or education environments, as well as the entertainment and technology industries. The implementation of the YOLO method in facial expression detection can help improve interaction between humans and machines, develop monitoring systems that are more responsive to human emotions, and be used in entertainment applications that adjust content based on user expressions and emotions. Thus, this research is expected not only to contribute to the development of science but also to have a real impact on various technological and social sectors.

Method

Type of Research

This study is a quantitative research that adopts an experimental approach. This research collects data from various online sources, and then conducts an analysis process to produce relevant conclusions.

Time and Place of Research

This research took approximately half a year and has been carried out starting January 2024 in the city of Medan, Indonesia.

Table 2

Research Schedule

Activities \ Time	January 2024	February 2024	March 2024	April 2024	May 2024	June 2024
Topic Discussion Research						
Collection Reference						
Determination Topic						
Manufacture Proposal						
Collection Data						
Data Analysis						
Testing Method						
Evaluation and Repair						
Manufacture Report						
Publication Journal						
Dissemination						

The following are the steps of the working procedure in this study:

1. Preliminary Stage

This research begins by looking for references from previous studies that are relevant to the research topic to be conducted.

2. Problem Determination Stage

Determine the formulation of problems that occur in the identification of types of facial expressions. In addition, the study also establishes problem boundaries to direct focus on the scope of the research.

3. Data Collection Stage

This stage is carried out by collecting data regarding the You Only Look Once (YOLO) method.

4. Analysis Stages

This process analyzes the working process of the method used and the method used in detecting facial expressions. After that, an application will be designed that implements the methods applied in this research.

5. Testing Stage

In this step, the collected data will be tested using a pre-prepared application.

6. Report Stage

The last stage of the report-making process is carried out by the guidelines for writing research that has been set.

The YOLO algorithm is a method in deep learning that uses a Convolutional Neural Network (CNN) to detect objects in images. This algorithm divides the image into grids of a certain size, where each grid predicts a bounding box and a class map for the objects it contains. If a grid predicts the existence of an object, then a bounding box will be predicted that surrounds the object. The confidence score is calculated for each bounding box and selection is made based on the confidence value. (Nasution & Kartika, 2022).

The data sources for this study come from journal literature and books that support the research topic. To ensure the success of the research, the tools used are as follows.

Hardware Requirements

In the process of designing and testing the system, a set of computer hardware is used that has the following specifications.

a. Intel Core i3.
b. 4 GB RAM.
c. Hard drive 2 TB.
d. LCD monitor with a minimum screen resolution of 1024 x 768.

Software Requirements

The software to design this program is:

a. Windows 10 operating system.
b. Visual Studio C# 2013
c. Microsoft SQL Server 2012.

Results and Discussion

The working process of the implementation of the YOLO method for detecting human emotions consists of two parts, namely:

The training process by entering the dataset

The user can select the expression type from the image file and click the Save button to save the data. The dataset used in this study is 455 pieces of data, with details of 315

pieces of training data (70%) and 140 pieces of data (30%), namely each expression has 45 pieces of training data and 20 pieces of testing data. The result of the training process is in the form of a confidence value for each training image. This value will be compared with the confidence value of the testing image. In this Expression Type Dataset Input form, there are seven types of expressions, namely angry, disgusted, fearful, happy, neutral, sad, and surprised. The display of the Expression Type Dataset Input form after data storage can be seen in the following image:

Figure 1 Input Dataset

The testing process to detect human emotions is based on facial mimicry.

Users can select the image file to be recognized. The trick is to click the Browse button so that the system will display the Browse dialog box. Then, the user can select the desired image file and click the Open button, so the system will read the contents of the selected image file and display it on the monitor screen. (Hutauruk et al., 2020). The display of the Human Face Expression Recognition form after data input can be seen in the following image:

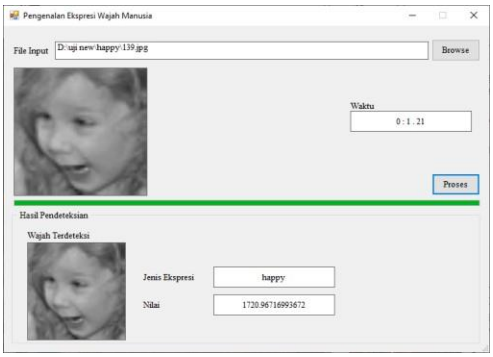

Figure 2 Detection of Human Emotion

The next test will be to detect human facial expressions against input images. In this test, there are seven types of human facial expressions with a breakdown of each facial expression has 140 dataset images. The results of the tests carried out can be detailed as follows:

Table 1
Angry Expression

Test Image	Types of Expressions	Detection Results	Value	Information
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Angry*	0	Succeed
	*Angry*	***Sad***	**1692.7865**	**FAIL**
	*Angry*	*Angry*	0	Succeed

	*Angry*	*Angry*	0	Succeed
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Neutral*	1710.5479	FAIL
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Fear*	1568.9315	FAIL
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Angry*	0	Succeed
	*Angry*	*Surprise*	1718.3687	FAIL
	*Angry*	*Neutral*	1730.9356	FAIL

The number of correct facial expression detections = 15 pieces and incorrect = 5 pieces

Table 2
Disgust Expression

Test Image	Types of Expressions	Detection Results	Value	Information
	*Disgust*	*Disgust*	1438.016	Succeed
	*Disgust*	*Disgust*	1161.1985	Succeed
	*Disgust*	*Disgust*	1515.1654	Succeed
	*Disgust*	*Disgust*	1609.9096	Succeed
	*Disgust*	*Disgust*	1245.4341	Succeed
	*Disgust*	*Disgust*	1634.6125	Succeed
	*Disgust*	*Disgust*	1760.2670	Succeed

	*Disgust*	*Disgust*	1504.8106	Succeed
	*Disgust*	*Disgust*	1579.0573	Succeed
	*Disgust*	*Disgust*	0	Succeed
	*Disgust*	*Disgust*	1413.7295	Succeed
	*Disgust*	*Disgust*	0	Succeed
	*Disgust*	*Disgust*	0	Succeed
	*Disgust*	*Disgust*	1773.9676	Succeed
	*Disgust*	*Disgust*	0	Succeed
	*Disgust*	*Disgust*	0	Succeed
	*Disgust*	*Disgust*	1280.9485	Succeed
	*Disgust*	*Disgust*	1759.4326	Succeed
	*Disgust*	*Disgust*	1613.7714	Succeed
	*Disgust*	*Disgust*	1669.0512	Succeed

The number of correct facial expression detections = 20 pieces and incorrect = 0 pieces

Table 3
Ekspresi Fear

Test Image	Types of Expressions	Detection Results	Value	Information
	*Fear*	*Fear*	1790.7884	Succeed
	*Fear*	***Sad***	**1608.5758**	**FAIL**
	*Fear*	*Fear*	1673.8898	Succeed
	*Fear*	*Fear*	1384.6595	Succeed
	*Fear*	***Sad***	**1696.3534**	**FAIL**
	*Fear*	***Sad***	**1667.2543**	**FAIL**

	*Fear*	*Fear*	1772.6923	Succeed
	*Fear*	*Fear*	0	Succeed
	*Fear*	***Neutral***	**1715.4285**	**FAIL**
	*Fear*	***Sad***	**1791.8585**	**FAIL**
	*Fear*	***Disgust***	**1590.0091**	**FAIL**
	*Fear*	***Neutral***	**1615.8116**	**FAIL**
	*Fear*	***Happy***	**1708.1165**	**FAIL**
	*Fear*	***Fear***	**1759.7**	**Succeed**
	*Fear*	***Surprise***	**1664.9078**	**FAIL**
	*Fear*	***Surprise***	**1688.6281**	**FAIL**
	*Fear*	***Disgust***	**1615.0486**	**FAIL**
	*Fear*	***Surprise***	**1686.7208**	**FAIL**
	*Fear*	***Happy***	**1743.3161**	**FAIL**
	*Fear*	***Sad***	**1711.2256**	**FAIL**

Jumlah deteksi ekspresi wajah yang benar = 6 buah dan salah = 14 buah

Table 4
Happy Expression

Test Image	Types of Expressions	Detection Results	Value	Information
	*Happy*	*Happy*	1720.9672	Succeed
	*Happy*	*Happy*	1727.6221	Succeed
	*Happy*	*Happy*	1668.1214	Succeed
	*Happy*	*Happy*	1675.4996	Succeed

	*Happy*	*Happy*	1728.2098	Succeed
	*Happy*	*Happy*	1795.8608	Succeed
	*Happy*	*Happy*	1761.3611	Succeed
	*Happy*	*Happy*	1639.1556	Succeed
	*Happy*	*Sad*	1742.6247	FAIL
	*Happy*	*Happy*	1700.9500	Succeed
	*Happy*	*Happy*	1665.5131	Succeed
	*Happy*	*Happy*	1751.5836	Succeed
	*Happy*	*Happy*	1706.7159	Succeed
	*Happy*	*Happy*	1706.9238	Succeed
	*Happy*	*Happy*	1674.6612	Succeed
	*Happy*	*Happy*	1767.5557	Succeed
	*Happy*	*Neutral*	1742.9228	FAIL
	*Happy*	*Happy*	1711.1560	Succeed
	*Happy*	*Happy*	1636.2408	Succeed
	*Happy*	*Happy*	1714.5682	Succeed

The number of correct facial expression detections = 18 pieces and incorrect = 2 pieces

Table 5
Neutral Expression

Test Image	Types of Expressions	Detection Results	Value	Information

	*Neutral*	*Neutral*	1581.9886	Succeed
	*Neutral*	*Neutral*	1750.1494	Succeed
	*Neutral*	*Neutral*	1725.4666	Succeed
	*Neutral*	*Neutral*	1675.0666	Succeed
	*Neutral*	*Neutral*	1529.7778	Succeed
	*Neutral*	*Neutral*	1784.1687	Succeed
	*Neutral*	*Neutral*	1801.3428	Succeed
	*Neutral*	*Neutral*	1721.6669	Succeed
	*Neutral*	*Neutral*	1619.4968	Succeed
	*Neutral*	*Neutral*	1746.587	Succeed
	*Neutral*	*Neutral*	1628.9739	Succeed
	*Neutral*	*Neutral*	1473.0811	Succeed
	*Neutral*	*Neutral*	1679.0822	Succeed
	*Neutral*	*Neutral*	1649.7373	Succeed
	*Neutral*	*Neutral*	1661.1144	Succeed
	*Neutral*	*Neutral*	1657.5032	Succeed
	*Neutral*	*Neutral*	1693.0118	Succeed
	*Neutral*	*Neutral*	1599.0063	Succeed

	*Neutral*	*Neutral*	1695.3749	Succeed
	*Neutral*	*Neutral*	1645.0729	Succeed

The number of correct facial expression detections = 20 pieces and incorrect = 0 pieces

Table 6
Sad Expression

Test Image	Types of Expressions	Detection Results	Value	Information
	*Sad*	*Sad*	1765.4651	Succeed
	*Sad*	***Angry***	**1733.5703**	**FAIL**
	*Sad*	*Sad*	1646.4437	Succeed
	*Sad*	*Sad*	1703.6881	Succeed
	*Sad*	*Sad*	1701.4940	Succeed
	*Sad*	*Sad*	1754.3153	Succeed
	*Sad*	*Sad*	1644.0298	Succeed
	*Sad*	*Sad*	1719.2821	Succeed
	*Sad*	*Sad*	0	Succeed
	*Sad*	*Sad*	1542.2908	Succeed
	*Sad*	*Sad*	1833.3347	Succeed
	*Sad*	*Sad*	1403.7877	Succeed
	*Sad*	*Surprise*	1657.6842	FAIL
	*Sad*	*Sad*	1799.6344	Succeed
	*Sad*	*Disgust*	1764.5966	FAIL

	*Sad*	*Fear*	1767.5811	FAIL
	*Sad*	*Neutral*	1577.8628	FAIL
	*Sad*	*Disgust*	1759.4655	FAIL
	*Sad*	*Happy*	1693.0921	FAIL
	*Sad*	*Sad*	1697.5989	Succeed

The number of correct facial expression detections = 13 pieces and incorrect = 7 pieces

Table 7
Expression of Surprise

Test Image	Types of Expressions	Detection Results	Value	Information
	*Surprise*	*Surprise*	1334.9906	Succeed
	*Surprise*	*Surprise*	1432.2234	Succeed
	*Surprise*	*Surprise*	1665.8268	Succeed
	*Surprise*	*Surprise*	0	Succeed
	*Surprise*	*Surprise*	1691.5824	Succeed
	*Surprise*	*Surprise*	1654.6921	Succeed
	*Surprise*	*Surprise*	1733.4696	Succeed
	*Surprise*	*Surprise*	1626.8012	Succeed
	*Surprise*	*Surprise*	1729.4933	Succeed
	*Surprise*	*Surprise*	1758.5832	Succeed
	*Surprise*	*Surprise*	1681.6037	Succeed
	*Surprise*	*Surprise*	1651.5372	Succeed
	*Surprise*	*Surprise*	1688.8908	Succeed
	*Surprise*	*Surprise*	1750.5942	Succeed
	*Surprise*	*Surprise*	1766.0045	Succeed
	*Surprise*	*Surprise*	1657.3292	Succeed
	*Surprise*	*Surprise*	1768.7086	Succeed
	*Surprise*	*Surprise*	1666.3007	Succeed
	*Surprise*	*Surprise*	1644.6495	Succeed
	*Surprise*	*Surprise*	1660.7420	Succeed

The number of correct facial expression detections = 20 pieces and incorrect = 0 pieces

The results of the test on human facial expressions can be summarized as seen in the following table:

Table 8
Test results on human facial expressions

Types of Expressions	Correct Amount	Number of Errors	Total
Angry	15	5	20
Disgust	20	0	20
Fear	6	14	20
Happy	18	2	20
Neutral	20	0	20
Sad	13	7	20
Surprise	20	0	20
Total	112	28	140

By using the confusion matrix method, the following can be obtained:

		Real Situation	
		+	-
Result Classification	+	TP = 112	FP = 0
	-	FN = 28	TN = 0

Accuracy = (TP+TN) / (TP+FP+FN+TN) = (112+0) / (112+0+28+0) = 112/140 * 100% = 80%. Precision = TP / (TP + FP) = 112 / (112 + 0) * 100% = 100%.
Recall = TP / (TP + FN) = 112 / (112 + 28) * 100% = 80%.
Error: (FP+FN) / (TP+FP+FN+TN) = (0+28) / (112+0+28+0) = 28/140 * 100% = 20%.

From the results of the tests carried out, the following information can be obtained:

1. The YOLO method can be used to detect human facial expressions with an accuracy rate of 80%.
2. The YOLO method is not able to detect fearful facial expressions, but the YOLO method has a good level of accuracy in detecting neutral, surprised, and disgusted facial expressions.
3. From the test results obtained, it can be seen that the application of the YOLO method has been able to solve the overfitting problem. This overfitting problem is often encountered in Convolutional Neural Network (CNN) model research, such as the Emotion Detection of Facial Expressions with Deep Learning study published in May 2024, where the test results obtained were only 62.44% of the test data.

Conclusion

From the description in the previous chapters, several important conclusions can be drawn in the research in this final project, including. The YOLO method can detect human facial expressions with an 80% success rate. Neutral, surprise, and disgust facial expressions have high accuracy, while fear facial expressions have low accuracy.

The YOLO method can recognize human facial expressions even when wearing accessories such as glasses.

Bibliography

Bamba, I., Yashika, Singh, J., & Chawla, P. (2022). Face Recognition Techniques and Implementation. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2022, Volume 2* (pp. 345–356). Springer.

Budiyanta, N. E., Mulyadi, M., & Tanudjaja, H. (2021). Sistem Deteksi Kemurnian Beras berbasis Computer Vision dengan Pendekatan Algoritma YOLO. *Jurnal Informatika: Jurnal Pengembangan IT*, *6*(1), 51–55.

Hartanto, H. (2019). Waktu Reaksi Dan Akurasi Dalam Pengenalan Ekspresi Wajah: Sebuah Eksperimen Psikofisik. *Jurnal Psikologi*, *17*(2), 131–142.

Husdi, H. (2016). Pengenalan Ekspresi Wajah Pengguna Elearning Menggunakan Artificial Neural Network Dengan Fitur Ekstraksi Local Binary Pattern Dan Gray Level Co-Occurrence Matrix. *ILKOM Jurnal Ilmiah*, *8*(3), 212–219.

Hutauruk, J. S. W., Matulatan, T., & Hayaty, N. (2020). Deteksi kendaraan secara real time menggunakan metode YOLO berbasis android. *Jurnal Sustainable: Jurnal Hasil Penelitian Dan Industri Terapan*, *9*(1), 8–14.

Jatmoko, C., Hartanto, D., Kurniawan, A. F., Rachmawanto, E. H., Sari, C. A., & Nilawati, F. E. (2020). Uji Implementasi Algoritma Viola-Jones Dalam Pengenalan Wajah. *Dinamik*, *25*(2), 68–76.

Nasution, S. W., & Kartika, K. (2022). Eggplant disease detection using Yolo algorithm telegram notified. *International Journal of Engineering, Science and Information Technology*, *2*(4), 127–132.

Prasetyawan, D. (2020). Penentuan Emosi pada Video dengan Convolutional Neural Network. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *5*(1), 23–35.

Putra, I. N. T. A., Kartini, K. S., Suyitno, Y. K., Sugiarta, I. M., & Puspita, N. K. E. (2023). Penerapan Library Tensorflow, Cvzone, dan Numpy pada Sistem Deteksi Bahasa Isyarat Secara Real Time. *Jurnal Krisnadana*, *2*(3), 412–423.

Redmon, J. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Rosiani, U. D., Choirina, P., Sumpeno, S., & Purnomo, M. H. (2018). Menuju Pengenalan Ekspresi Mikro: Pendeteksian Komponen Wajah Menggunakan Discriminative Response Map Fitting. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *7*(2), 204–211.

Wijaya, I. G. P. S., Firdaus, A. A., Dwitama, A. P. J., & Mustiari, M. (2018). Pengenalan Ekspresi Wajah Menggunakan DCT dan LDA untuk Aplikasi Pemutar Musik (MOODSIC). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *5*(5), 559–566