Realtime Detection of Masks and Distance in an Effort to Control Physical Distancing based on Faster R-CNN

Agus Khumaidi, Intan Puspita Sari, Joko Endrasmono and Ryan Yudha Adhitya
Department of Marine Elctrical Engineering, Politeknik Perkapalan Negeri Surabaya, Surabaya, Indonesia

Keywords: Physical Distancing, Distance, Mask, Faster Regional Convolutional Neural Network (Faster R-CNN), Euclidean Distance, Object Tracking.

Abstract: According to the Circular Letter of the Governor of East Java in 2020 concerning Control, Supervision, and Law Enforcement in the Implementation of Large-Scale Social Restrictions in East Java in point 1b, it explains that everyone is required to wear a mask and maintain a distance of at least 1 meter when in outside the home, while point 2b explains that the person in charge of a restaurant / restaurant / similar business is obliged to maintain a distance in the queue of at least 1 meter between customers. In this study, The Faster R-CNN method has been applied to classify objects; people, masks, and no masks. Before the classification process is carried out, the dataset is collected and trained first. This classification applies to the queuing conditions in the room. From the results of real-time trials, the success of the model when classifying objects in the form of masks, no masks, and people has an average success of 92,67% with a safe detection distance of 400 cm. Based on the tests that have been carried out, the distance calculation using Euclidean Distance produces an average error of 4,591 % with the largest distance error reaching 7,32 cm.

1 INTRODUCTION

Humans are actually social creatures who always need the help and presence of others. However, the COVID-19 Virus requires each individual to wear a mask and perform Physical Distancing by keeping a distance of more than 1 meter from anyone to reduce the risk of spreading the virus.

According to the 2020 East Java Governor's Circular on Control, Supervision and Law Enforcement in the Implementation of Large-Scale Social Restrictions in East Java, point 1b explains that everyone is required to wear a mask and maintain a distance of at least 1 meter at all times. Outside the home, while point 2b explains that the person in charge of a restaurant / similar business is obliged to maintain a distance in the queue of at least 1 meter between customers. However, the rules for maintaining a safe distance and the use of masks are often violated in the application of Physical Distancing, especially in crowd locations such as queues at malls and restaurants(Timur, 2020).

With the background of these problems, the authors have innovations to overcome Physical Distancing violations, namely by implementing a Distance Detection System and Masks as Prevention of Physical Distancing Violations in Queues Using the Faster R-CNN Method. This study uses a camera as a sensor that functions like the human eye. Then used video processing to detect objects using OpenCV. With the use of OpenCV, videos can be processed in real-time and can be classified into several objects using the Faster R-CNN Method (Salim, 2020). And can predict the distance between human objects using the Euclidean Distance measurement method (Nishom, 2019).

2 METHODOLOGY

2.1 Identification of Problems

In this system, the problem raised is an effort to reduce violations in Physical Distancing. The purpose of this research is to reduce the spread and risk of being exposed to the Covid-19 virus. The problem formulation of this research is how the system can detect people and objects using Faster R-CNN and efficiently estimate the distance between objects using Euclidean Distance.

Khumaidi, A., Sari, I., Endrasmono, J. and Adhitya, R.
Realtime Detection of Masks and Distance in an Effort to Control Physical Distancing based on Faster R-CNN.
DOI: 10.5220/0010958400003260
In *Proceedings of the 4th International Conference on Applied Science and Technology on Engineering Science (iCAST-ES 2021)*, pages 1034-1038
ISBN: 978-989-758-615-6; ISSN: 2975-8246

2.2 Study of Literature

At this stage, the authors seek as much information as possible about the concepts that will be used in the study. The search was carried out related to information about the Convolutional Neural Network (Adhitya et al., 2020), Faster R-CNN, Centroid Tracking, Euclidean Distance, OpenCV (Rinanto & Khumaidi, 2015). It is hoped that this information will be able to support the completion of this research.

2.3 System Planning

System design is the stage that is used to provide an overview of the system used in research. This research uses equipment, namely a notebook with GPU (Ren et al., 2016), Logitech C922 Pro webcam with 640 x 480 camera resolution as image input, and speaker as sound notification output. The system flowchart in this study is shown in Figure 1.

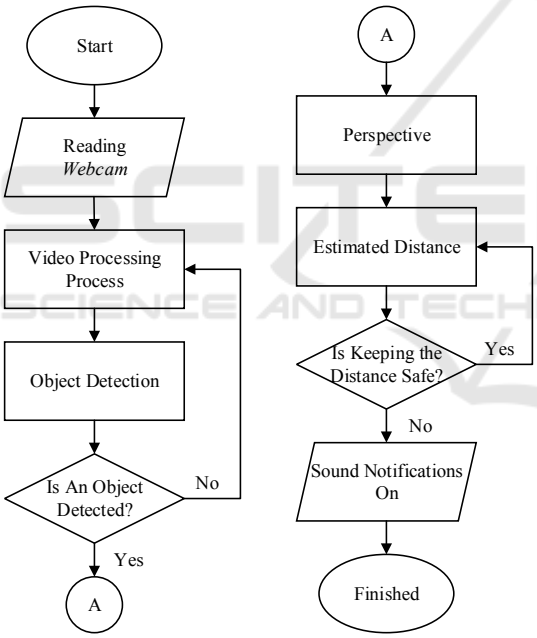


Figure 1: System Flowchart.

2.4 Hardware Design

Hardware design aims to provide an overview of the tools used. In Figure 2, is the application of the tool in the queuing room where humans who are objects will be detected, webcams that function to capture images are placed at the position of the upper end of the room, with the aim of being able to reach a wider area of a room. If the "people" is too close or there is a "people" that is not wearing a mask, then the speaker will issue a sound notification.

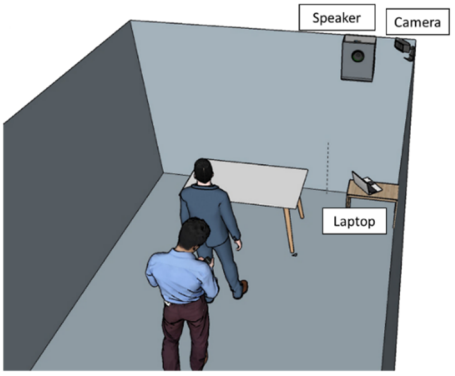


Figure 2: Hardware Design.

2.5 Software Design

In this research, the system uses the Faster R-CNN method. The steps taken to detect objects are as shown in Figure 3.

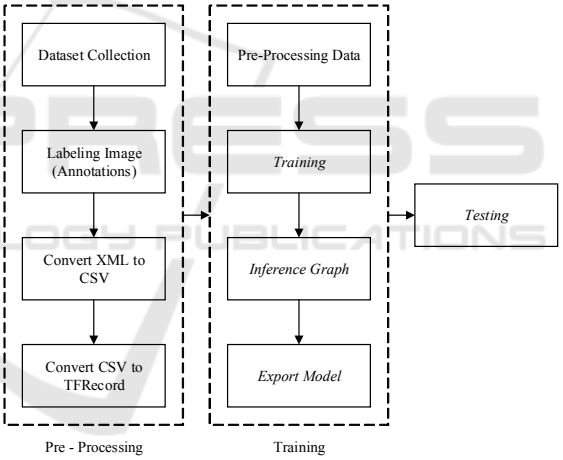


Figure 3: Software Design.

2.6 Centroid Euclidean Distance

Centroid and Euclidean Distance are used to determine the distance between objects by determining the coordinates of the Centroid value as shown in Figure 5 from each bounding box in each frame using equation 1.

$$D_{ij} = \sqrt{(X_j - X_i)^2 + (Y_j - Y_i)^2} \qquad (1)$$

description:

D : Distance result

X_{ij} : Distance x is measured from the distance measurement based on pixel value variation

Y_{ij} : Distance y is measured from the distance measurement based on pixel value variation

3 RESULT

3.1 Dataset

The dataset was obtained from the collection of photos with the provision of faces using masks, without masks, and people in a standing position, the number of datasets used were 1580 images. The dataset is then converted into an XML file and divided into 2 parts, namely 80% train data and 20% test data. Figure 4 is a few samples from the dataset used.

Figure 4: Dataset Class (A) Mask (B) No Mask (C) People.

3.2 Inference Graph Tensorboard

Tensorboard is used because the neural network is a process known as a black box, so it cannot be observed in detail what processes occur in the neural network system. Training process is carried out until step 200,000 and generate total loss below 0,015. The total Loss graph is shown in Figure 5.

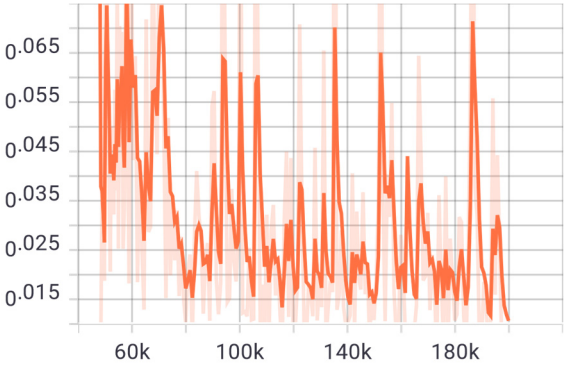
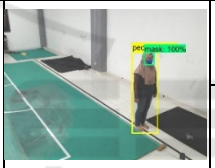
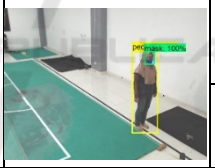


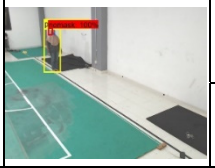
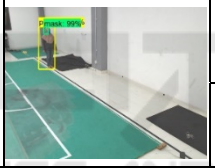
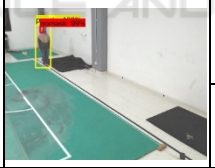
Figure 5: Total Loss Graph.

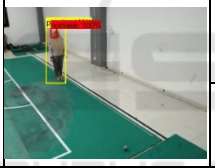
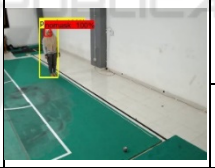
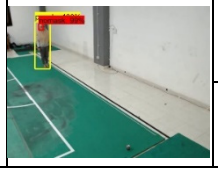
3.3 Testing Object Detection

Detection is divided into 3 according to the predetermined class, namely Mask, No mask, and People. The test was carried out with the camera position being 220 cm from the floor and the first object at a distance of 347.27 cm from the camera.

Table 1: Testing Object Detection (24 data from 140).

No	Dis-tance	Prediction Image	Results	State ment
1	0 cm		Mask	True
			People	True
2	50 cm		Mask	True
			People	True
3	100 cm		Mask	True
			People	True
4	150 cm		Mask	True
			People	True
5	200 cm		Mask	True
			People	True
6	250 cm		Mask	True
			People	True
7	300 cm		Mask	True
			People	True

No	Dis-tance	Prediction Image	Results	State ment
8	350 cm		Mask	True
			People	True
9	400 cm		Mask	True
			People	True
10	450 cm		Mask	True
			People	True
11	500 cm		No mask	False
			People	True
12	550 cm		Mask	True
			People	True
13	600 cm		No mask	False
			People	True
14	0 cm		No mask	True
			People	True
15	50 cm		No mask	True
			People	True
16	100 cm		No mask	True
			People	True

No	Dis-tance	Prediction Image	Results	State ment
17	150 cm		No mask	True
			People	True
18	200 cm		No mask	True
			People	True
19	250 cm		No mask	True
			People	True
20	300 cm		No mask	True
			People	True
21	350 cm		No mask	True
			People	True
22	400 cm		No mask	True
			People	True
:	:	:	:	:
			:	:
140	500 cm		No mask	True
			People	True

Based on the test data in Table 1 from 140 data, the percentage of detection success was 92.67%.

3.4 Testing Object Tracking

In object tracking testing, the system can perform object tracking in the form of object IDs precisely

according to the existing queue conditions, the test results are as shown in Figure 6.

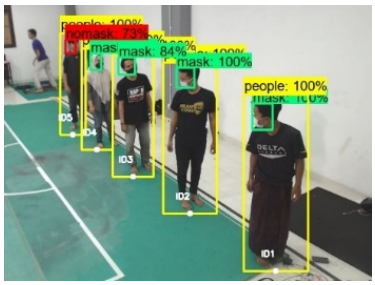


Figure 6: Object Tracking.

3.5 Testing Distance Object

Distance measurement is done using the perspective model. Then, for distance estimation, Euclidian Distance calculation is used to get the distance between "People" objects. Based on the test results as shown in Figure 7 using 12 distance data, the largest error is 1.86% with a distance difference of 0.93 cm.

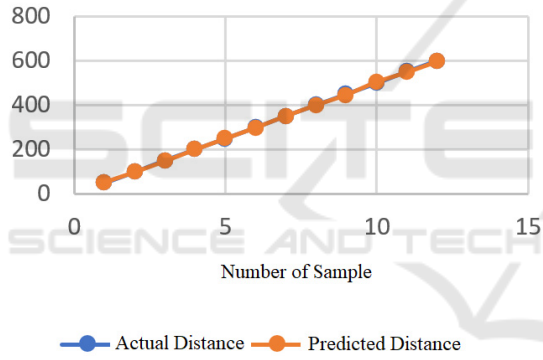


Figure 7: Distance Testing Graph.

4 CONCLUSION

From the testing that has been done, the results of this research can be concluded that:

1. Based on the test results from 140 data, the success of the system when classifying objects in the form of masks, no masks, and people has an average success of 92.67% with a safe detection distance of 400 cm.
2. Based on the tests that have been carried out, the distance calculation using the Euclidean Distance calculation produces an average error of 4.591 % with the largest distance error reaching 7.32 cm.

REFERENCES

Adhitya, R. Y., Khumaidi, A., Sarena, S. T., Kautsar, S., Widiawan, B., & Afriansyah, F. L. (2020). Applied Haar Cascade and Convolution Neural Network for Detecting Defects in the PCB Pathway. *CENIM 2020 - Proceeding: International Conference on Computer Engineering, Network, and Intelligent Multimedia 2020*, *Cenim*, 408–411. https://doi.org/10.1109/CENIM51130.2020.9297996

Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, *4*(1), 20–24. https://doi.org/10.30591/jpit.v4i1.1253

Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Rinanto, N., & Khumaidi, A. (2015). Aplikasi Android-Raspberry Pi Pada Kapal Tanpa Awak Untuk Pencarian Korban Kecelakaan Laut. *Jurnal SISFO : Inspirasi Profesional Sistem Informasi*, *05*(04). https://doi.org/10.24089/j.sisfo.2015.09.003

Salim, A. (2020). *Estimasi Kecepatan Kendaraan Melalui Video Pengawas Lalu Lintas Menggunakan Parallel Line Model*.

Timur, G. J. (2020). *Surat Edaran Pengendalian, Pengawasan Dan Penegakan Hukum Dalam Pelaksanaan Pembatasan Sosial Berskala Besar Di Jawa Timur*. http://files.bpbd.jatimprov.go.id/KEDARURATAN/COVID19/PRODUK HUKUM DAN KEBIJAKAN/PEMERINTAH PROVINSI/SEGUB PSBB.pdf

A. Khumaidi, E. M. Yuniarno and M. H. Purnomo,. (2017). "Welding defect classification based on convolution neural network (CNN) and Gaussian kernel," 2017 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, 2017, pp. 261-265, doi: 10.1109/ISITIA.2017.8124091

A. Budianto et al. (2017). "Analysis of artificial intelligence application using back propagation neural network and fuzzy logic controller on wall-following autonomous mobile robot," 2017 International Symposium on Electronics and Smart Devices (ISESD), Yogyakarta, pp. 62-66, doi: 10.1109/ISESD.2017.8253306

R. Y. Putra et al.(2016). "Neural network implementation for invers kinematic model of arm drawing robot," 2016 International Symposium on Electronics and Smart Devices (ISESD), Bandung, pp. 153-157, doi: 10.1109/ISESD.2016.7886710

S. R. Dewi, (2018). “Deep Learning Object Detection Pada Video Menggunakan Tensorflow dan Convolutional Neural Network,”