Real-Time Face Mask Detection using Deep Learning

Pranad Munjal*, **Vikas Rattan, Rajat Dua and Varun Malik**

Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab-140401, India

*munjalpranad99@gmail.com (Corresponding Author)

ARTICLE INFORMATION

ABSTRACT

The outbreak of COVID-19 has taught everyone the importance of face masks in their lives. SARS-COV-2(Severe Acute Respiratory Syndrome) is a communicable virus that is transmitted from a person while speaking, sneezing in the form of respiratory droplets. It spreads by touching an infected surface or by being in contact with an infected person. Healthcare officials from the World Health Organization and local authorities are propelling people to wear face masks as it is one of the comprehensive strategies to overcome the transmission. Amid the advancement of technology, deep learning and computer vision have proved to be an effective way in recognition through image processing. This system is a real-time application to detect people if they are wearing a mask or are without a mask. It has been trained with the dataset that contains around 4000 images using 224×224 as width and height of the image and have achieved an accuracy rate of 98%. In this research, this model has been trained and compiled with 2 CNN for differentiating accuracy to choose the best for this type of model.It can be put into action in public areas such as airports, railways, schools, offices, etc. to check if COVID-19 guidelines are being adhered to or not.

1. Introduction

COVID-19 has compelled people to wear face mask all over the world (Chavez, Long, Koyfman & Liang, 2020) and it has also dropped down the economic growth of the country (Yu, Zhu, Zhang & Han, 2020). People nowadays have been bounded by certain set of enforced protocols to wear face masks in public areas. Even researchers have proved that wearing a face mask can protect the transmission of this virus. The research department has stated the difference in the effectiveness of N95 and surgical masks i.e., 91% and 68% respectively (Feng, Shen, Xia, Song, Fan & Cowling, 2020). However, potency to prevent the transmission of this disease has lessened due to improper usage of face masks in public areas (Cowling et al., 2009). It is essential to automate the process of detection of wearing facemask in public areas which will provide protection to the individual and prevent the local pandemic.

Machine Learning is one of the most exciting sub-set of artificial intelligence in which machines are trained for new data independently by iterations. It is trained multiple number of times till the desired output is determined. Typically, machine learning can be used in image processing, real-time ads, spam filtering techniques etc.

Deep learning is a part of machine learning methods based on ANN (Sandler, Howard, Zhu, Zhmoginov & Liang-Chieh, 2018) which is made to learn abstractions in data using hierarchal architecture. Deep learning is comprised of various neural networks which uses the cores of a processor to manage the neural networks (Militante, 2019). It is an emerging approach which is widely being used in artificial intelligence domains, computer vision is one of those domains.

The Convolutional Neural Network (CNN) is a significant deep learning approach which performs training in a robust manner (LeCun, Bottou, Bengio & Haffner, 1998). In this research, we have used CNN i.e., MobileNetV2 (Bengio, Courville & Vincent, 2013) and VGG16 for comparison of accuracy between the two, with some popular libraries i.e.,TensorFlow, Keras, Imutils, OpenCV. The system has been balanced with limited dataset and good recognition accuracy.

The proposed system is a real-time application based on automation of the process of screening performed in the public areas to enforce the people to wear masks and to contribute in combatting with this virus. This system can be used on real-time video surveillance cameras, drones, or any mobility cameras to monitor public places to detect people that are without mask.

2. Literature Review

Militante & Dionisio (2020) 0have presented their research paper in which dataset of 25,000 images is being used with pixel resolution of 224×224 and have achieved accuracy rate of 96%. They have used ANN i.e., Artificial Neural Network to replicate the stimulation of human brain. In their research, raspberry pi is being used for alarming in public area if someone is entering without mask.

Loey, Manogaran, Taha & Khalifa (2021a) presented their research paper for medical face mask detection. In this research, they have used 2 datasets to find the accuracy between the two. They calculated average precision of both datasets, in addition to this they believe that using YOLO-v2 in combination with ResNet-50 model can help to achieve high average precision. The experimental results of this research are based on validation accuracy and loss on each epoch, iteration and variable learning rate.

Guillermo et al. (2020) presented their research paper which has been used for face mask detection. In this research paper, they have implemented this model to create an artificial dataset by their own. They have used around 600 raw without mask images and have obtained an artificial dataset by putting mask on the face by coordinates gathered using machine learning algorithms.

Boyko, Basystiuk & Shakhovska (2018) 0have presented their research paper using Dlib and OpenCV library. This research is based on hog method for searching and subsequent recognition to compare the performance of these two most popular computer vision libraries. OpenFace library has been used to get the coordinates of face boundary. They have used 128 facial feature extraction to divide the facial features into groups which will lead them with better accuracy.

Loey, Manogaran, Taha & Khalifa (2021) have presented his research paper in which they have proposed to use 3 datasets to differentiate the accuracy of those datasets by passing them through the set of same algorithms. The three datasets that they have used are RMFD, SMFD, LFW. In this research, they have employed deep transferring learning (ResNet50) with some classical machine learning (SVM) and they believe that ResNet-50 achieves better results when used as feature extractor. So, ResNet-50 is used as a feature extractor while the SVM (Support Virtual Machine) is used in training, validation and testing phase. Using these technologies, the research achieved the accuracy of 99.64%, 99.49%, 100% in RMFD, SMFD, LFW respectively.

Mohan, Paul & Chirania (2021) have presented their research paper which is used to detect a face mask using a microcontroller i.e., ARM-Cortex M7 which has been clocked at 480MHz with 496kb framebuffer. They proposed this model for 138kb post-quantization to run at inference speed of 30fps and have deployed 3 datasets, 2 from Kaggle with 12,232 images and 3rd dataset with 1979 was generated by the author through OpenMV cam H7 controller camera. In this research, the dataset has been augmented to 1,31,055 and then resized all images to 32×32 pixels as is the optimal size for the set frame buffer of the microcontroller i.e., 496kb. At last, it turned out with 99.79% accuracy and was best fit model for RAM constraint microcontrollers.

Lin et al. (2020) have presented their research paper in which they have proposed a framework which is able to detect faces correctly and precisely segmenting each face from each image. Lin et al. named this framework as G-Mask; ResNet-101 is being used as a feature extractor they believed that it works more precisely than other bounding box algorithms. In this research, they have adopted Keras framework to train the G-Mask model with 5115 samples of images with 3000 steps_per_epoch, 0.001 as the learning rate of the model. The dataset images have been resized to 1024×800 by height and width resp.

Pandiyan (2020) has presented this research paper in which Pandiyan has implemented SMS alert system for non-mask wearers which are being screened through the CCTV cameras at the public places. In this research, CNN layers are being used for the detection of mask and to capture the images of people not wearing mask and AWS (Amazon Web Services) is being used to keep the captured images on-the-fly. Twilio messaging i.e., is an API to send and receive SMS that is being used to send an SMS alert to the person whose image has been captured and is residing in the AWS database.

Chen et al. (2020) have presented their research paper in which they have proposed to identify whether he/she is wearing the face mask. This research is an amalgamation of deep learning and machine learning techniques. They have concluded this model with 7 steps that are: input face mask dataset, Train the dataset with python libraries, rearrange the dataset to disk, load dataset from disk, detect faces from image/video stream, determine with or without mask and output the result. They believe that this particular model could also be utilized as a utilization case for edge examination.

Das, Ansari & Basak (2020) have presented their research paper in which they have deployed 2 datasets to compare the accuracy and loss results between the 2 datasets. This research is based on OpenCV, TensorFlow, Keras libraries to get the result. Das et al. have defined their own dataset with people facing in front with total 1376 images, 690 with white mask and 686 without mask; 2nd dataset is from Kaggle with total of 853 images which has been classified to two classes i.e., with mask and without mask. They have trained their model with 20 epochs and with test_split of 90% training and 10% validation data

which concluded them to get an accuracy rate of 95.77%, 94.58% of dataset1 and dataset2 respectively.

Salihbasic & Orehovacki (2020) have presented their research paper in which they have used OpenCV to detect face from an image and detect the gender from the features of the face. For feature classification, Salihbašić et al. have used LBPH model which detects the features and it inputs it to the 3 CNN layers. The first layer of CNN holds 96 filters, the second holds 256 filters and the last layer holds 384 filters. But the accuracy of their model concluded to be very less if the face of the person is illuminated, pose is different, camera features of the mobile phone and also the performance of the mobile phone.

Kalas (2019) has presented this research paper in which the author has worked on detecting the face of a person from a video stream. In this research, 3 technologies are being used for face detection i.e., adaboost, haar like algorithm and OpenCV. According to the research paper, OpenCV is being used for object-detection, Adaboost is being used for the training of the sample images as adaboost doesn't overfit the model, and Haar-like Algorithm are being used to extract out the bounding coordinates of the face by visualizing the features of the face.

3. Methodology

The process of Face Mask Detection in real-time involves 2 phases i.e., Training and then Detection. Further, these 2 phases involve sub-phases that are described below:

Phase 1: Training

Load Dataset

For training the model, we have the deployed dataset from Kaggle (Chakraborty, 2020) which contains around total 3833 images, i.e., 1918 with mask and 1915 without mask as shown in Figure 1 and Figure 2. We have employed this dataset as it contained horizontal and vertical tilted images.
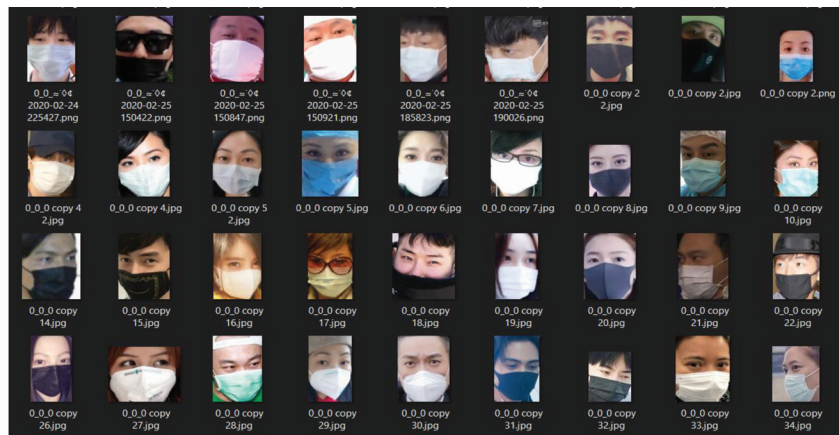

Figure 1: Dataset(with mask).

Figure 2: Dataset (without mask).

b. Pre-Processing Image

For better accuracy in the model, pre-processing the image is important step before the actual training. As in this paper, we have used Convolutional Neural Networks, so we have resized the dataset images to 224×224-pixel resolution as shown in Figure 3. After resizing the images, they have been pixelized (i.e., converted into array) with OpenCV library functions. After they have been pixelized, the images are labelled as 'with_mask' or 'without_mask' with TensorFlow and Keras library functions. A master array containing all pixelized images with labels has been generated for training.
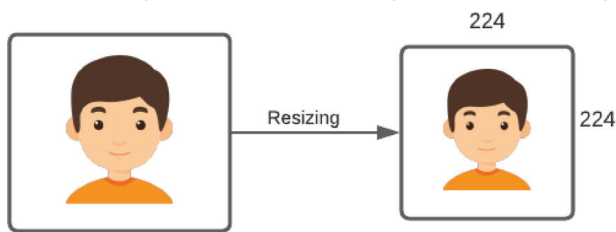
Figure 3: Diagram to depict Resizing Image.

c. Training the Model

The final step of training the model is using the Convolutional Neural Network as shown in Figure 4 i.e., MobileNetV2 and VGG16. Both the models have been passed with same parameters for training as shown in Figure 4. We have divided the dataset into 2 proportions: 80% for training and 20% for validation. For validation of the data, accuracy metrics is used for both models as it tests the model after each epoch. In this research, we have used images with batch size of 32 and the input size of the images to be 224×224 as width and height resp. for the CNN model. The parameters for ADAM optimizer were fixed according to the accuracy after testing, which turned out to be best at 20 epochs, with learning rate of 1e-4. Consequently, Image Data Generator is applied to increase the dataset by simply rotating, rescaling, zooming, flipping the actual images. As when the features are extracted the image resolutions are reduced, so to avoid over fitting of the model the dropout rate is fixed at 50 for both models.
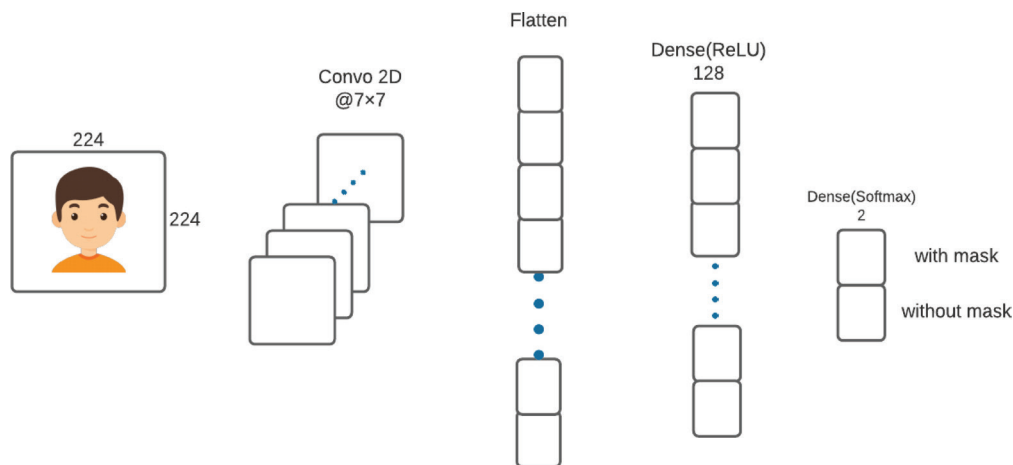

Figure 4: Diagram for Depiction of CNN.

Phase 2: Detection

a. Loading Trained Model from Disk

Training will eventually lead to generating a model extension file, which will contain trained model classifier. This trained will help in detection of mask from detected faces.

b. Detect Faces in Video Stream

When an object(person) passes by the real-time camera, that frame is captured and processed, and the processing of the frame will lead to detection and extraction of faces from it as shown in Figure 5.
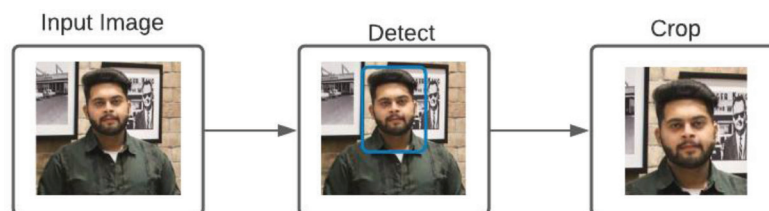

Figure 5: Detecting and Cropping Face.

c. Feature Extraction from Each Face

In this step, the features are extracted from the cropped face image which has been detected by the camera. These features help in building up the accuracy rate of mask detection.

d. Output (With/Without Mask)

The final step is compiling all results from face extraction to detection, whether he/she is wearing a mask.

4. Results

The validation accuracy achieved during the training of this model after 20 epochs and with batch size of 32 is 99.2% and 98.6% for MobileNetV2 and VGG16 resp. In this research,

we have trained this model with 5, 10, 15, 20 epochs and the validation accuracy has been illustrated in Table 1.

Table 1: Comparison table for MobileNetV2 vs VGG16.

Epochs	MobileNetV2	VGG16
Accuracy on 5 epochs	89.57%	95.83%
Accuracy on 10 epochs	89%	98%
Accuracy on 15 epochs	90.74%	98.5%
Accuracy on 20 epochs	99.2%	98.6%

The model's output as shown in

Table 1 at 20 epochs is nearly equal but as shown in Figure 6 the training and validation loss is much more when trained by VGG16. Due to this, the accuracy of detecting the mask is being affected in VGG16.
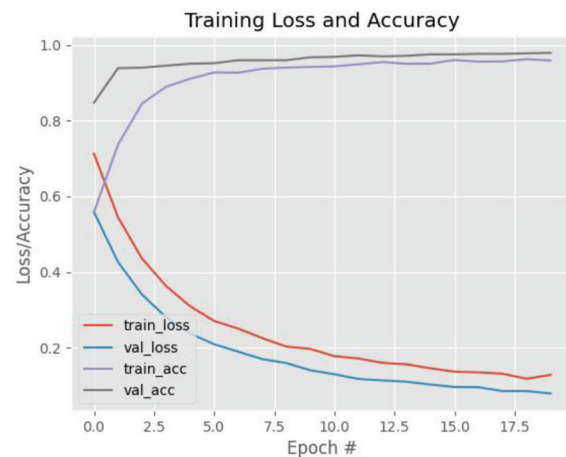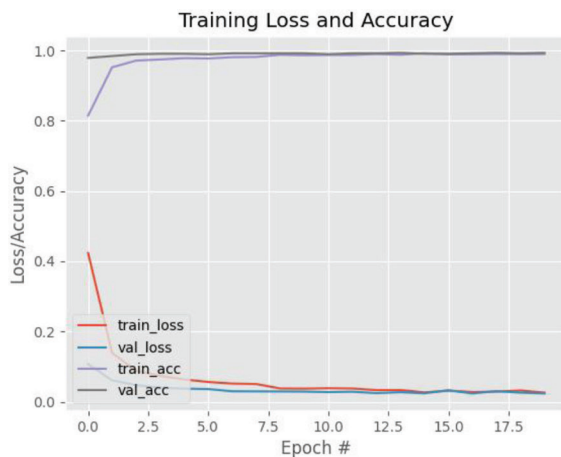

Figure 6: Training Accuracy/Loss Plot Graph of MobileNetV2 vs VGG16.

Therefore, Figure 8 displays the result of model trained with VGG16, the accuracy attained in Figure 8 is less than the accuracy attained in Figure 7 which displays the result of model trained with MobileNetV2.
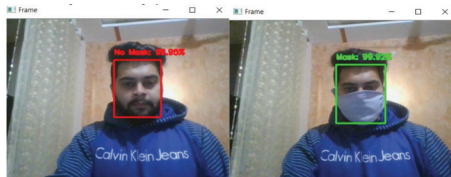

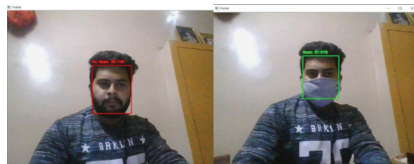**Figure 7:** Result Accuracy with MobileNetV2 (with/without mask).

Figure 8: Result accuracy with VGG16(with/without mask).

Conclusions

This paper manuscript is a comparison study of two Convolutional Neural Networks i.e., MobileNetV2 and VGG16 on mask detection. The study clearly represents that MobileNetV2 was able to attain maximum accuracy, minimum loss with less dataset and less epochs. VGG16 can also get maximum accuracy if trained the same model with 70-80 epochs. Moreover, this trained model can be implemented in public areas with the help of IP cameras to detect if he/she is wearing a mask or not and it is a useful tool in combatting with COVID-19 virus.

For further research, we can use another dataset with much more images and can be trained with VGG16 with more epochs. We can integrate this model with alarming system, social distancing system, SMS alert system. Also, this model can be tested with other optimizers, and can use adaptive learning techniques.

References

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798-1828. https://doi.org/10.1109/TPAMI.2013.50

Boyko, N., Basystiuk, O., & Shakhovska, N. (2018). *Performance Evaluation and Comparison of Software for Face Recognition, Based on Dlib and Opencv Library*. 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine. https://doi.org/10.1109/DSMP.2018.8478556

Chakraborty, A. (2020). *Face Mask Detection Data*. Retrieved from: https://www.kaggle.com/aneerbanchakraborty/face-mask-detection-data/activity

Chavez, S., Long, B., Koyfman, A., & Liang, S. Y. (2020). Coronavirus Disease (COVID-19): A primer for emergency physicians. *The American Journal of Emergency Medicine, 44,* 220-229. https://doi.org/10.1016/j.ajem.2020.03.036

Chen, Y., Hu, M., Hua, C., Zhai, G., Zhang, J., Li, Q., & Yang, S. X. (2020). *Face Mask Assistant: Detection of Face Mask Service Stage Based on Mobile Phone*. Preprint arXiv:2010.06421. https://arxiv.org/abs/2010.06421

Cowling, B. J., et al., (2009). Facemasks and Hand Hygiene to Prevent Influenza Transmission in Households: A Cluster Randomized Trial. *Annals of Internal Medicine*, 151(7), 437-446. https://doi:10.7326/0003-4819-151-7-200910060-00142

Das, A., Ansari, M. W., & Basak, R. (2020). *Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV*. 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India. https://doi.org/10.1109/INDICON49873.2020.9342585

Feng, S., Shen, C., Xia, N., Song, W., Fan, M., & Cowling, B. J. (2020). Rational use of face masks in the COVID-19 pandemic. *The Lancet Respiratory Medicine*, 8(5), 434–436. . https://doi.org/10.1016/S2213-2600(20)30134-X

Guillermo, M., Pascua, A. R. A., Billones, R. K., Sybingco, E., Fillone, A., & Dadios, E. (2020). *COVID-19 Risk Assessment through Multiple Face Mask Detection using MobileNetV2 DNN*. The 9th International Symposium on Computational Intelligence and Industrial Applications (ISCIIA2020), Beijing, China. https://isciia2020.bit.edu.cn/docs/20201114082420135149.pdf

Kalas, M. S. (2019). Real Time Face Detection and Tracking using OpenCV. *International Journal of Soft Computing and Artificial Intelligence*, 2(1), 41-44.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceeding of IEEE*, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Lin, K., Zhao, H., Lv, J., Li, C., Liu, X., Chen, R., & Zhao, R. (2020). Face Detection and Segmentation Based on Improved Mask R-CNN. *Discrete Dynamics in Nature and Society*, 2020, 9242917. https://doi.org/10.1155/2020/9242917

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. https://doi.org/10.1016/j.measurement.2020.108288

Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021a). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustainable Cities and Society*, 65, 102600. https://doi.org/10.1016/j.scs.2020.102600

Militante, S. (2019). Fruit Grading of GarciniaBinucao (Batuan) using Image Processing. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(2), 1829-1832.

Militante, S. V., & Dionisio, N. V. (2020). *Real-Time Face Mask Recognition with Alarm System using Deep Learning*. 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia. https://doi.org/10.1109/ICSGRC49013.2020.9232610

Mohan, P., Paul, A. J., & Chirania, A. (2021). A Tiny CNN Architecture for Medical Face Mask Detection for Resource-Constrained Endpoints. In: Mekhilef, S., Favorskaya, M., Pandey, R. K., Shaw, R. N. (eds.) *Innovations in Electrical and Electronic Engineering. Lecture Notes in Electrical Engineering*, vol 756. Springer, Singapore. https://doi.org/10.1007/978-981-16-0749-3_52

Pandiyan, P. (2020, December 17). *Social Distance Monitoring and Face Mask Detection Using Deep Neural Network*. Retrieved from: https://www.researchgate.net/publication/347439579_Social_Distance_Monitoring_and_Face_Mask_Detection_Using_Deep_Neural_Network

Salihbasic, A., & Orehovacki, T. (2019). *Development of Android Application for Gender, Age and Face Recognition Using OpenCV*. 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia. https://doi.org/10.23919/MIPRO.2019.8756700

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A.,& Liang-Chieh, C. (2018). *MobileNetV2: Inverted Residues and Linear Bottlenecks*. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA (pp. 4510-4520). https://doi.org/10.1109/CVPR.2018.00474

Yu, P., Zhu, J., Zhang, Z., & Han, Y. (2020). A Familial Cluster of Infection Associated With the 2019 Novel Coronavirus Indicating Possible Person-to-Person Transmission During the Incubation Period. *The Journal of Infectious Diseases*, *221*(11), 1757-1761. https://doi.org/10.1093/infdis/jiaa077