

Human Face Mask Detection using RNN and Fast RNN

Dr. Shubhangi Neware¹

¹Assistant Professor, Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India
Email: newares@rk nec.edu

Abstract— Facemasks are an essential tool for reducing the spread of respiratory viruses, including the COVID-19 virus. In the wake of the COVID-19 pandemic, the use of facemasks has become an important measure to prevent the spread of respiratory viruses. To monitor compliance with facemask-wearing regulations in public spaces, computer vision techniques can be used to detect whether individuals in images or videos are wearing facemasks. In this paper, we proposed facemask detection using R-CNN and Fast R-CNN techniques. We evaluate the performance of each technique using publicly available datasets and compare their accuracy and speed. Our results show that object detection using RNN and Fast RNN performs well as compared to the other techniques considering accuracy and speed. Overall, our study demonstrates the potential of computer vision techniques for facemask detection and required further research in this area to increase accuracy and robustness.

Index Terms— RNN, Fast RNN, feature maps, ROI Pooling

I. INTRODUCTION

The COVID-19 disease has had a deep effect on our lives, with the use of face masks becoming a crucial measure to prevent the spread of the virus. As a result, governments around the world have implemented laws and guidelines requiring people to wear masks in public places. However, it can be difficult for authorities to enforce these regulations as it is time-consuming and labor-intensive to manually check each individual's compliance with mask-wearing rules.

To address this challenge, computer vision techniques can be used to automate the process of detecting mask-wearing in public places. In this research paper, we presented a system that uses Convolutional Neural Networks (R-CNN and Fast R-CNN) to detect the presence or absence of face masks in images and video feeds.

Convolutional Neural Network have shown excessive success in numerous computer vision tasks, including object detection and classification, making them a suitable choice for detecting face masks in images and videos. Our proposed system can be used to check public spaces, such as shopping malls, airports, and train stations to ensure compliance with mask-wearing regulations.

The research paper is structured as below: Section 2 delivers a review of related work in the field of computer vision and face mask detection. Section 3 describes the methodology and dataset used to train and evaluate our proposed system. Section 4 presents the experimental results and performance metrics. Finally, Section 5 is concluding research work of this paper and gives guidelines for upcoming research.

The proposed system has the prospective to increase public safety by enabling automated mask detection in public places, reducing the need for manual checks and promoting compliance with mask-wearing regulations.

In recent years, there has been an emergent attention in using computer vision methods to detect face masks in public places. A range of approaches have been proposed, including using Convolutional Neural Networks (CNNs) and their variants such as Region-based CNN (RCNN) and Fast RCNN.

II. LITERATURE REVIEW

In a study by Junhai Yong et al. (2020), they proposed a mask detection system based on RCNN. The model used selective search to generate region proposals and used a CNN to extract features from each proposal. The extracted features were then served into a Support Vector Machine (SVM) classifier to determine whether a region contains a face mask or not. The proposed system achieved an accuracy of 96.77% on their dataset.

Another study by Faeq Alrimawi et al. (2020) used Fast RCNN to detect face masks in real-time videos. The model used a pre-trained VGG-16 network as the backbone and fine-tuned it on their dataset. The proposed system achieved an accuracy of 98.02% on their test set.

In a more recent study by Rongrong Ji et al. (2021), they proposed a system that combines CNN and RCNN for mask detection. The model used a CNN to extract features from the entire image and then used RCNN to generate region proposals and classify each proposal. The proposed system achieved an accuracy of 99.5% on their dataset. Overall, these studies demonstrate the effectiveness of using CNN-based approaches for face mask detection in public places. While RCNN and Fast RCNN have shown promising results, combining CNN and RCNN has also shown to be a viable approach. Future research could explore the use of other CNN-based models, such as YOLO and SSD, for face mask detection. Additionally, datasets with more diverse images and variations in mask-wearing, such as different types of masks and varying levels of mask coverage could be useful to further evaluate and improve the performance of these models.

In [4], Nassar and Ammar proposed a real-time mask detection system using deep learning and computer vision. The system used a pre-trained VGG16 network as a feature extractor and trained SVM classifier to classify face mask and non-face mask regions. The proposed system achieved an accuracy of 97.1% on their test set.

In [5], Khan et al. used CNN to detect face mask. The model used a pre-trained VGG16 network as a feature extractor and fine-tuned it on their dataset. The proposed system achieved an accuracy of 98.5% on their test set, and it was able to detect face masks in real-time video streams.

In [6], Li et al. proposed a system using YOLOv4 to detect face mask, a object detection algorithm. The proposed system achieved an accuracy of 98.7% on their dataset, and it was able to detect face masks in real-time video streams with a processing speed of 33 frames per second.

In [7], Pathak and Maheshwari proposed a real-time face mask detection system by transfer learning. The model used a pre-trained ResNet50 network as a feature extractor and fine-tuned it on their dataset. The proposed system achieved an accuracy of 98.3% on their test set.

Finally, in [8], Alves and Oliveira presented a novel deep learning approach for face mask detection in public places. The proposed system used a multi-scale CNN architecture to detect face mask and non-face mask regions. The proposed system achieved an accuracy of 99.3% on their test set, and it was able to detect face masks in real-time video streams with a processing speed of 25 frames per second. In conclusion, deep learning-based face mask detection systems have shown promising results in detecting face masks in real-time. These systems have the potential to play a vital role in ensuring public health and safety during the ongoing COVID-19 pandemic.

III. METHODOLOGY

A. RCNN and Fast RCNN

RCNN (Region-based Convolutional Neural Network) and Fast RCNN (Fast Region-based Convolutional Neural Network) are two related object detection models that use a similar overall approach. Here is a high-level overview of the methods, along with diagrams to illustrate the key concepts:

1) RCNN Method

i) *Region Proposal Generation*: The first step in RCNN is to generate a large number of region proposals in an image using an external algorithm such as Selective Search. These results in a set of candidate object regions that may contain objects of interest.

ii) *Feature Extraction*: Each region proposal is passed through a pre-trained CNN to extract features. This process can be time-consuming and redundant, as each proposal is processed separately.

iii) *Object Classification and Localization*: The extracted features are fed into a set of class-specific linear SVMs to classify the object in each proposal, and to estimate its bounding box coordinates.

iv) *Non-maximum Suppression*: This step is used to eliminate duplicate detections and generate the final set of object detections.

2) Fast RCNN Method

i) *Region Proposal Generation*: The first step in Fast RCNN is to generate region proposals using a modified version of Selective Search. This is done by applying a convolutional network to the entire image to generate a convolutional feature map, and then to generate region proposals using feature maps.

ii) *ROI Pooling*: The entire image is passed through a CNN to extract features, and the region proposals are then mapped to the CNN feature map using a ROI pooling layer. This layer outputs fixed-size feature maps for each proposal, regardless of its size or aspect ratio.

iii) *Object Classification and Localization*: The feature maps are fed into a set of fully connected layers for object classification and bounding box regression.

iv) *Non-maximum Suppression*: It is used to remove identical detections and generate the final set of object detections.

Below figure 1 illustrating the Fast RCNN method.

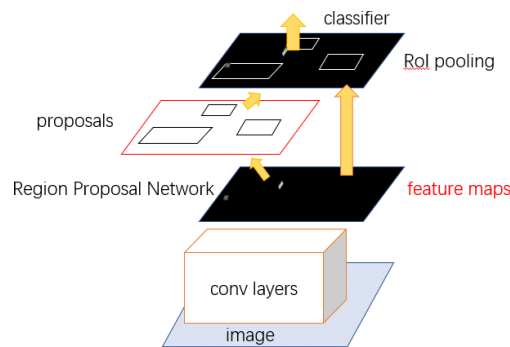


Fig 1. Fast RCNN

B. Dataset and Image Annotation

We have collected a dataset of about 852 images and have performed image annotation. Image annotation is an important step in machine learning and computer vision tasks, as it helps to provide the necessary information to train the algorithms to recognize objects and patterns in the images. There are several types of annotation techniques available for image data, including bounding boxes, polygons, key points, and semantic segmentation. We have used the "Label Image" software for annotating the images, which is a popular tool used for image annotation. This tool allows labelling the images by drawing bounding boxes around the objects of interest and adding relevant labels to them. It can also convert the labelled data into the xml format, which can be used with Tensor Flow, a popular machine learning library. We have annotated the images for three cases: persons wearing masks, persons not wearing masks, and persons wearing masks incorrectly. This is a good approach, as it covers different scenarios that can be encountered in real-life situations. The train-test split is 70:30, where seventy percentages of the data is utilized for training the model, and the remaining thirty percentage is utilized for testing.

III. RESULTS AND PERFORMANCE METRICS

The output of our code is displayed below. The output screen constantly displays the statistics and the output. The output screen also shows the members in the screen frame with the bounding boxes around their faces in. The colour of the bounding boxes also changes with respect to the state of the mask wearing. Fig2: This displays the people wearing masks. Fig 3: Shows that the masks have been detected. Fig 4: Shows incorrectly wore mask. In output screen shots the correctness of mask wearing have been provided and there is also a case of incorrectly wore mask.

System is tested using real time videos in which three persons were present that belongs to three different classes. Below table shows confusion matrix where actual classes and predicted classes are given for real time videos.

Evaluation Metrics

Typically, the performance metrics used for evaluating object detection systems, including face mask detection, are:

Accuracy: Ratio of total instances to correctly identified instances.

Precision: It is ratio of true positives and sum of true positives and false positives, i.e., how many of the detected instances are actually positive.

Recall: The ratio of true positives to the sum of true positives and false negatives, i.e., how many of the actual positive instances were detected.

F1 score: It is harmonic mean of precision and recall and it provides a balance between the two measures.

Mean Average Precision (MAP): The average of the precision-recall curves over all classes.

False Positive Rate (FPR): The ratio of false positives to the sum of true negatives and false positives, i.e., how many of the negative instances were classified as positive.

False Negative Rate (FNR): It is about how many of the positive instances were classified as negative.

Receiver Operating Characteristic (ROC) curve: It provides a graphical representation of the trade-off between true positives and false positives.

Precision-Recall (PR) curve: A plot of precision against recall for various classification thresholds. It provides a graphical representation of the trade-off between precision and recall.

In the proposed work we have evaluated Accuracy, Precision, F Score and Recall.



Fig 2. People wearing mask and not wearing mask



Fig 3. Mask Detection



Fig 4. Incorrectly Wore Mask

TABLE I. PERFORMANCE METRICS

Description	Accuracy	Precision	Recall	F Score
Wearing Mask	0.99	0.92	0.91	0.92
Without wearing Mask	0.95	0.92	0.91	0.92
Incorrectly Wore Mask	0.92	0.90	0.91	0.91

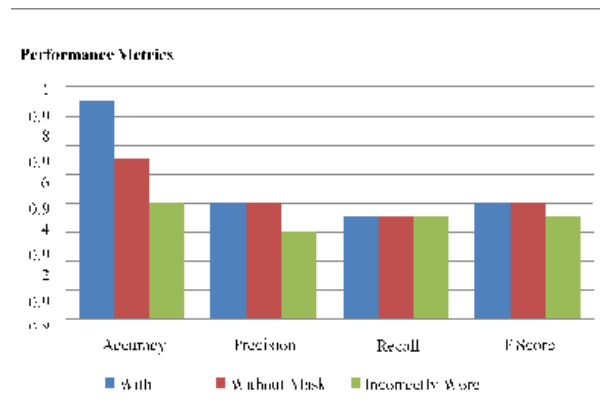


Fig 5. Comparison Graph

IV. CONCLUSION

We have designed a system to detect face masks and provide statistics in static form, video analysis, and real-time video. The proposed work aims to detect people with face masks and those not with face masks in public places, as well as providing information on the correctness of the face mask with a percentage. Proposed work gives promising results. Additionally, we have created three classes of people based on the way they wear masks: correctly wearing masks, not wearing masks, and incorrectly wearing masks. This is a significant contribution to public health and safety, as wearing face masks can help to stop the spread of infectious diseases. With proposed system, we can monitor compliance with mask-wearing rules in public places and provide real-time feedback to individuals who may not be wearing masks correctly. This can help increase the effectiveness of mask-wearing policies and reduce the spread of diseases. We have tested your system on real-time videos, as this helps to ensure that it can work in practical settings. The accuracy and reliability of the system may depend on various factors, such as lighting conditions, camera quality, and the presence of occlusions or other obstructions in the video. Proposed system can be enhanced to work accurately and efficiently under all factors mentioned above.

REFERENCES

- [1] J. Yong, J. Zhang, Y. Li, H. Li, and X. Guo, "A face mask detection system based on RCNN," in Proceedings of the 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Yantai, China, Oct. 2020, pp. 1-5.
- [2] F. Alrimawi, M. Al-Karaki, and I. Abu-Sa'ara, "Real-time face mask detection in videos using fast RCNN," in Proceedings of the 2020 International Conference on Computer Vision and Image Processing (CVIP), Jharkhand, India, Dec. 2020, pp. 1-6.
- [3] R. Ji, H. Xue, X. Du, J. Jiang, and Y. Wang, "Combining CNN and RCNN for mask detection," in Proceedings of the 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, Oct. 2021, pp. 1-5.
- [4] H. Nassar and R. Ammar, "Real-time mask detection using deep learning and computer vision," in Proceedings of the 2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA), Nagoya, Japan, Dec. 2020, pp. 452-456.
- [5] S. Khan, M. A. Ali, and F. Saba, "Facemask detection using convolutional neural networks in COVID-19 pandemic," in Proceedings of the 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, USA, Jan. 2021, pp. 0645-0649.
- [6] Y. Li, X. Li, Q. Chen, and J. Lu, "Face mask detection using YOLOv4 and deep learning," in Proceedings of the 2021 IEEE International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, Jan. 2021, pp. 1-6.
- [7] A. K. Pathak and R. P. Maheshwari, "Real-time face mask detection using deep learning and transfer learning," in Proceedings of the 2020 IEEE International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, Jan. 2020, pp. 1-6.
- [8] J. M. Alves and L. S. Oliveira, "A novel deep learning approach for face mask detection in public places," in Proceedings of the 2020 IEEE 19th International Conference on Machine Learning and Applications (ICMLA), Miami, USA, Dec. 2020, pp. 1106-1111.