Face Mask Recognition Based on Two-Stage Detector

Hewan Shrestha[1](✉), Swati Megha[2], Subham Chakraborty[3], Manuel Mazzara[2], and Iouri Kotorov[4,5]

[1] Saarland University, Saarbrücken, Germany
shresthahewan12@gmail.com
[2] Innopolis University, Innopolis, Russia
[3] Örebro University, Örebro, Sweden
[4] Department of International Business, North Karelia University of Applied Sciences, Karjalankatu 3, 802 00 Joensuu, Finland
[5] Institut de Recherche en Informatique de Toulouse (IRIT), Université de Toulouse, 31062 Toulouse, France

Abstract. With the number of positive cases of Covid-19 infection is increasing, it is essential for everyone to wear a face mask and prevent the spread of Covid. As people are gathering in a large number at different locations, it is quite important for everyone to wear a face mask and prevent the covid spread. With the increase in the crowd gathering, it is often hard to see who is not wearing a mask. Although various techniques have been proposed earlier for face mask detection, the results have not been effective. This paper proposes region-based deep learning detection techniques for face mask detection using Faster R-CNN. The proposed model uses ResNet-50 as RPN which generates anchors and output region proposals. Later, ROI pooling is used to map the feature map in proposal to target dimensions. Finally, a classifier is used to output the final class and bounding box around the face. The proposed work attained a final mean average precision (mAP) of 45% over 30 epochs and achieved satisfactory performance.

Keywords: Region proposal networks · Anchors · Non-maximum supression · COVID-19 · Face mask detection

1 Introduction

As Covid-19 pandemic upsurged, there have been various complications and casualties all around the world. Covid-19 pandemic has been widely spreading throughout the world via various mediums. More precisely, air medium is considered to be one of the most influential medium for infection transmission. With the rapid growth in human population all around, it becomes quite difficult to control the infection spread. Although many countries have made it mandatory to wear face masks at all times to prevent the spread of pandemic. However,

A. Abraham et al. (Eds.): ISDA 2022, LNNS 715, pp. 576–585, 2023.
https://doi.org/10.1007/978-3-031-35507-3_56

people are getting careless and seem to be not following the instructions provided by the governments of individual nations. Having some sort of detection systems to localise faces whether the people are wearing masks properly or not, makes it convincing to help prevent the pandemic spread [1].

Wearing face mask is proved to be efficient to reduce the risk of viral transmission of any disease [2]. With the inclination in Covid cases all around, it becomes important to have every individual wear face mask when going out from their respective houses. Detecting face masks in crowd is quite challenging task. Computer Vision is a possible way to help prevent the Covid pandemic spread around the world. As the pandemic spreads, few face mask detection techniques have been proposed using deep learning algorithms. Detection models proposed for face mask detection are based on single-stage detector algorithms like SSD and YOLO.

We present a two-stage detector based on Faster R-CNN with MobileNet_V2 where first stage proposes regions of interests and the second stage localizes the face masks out of those regions generated by the first stage. The second stage localizes and classifies the face images with either 'with_mask', 'without_mask' or 'mask_worn_incorrectly'.

2 Related Work

With the coronavirus outbreak all around the world, it has become difficult for everyone to return to their normal life. Wearing a mask prevents covid spread and gives a sense of protection all around [1]. With two-stage Convolution Neural Network (CNN) architecture, it is compatible to embed in CCTV cameras and efficient in detecting face masks in a safe working surrounding [3].

As the gathering increases, people come together and the virus spreads all around through the air medium. Although it has been compulsory to wear face mask, monitoring everyone manually is a challenging task. MobileNet with a global pooling system [4] has been proposed which flattens the feature vector and the softmax layer in the fully connected layer classifies the objects.

Various face mask detection techniques have been proposed with Haar cascades, Histogram-of-Gradients (HOG) and neural networks as well. Haar cascades and HOG approaches are referred to as feature-based approaches, whereas neural network approaches refer to Multitask Cascade Convolution Neural Network, Max Margin Object Detection and TinyFace [2,5]. Among feature-based and neural network based approaches, TinyFace performs well on face mask detection.

Masked face detection is often quite challenging to implement in real-time scenarios. However, having some method to detect masked faces helps in realtime tracking of objects in crowdy places specially during the current situation of Covid pandemic. Three different layers of cascaded Convolution Neural Network (CNN) architecture implemented with the MASKED FACE dataset [6] results in satisfactory detection.Although the model trained on MASKED FACE dataset is capable enough to detect medical face mask, it may not be useful in effective detection and surveillance as the model maynot be able to generalize masked face and a face mask separately.

Analysing detection algorithms on occluded faces are often not performed. Few algorithms like MTCNN, RetinaFace and DLIB are well applicable to train datasets with occluded faces. To evaluate how well the model is trained with the dataset, different parts of the face are removed as pixel cells such as the landmarks of eyes, nose and mouth [7]. Also, face inside another face cannot be detected with the previously mentioned MTCNN algorithm.

Lightweight and robust monitoring system capable of surveillance 24×7 is required for effective control of coronavirus spread. Object detection, clustering and CNN based effective system has been proposed to monitor person detection, face mask detection and social distancing violation detection [8–10]. DBSCAN, DBSD and MobileNetV2 based classifiers have been employed in the system.

Covid spread has resulted in numerous infected and deaths in several parts of the world. Face mask can be one way to prevent the pandemic spread. Three different machine learning algorithms namely KNN, SVM and MobileNet has been studied to find the best algorithm for detecting face mask in real-time situation [11]. It is found that MobileNet performs well with both images and real-time video.

Mask can be a way to epidemic control and also filter pathogenic particles present in the air. In order to reduce the infection rate of viruses, it is essential to have a monitoring system to observe people wearing masks in public places. SSD-Mask algorithm [12–14] has been proposed which utilises attention mechanism to retain information of different feature levels and also optimize the loss function.

Different organizations have been desperate to overcome the viral coronavirus spreading all around the world. It is important to have a safety shield to prevent human society from coronavirus infection [12]. CNN and VGG16 based models have been proposed to enforce AI techniques for effective face mask detection on the Simulated Mask Face Dataset (SMFD) [15].

Maintaining social distancing and personal hygiene is crucial in this pandemic situation with the use of face masks in public places. Few object detection methods like convolution neural networks can be applied for effective face mask detection. Tiny-YOLOV4 is one of the lightweight method to detect face masks with low resources [16,17].

3 Methodology/Pipeline

3.1 Dataset

The input dataset used in this study is a face mask detection dataset obtained from the Kaggle repository which is maintained by Larxel. The dataset contains pictures of people wearing medical masks and XML files containing their descriptions and masks. There are 853 images belonging to 3 classes, as well as their bounding boxes in PASCAL VOC format [18]. The classes available in this dataset are: with mask, without mask and mask worn incorrectly. The dataset contains annotations for all the images in 3 different classes. Having images in 3 different classes and variation in the types of images make it better for the model to classify unknown image and so, we have picked to work on this dataset.

3.2 Data Preprocessing

The dataset obtained from the Kaggle repository contained a lot of noise, which required pre-processing before training the model. As the accuracy of any trained model depends on the dataset chosen, the above mentioned dataset was used for this project. The images in the dataset were then processed before feeding the model to train. This part describes the preprocessing process of the dataset and how the training of data. First, we define a function which takes the dataset as input and converts the XML input annotations to dictionaries and also resizes the images present in the dataset. Later, the images are converted into tensors. After this, image augmentation technique is applied to increase the accuracy during the model training [11,12].

3.3 Data Augmentation

The training of Faster R-CNN model requires a huge amount of data to perform model training effectively. Data augmentation is a technique used to provide sufficient amount of data for the model to train so that the model can generalize unseen data more effectively. The images undergo various methods like rotation, zooming, shifting, shearing and flipping to generate a large number of versions of the same image. In this project, image augmentation has been implemented for the data augmentation process and it is included in the preprocessing stage.

3.4 Classification of Images Using Faster R-CNN

Faster R-CNN is a deep neural network algorithm that is implemented for object detection problems. Pretrained ResNet50 FPN has been loaded into Faster R-CNN model. After getting the input features of the classifier from the pretrained model, the pretrained head is replaced with a new one [18–20]. The new trainable layers are added, and these layers are trained on the input dataset provided so that the model can decide the features to classify the faces with mask, without mask or mask worn incorrectly. The model is then fine-tuned and the relative weights are saved. Computational costs are saved and already learned features are saved when pretrained models are applied (Fig. 1).

3.5 Building Blocks of Faster R-CNN

Faster R-CNN is a deep learning model which is based on Convolutional Neural Network and consists following layers and functions.

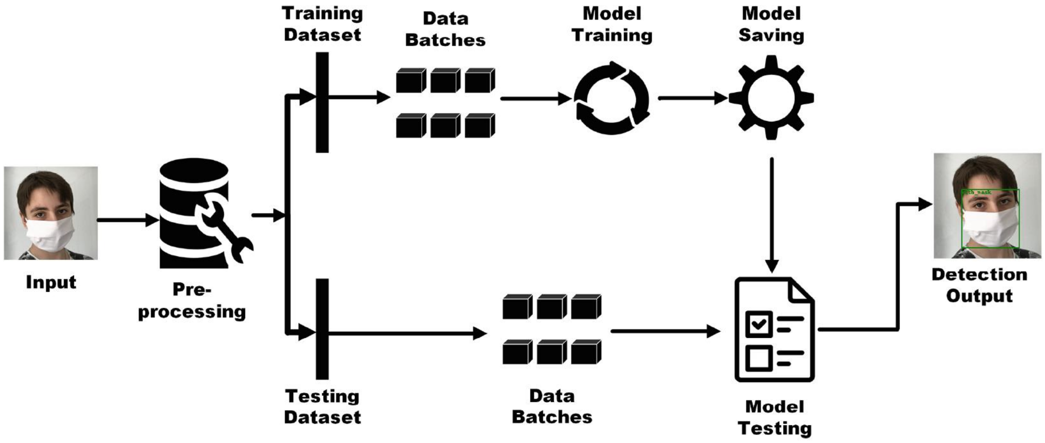

Fig. 1. Flow diagram of proposed model

- ***Convolutional Layer***
  The fundamental element of the Convolutional Neural Network is a convolution layer. It works on the principle of sliding window mechanism. Convolutions are basically a filter that transforms input image into feature map. The filter is often known as a kernel or a feature detector. The kernel is generally 3 × 3 matrix. However, it can be 5 × 5 or 7 × 7 matrix as well. The kernel is mapped over the input image which convolves all the pixels in the input image and outputs feature maps of the input image. Mathematically, the feature map is obtained from input image X with the convolution kernel Y [13] as follows:
  $$(X * Y)(t) = \int_{-\infty}^{\infty} X(T) * Y(T-t)dT \tag{1}$$
- ***Region Proposal Network***
  A good number of proposals are generated from the anchor boxes using Region Proposal Network (RPN). The feature maps generated by the kernel are taken as input by the RPN and outputs two convolution layers. One of the layers provides bounding box regressor outputs which predicts the scales to apply over the anchors and improve predictions [18]. The other layer gives classification outputs as the probability of bounding boxes belonging whether to the background or foreground. The loss function of RPN is a combination of both the classification loss and regression loss.
- ***Anchors***
  Anchors are bounding boxes of different sizes and ratios that are predefined and used as a reference for object detection using RPN. Anchor boxes are mainly used to find the scale and aspect ratio of objects required to be detected and it depends on the sizes of objects in the training set [14,15,17]. Originally, there are 3 scales and 3 aspect ratios which makes k=9.
- ***Non-Maximum Suppression***
  Non-Maximum Suppression (NMS) is a filtering procedure for predicting object detectors. This technique focuses on eliminating bounding boxes that are overlapping by filtering on the basis of IOU threshold of each bounding

box. IOU is mainly taken to find out the overlap between any two proposals generated by the RPN.

- ***RoI Pooling***
  Region of Interest (RoI) pooling is a feature mapping layer in region based deep convolutional neural networks. The proposed regions obtained from RPN vary in different sizes and thus RoI pooling is applied over the proposed regions. Having different sizes of regions result in different feature maps. RoI pooling reduces all the feature maps into same size [14,18,20]. The input feature map is divided into n number of equal regions and Max-Pooling is applied over the divided regions which results in same number of n despite the variation in size of input.

- ***Fully-Connected Layer***
  These layers are usually added towards the final layer of a deep neural network and they have connections to the activation layers as well. In Faster R-CNN, there are two fully-connected layers in the final layer. One of the fully-connected layer gives probability of the bounding boxes while the other layer gives probability of the object category. For object categorisation, the second fully-connected layer uses *softmax* activation function. Mathematically, softmax activation function can be given as:

$$softmax(x_i) = \frac{exp(x_i)}{\sum_j exp(x_j)} \tag{2}$$

4 Implementation

This project has been implemented with a cloud computing platform. GPU (1xTeslaK80) with 2496 CUDA cores and 12GB GGGR5 VRAM has been utilised to implement the the proposed approach in Google Colab. Parallel processing of all the images was achieved with the help of GPU which reduced the training time. The hyperparameters provided for building the model can be found in Table 1.

Table 1. Hyperparameters for the model

Hyperparameters	Value
Epochs	30
Batch size	4
Optimizer	Adam
Initial learning rate	1e-4

4.1 Performance Metrics

Once a model has been trained, it is equally important to evaluate how the model performs on unseen data. Among various model evaluation techniques, selection of suitable metrics depend on the category of task to be performed. Few terms used in model evaluation are:

- ***Intersection Over Union (IOU):*** It is based on jaccard index evaluating the overlap between two bounding boxes. However, both the ground truth bounding box and predicted bounding box is required. It determines whether a detection is positive or negative [20]. If A_p and A_{gt} are area of predicted box and ground truth bounding box respectively, then IOU can be given as:

$$IOU = \frac{A_p \cap A_{gt}}{A_p \cup A_{gt}} \tag{3}$$

- ***True positives (TP):*** IOU Detection $\geq$ *threshold* is referred as a true positive.

- ***True negatives (TN):*** All possible bounding boxes that are incorrectly detected are referred to as true negatives and so, it is not considered by the metrics.

- ***False positives (FP):*** IOU Detection $<$ *threshold* is referred as a true negative.

- ***False negatives (FN):*** Whenever a ground truth is not detected, it is referred as a false negative.

- ***Precision:*** Precision is the ratio of true positives to all detections. Mathematically, it can be given as:

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{all\ detections} \tag{4}$$

- ***Recall:*** Recall is the percentages of true positives detected among all ground truths. It can be calculated as:

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{all\ ground\ truths} \tag{5}$$

- ***Average Precision:*** Average precision is the average of accuracy over recall values from 0 to 1 differing in IOU thresholds.

- ***Mean Average Precision:*** Mean average precision is the mean value of all the average precision taken over all classes and IOU thresholds.

5 Results

Face mask detection is an important procedure to be followed in this new normal of coronavirus pandemic. The testing set created from the face mask dataset containes 150 images and has been used for the model evaluation in this project. For the detection, our model achieved a mean Average Precision (mAP) of 44.9% at 30 epochs on the test dataset. Table 2 shows the loss values and loss classifier values at different epochs from 5 till 30.

The mean Average Precision (mAP) and Average Precision (AP) at three different thresholds of 50%, 75% and 90% have been noted and shown in Table 2.

Table 2. mAP and AP values with loss classifier results

Epochs	Loss	Loss Classifier
5	0.1427	0.0391
10	0.0906	0.0274
15	0.0694	0.0264
20	0.0765	0.0257
25	0.0619	0.0232
30	0.0563	0.0221

(a) Loss and loss classififer values at different epochs

Epochs	mAP	AP_{50}	AP_{75}	AP_{90}
5	0.461	0.683	0.546	0.053
10	0.459	0.659	0.581	0.059
15	0.453	0.660	0.532	0.036
20	0.454	0.660	0.451	0.043
25	0.455	0.660	0.544	0.043
30	0.455	0.660	0.544	0.044

(b) mAP and AP at thresholds of 50,75 and 90 at different epochs

Few prediction outputs of the implemented model can be seen in Fig. 2.

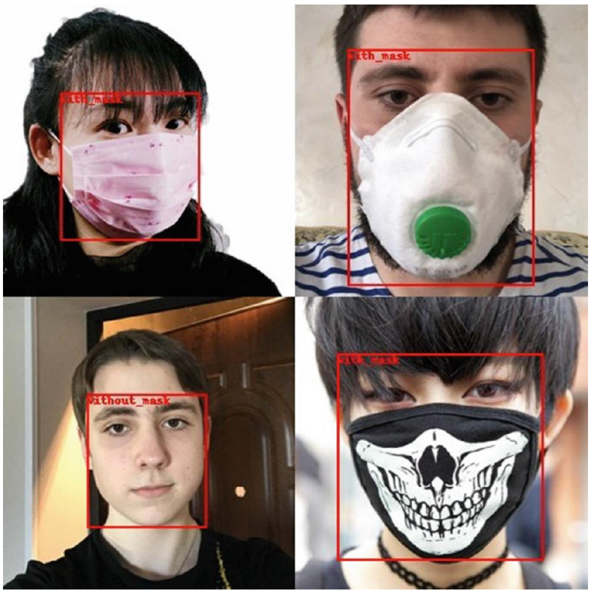

Fig. 2. Prediction outputs

6 Future Work

As different modern methods and algorithms are being implemented for better detection and recognition to reduce the covid pandemic spread, deep learning based segmentation techniques could be used with this dataset to test further.

Image forgery techniques could also be used for better detection of face masks in public places. Few algorithms like Generative Adversial Networks (GAN) and Granulated R-CNN (GRCNN)can be tested to detect face masks more efficiently. Algorithms like GAN and GRCNN could be implemented for better object detection and may enhance the efficiency of face mask detection in this covid pandemic scenario all around the world.

7 Conclusion

With the rise in number of cases of SARS-CoV-2 virus, also called the Coronavirus, wearing a face mask and maintaining social distancing is a new normal. However, gathering of people in public places makes it more easy for the virus to spread all around. Face mask detection is an important factor to reduce the infection of coronavirus. ResNet-50 used as a RPN in this project generates better region proposals for detection and softmax activation function performs classification at the final fully connected layer. The Faster R-CNN algorithm used on this dataset [18] provides a good mean average precision (mAP) of 45%. The implemented model achieves satisfactory performance in face mask detection. Possible future work on this project is to extend the project with GAN-based and GR-CNN algorithms in detecting multiface images in real-time.

References

1. Pooja, S., Preeti, S.: Face mask detection using AI. In: Khosla, P.K., Mittal, M., Sharma, D., Goyal, L.M. (eds.) Predictive and Preventive Measures for Covid-19 Pandemic. AIS, pp. 293–305. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4236-1_16
2. Chavda, A., Dsouza, J., Badgujar, S., Damani, A.: Multi-Stage CNN architecture for face mask detection. In: 2021 6th International Conference for Convergence in Technology (I2CT), pp. 1–8 (2021). https://doi.org/10.1109/I2CT51068.2021.9418207
3. Dhaya, R.: Efficient two stage identification for face mask detection using multiclass deep learning approach. J. Ubiquit. Comput. Commun. Technol. **3**(2), 107–121 (2021)
4. Venkateswarlu, I.B., Kakarla, J., Prakash, S.: Face mask detection using mobilenet and global pooling block. In: 2020 IEEE 4th Conference on Information & Communication Technology (CICT) (pp. 1–5). IEEE (2020)
5. Prinosil, J., Maly, O.: Detecting faces with face masks. In: 2021 44th International Conference on Telecommunications and Signal Processing (TSP), pp. 259–262 (2021). https://doi.org/10.1109/TSP52935.2021.9522677
6. Bu, W., Xiao, J., Zhou, C., Yang, M., Peng, C.: A cascade framework for masked face detection. In: 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), pp. 458–462 (2017). https://doi.org/10.1109/ICCIS.2017.8274819
7. Hofer, P., Roland, M., Schwarz, P., Schwaighofer, M., Mayrhofer, R.: Importance of different facial parts for face detection networks. IEEE Int. Workshop Biometrics Forensics (IWBF) **2021**, 1–6 (2021). https://doi.org/10.1109/IWBF50991.2021.9465087

8. Srinivasan, S., Rujula Singh, R., Biradar, R.R., Revathi, S.: COVID-19 monitoring system using social distancing and face mask detection on surveillance video datasets. In: 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), pp. 449–455 (2021). https://doi.org/10.1109/ESCI50559.2021.9396783
9. Dondo, D.G., Redolfi, J.A., Araguás, R.G., Garcia, D.: Application of deep-learning methods to real time face mask detection. IEEE Latin America Trans. **19**(6), 994–1001 (2021)
10. Varshini, B., Yogesh, H.R., Pasha, S.D., Suhail, M., Madhumitha, V., Sasi, A.: IoT-Enabled smart doors for monitoring body temperature and face mask detection. Global Transitions Proc. **2**(2), 246–254 (2021)
11. Vijitkunsawat, W., Chantngarm, P.: Study of the performance of machine learning algorithms for face mask detection. In: 2020–5th International Conference on Information Technology (InCIT), pp. 39–43 (2020). https://doi.org/10.1109/InCIT50588.2020.9310963
12. Xu, M., Wang, H., Yang, S., Li, R.: Mask wearing detection method based on SSD-Mask algorithm. In: International Conference on Computer Science and Management Technology (ICCSMT), vol. 2020, pp. 138–143 (2020). https://doi.org/10.1109/ICCSMT51754.2020.00034
13. Goyal, H., Sidana, K., Singh, C., Jain, A., Jindal, S.: A real time face mask detection system using convolutional neural network. Multimedia Tools Appl. **81**(11), 14999–15015 (2022)
14. Lodh, A., Saxena, U., Motwani, A., Shakkeera, L., Sharmasth, V.Y.: Prototype for integration of face mask detection and person identification model-COVID-19. In: 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1361–1367). IEEE (2020)
15. A. Negi, K. Kumar, P. Chauhan and R. S. Rajput, "Deep Neural Architecture for Face mask Detection on Simulated Masked Face Dataset against Covid-19 Pandemic," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 595–600, https://doi.org/10.1109/ICCCIS51004.2021.9397196
16. Sathyamurthy, K.V., Shri Rajmohan, A.R., Ram Tejaswar, A., Kavitha, V., Manimala, G.: Realtime face mask detection using TINY-YOLO V4. In: 2021 4th International Conference on Computing and Communications Technologies (ICCCT), pp. 169–174 (2021). https://doi.org/10.1109/ICCCT53315.2021.9711838
17. Kumar, T.A., Rajmohan, R., Pavithra, M., Ajagbe, S.A., Hodhod, R., Gaber, T.: Automatic face mask detection system in public transportation in smart cities using IoT and deep learning. Electronics **11**(6), 904 (2022)
18. Mask dataset. https://makeml.app/datasets/mask
19. Sethi, S., Kathuria, M., Kaushik, T.: Face mask detection using deep learning: an approach to reduce risk of Coronavirus spread. J. Biomedical Informat. **120**, 103848 (2021)
20. Shrestha, H., Dhasarathan, C., Kumar, M., Nidhya, R., Shankar, A., Kumar, M.: A deep learning based convolution neural network-DCNN approach to detect brain tumor. In: Gupta, G., Wang, L., Yadav, A., Rana, P., Wang, Z. (eds) Proceedings of Academia-Industry Consortium for Data Science. Advances in Intelligent Systems and Computing, vol 1411. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-6887-6_11