Eduvest – Journal of Universal Studies
Volume 4 Number 10, October, 2024
p- ISSN 2775-3735- e-ISSN 2775-3727

MODELING OF MASK DETECTION SYSTEMS, DISTANCE BETWEEN OBJECTS AND FACIAL RECOGNITION USING THE TINY-YOLOV4 METHOD, *CONVOLUTIONAL NEURAL NETWORK*, AND VIOLA JONES

Erby Virta Joseph Paays[1], Henry Candra
Universitas Trisakti, Jakarta, Indonesia[1,2]
Email: erbyvirtajp91@gmail.com

ABSTRACT

The development of technology is currently increasing very rapidly, the latest technology such as Face Detection and Face Recognition in public facilities has been widely applied, such as cameras for attendance, facial detection devices as mobile phone security and various other new innovations in the field of technology. The purpose of creating this model is to determine the distance, and the use of masks in public places and to determine the performance when the two algorithms between Tiny-YOLOV4 and Viola Jones are combined. This method uses CPU testing with a clock speed of 2.9 GHz and GPU with a clock speed of 1590 MHz with an accuracy level of 92.6% for mask object detection and 90.67% for face recognition with a maximum distance in front of the camera for human detection is 830 centimeters, mask detection 730 centimeters, and face recognition 530 centimeters. The results of the study produced an accuracy of 92.6% for mask detection and an accuracy of 90.67% for face recognition at maximum distances of 730 cm and 530 cm respectively. The conclusion shows that this system is effective for detecting masks, faces, and distance, providing a significant tool for monitoring compliance in public spaces. Implementing this system can serve as a preventive measure for potential future pandemics.

KEYWORDS *Tiny-YOLOV4, CNN, mask detection, distance detection, face recognition*

INTRODUCTION

In the last year, the whole world has been shocked by a new virus that is spreading rapidly known as the COVID-19 virus. It is known that at the beginning of 2020, the world was shocked by a new protein virus so that there is no antibiotic drug to treat the disease, namely Protein S, more familiarly called spike protein, which is often called Corona Virus or COVID 19 (Coronavirus Disease). The

How to cite: Erby Virta Joseph Paays, et.al. (2024). Modeling Of Mask Detection Systems, Distance Between Objects And Facial Recognition Using The Tiny-YOLOV4 Method, *Convolutional Neural Network*, And Viola Jones. *Journal Eduvest*. *4*(10), 9771-9790
E-ISSN: 2775-3727
Published by: https://eduvest.greenvest.co.id/

Coronavirus can certainly spread from human to human, which was initially only suspected to be an ordinary pneumonia disease, but as time goes by, the spread of COVID-19 cases is increasing. According to WHO data, as of March 2, 2020, the number of patients is 90,308 people infected with the coronavirus around the world, including Indonesia. The case in Indonesia began with a patient who had physical contact with a Foreign Citizen (WNA) from Japan living in Malaysia, after which the patient immediately complained of shortness of breath, cough, and fever (Yuliana, 2020).

In general, the most effective transmission of the Coronavirus between humans is droplets or fluids that are released when coughing or sneezing and that stick to surrounding objects (Mao et al., 2020). Massive human-to-human transmission. Fluids containing the Coronavirus that come out through coughing or sneezing can stick to a person's mouth or nose, then inhale when taking a breath and enter the lungs. Even the WHO (World Health Organization) announced that the coronavirus can spread through the air. These splashes are heavy, so they will fall not far from the source and can stick to objects and other surfaces around people, such as tables, doorknobs, and handrails. People can become infected by touching such objects or surfaces and then touching their eyes, nose, or mouth.

Technology that continues to develop can make it easier to control the spread of the COVID-19 virus, for example, the idea of this research uses CCTV (Closed Circuit Television) with a mixture of algorithms in it (Costa & Peixoto, 2020). The purpose of using algorithms in CCTV is to integrate the system so that it can recognize special objects such as humans, faces, and masks. As well as real-time distance parameters, so social distancing is easier to monitor.

Many researchers have discussed CCTV by integrating algorithms into the system but cannot recognize a person's face, so the model of this study will be different from previous researchers (Chellappa et al., 2010). With the title "*A deep learning-based social distancing monitoring framework for COVID-19*", the algorithm used is Yolo V3 with object detection only humans and distance between objects as parameters with an accuracy level of 95%, the test is carried out using datasets and *overhead views.* From this study, there is still false *detection,* and testing is carried out two times with pre-trained learning data and trained learning data (Ahmed et al., 2021).

"*Real-Time Social Distancing Detector Using Socialdistancing-19 Deep Learning Network*" in this study uses the Yolo V3 algorithm with real-*time* testing and *trained images*, images, and videos with an accuracy of 92.8% (Keniya & Mehendale, 2020).

"*Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques*" in this study compared Yolo V3, Faster R*CNN*, and Single Shot Detection, where Yolo V3 got the best results with FPS of 23, mAP 0.846, *Training Time* 5659 seconds, with *a total loss* of 0.87 testing using datasets and *overhead view* (Punn et al., 2020).

"*A Vision-based Social Distancing and Critical Density Detection System for COVID-19*" has used Yolo V4 with dataset testing and mAP overhead view, which is quite small, 0.41, and the time required is very short of 0.04 (Yang et al., 2021).

"*Deep Learning Implementation of Facemask and Physical Distancing Detection with Alarm Systems*" In this study, detecting distance and masks, the tests were carried out for those who wore masks and those who did not wear masks. For the test without a mask from the dataset, the result was 95.15% accuracy (Militante & Dionisio, 2020).

The algorithms used are YOLO V4 Tiny *CNN* algorithms and Viola Jones. Yolo V4 Tiny will function to detect objects, namely humans, as well as distance, assisted *by CNN* to detect faces and facial recognition, while Viola Jones will detect mask objects. The target of this study is to find out the development limitations of the use of the Yolo *CNN algorithm* because it is combined with other algorithms and also to know the level of accuracy that will be obtained at the distance parameter with the detection of other objects.

Although now Covid-19 has subsided and is no longer at risk due to vaccination from the government, social distancing research is still good to be done as one of the preventive measures in the event of a new pandemic, even in some countries Covid-19 is still an epidemic disease, and there are still many confirmed positive (Yadav et al., 2022). In Indonesia itself, there are still many daily cases confirmed positive for COVID-19, and there are still those who have died. Although vaccination does not seem to be the definitive answer to becoming immune to COVID-19, the implementation of social distancing remains the answer to the COVID-19 pandemic.

According to research (Loey et al., 2021) entitled A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. This study combines YOLOv2 and ResNet50 for real-time detection of face masks, achieving high accuracy in various lighting conditions and facial positions.

The advantage of this research is that it can detect a smaller object in an object detection result, in other words, it can create a *bounding box* in a *bounding box* using GPU and CPU specifications that are not too high but by utilizing the concept and efficiency of YOLOV4-Tiny. In addition, the testing in this study was carried out in *real-time*, not using existing datasets.

The purpose of modeling mask detection, distance, and facial recognition is as follows:

1. Modeling a system that can detect distance, masks, faces, and facial recognition
2. Simulate system modeling results
3. Analyze and evaluate the results of simulation and system testing

The benefits of this research are:

1. Help remind everyone to wear masks.
2. Help remind everyone to keep their distance.
3. Reducing the potential for the spread of the COVID-19 virus during the pandemic.
4. Generate models for face detection recognition with masks in *real-time.*

RESEARCH METHOD

Problem Identification

Finding a problem that is then raised as a topic and given a solution to the problem, where the ideas and ideas are outlined in the Thesis book.

Information collection

This is necessary because this research is based on a system that can find out *contact events* or *contact tracing* that occur at a location. Later, if someone is infected with the COVID-19 virus, their movement can be monitored, and it can be seen who does not apply *social distancing* with victims/patients during the incubation period.

Regulatory data

The regulatory data referred to here is for those who know about the regulations that apply at a work site regarding social distancing or the definition of contact tracing (Ferretti et al., 2020). Even if the design of this system is successful and wants to be applied to other locations, then at least the researcher and *the client* have 1 page about the definition of *contact tracing.*

Finding a Solution

The solution studied by researchers today is to utilize CCTV to monitor and ensure that everyone has adhered to contact tracing, especially in areas that are prone to crowds, such as eateries, gyms, smoking areas, and *minimarket areas.*

Learning the Theoretical Foundations

At this stage, researchers will look for information about previous research that has the same goal. Then, the researcher will study and develop or improve a variable in previous research, such as accuracy, component efficiency, and speed when *running* the system. At that point, researchers should have been able to confirm whether this research can be carried out or not. Researchers should already know what methods are needed, such as YOLO, *CNN*, and Viola Jones.

Software *and* hardware *design*

When you have a goal and study the research, the researcher should be able to determine what software and hardware can be used. In this study, the software used is Thonny as a Python IDE, and for hardware, at least there is a webcam or camera connected to a laptop or PC with a full HD resolution of 1080 to support the research, with the higher specifications of the camera and laptop, the faster the process of executing mask recognition and distance.

Making a plan

In this process, the researcher should have created coding in Pycharm and prepared *the hardware.*

Testing and Conclusion

Calibration

Calibration is an activity carried out to determine the analog threshold value in the system (*threshold*) as a basis for comparing the actual distance value with the distance read in the system because the system cannot recognize the distance value but can only detect analog and digital values (Winjaya, 2017). What the researcher expects is that when the *threshold value* has been obtained, there will be no more influence on the detection of the distance between objects to the distance between objects and the camera besides that, the researcher has also determined the height of the camera, which is 110 centimeters above the flat ground surface, the researcher will apply this to all types of tests, the calibration in this study is carried out in 2 ways, namely horizontal calibration as in Figure 1.

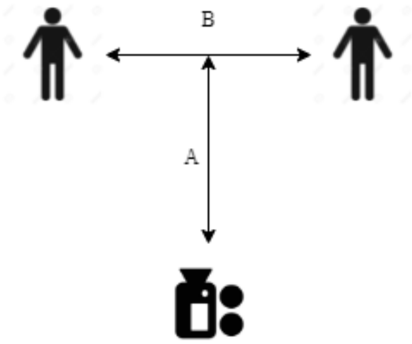

Figure 1. Horizontal Calibration

The diagonal calibration, as shown in Figure 1.

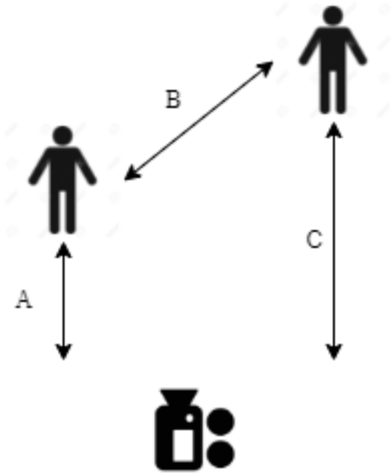

Figure 2. Diagonal Calibration

However, diagonal calibration is carried out when the horizontal calibration test has been successful in testing the threshold value obtained in the horizontal calibration when the state of the distance between the two objects is different from the camera. If the test results are still the same, then it can be concluded that the threshold value obtained is quite precise.

Comparison of actual distance with system result distance

In the test comparing the actual distance with the distance detected with the system, several steps are needed, namely by setting a minimum distance of 2 testing personnel as far as more than 2 meters (B), for example, 2 meters 50 centimeters.

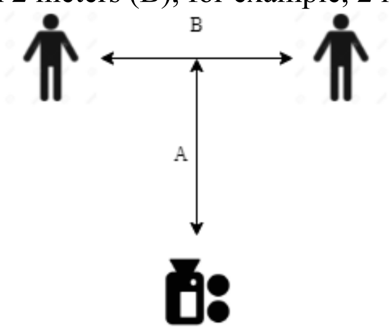

Figure 4. Distance Testing Scenarios

Then, the distance of the camera to the volunteer (A) will be changed until the error indicator (alarm) does not sound, which indicates the distance of the value *threshold,* which is the difference *pixels* set on the Manhattan distance method (in this study 100 *pixel*) is already true and no longer needs to be changed. After that, the distance between volunteers will be changed (B) from 0 meters (very far) to the farthest distance, which is the limit *Point of view* (POV) from the camera. If the alarm sounds when the distance between volunteers is more than 2 meters, there is an error or error, and the experiment must be retested from the earliest step. This test will also be tried by adding volunteers to 3 people, then the researcher will record whether there is a change in fps by adding volunteer personnel, if there is a change, it will be recorded by the researcher the fps value. By changing the A value, the researcher expects that the distance of the camera from the object will not affect the accuracy level of the method *Manhattan Distance* in work (Saleh et al., 2021).

A = Distance from the camera to the object
B = Distance between objects

Maximum distance testing for object detection

In this test, the target that the author wants to achieve is to find out the optimal distance or the farthest distance from human detection and mask detection. In the image, you will see a camera that detects volunteers with a distance (A) that will be changed further and further away, as shown in Figure 3.5. The researcher also wanted to find out if there was an effect of FPS on the distance changes made in this test.

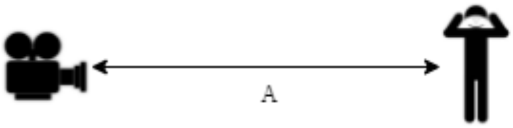

Figure 5. Maximum Distance Test Scenario on Object

During this test, the thing to make sure of is whether the volunteers use masks well and whether the cameras used in this study are good enough so that the detection of masks (small objects) will have the same maximum visibility (cutoff) as humans (large objects)

Maximum distance testing for facial recognition

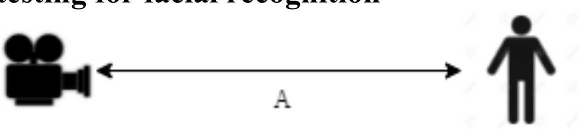

Figure 6. Facial Recognition Scenarios

This test will be more or less the same as the mask detection and face detection tests. However, there is something that needs to be ensured that the volunteer's face is already included in the system so that later, the camera can recognize the volunteer's face, as shown in Figure 3.6. The difference is that if volunteers no longer wear masks, then the distance (A) will be changed from very close to very far. The researchers wanted to find out the maximum distance from the camera that could recognize the volunteers' faces and how many fps they got.

Overall system testing

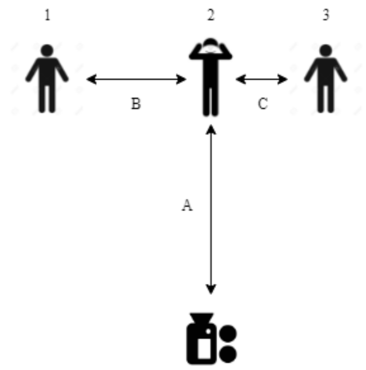

Figure 7. Error Indicator Testing Scenario

In Figure 3.7, the test will be carried out by arranging volunteers to line up horizontally as in the Figure, and each volunteer has entered their face data into the system. Later, there will be a minimum of 3 volunteers, and each volunteer makes mistakes in terms of *social distancing.* For example, not keeping a distance, not wearing a mask, or maybe both. The roles of each volunteer are as follows:

1. The first person will not wear a mask and still keep a distance from the second person (B value is more than 2 meters)
2. The second person will keep a distance from the first person (B value is more than 2 meters) and wear a mask, but not keep a distance from the third person (C value is less than 2 meters)
3. The third person did not keep a distance from the second person (The c value was less than 2 meters) and did not wear a mask.

Later, the researcher wants to listen to what kind of violation indicators will be sounded and how many fps will be obtained when a violation occurs, and there are three people in the *frame*.

Mask Detection Accuracy Testing

Researchers will test the system by trying to detect the medical masks at an optimal distance, in which case optimal means that the camera can detect masks and objects at the same time. The value of the distance between the object and the camera is 1 meter.

Facial Recognition Accuracy Testing

Researchers will test the system by trying the facial recognition feature on the system. Later, the dataset will be filled with several images of people's faces; then, there will be several people whose faces appear in 1 *GUI frame*, the researcher wants to know if the system can distinguish each person who appears in the *GUI*, and what if the faces that appear are indeed not in the dataset. This test was carried out at an optimal distance, and volunteers who appeared in the *frame* would not wear masks.

Hardware *and* Software *System Design*

The design of this system can be seen in Figure 3.8, where the system will receive input from a high-resolution Logitech C920 camera, which will then be processed with YOLO, Viola Jones, and *CNN* algorithms with Python programming, as discussed in the previous section. Then, the output of the results of the process is a camera that can distinguish between people wearing masks or not, the distance between each person (object), and facial recognition if there is an object that is not wearing a mask.

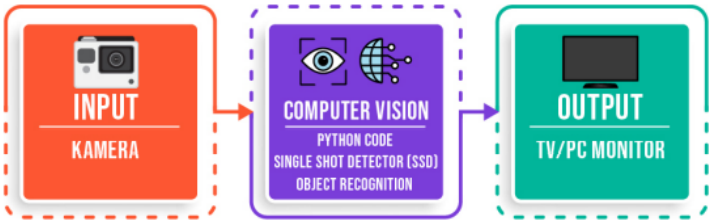

Figure 8. Hardware Design Model

The results of the pose will be displayed on the monitor so that everyone can see the results of the image management. After being seen by everyone, the distance and status of their mask use (mask/unmasked), there is a *speaker/buzzer* that will sound as an indicator to give directions if anyone does not wear a mask or does not maintain distance.

How the Overall System Works:

Image Input

This system is applied to an attraction where there is a camera that can detect the distance and mask on every person who enters the camera *frame*.

YOLO Process

The image captured by the camera is then processed to create *a bounding box that* aims to identify the object. Then, a box will appear around the object called *a bounding box. Bounding boxes* aim to find out the distance between objects through the difference *in pixels* between *bounding boxes*. *The output* of this process is that the system can distinguish which objects should be processed and which ones should not, and then it can know the distance between objects.

The Viola-Jones Process

Images that already have *a bounding box* will be further processed so that the camera can recognize which part of the object is the face of the object recognized by *the bounding box*. This is very important for the next process because, in the *CNN process*, a comparison of the facial images in the dataset with faces that are not in the dataset will be processed, the goal is to be able to distinguish recognized and unrecognized faces.

CNN Process

The *CNN* algorithm has a dataset that stores Images/photos, which contain the following:

Photos Without Masks with external photos

Photos of people without masks, left side, front, and right side. The more photos, the longer the training process on the dataset will be, and the more photos will also be compared by *CNN* during the facial recognition process. Examples of datasets in this study taken internally or externally by the camera can be seen as shown in Figures 9 and 10.

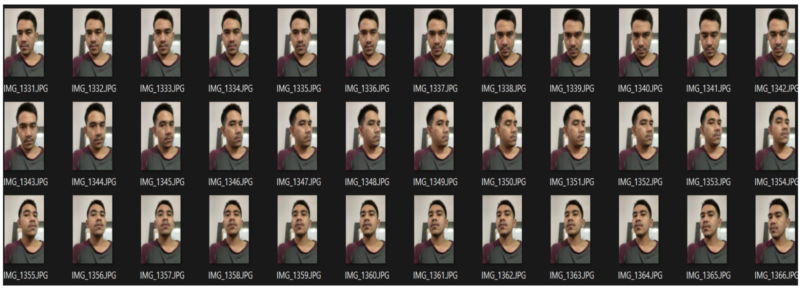

Figure 9. Faces on Dataset Without Masks *Input* External

Unmasked photo taken by webcam

Photos of people without masks will be taken at an angle of 45° to the left and right and front view. The results of the image input will be directly entered into the dataset and grayscale using a webcam, as shown in Figure 3.4.

Figure 10. Faces on Dataset Without Masks *Input* Webcam

At this stage, *the expected output* of the researcher is that the system can distinguish faces that should be recognized (stored in the dataset) from face data that is not recognized (not in the dataset)

***Image Detection* Output**

The output of this detection system is in the form of an image that has been given *a bounding box* as an indicator for each person detected along with their distance(Mohan et al., 2001).

RESULT AND DISCUSSION

Results of mask and object detection

In this study, to detect masks and objects, they successively employed the results of training data from Viola Jones and YOLO V4 Tiny. Viola Jones by taking the haar cascade method for mask detection while YOLO V4 Tiny will detect moving objects in this case only humans are detected, the essence of using these two methods is to separate the object from the *background*. The results of mask and object detection on the *GUI* can be seen in Figure 13.

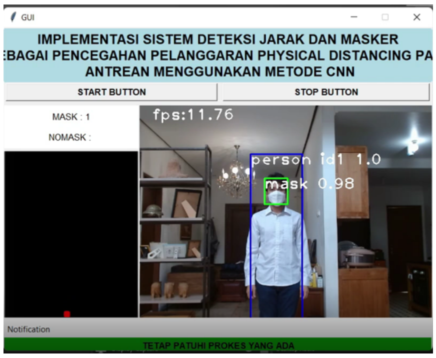

Figure 13. Results of Detection of Masks and Objects on *GUI*

Facial recognition results

This study is also equipped with facial recognition using *the CNN* method. When *the GUI* is running, the camera will take an image and *the CNN* method will continue to detect faces, when there are faces detected the system will compare the faces that appear in the *frame*, with the faces in the dataset. If the face is recognized in the data set, a *bounding box* will appear with the name of the face detected by *CNN*, but if the results of the comparison process with the dataset do not match, a *bounding box* will appear on the face with the word "Unknown" which means it is not recognized by the system. In addition, the alarm will sound by mentioning the name of the person who does not wear a mask, and reminding the person to wear a mask, the alarm will not stop until the person is wearing a mask even if the person is not recognized by the system". The results of the process on the *GUI* can be seen in Figure and Figure 14.

Figure 14. Correct Object Detection and Face Recognition Results

Image the result of the facial recognition detection correctly on the *GUI*. If the face is recognizable, the name of the violator will appear on the NOMASK table as shown in Figure 15.

Figure 15. Examples of Failed Face Recognition

Distance detection results

In this study, the camera can also detect the distance between people by utilizing analog values measured if 2 or more objects are detected in 1 *frame*. Each object will be marked with 1 red dot which will be measured by the value of the distance between the red dots. If the analog value of the distance detection result is less than the *threshold value*, the alarm will sound and a warning "DON'T FORGET TO KEEP THE DISTANCE!!" will appear on the *GUI*. The results of the distance test in this study can be seen in Figure 16.

Figure 16. Display *GUI* On Distance Detection Testing

Test results

The testing stages are carried out several times starting from determining the analog value which will be the *threshold value* in the distance test, the calibration generator is carried out 2 times, namely horizontal calibration testing and diagonal calibration, then continued with object detection testing to find out the maximum distance from object detection (humans and masks) then the maximum distance from the face recognition, then continued with testing the comparison of the actual distance with the results analog values detected from the computational results of the YOLOV4-Tiny algorithm, and the latter tests the accuracy of object detection and facial recognition accuracy

Calibration

Calibration is one of the tests to determine the analog threshold value which will be the maximum value of the distance detection results read by the YOLO method. Of course, the *threshold* value will not be the same as the actual distance value because 1 analog value does not mean 1 centimeter. In social *distancing,* everyone must maintain at least 2 meters or more to reduce the potential for transmission of the covid-19 virus, so in this test an analog value will be sought that is equal to 2 meters, if the distance of 2 meters is met, the YOLO method will be able to distinguish the distance between objects which then YOLO will trigger an alarm to be active if the result of the analog value is lower than the analog threshold value.

The test for calibration is divided into two according to the position of the person in front of the camera. The following is a threshold test table for horizontal calibration whose results can be seen in table 1.

Table 1. Horizontal Calibration Testing

It	*Threshold*	A (meter)	B (meter)	Alarm	Information
1	100	2	0.5	On	Right
2	100	2	1	Off	Wrong
3	100	2	1.5	Off	Wrong
4	100	2	2	Off	Wrong
5	200	2	0.5	On	Right
6	200	2	1	On	Right
7	200	2	1.5	On	Right
8	200	2	1.6	On	Right
9	200	2	1.7	On	Right
10	200	2	2	On	Wrong
11	150	2	0.5	On	Right
12	150	2	1	On	Right
13	150	2	1.5	On	Right
14	150	2	2	Off	Right
15	150	2	1.9	On	Right
16	150	2	1.8	On	Right
17	150	2.8	0.5	On	Right
18	150	2.8	1	On	Right
19	150	2.8	1.5	On	Right
20	150	2.8	2	Off	Right
21	150	2.8	1.9	On	Right
22	150	2.8	1.8	On	Right
23	150	1.5	0.5	On	Right
24	150	1.5	1	On	Right
25	150	1.5	1.5	On	Right
26	150	1.5	2	Off	Right
27	150	1.5	1.9	On	Right
28	150	1.5	1.8	On	Right

From the horizontal calibration test table above, it can be concluded that the analog threshold *value* obtained is 150 to be able to detect a distance of 2 meters. Researchers have also changed the distance between the object and the camera and the results are consistent and sensitive when the distance between the objects is 1.9 meters, the analog value will decrease below 150 and the alarm is active. Next is the test when the object is diagonally and the results can be seen in table 1. The intention in the Description column is to indicate that the alarm sounds according to the desired expectations or not, the researcher's hope in this section is that the alarm sounds when the distance is less than 2 meters and does not sound when the distance between objects is more than 2 meters. When *the threshold* value is 200, at a distance of 2 meters the alarm still does not sound, this means that the value of 200 is still too large so the researcher tries with a *threshold value* of 100, with that value the alarm sounds only at a distance of 50 centimeters between objects, and does not sound when the object is more than 1 meter away, it means the *threshold* valueIt was still too small so the researcher made the threshold value to 150 and the GUI succeeded in distinguishing the distance of 50 centimeters to 190 centimeters in the condition of an alarm sounding, and a distance of 200 centimeters when the alarm did not sound.

Table 2. Diagonal Calibration Testing

It	*Threshold*	A (meter)	C (meter)	B (meter)	Alarm	Information
1	150	1.2	2.8	0.5	On	Right
2	150	1.2	2.8	1	On	Right
3	150	1.2	2.8	1.5	On	Right
4	150	1.2	2.8	1.8	On	Right
5	150	1.2	2.8	1.9	On	Right
6	150	1.2	2.8	2	Off	Right

In this test, the researcher equalized the analog threshold values that had been obtained in the horizontal calibration test. From the results of the test table above, it can be concluded that, YOLO can distinguish objects that are closer and farther away from the camera and the distance value between objects also remains sensitive, as long as the analog threshold value is below 150 then the alarm will sound and the analog threshold value will be above 150 when the actual distance is ≥ 2 meters. With A, B, C in a row is the distance between the first person and the camera, B is the distance between people, and C is the distance of the second person with the camera.

Distance Detection and Alarm Testing

In the distance detection test, the *threshold value* that has been obtained can be calibrated test, which is 150. This test is to find out the analog value that appears when the distance is changed and to find out the alarm response that should sound when the analog value that appears is lower than 150. The results of the distance detection and alarm tests can be seen in table 4.3

Table 3. Distance Detection and Alarm Testing

It	*Threshold*	A (meter)	B (meter)	Alarm	Analog Values Detected	Information
1	150	2	0 (very close)	On	80-99	Right
2	150	2	1≤ B ≤1.9	On	100-140	Right
3	150	2	≥2	Off	160-200	Right
4	150	2.8	0 (very close)	On	80-99	Right
5	150	2.8	1≤ B ≤1.9	On	100-140	Right
6	150	2.8	≥2	Off	160-200	Right
7	150	1.5	0 (very close)	On	80-99	Right
8	150	1.5	1≤ B ≤1.9	On	100-140	Right
9	150	1.5	≥2	Off	160-200	Right

In this test, if the actual distance is less than 2 meters, the alarm will sound "Queue number 2 and 1 do not get too close" in the column of Analog values detected, the values vary and fluctuate because the test is carried out in *real time* and the object is detected moving.

Mask and Object Detection Testing

In the test to be carried out with 1 volunteer wearing a mask, the researcher wanted to know the maximum distance from the detection of the mask and the object from the Logitech C920 camera used in this study, and how sensitive the YOLO method to detect objects and also Viola Jones to detect masks. The results of this test can be seen in table 4.

Table 4. Results of Maximum Distance Testing of Mask and Object Detection

No.	Distance (centimeter)	Test 1		Test 2		Test 3	
		Mask	Object	Mask	Object	Mask	Object
1	100	✓	✓	✓	✓	✓	✓
2	200	✓	✓	✓	✓	✓	✓
3	300	✓	✓	✓	✓	✓	✓
4	400	✓	✓	✓	✓	✓	✓
5	500	✓	✓	✓	✓	✓	✓
6	600	✓	✓	✓	✓	✓	✓
7	700	✓	✓	✓	✓	✓	✓
8	730	✓	✓	✓	✓	✓	✓
9	760	×	✓	✓	✓	✓	✓
10	790	×	✓	×	✓	×	✓
11	800	×	✓	×	✓	×	✓
12	830	×	✓	×	✓	×	✓
13	860	×	×	×	✓	×	×
14	890	×	×	×	×	×	×
15	900	×	×	×	×	×	×
16	1000	×	×	×	×	×	×

Information:

✓ : Object detected

× : Object not detected

In this test, it can be seen that for the maximum distance to detect masks and objects that are not the same, the researcher conducted a sensitivity test on mask detection at a distance of 7 meters from the camera, because at a distance of 8 meters or more from the camera the mask can no longer be detected, therefore the researcher added a test at a distance of 730-790 cm from the camera. It can be seen that the camera can still detect masks consistently at a distance of 730 cm from the camera. At a distance of 760 cm, the camera can still detect the presence of a mask but is no longer consistent because from 3 tests at a distance of 760 cm from the camera, only 1 mask is detected. At a distance of 790 cm from the camera, the mask can no longer be detected by the camera (*GUI).*

The researcher then continued for the detection of humans (objects), the researcher also conducted a sensitivity test on object detection at a distance of 8 meters from the camera, because at 9 meters or more the camera could not detect objects anymore, therefore the researcher added a test at a distance of 830-890 cm from the camera. It can be seen that the camera can still detect objects consistently at a distance of 830 cm from the camera, while at a distance of 860 cm, out of 3 tests the object is detected only 1 time, and at a distance of 890 cm the object can no longer be detected by the camera (*GUI*).

Facial Recognition Testing

The facial recognition test was carried out with 1 volunteer and did not wear a mask. Researchers wanted to know the maximum distance from the Logitech C920 camera and *the CNN method* for recognizing faces. The number of datasets for 1 user is 400 so that what is expected is always accurate for the results of facial recognition. The results of the facial recognition test can be seen in table 5.

Table 5. Facial Recognition Test Results

Distance (centimeter)	Test 1		Test 2		Test 3	
	Resul t	Trust Score	Result	Trust Score	Result	Trust Score
100	✓	1	✓	0.99	✓	1
200	✓	0.99	✓	0.99	✓	0.99
300	✓	0.99	✓	0.95	✓	0.96
400	✓	0.91	✓	0.94	✓	0.87
500	✓	0.88	✓	0.81	✓	0.78
530	✓	0.81	✓	0.77	✓	0.71
560	✓	0.77	✓	0.73	×	-
590	×	-	×	-	×	-
600	×	-	×	-	×	-
700	×	-	×	-	×	-
800	×	-	×	-	×	-
900	×	-	×	-	×	-
1000	×	-	×	-	×	-

In the facial recognition test, the researcher conducted a sensitivity test at a distance of 5 meters, because at a distance of 6 meters the face could no longer be recognized. The researcher added that the test was at a distance of 530-590 cm. And the results are that the camera is still consistent at a distance of 530 cm to recognize faces, but at a distance of 560 cm from 3 tests, only 2 times can recognize faces well, while at 590 cm the face can no longer be recognized by the camera (*GUI).* The researcher also recorded the value of the confidence score that appeared on the *bounding box* when the camera managed to recognize the face that appeared, if you pay attention to the value of the trust score decreases along with the distance between the object and the camera.

Accuracy Testing

The accuracy test in this study is a test that intends to find out how accurate the use of the YOLO V4 tiny method is when combined with other algorithm methods, in this case the researcher intends of course with *CNN* and Viola Jones. The accuracy test was carried out with the optimal distance between the object and the camera, which is 50-200 cm. The distance of 50 centimeters also means that the quality factor of the camera used is expected not to affect the accuracy value of this study too much. The accuracy tests to be tested are mask detection accuracy testing and facial recognition accuracy.

Actual in front of the camera			
GUI Predictions	(+,-)		
		19	2
		0	6

Figure 24. Result *Confusion Matrix* Testing for Correct Mask Detection

So with the *Confusion Matrix* above, the values of accuracy, precision, *recall, f-1 score,* and *error* can be calculated using equations 2.2 to 2.6, then the calculation results are as follows

1. $\text{Akurasi} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} x\ 100\% = \frac{19+6}{19+2+0+6} x 100\% = 92.6\%$
2. $\text{Presisi} = \frac{\text{TP}}{\text{TP+FP}} = \frac{19}{19+2} = 0.905$
3. $\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} = \frac{19}{19+0} = 1$
4. $\text{F} - 1\ \text{Score} = \frac{2\text{ x Recall x Presisi}}{\text{Recall+Presisi}} = \frac{2x1x0.905}{1+0.905} = 0.95$
5. $\text{Error} = \frac{\text{FP}}{\text{TP}} \text{x}100\% = \frac{2}{19} x 100\% = 10.52\%$

With the accuracy value above, it is much better than testing the use of masks incorrectly. With 92.6% already showing a good number even with modest hardware, and a smaller error rate of 10.52%.

Facial Recognition Accuracy Testing

Table 7. Faces on the dataset

No.	User.id	Name	Photos on datasets	How to Input
1	User.1	V		Webcam
2	User.2	W		External Photos

The following is table 7. of the results of the Face Recognition Accuracy Test with a distance of 1-2 meters and displaying several people at once in *a frame*

Table 8. Facial Recognition Accuracy Test Results

Faces that appear (*frames*)	Distance (meters)	Test 1	Test 2	Test 3

V	1	V	X	V
W	1	W	W	W
X	1	X	X	X
Y	1	Y	Y	Y
Z	1	Unknown	Unknown	Unknown
Z	1	Z	Z	Z
V and X	1	V and X	V and X	V and X
V and Y	1	V and Y	V and Y	V and Y
V&W	1	V&W	V&W	V&W
V	2	V	V	V
W	2	W	W	W
X	2	X	X	X
Y	2	Y	Y	Y
Unknown	2	W	W	Unknown
V and X	2	V and X	V and X	V and X
V and Y	2	V and Y	V and Y	V and Y
V&W	2	V&W	V&W	V&W
V & Unknown	2	V&W	V&W	V & Unknown
V, X, and Unknown	2	V, X, & W	V, X, & W	V, X, & Unknown

Unknown means an unrecognized face, which means the face is not in the dataset. Faces can be misdetected or indeed not in the dataset. In the table of the results of the facial recognition accuracy test above, it can be concluded that there is a decrease in the accuracy level when the distance is added, especially on face objects that are not recognized in the dataset. So that this will decrease the overall accuracy value.

***Confusion Matrix* Face Recognition**

In this study, the *Confusion Matrix* was used to determine the accuracy value of the facial recognition test, and determine how reliable the system has been created by knowing other parameters. The results of the *Confusion Matrix* referring to table 8 can be seen in Figure 24.

Actual in front of the camera			
GUI Predictions	(+,-)		
		68	7
		0	0

Figure 25. *Confusion Matrix* On Facial Recognition Testing

By looking at the results *of the Confusion Matrix* above, it can be concluded that the true positive value is 68 which means that the *GUI* prediction results are in accordance with the face that actually appears in front of the camera, and the false positive value is 7 which means *that the GUI* prediction results do not match the face whose role appears in front of the camera. So by knowing these values, the values of accuracy, precision, recall, F-1 Score and error can be calculated using equations 6.2 to 6.6, then the calculation results are as follows:

1. $\text{Akurasi} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} x\ 100\%$ ==90.67% $\frac{68+0}{68+7+0+0} x\ 100\ \%$
2. $\text{Presisi} = \frac{\text{TP}}{\text{TP+FP}} = \frac{68}{68+7} = 0.9067$
3. $\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} = \frac{68}{68+0} = 1$
4. $\text{F}-1\ \text{Score} = \frac{2\text{ x Recall x Presisi}}{\text{Recall+Presisi}} = \frac{2x0.9x1}{1+0.9} = 0.95$
5. $\text{Error} = \frac{\text{FP}}{\text{TP}}\text{x}100\% = \frac{7}{68}x100\% = 10.3\%$

Best Scenarios For System Usage

Based on the results of the tests that have been carried out, and after knowing some of the limitations that this system has. Researchers hope that the system that has been created can be applied to various conditions such as class attendance that applies covid-19 protocols such as the use of masks and social distancing or maybe it can be added to open access to certain rooms that require facial recognition as the key (Yusriani, 2022).

CONCLUSION

Based on the results of the tests that have been carried out, it can be concluded that the modeling for object detection, distance, mask, and facial recognition by combining the YOLO V4 Tiny, Viola Jones, and *CNN* methods combined with hardware in the form of a Logitech C90 camera and a 1660ti GPU (80 watts) is as follows:

1. YOLO V4 Tiny, Viola Jones, and *CNN* managed to combine as a single system for object detection, distance, masks, and facial recognition.
2. To be able to detect the distance properly, YOLO V4 Tiny needs an analog value as an approximate threshold value to be able to determine whether the alarm needs to be activated or not, as one of the social distancing rules, namely complying with a maximum distance of 2 meters, and the analog threshold value is 150.
3. Researchers managed to activate the alarm if there is a social distancing violation, in this study only includes keeping a distance of 2 meters and using masks
4. The maximum distance for mask detection in this study was 730 centimeters with an *error margin* value of 30 centimeters, while the maximum distance for object detection (humans) in this study was 830 centimeters with a *margin of error* of 30 centimeters.
5. The maximum distance for facial recognition in this study was 530 centimeters with a *margin of error value* of 30 centimeters.
6. The accuracy value of mask detection is 92.6% respectively.

7. The accuracy value of facial recognition is 1-2 meters with how many faces are recognized or unrecognized in the dataset and the number of faces detected in 1 *frame* is 90.67%.

REFERENCES

Ahmed, I., Ahmad, M., Rodrigues, J. J. P. C., Jeon, G., & Din, S. (2021). A deep learning-based social distance monitoring framework for COVID-19. *Sustainable Cities and Society*, *65*, 102571.

Chellappa, R., Sinha, P., & Phillips, P. J. (2010). Face recognition by computers and humans. *Computer*, *43*(2), 46–55.

Costa, D. G., & Peixoto, J. P. J. (2020). COVID-19 pandemic: a review of smart cities initiatives to face new outbreaks. *IET Smart Cities*, *2*(2), 64–73.

Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., & Fraser, C. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, *368*(6491), eabb6936.

Keniya, R., & Mehendale, N. (2020). Real-time social distancing detector using socialdistancingnet-19 deep learning network. *Available at SSRN 3669311*.

Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*, *167*, 108288.

Mao, N., An, C. K., Guo, L. Y., Wang, M., Guo, L., Guo, S. R., & Long, E. S. (2020). Transmission risk of infectious droplets in physical spreading process at different times: a review. *Building and Environment*, *185*, 107307.

Militante, S. V, & Dionisio, N. V. (2020). Deep learning implementation of facemask and physical distancing detection with alarm systems. *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, 1–5.

Mohan, A., Papageorgiou, C., & Poggio, T. (2001). Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(4), 349–361.

Punn, N. S., Sonbhadra, S. K., Agarwal, S., & Rai, G. (2020). Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques. *ArXiv Preprint ArXiv:2005.01385*.

Saleh, A., Sibero, A. F. K., & Manurung, I. H. G. (2021). Pengenalan Tanaman Herbal Menggunakan Algoritma Learning Vector Quantization Dan Manhattan Distance. *JURNAL TEKNOLOGI KESEHATAN DAN ILMU SOSIAL (TEKESNOS)*, *3*(2), 271–276.

Winjaya, F. (2017). Rancang bangun mesin pemanggang biji kopi berbasis image processing dan akustik. *Department of Electrical Engineering, Faculty of Industrial Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*.

Yadav, S., Gulia, P., Gill, N. S., & Chatterjee, J. M. (2022). [Retracted] A Real-Time Crowd Monitoring and Management System for Social Distance Classification and Healthcare Using Deep Learning. *Journal of Healthcare Engineering*, *2022*(1), 2130172.

Yang, D., Yurtsever, E., Renganathan, V., Redmill, K. A., & Özgüner, Ü. (2021). A vision-based social distancing and critical density detection system for COVID-19. *Sensors*, *21*(13), 4608.

Yuliana. (2020). Corona virus diseases (Covid -19); Sebuah tinjauan literatur. *Wellness and Healthy Magazine*, *2*(1), 187–192.

Yusriani, Y. (2022). *Monograf Peran Kader Kesehatan Dalam Meningkatkan Imunitas Tubuh Ibu Hamil Sebagai Upaya Pencegahan Covid-19*. Tahta Media Group.