

Digitimes #37
JANUARI 2022

FACE MASK DETECTION WITH OBJECT TRACKING AND PERSON REIDENTIFICATION FOR PANDEMIC SURVEILLANCE

Wearing Mask

Wearing Mask

Wearing Mask

Photo by Pourab Karchaughuri on Towards Data Science



CfDS
CENTER FOR DIGITAL SOCIETY

Room BC 201 - 202
Faculty of Social and Political Sciences
Universitas Gadjah Mada
Jalan Sosio Yustisia 1, Bulaksumur, Yogyakarta
(0274) 563362, Ext. 116 | cfds.fisipol@ugm.ac.id



RESEARCHERS AND CONTRIBUTORS



Penulis

Nadia Elaesiana
Vidiskiu Fortino Kurniawan
Josia Paska Darmawan

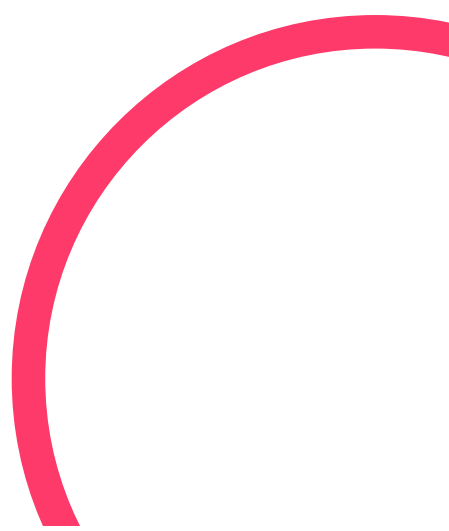
Editor

Treviliana Eka Putri

Desain dan Tata Letak

Rafa Ahamada Wijaya

Januari 2022



■ TABLE OF CONTENTS

2	Introduction
3	Previous Works
6	Datasets
7	Proposed Model
9	Results
11	Conclusion
12	References





■ INTRODUCTION

The COVID-19 pandemic has disrupted in-person activities all around the globe due to the highly contagious characteristic of the disease. SARS-CoV-2, commonly known by the public as the coronavirus, can be spread to other individuals who are in close contact with an infected person. As a result, many governments implement various measures to curb the spread of the virus. One of them is by mandating the use of face masks in public settings. Previous studies have shown that the use of face masks is effective in preventing transmission of the coronavirus (Chu et al., 2020; Wang et al., 2020). Widespread use of face masks has also been proven to mitigate the spread of COVID-19 in multiple countries, such as Hong Kong (Cheng et al., 2020) and the United States (Lyu & Wehby, 2020).

Unfortunately, there are groups of people who refuse to wear a face mask despite government mandates. Researchers from Brookings found that 20% of Americans do not wear masks in public (Vargas & Sanchez, 2020). Out of these respondents, approximately 40% believe that it is their right as a citizen to not wear a mask. The other 24% of the respondents do not wear a mask because it is uncomfortable. Indonesia also faces a similar problem, albeit to a lesser degree. In Bali, for instance, foreign visitors are often found maskless despite the mandatory rule by the government (Sanders, 2021).

Therefore the government needs a mechanism to monitor the use of face masks in public places. The existing rules will not be effective in keeping the population safe during the pandemic if the government cannot detect those who do not obey the health protocols. Several countries have implemented different mechanisms to monitor the use of face masks. For example, Dubai uses drones to detect various types of violations, including not wearing a mask in public. Until May 2021, the Dubai Police have fined 518 people for not complying with the mask mandate (Al Shouk, 2021). China has also installed sensors at subway entrances to detect passengers who are not wearing masks (Bloomberg, 2020).

We can see that many governments have started using face-mask recognition to help them monitor compliance. However, unfortunately, this technology is still not widely used in Indonesia. Moreover, when the detection system is implemented, there is still much room for improvement. The city of Semarang, for instance, has implemented face mask detection technology on multiple CCTVs. The government also provides statistics on how many people are detected for not wearing masks, among many others. However, the statistics raise a concern on its accuracy as we found many consecutive days where the number stays at zero.



Therefore, this research aims to build a model that can detect the use of face masks more accurately. We have considered several areas of improvement in building our model. First, we want to minimize duplicate identifiers by using person re-identification. This would address the problem of inflated statistics due to multiple counting of one person who re-enters the frame. Second, we also want to detect improper use of face masks. In this case, people who do not wear a mask properly (i.e., not covering their nose) will be labeled as "not wearing a mask". Lastly, we aim to build a model that can be implemented on low-resolution images. Public surveillance cameras often produce small and blurry images,

making it more challenging to identify faces, let alone face masks. Therefore, it is crucial to ensure that the model can be applied to low-resolution images and still produce an accurate result.

The next part of this paper will discuss previous works on face mask detection. The next part will elaborate on the datasets that are being used for this research. The proposed model, which includes face detection, face mask detection, object tracking, and re-identification will be explained afterward. The accuracy rate and the limitations will then be discussed to evaluate the proposed model. The last part will summarize the result and provide recommendations for future research.

■ PREVIOUS WORKS

In populated areas, the use of automation and machine learning would ease the monitoring of health protocol. There have been numerous attempts in the research on face mask detection since WHO declared Covid-19 as a pandemic in 2020. Previously, face mask detection technology usage was focused on improving the machine's ability to detect faces with occlusion. In their research, Wen et al. (2005) used Gabor filters to produce cues of facial features which are then processed using geometric analysis algorithms to detect the existence of face

masks. The research aimed to give an early warning system through surveillance to heighten awareness upon having such detections.

The research on face mask detection is accelerated by improving machine learning technologies and the urgency of such research during the pandemic. Due to the importance of processing speed in live detection systems, Singh et al. (2021) use YOLOv3 and faster R-CNN in their model to classify two categories for images of faces that are with or without a mask in videos.



The authors expressed the research difficulties due to the existence of multiple camera angles that causes misclassification from the model. Hence, in their research, they used ~7500 images from MAFA, WIDER FACE to train over the model with an additional custom dataset taken from the web. The result shows that YOLOv3 has an average precision of 55 and inference time of 0.045 s, while Faster R-CNN has an average precision of 62 and inference time of 0.15 s. The tradeoff between precision and time spent is a recurring theme where this will become relevant depending on the implementation of the technology, e.g., live monitor or post-event monitoring. The research highlighted the use of bounding boxes from the raw surveillance data and then classifying the retrieved facial image using models with short classification time.

Additionally, further progress has been launched by Rios et al. (2020) from Tryolabs where 2 other models were tried, namely the MobileNetV2 and BiT-M model; both are pre-trained models from Google. Their research also uses a similar workflow which is to find the bounding box to detect faces from the surveillance image and then classify them using the models. Additionally, this research addressed the common friction in public surveillance situations where the images used as an input are of low quality, varying sizes, and at different angles. Realizing the difference between the available dataset and practical solutions, they collected another 200 images for each category of mask and no mask. The main procedure is also reinforced with the tracking ability of the model using Kalman filter that allows it to detect the existence of the same person within the video taken. Their research concluded that using the existing pre-trained model resulted in a validation accuracy below 75%, which in the case of a binary classifier is still close to randomness, i.e., 50% accuracy.

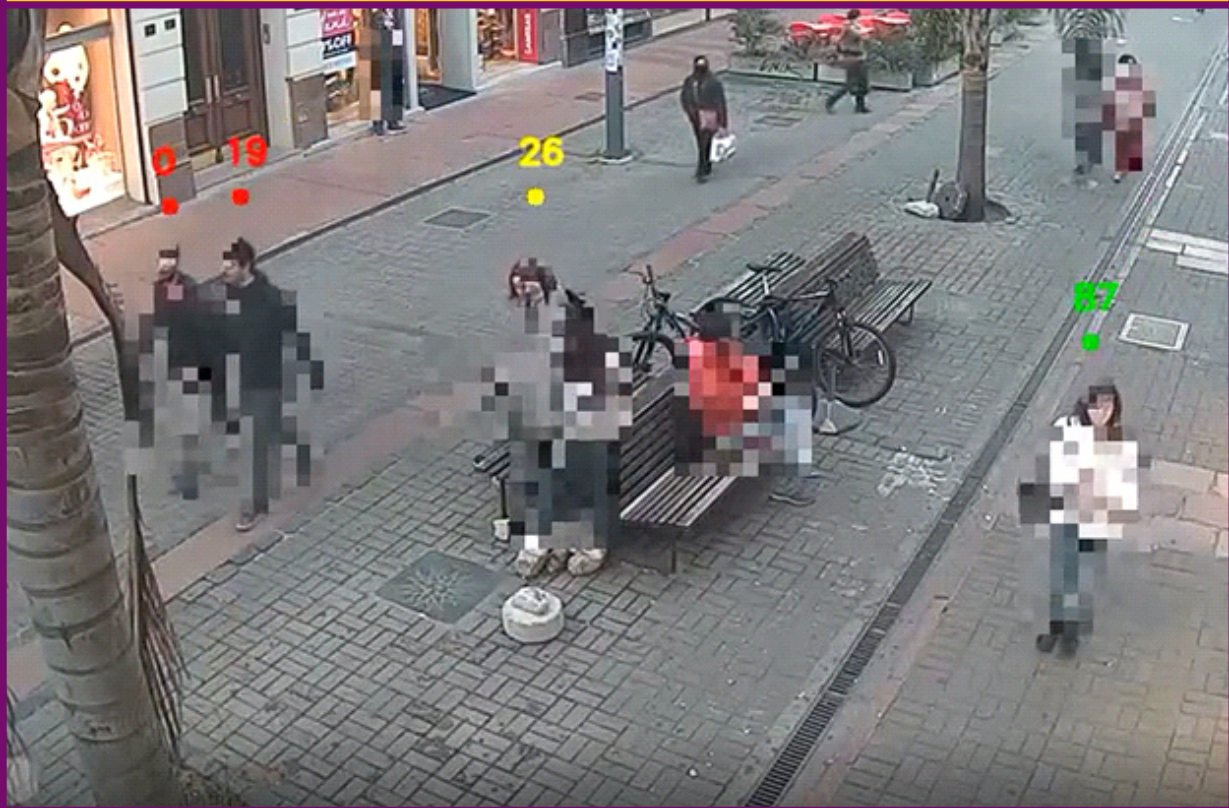


Sample augmented dataset used to train the model by Rios et al. (2020) which contains images which are present in practical conditions.

Source: <https://tryolabs.com/blog/2020/07/09/face-mask-detection-in-street-camera-video-streams-using-ai-behind-the-curtain>

Rios et al. then uses the CIFAR dataset along with the BiT-M model since its images have similar characteristics as the practical condition, being a dataset that contains low-resolution images. Additionally, the researchers use aggressive augmentations on the train and test dataset through means of random rotations, translations, and color jittering. The research concluded that the method above "generalizes much better to other unseen images", hence the final test accuracy of ~96%. The limitation of the model is that it shows better results when both eyes of the detected face are visible. Possible improvements mentioned by the researchers include a more robust model towards head rotation, increasing a category towards faces with misused masks, and person re-identification given the same person re-enters the video. These are the features that we will try to improve in our research.

Sample result from Rios et al. (2020) in which the classifier score tends to 0 for no masks, 50 for undecided, and 100 if mask is detected.



Source: <https://tryolabs.com/blog/2020/07/09/face-mask-detection-in-street-camera-video-streams-using-ai-behind-the-curtain>

■ DATASETS

In this research, we use 5 different datasets to train our model :

1. CrowdHuman

This dataset consists of 24.370 images and 470.000 human instances from highly crowded environments with various kinds of occlusions (Shao et al., 2018). This dataset is used to train the YOLO model to detect the face and human body.
2. Flickr-Faces-HQ (FFHQ)

This dataset is a high-quality image dataset of human faces collected from Flickr and consists of 70.000 images. This dataset offers a lot of variation in terms of age, ethnicity, and image background (Karras et al., 2019).
3. Real-World Masked Face Dataset (RMFD)

This dataset consists of 5000 faces with masks and 90.000 faces without masks that were crawled from the website (Wang et al., 2020). In this research, we only use the dataset of masked faces.
4. Realtime Medical Mask Detection

This dataset is collected from different resources and applied data augmentation, consisting of 5521 mask and no mask images, respectively.
5. Neuralet

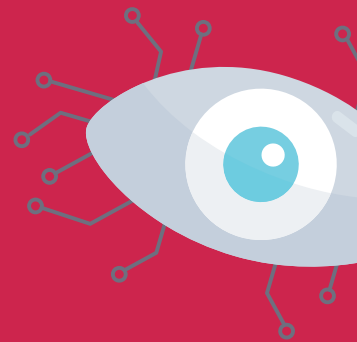
Neuralet provides two kinds of datasets. The first one is Extended Synthetic Dataset, consisting of 10.881 mask images and 11.560 no mask images, while the second one is 50.000 synthesized mask images from the CelebA dataset. These datasets are used to train the face mask model.

Table 1. Total number of images based on the dataset and image type

Datasets	Face Images	Mask Images	No Mask Images	Total
CrowdHuman	24,370			24,370
Neuralet		60,881	11,560	72,441
RMFD		5,000		5,000
Real time medical mask		5,521	5,521	11,042
FFHQ			70,000	70,000
Total	24,370	71,042	87,081	182,853

PROPOSED MODEL

The proposed model includes 4 main components: face detection, face mask detection, object tracking, and person re-identification.

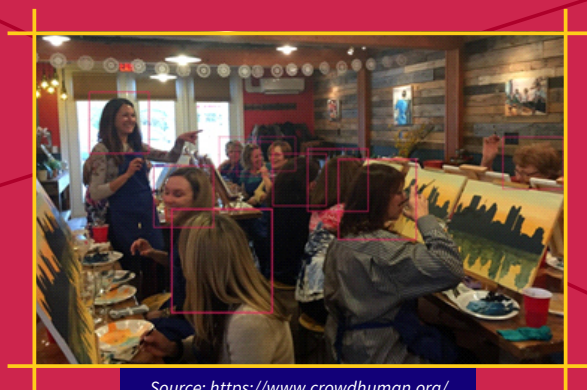


1. Face Detection



We use the YOLO v4 model to detect faces and the human body. This YOLO v4 model was trained on a customized Crowdhuman dataset. Crowdhuman dataset originally contains 3 different types of annotations: head bounding box, human-visible region bounding box, and human full-body bounding box. Our proposed model requires a face bounding box and human full-body bounding box. Since the Crowdhuman dataset does not have annotation for the face bounding box, we replaced the head bounding box with a face bounding box by detecting faces in the Crowdhuman dataset with TinaFace model (Zhu et al., 2021).

Image with head annotation



Source: <https://www.crowdhuman.org/>



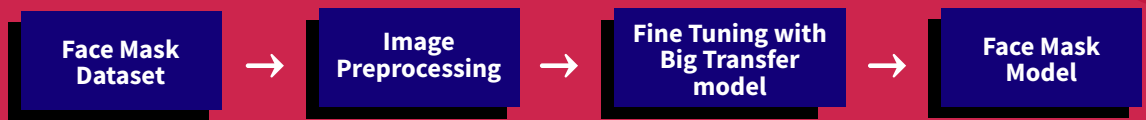
Image with face annotation from TinaFace



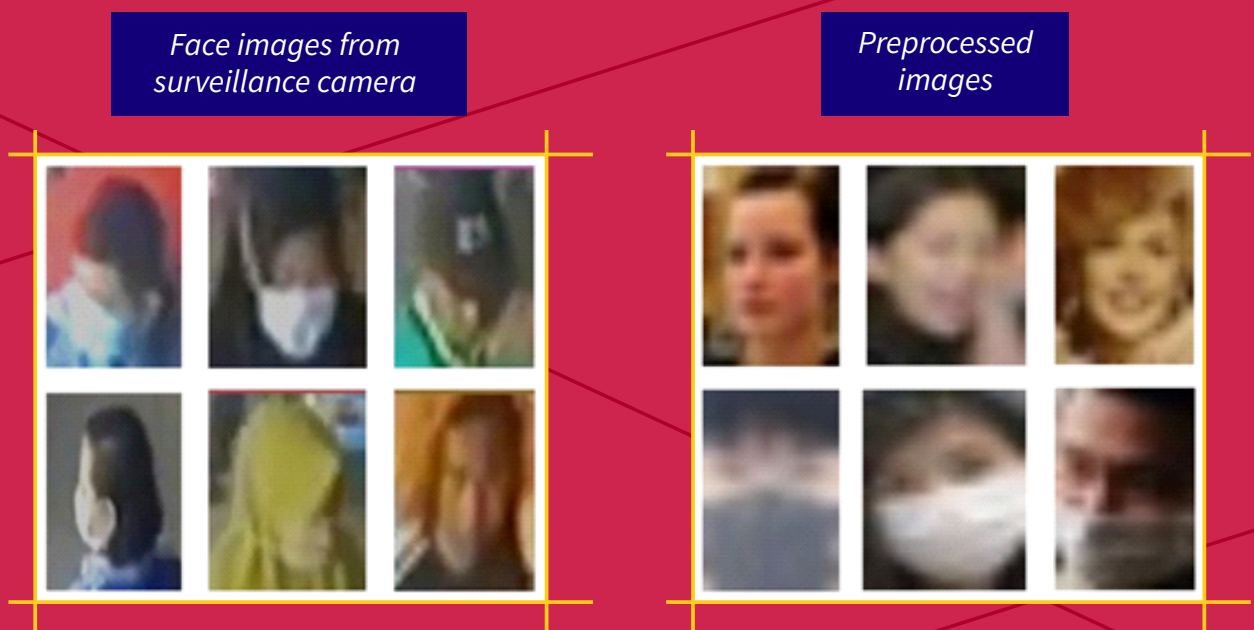
Source: <https://www.crowdhuman.org/>



2. Face Mask Detection



For the face mask detection, we fine-tuned the Big Transfer Model (Kolesnikov et al., 2020) on a dataset from different sources. This dataset was preprocessed by resizing the images into smaller ones to replicate the images from a surveillance video.



3. Object Detection

We used the Simple Online and Realtime Tracking (SORT) model to do object-tracking. SORT is a pragmatic approach to multiple object-tracking with a focus on simple, effective algorithms (Bewley et al. 2016). Bounding box position and size are used in SORT for motion estimation and data association.





4. Re-Identification

The SORT model has limitations, such as ID switching. In order to overcome this problem, we also applied person re-identification. We are using Torchreid for person re-identification.

With those being said, the main contributions of this work are:

- 1 We have built a face mask detection system for low resolution images.
- 2 In addition to face detection, we have also implemented object tracking and person re-identification into our model.



RESULTS

1 Data Sample

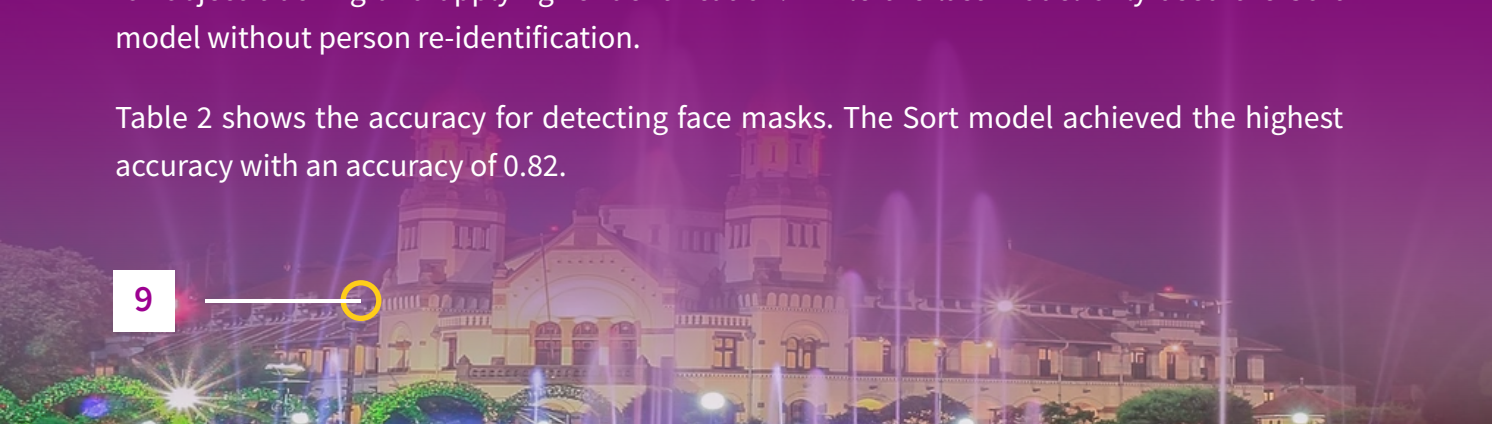
We tested our model on a public surveillance camera from Semarang. In particular, we were using the surveillance camera in Pasar Pedurungan (Pedurungan Market). There are 16 cameras installed in Pasar Pedurungan, Semarang. Based on our consideration of the camera angle, the amount of people passing by, and clarity of the image, we decided to choose the Pedurungan 12 camera. The video was taken on September 10th 2021, with the duration of 10 minutes.

2 Comparisons of accuracy rate

For the baseline model, we used Dlib for face detection, Big Transfer model for face mask detection, and Deep-Sort for object tracking.

We compared the baseline model with 3 other models. The first model used Deep-Sort as the object tracker and did not apply person re-identification. The second model employed Sort for object tracking and applying re-identification. While the last model only used the Sort model without person re-identification.

Table 2 shows the accuracy for detecting face masks. The Sort model achieved the highest accuracy with an accuracy of 0.82.



The Dlib Deep-Sort model has the lowest accuracy with 0.59. While the other two models have a slight difference in accuracy with 0.76 and 0.77 for Deep-Sort model and Sort with Re-ID, respectively. Except for the baseline model, all models utilized the same model for face detection and face mask detection. Choosing the right face detection model will affect the overall accuracy. As shown in the results, the Dlib Deep-Sort model and Deep-Sort model have a 0.17 difference in accuracy.

Table 2. Face Mask Detection Accuracy

Model	Accuracy
DLIB Deep Sort	0.59
Deep Sort	0.76
Sort with Re-ID	0.77
Sort	0.82

In addition to detecting whether a person is wearing a mask or not, we also want to count the number of people who wear masks by applying object tracking. We evaluated two different algorithms for object tracking; Deep-Sort and Sort. These algorithms track a person based on the bounding box coordinates and by not considering the object in the bounding box. Thus, it has several limitations, such as assigning the same person with multiple IDs and assigning one ID to multiple persons. In order to overcome these problems, we also applied person re-identification.

Table 3. Person Identification Accuracy

Model	Actual ID	Predicted ID	Accuracy
DLIB Deep Sort	27	37	0.73
Deep Sort	35	66	0.53
Sort with Re-ID	31	42	0.74
Sort	38	55	0.69

In Table 3, Predicted ID is the number of people tracked by the object tracker and Actual ID is the actual number of people detected. The DLIB Deep-Sort model has the lowest number of Actual ID, with 27 persons detected, while the Sort model is the highest with 38 persons detected. Except for the DLIB Deep-Sort model, the other models used the same face detector, but they have different numbers of Actual ID. This is because not all of the detected persons are assigned to an ID by the object tracker.

As shown in Table 3, the Deep-Sort model has the highest number of Predicted ID, but is the lowest in accuracy. It means that many people are assigned with multiple IDs.

The Sort model is better than the Deep Sort model since it has a lower number of Predicted ID, with 55 IDs and accuracy of 0.69. By applying person re-identification, the accuracy of the Sort model is increased from 0.69 to 0.74. DLIB Deep Sort and Sort with Re-ID models have a slight difference in accuracy, but the DLIB Deep Sort model has the lowest number of Actual ID and Predicted ID. Based on Table 2 and Table 3, there is a trade off between the face mask detection task and the people counter task. Among four models that have been evaluated, none of the models have the best performance in both tasks. Sort model achieved the best performance in face mask detection, but has lower performance than Sort with Re-ID for the people counter task. In contrast, Sort with Re-ID has the highest accuracy in counting the number of people but has lower accuracy in face mask detection.

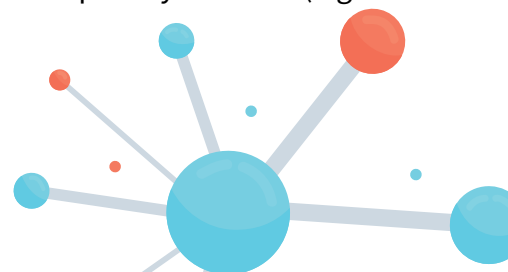
3 Limitations

There are several limitations in our proposed model. Our model has four main components: face detection, face mask detection, object tracking, and person re-identification. It has a long-running time and is not suitable to run in a realtime scenario. The second limitation, even though we have added person re-identification to reduce the number of ID switching and multiple IDs, improvement in object tracking and person re-identification is still needed to count the number of people wearing or not wearing a mask.

CONCLUSIONS

The models that we offer in this paper, especially Sort and Sort with Re-ID, could become an option for face mask monitoring. We considered several elements, such as face instead of head detection, the differentiation between proper and improper use of face mask, object tracking, and person re-identification, and the applicability of the model for low-resolution images. We also measured the accuracy of our model as we want the detection system to provide reliable statistics. The accurate number of people who comply and do not comply with the health protocol, in this case face mask use in public, will be useful for the government to evaluate their policy (Descoins in Yan, 2020). Therefore, face mask recognition technology may become an important tool to mitigate the spread of a pandemic.

However, as mentioned above, our model still has some limitations. Future research may improve the model by decreasing the running time and improving the object tracking and person re-identification. Another area of improvement would be the accuracy of the detection, as we still identify some inaccurate detections in our data. Researchers may also discuss the downsides of face mask recognition, especially in regard to privacy and bias (e.g. gender, racial, and cultural) in the algorithm.



■ REFERENCES

- Al Shouk, A. (2021, May 26). Drones catch 518 people without masks in Naif, Dubai Police issue fines. Gulf News. <https://gulfnews.com/uae/drones-catch-518-people-without-masks-in-naif-dubai-police-issue-fines-1.79471868>
- Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. Proceedings - International Conference on Image Processing, ICIP, 2016-August, 3464–3468. <https://doi.org/10.1109/ICIP.2016.7533003>
- Bloomberg. (2020, Aug 19). SenseTime, the Chinese AI giant blacklisted by Trump, sees demand surge during coronavirus. South China Morning Post. https://www.scmp.com/tech/start-ups/article/3097886/chinese-ai-giant-blacklisted-trump-mints-money-coronavirus?module=perpetual_scroll_0&pgtype=article&campaign=3097886
- Cheng, V. C.-C., Wong, S.-C., Chuang, V. W.-M., So, S. Y.-C., Chen, J. H.-K., Sridhar, S., To, K. K.-W., Chan, J. F.-W., Hung, I. F.-N., Ho, P.-L., & Yuen, K.-Y. (2020). The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. Journal of Infection, 81(1), 107–114. <https://doi.org/10.1016/j.jinf.2020.04.024>
- Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H. J., Chu, D. K., Akl, E. A., El-harakeh, A., Bognanni, A., Lotfi, T., Loeb, M., Hajizadeh, A., Bak, A., Izcovich, A., Cuello-Garcia, C. A., Chen, C., Harris, D. J., Borowiack, E., ... Schünemann, H. J. (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: A systematic review and meta-analysis. The Lancet, 395(10242), 1973–1987. [https://doi.org/10.1016/S0140-6736\(20\)31142-9](https://doi.org/10.1016/S0140-6736(20)31142-9)
- Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019-June, 4396–4405. <https://doi.org/10.1109/CVPR.2019.00453>
- Kolesnikov, A., Beyer, L., Zhai, X., Puigcerver, J., Yung, J., Gelly, S., & Houlsby, N. (2020). Big Transfer (BiT): General Visual Representation Learning. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12350 LNCS, 491–507. https://doi.org/10.1007/978-3-030-58558-7_29
- Lyu, W., & Wehby, G. L. (2020). Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US. Health Affairs, 39(8), 1419–1425. <https://doi.org/10.1377/hlthaff.2020.00818>
- Rios, B., Toscano, M., & Descoins, A. (2020, Jul 9). Face mask detection in street camera video streams using AI: behind the curtain. Tryolabs. <https://tryolabs.com/blog/2020/07/09/face-mask-detection-in-street-camera-video-streams-using-ai-behind-the-curtain>
- Sanders, L. (2021, January 20). Tourists in Bali are being made to do push-ups for not wearing masks. Euronews. <https://www.euronews.com/travel/2021/01/20/tourists-in-bali-are-being-made-to-do-push-ups-for-not-wearing-masks>
- Shao, S., Zhao, Z., Li, B., Xiao, T., Yu, G., Zhang, X., & Sun, J. (2018). CrowdHuman: A Benchmark for Detecting Human in a Crowd. 1–9. <http://arxiv.org/abs/1805.00123>
- Singh, S., Ahuja, U., Kumar, M., Kumar, K., & Sachdeva, M. (2021). Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. Multimedia Tools and Applications, 80: 19753–19768. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7917166/>
- Vargas, E. D., & Sanchez, G. R. (2020, August 31). American individualism is an obstacle to wider mask wearing in the US. Brookings. <https://www.brookings.edu/blog/up-front/2020/08/31/american-individualism-is-an-obstacle-to-wider-mask-wearing-in-the-us/>
- Wang, Y., Tian, H., Zhang, L., Zhang, M., Guo, D., Wu, W., Zhang, X., Kan, G. L., Jia, L., Huo, D., Liu, B., Wang, X., Sun, Y., Wang, Q., Yang, P., & MacIntyre, C. R. (2020). Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: A cohort study in Beijing, China. BMJ Global Health, 5(5), e002794. <https://doi.org/10.1136/bmjgh-2020-002794>
- Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., Pei, Y., Chen, H., Miao, Y., Huang, Z., & Liang, J. (2020). Masked Face Recognition Dataset and Application. 1–3. <http://arxiv.org/abs/2003.09093>
- Wen, C. Y., Chiu, S. H., Tseng, Y. R., & Lu, C. P. (2005). The mask detection technology for occluded face analysis in the surveillance system. Journal of Forensic Sciences, 50(3): 593–601. <https://pubmed.ncbi.nlm.nih.gov/15932092/>
- Yan, W. (2020, Sep 11). Face-mask recognition has arrived—for better or worse. National Geographic. <https://www.nationalgeographic.com/science/article/face-mask-recognition-has-arrived-for-coronavirus-better-or-worse-cvd>
- Zhu, Y., Zhang, S., & Wang, C. (2021). TinaFace: Strong but Simple Baseline for Face Detection. <https://arxiv.org/abs/2011.13183v2>

ABOUT CfDS

Center for Digital Society (CfDS) is the research center established by the Faculty of Social and Political Sciences, Universitas Gadjah Mada. The institution is created under the concern over the contemporary dynamics of socio-political condition of the world that is accentuated by the impeccable influence of information technology. The phenomenon triggers the new patterns and complexities in the society, and thus requires new approaches in managing such complexities.

CfDS pledged to delve more on the study of contemporary digital society, including related issues surrounding the topic; such as the issues of smart city and urban development. The emphasis is then put on the utilization of technology to shape the society in becoming digitalized, as well as to bring solvency to social issues.

Vision

Solving Social Problem and Accelerating Prosperity through Indonesia Digital Society

Activities

1. Research and Development
2. Dissemination and Publication
3. Education and Policy Advocacy

Research Scopes

1. Digital Governance
2. Digital Economy
3. Future Technology



 facebook.com/cfdsugm

 [@cfds_ugm](https://twitter.com/cfds_ugm)

 [Center for Digital Society \(CfDS\)](https://www.linkedin.com/company/center-for-digital-society-cfds)

 [@cfds_ugm](https://line.me/tv/cfds_ugm)

 cfds.fisipol.ugm.ac.id

 [cfds_ugm](https://www.instagram.com/cfds_ugm)

 [CfDS UGM](https://www.youtube.com/CfDS_UGM)