Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11190
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Identification of Face Mask with YOLOv4 Based on Outdoor Video

Mawaddah Harahap[1)*], LeonardoKusuma[2)], Melva Suryani[3)], Candra Ebenezer Situmeang[4)], Juniven Francisco Purba[5)]
[1)2)3,4,5)]Universitas Prima Indonesia, Indonesia
[1)]mawaddah@unprimdn.ac.id, [2)]leonardokusuma28@gmail.com, [3)]melvasuryani12@gmail.com, [4)]chandraanggara562@gmail.com, [5)]junivenf@gmail.com

Submitted : Oct 13, 2021 | **Accepted** : Oct 16, 2021 | **Published** : Oct 19, 2021

The use of face masks in the current era is one of the special regulations in many countries including Indonesia to prevent the spread of coronavirus. However, not all people strongly agree to wear masks because they feel uncomfortable to wear even in crowded places require the use of masks such as shopping malls, hospitals, factories, stations and others by checking manually. Therefore, in the study proposed automatic detection of masks with YOLOv4 with the stage of data collection recording community activities in crowded places, labeling images of masks and non- masks. The labelling results were conducted in training that resulted in 90.3% accuracy in the 2000 iteration, the last of which was video testing in three different crowd locations: taxes, city parks and highways. Based on the test results, YOLOv4 can detect masks and non-masks on videos with different obstruction conditions such as people wearing helmets, hand obstacles. However, for the detection of people with tissue obstruction conditions and improper position of wearing masks has not resulted in good detection.

Keywords: Automatic Mask Detection; Corona Virus; Mask Detection; YOLOv4

INTRODUCTION

The use of face masks in the current era is one of the special regulations in many countries including Indonesia to prevent the spread of the coronavirus. The use of masks to cover the mouth and nose, limiting transmission is very efficient in limiting the spread of Covid-19 (I. Buciu, 2020). However, not all people strongly agree to wear masks because they feel less comfortable to wear even in crowded places require the use of masks such as shopping places, hospitals, factories, stations and others by checking manually. However, people only use masks when there is an examination. So it is necessary to build object detection technology by using a camera for the detection of face masks so that the goal of non-contact automatic detection is achieved.

Detection of the use of face masks using object detection technology has recently been proposed by (C. Z. Basha, 2021) using the Modified Resnet Convolution Neural Network (CNN) network to improve mask and non-mask detection accuracy, then work(W. Hariri, 2021) using 3 trained models namely VGG-16, AlexNet, and ResNet-50 to extract deep facial features. Furthermore, the work(Y. Li, 2021) applies a new Convolutional Block Attention Module (CBAM) croping-based method in which the authors report that the proposed method could improve recognition accuracy by 0.104% tested on 4 different datasets.

LITERATURE REVIEW

A one-stage face detection approach such as YOLOv2 was proposed by (M. Loey, 2021) for facial maser detection; the authors used two approaches. First based on the ResNet-50 deep transfer learning model and second directly apply to the YOLOv2 network on 2 different datasets. In (M. Jiang, 2020) it proposes RetinaFaceMask a one-stage detector, consisting of a network of feature pyramids to combine high-level semantic information with multiple feature maps, and a new context attention module to focus on the detection of face masks.

Some literature has reported promising results to be applied in the real world as an automatic mask detection system, but it is still difficult to determine the most accurate model because it has differences at several stages of training and datasets. In this study the YOLOv4 network(A. Bochkovskiy, 2020) was proposed because it has a very high potential in the problem of detection of mask use on video.

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11190
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The Computer Vision variable in the last decade has produced many models and methods, but research on the detection of the use of automatic masks is still minimal. One of the single-stage detection architectures reported in 2020 by (M. Jiang, 2020) which is one of the models that produces accuracy and time efficiency in detecting masks, then work (F. Boutros, 2021) proposes the Embedding Unmasking Model (EUM) which is operated on top of the existing facial recognition model. with a promising level of accuracy and research (M. S. Islam, 2020) using the Convolution Neural Network (CNN) model.

METHOD

This research uses this type of verification research and experimentation with the aim of creating new knowledge from existing theories. The YOLOv4 framework is one of the most accurate methods in the field of computer vision with deep learning and will be applied to this research where the dataset used for testing is sourced from observations in several places in the city of Medan. The dataset was obtained from observations in 3 (three) places, namely Belawan Market (Medan Belawan), Merdeka Field (Medan Kota), USU Market (Medan Baru), then tested the dataset on the YOLOv4 method to see the accuracy of object detection in eachplace.

The following is the work procedure carried out in this study:

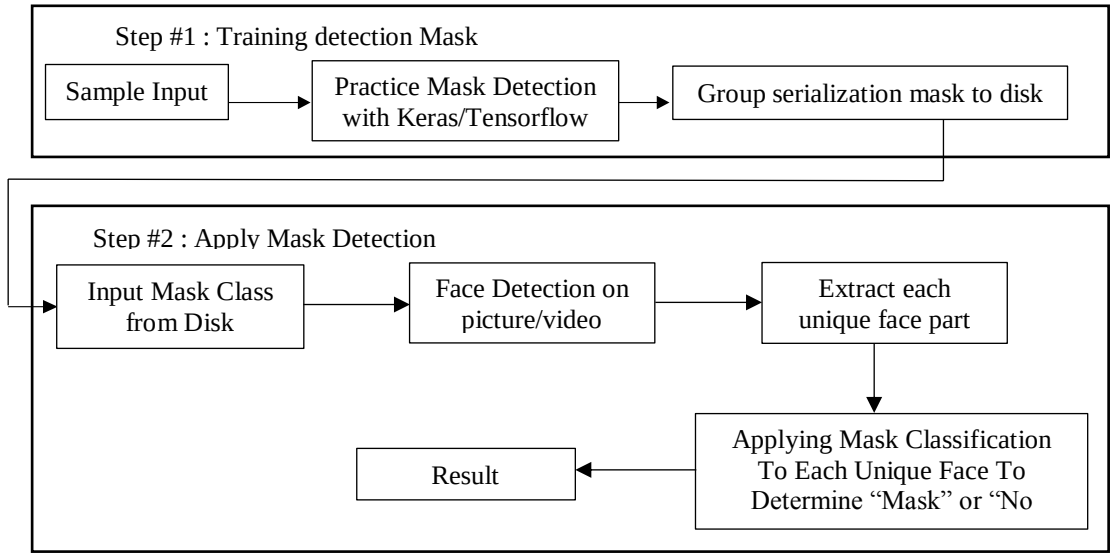


Fig. 1 Work procedure

The data used in this study is observational data from 3 places in the city of Medan are using primary data in research where data obtained from observations in the city of Medan. And using secondary data in the study sourced from a bibliography in the form of a journal, namely the YOLO method.

There are several samples of place observation datasets in the city of Medan, namely picture (a) : USU Market, picture (b) : Belawan Market, and picture (c) : Medan Merdeka Walk which consists of 4 pictures.

(a) (b)

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11190
e-ISSN : 2541-2019
p-ISSN : 2541-044X

(c)

Fig. 2 Sample Place Observation

RESULT

In this study will be outlined the results of mask user detection testing on dataset images and videos that are specifically recorded in various places. All experimental results were conducted on the specifications of the Intel Core i5-7300HQ quad-core 2.5GHz TurboBoost, 8GB RAM, Nvidia GeForce GTX 1050 4 GB GPU and AMD Ryzen 3 4300U 2.7 GHz CPU device, 8GB RAM, Windows 10Home.

DATA SET

In this study, the source of the dataset was specifically carried out by several processes of taking and recording community activities in various crowded places in Medan where there were 3 places recorded, namely USU Tax, Belawan Market, and Merdeka Walk Square. Activities at USU Tax were recorded on June 24, 2021 using OPPO A54 Mobile Phone, Belawan Market was conducted at 3 different times, namely June 18, 2021, June 10, 2021, and July 21, 2021 using VIVO V15 Pro, then at Merdeka Walk Field recorded on June 13, 2021 with OPPO A54 and Canon DSLR Camera. The total video recording amounted to 37 videos, then several changes were carried out and combined videos in the same location, bringing the total video to 11 with dimensions of 1280x720.

Table 1 Dataset Set

Location	Sum	Capacity
USU Taxes	3	562 MB
Merdeka Walk	2	681 MB
Belawan Market	6	1.7 GB

In table 1 describes the details of the results of data collection of community activities used to detect masks and non-masks with a total capacity of 3.5 GB. From this dataset set will be done extracting videos into images for labeling people with masks and non-masks. The total number of images is 1,326. some images of people using masks and non masks in the images used can be seen in fig. 3

Fig. 3 Image dataset of masks and non-masks

*name of corresponding author

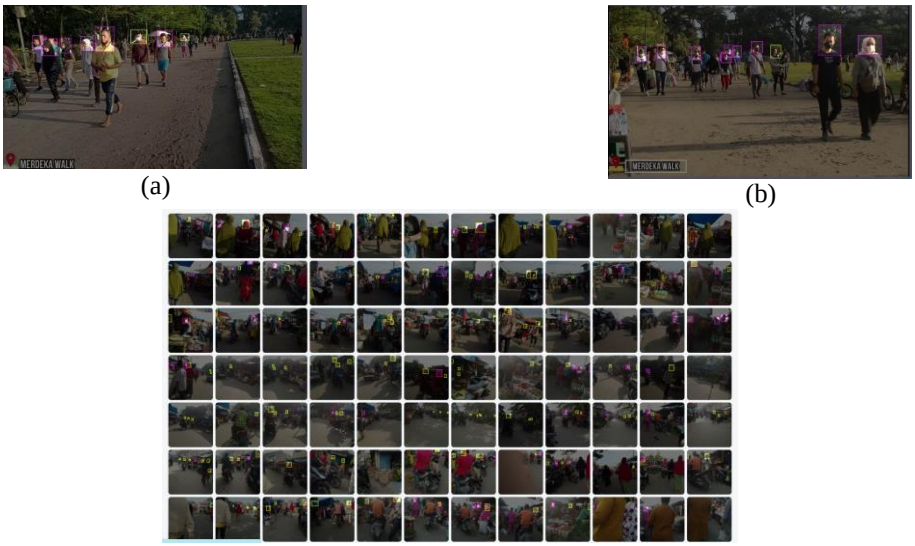
This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11190
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The results of mask and non mask image labeling will be divided by 70% of train data, 20% valid data and 10% of test data. For more details of the division of the label dataset see table 2 and label results in fig. 3

Table 2. Dataset Sharing

Source	Dataset sharing		
	Train	Valid	Test
VIVO Y51	163	47	23
VIVO Y51	153	43	22
VIVO Y51	169	48	24
HP OPPO A54	164	47	23
Canon 400D Camera	170	49	24
Canon 400D Camera	110	31	16



(a) (b)

(c)

Fig. 4 Results Label image of mask and non mask

Data Training

In the training process is carried out on the specifications of Nvidia GeForce GTX 1050 GPU devices by doing some paremeter settings on the data on the training file for YOLOv4 i.e. yolov4.conv.137 and training paremeter stored on yolov4.custom file.cfg with changes to learning_rate=0.001, batch=64. subdivisions=64, steps=4800.5400, max_batches = 6000 and epochs = (6000*64)/700 = 548 with a training yield with mAP (Mean Average Precision) of 90.3% as shown in figure 3

In fig. 5 is a chart of the results of the training dataset where in the 2000 mAP iteration has reached 90.3%, this proves that training in iteration 2000 can already be tested on different datasets, along with details of image dataset training for data trains, and valid tests.

Train	Valid	Test	FPS
94.65	88.45	92.12	22 FPS

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11190
e-ISSN : 2541-2019
p-ISSN : 2541-044X

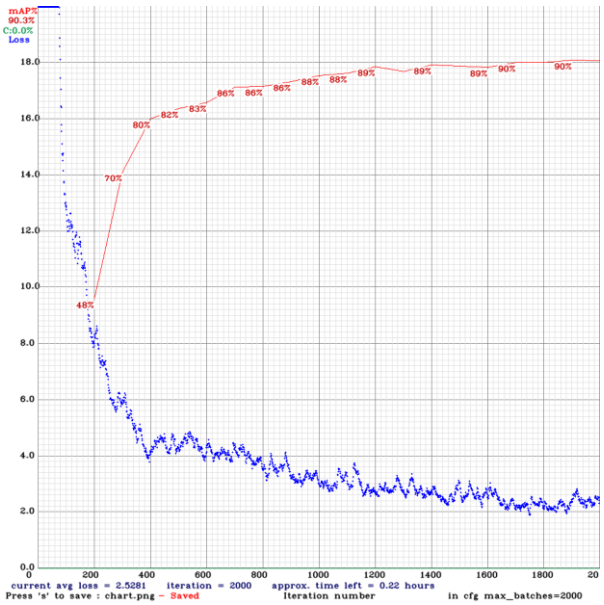


Fig. 5 mAP 2000 iteration Chart

In table 3 is the result of the details of the train dataset test, valid and test of a total of 1,326 images used, it is seen that for train data produces accuracy of 94.65%, valid data 88.45% and tests of 92.12%. From these results prove that the data weights in the 2000 iteration has produced accuracy above 90%, then the next step is to test new image data sourced from video data that is not used as a source of label data. The results of image detection can be seen in fig. 6.

Fig. 6. Image detection results

DISCUSSIONS

In this section will be outlined the results of mask and non mask detection testing in the video dataset using YOLOv4. The application will read the frame per frame for each video, the detected object wearing a mask and non mask will be calculated and displayed directly in the video, after the detection process is complete for all video testing frames will be stored with the file extension .avi so that it can be reopened for evaluation materials. Some of the results of frame detection per frame in video testing will be visible in fig. 7

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

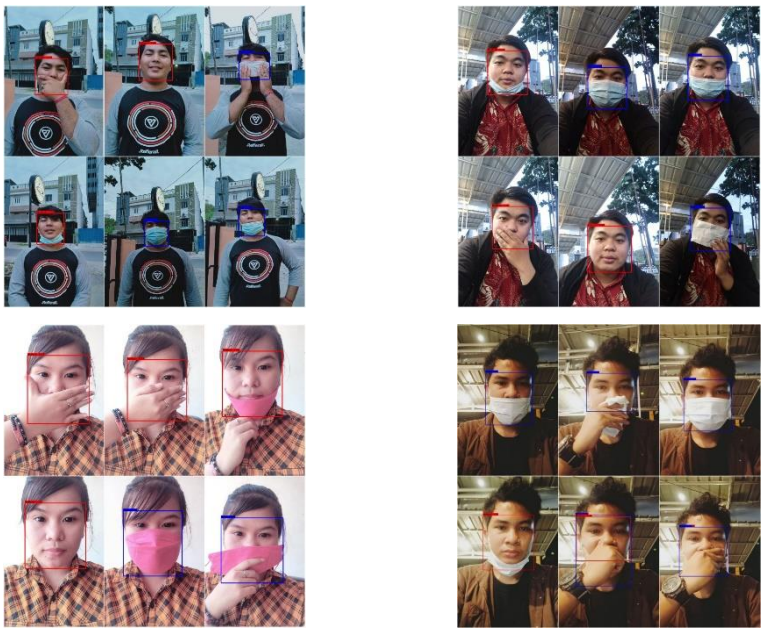
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11190
e-ISSN : 2541-2019
p-ISSN : 2541-044X

(a) Original (b) Detection

Fig. 7 Video detection results (a) Original, (b) Detection

In fig.8 is the result of detection of the mask directly on the video, in fig. 8 (a) is the original image per frame and (b) is the result of detection. The app will count the number of people who use masks and non-masks automatically. Further testing is carried out to detect image objects with different obstacle conditions such as covering the face by hand, opening the mask, covering with paper and the results of this experiment can be seen in fig. 8.

Based on fig. 8 YOLOv4 can detect faces that do not wear masks with hand barriers, open masks, but for obstructions by using masks that are used in inappropriate positions and face barriers using wipes detected as people masks, in this case there is a detection error where in this position is not in the training image. Further testing is done by recording the scene of several people in the video with the initial position not using a mask, then everyone using amask.



*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11190
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Fig. 8 Detection mask with barrier

Based on fig.9 YOLOv4 can detect faces that do not wear masks with hand barriers, masks open, but for obstructions by using masks that are used in inappropriate positions and face barriers using wipes detected as a mask person, in this case there is a detection error where in this position is not in the training image. Further testing is done by recording the scene of several people in the video with the initial position not using a mask, then everyone uses a mask. This test aims to measure the accuracy and speed of the YOLOv4 model in detecting masks and the test results are presented in fig. 9.

Fig. 9. Video Detection

Next reported detection comparison results using GPU and CPU on the same videodataset.
Table 3. Comparison of GPUs with CPUs

Time		
CPU	GPU	Difference
00:30:39	00:05:19	00:25:20
01:05:02	00:10:36	00:54:26
01:26:08	00:14:50	01:11:18

Based on the results of several tests conducted on video footage of activity in crowded places, YOLOv4 was able to detect people wearing masks and non-masks, in addition some images of people using helmets as a barrier in video sequence inference can also be detected. However, for images of people who use tissue barriers or improper use of masks in their position will be detected as people wearing masks. In this case it becomes an evaluation material for future research. Then, the results of testing on the video with the initial position do not use a mask, then in the next frame using a YOLOv4 mask can detect with FPS 15.44 ms, it proves that YOLOv4 has

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

e-ISSN : 2541-2019
p-ISSN : 2541-044X

a high speed level. Furthermore, the results of the comparison of testing using the CPU and GPU resulted in testing using the CPU resulting in a +- difference of 5 minutes longer.

CONCLUSION

In this study specifically conducted the collection of datasets in different locations in medan city with the aim of testing YOLOv4 for the detection of people wearing masks with non-masks. From the results of several tests conducted, YOLOv4 can detect multiple images of single people as well as with many people in an image with an accuracy above 90%. In addition, the process of detection people in the initial video sequence does not use the mask produces a high detection speed with FPS of 15.83 ms, but detection of images with several barriers such as tissues and improper mask position detection errors as the mask wearer will be an evaluation for future research and last detection results using a GPU faster 5 minutes compared to testing with the CPU.

REFERENCES

A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv*. 2020.

C. Z. Basha, B. N. L. Pravallika, and E. B. Shankar, "An efficient face mask detector with pytorch and deep learning," *EAI Endorsed Trans. Pervasive Heal. Technol.*, vol. 7, no. 25, pp. 1–8, 2021, doi: 10.4108/eai.8-1-2021.167843.

F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Unmasking Face Embeddings by Self-restrained Triplet Loss for Accurate Masked Face Recognition," Mar. 2021, Accessed: Jun. 15, 2021. [Online]. Available: http://arxiv.org/abs/2103.01716.

I. Buciu, "Color quotient based mask detection," *2020 14th Int. Symp. Electron. Telecommun. ISETC 2020 - Conf. Proc.*, pp. 12–15, 2020, doi: 10.1109/ISETC50328.2020.9301079.

M. Jiang and X. Fan, "Retinamask: A Face Mask Detector," *arXiv*, 2020.

Y. Li, K. Guo, Y. Lu, and L. Liu, "Cropping and attention based approach for masked face recognition," *Appl. Intell.*, vol. 51, no. 5, pp. 3012–3025, May 2021, doi: 10.1007/s10489-020-02100-9.

M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," *Sustain. Cities Soc.*, vol. 65, p. 102600, Feb. 2021, doi: 10.1016/j.scs.2020.102600.

M. S. Islam, E. Haque Moon, M. A. Shaikat, and M. Jahangir Alam, "A novel approach to detect face mask using CNN," *Proc. 3rd Int. Conf. Intell. Sustain. Syst. ICISS 2020*, pp. 800–806, 2020, doi: 10.1109/ICISS49785.2020.9315927.

W. Hariri, "Efficient Masked Face Recognition Method during the COVID-19 Pandemic," 2021, doi: 10.21203/rs.3.rs-39289/v1.

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.