Deep Learning Skeleton for Detecting Face Masks from Video Footage

Dr. Savya Sachi[1],Dr Santosh Kumar[2], Mrs. Atmaprabha[3]

[1]Assistant Professor, Department of Information Technology
L. N. Mishra College of Business Management, Muzaffarpur, Bihar, India
[2]Associate Professor, CSE- Internet of Things
Noida Institute of Engineering and Technology, Greater Noida,India
Assistant professor, Department of information technology,
L. N. Mishra College of Business Management, Muzaffarpur, Bihar

savyasachilnmcbm@gmail.com

Abstract- since the COVID-19 global epidemic, the usage of facial masks in public places has become a societal duty, and the identification of facial masks might be critical to protect public safety. Face mask recognition in video footage is a difficult task, owing to the fact that the masks themselves act as occlusions to face detection algorithms due to the lack of facial landmarks in the masked regions. In this paper, we offer a deep learning-based method for recognizing facial masks in films. The suggested framework makes use of the MTCNN face detection model to recognize faces and the facial landmarks that correspond to them that are visible in the video frame. A neoteric classifier uses the MobileNetV2 architecture as an object detector for finding masked regions to analyze these facial images and cues. The suggested framework was evaluated using a dataset made up of films documenting how individuals move about in public areas while adhering to COVID-19 safety standards. By achieving excellent precision, recall, and accuracy, the suggested methodology demonstrated its efficacy in identifying facial masks.

Keywords – *Deep Learning, Skeleton, Detecting Face, Masks, Video Footage,* COVID-19, *Global Epidemic.*

INTRODUCTION

Face identification has been greatly aided in recent years by the ever-rapid development of machine learning algorithms and approaches. For instance, even when detecting several faces, the face detection algorithm suggested in achieves a precision of 93%. Numerous applications, including real-time face recognition systems, security surveillance systems, etc., have been created as a result of the development of facial detectors. Even if these methods are effective, there is a growing need for the creation of reliable and effective face detection models. For existing face detection techniques, it is very difficult and difficult to recognise masked faces for a variety of reasons. First of all, classic face detection algorithms rely on manually extracted information. The Viola Jones face detector extracts facial features by combining Haar features with the integral pictures approach. Utilising the Histogram of Gradients (HOG), Fast Fourier Transform (FFT), and Local Binary Patterns (LBP) are other feature extraction strategies. Deep learning has advanced to the point where neural networks can now create feature extractors like the You Only Look Once (YOLO) method by learning features without using prior information. The pressing concern with the aforementioned approaches when it comes to face mask detection is that the face masks, with their visual diversity and various orientations behave as occlusions and variable noise to the models. This leads to a lack of local facial features, resulting in the

failure of even state-of-the-art face detection models. Moreover, there is a lack of large datasets with labeled images of faces with facial masks required in order to analyse the vital characteristics common to masked faces, thus accounting for the low accuracy of existing models. These factors together justify the challenging nature of masked face detection in the field of image processing. During the COVID-19 pandemic, everyone is advised to wear face masks in public. According to the World Health Organization (WHO), masks can be used for source control (worn by an infected individual to inhibit further transmission) or for the protection of healthy people. At the time of writing, the global pandemic has infected over 11 million people worldwide and has led to over half a million casualties. The wide-scale usage of face masks poses a challenge on public face detection based security systems such as those present in airports, which are unable to detect facial masks. Since the improper removal of masks can lead to contracting the virus, it has become essential to improve facial detectors that rely on facial cuts, so that detection can be performed accurately even with inadequately exposed faces.

LITERATURE REVIEW

We review some similar works done in this domain. although research on face detection has been going on for decades and has achieved great success, algorithms and methodologies that are earmarked for face mask detection are limited. Ge et al. developed a deep learning methodology to detect masked faces using LLE-CNNs, which outperforms state-of-the-art detectors by at least 15%. In the given work, the authors introduced a new dataset called MAsked Faces (MAFA), containing 35,806 images of masked faces having different orientations and occlusion degrees. The proposed LLE-CNNs consist of three modules - proposal module, embedding module and verification module. The proposal module first combines two CNNs to extract candidate facial regions from the input image and represents them with high dimensional descriptors. After that, the embedding module is turns these descriptors into similarity based descriptors using Locally Linear Embedding algorithms and dictionaries trained on a set of faces, comprised of masked and unmasked images. Finally, the verification module is used to identify candidate facial regions and refine their positions with the help of classification and regression tasks. Nair et al. utilised the Viola Jones object detection framework to detect masked faces in surveillance videos. The authors argued that detecting cosmetic components such as face masks takes a significantly longer period than face detection. The framework uses the Viola Jones face detection algorithm to detect the eyes and face of subjects. If eyes are recognised and later the face is recognised as well, it signifies that no face mask was used. However, if eyes are recognised but the face is not, it signifies that a face mask was worn by the person in consideration. Bu et al. built a CNN-based cascaded face detector framework, consisting of three convolutional neural networks. The first CNN, Mask-1 is a very shallow fully convolutional layer network with 5 layers that gives a probability of being a masked face for each detection window, followed by a Non-maximum Supression (NMS) to merge overlapping candidates. Mask-2 is a deeper CNN with 7 layers, which resizes the candidate windows and also sets a detection threshold from the previous CNN. Mask-3 is also a 7 layer CNN which resizes the input windows it receives and gives a likelihood of whether it belongs to a masked face based on a preset threshold. After NMS, the remaining detection windows are the predicted detection results. Coming to more recent methodologies, Jiang et. al. developed Retina Face Mask, which is a novel framework for accurately and efficiently detecting face masks. The proposed framework is a one-stage detector

which consists of a feature pyramid network to combine high-level semantic data with numerous feature maps. The authors propose a novel context attention module for the detection of face masks in addition to a cross-class object removal algorithm that discards predictions with low confidence values. The authors state that their model performs 2.3% and 1.5% more than the baseline result in face and mask detection precision respectively, and 11.0% and 5.9% higher than baseline for recall.

THE METHOD'S DESCRIPTION

(i) Deep Neural Network Algorithm- DNN algorithm moves data through a sequence of "layers" of neural network models, with each layer passing a simplified summary of the data to the next layer. Several computer vision algorithms work well on datasets with a few hundred features or columns. An unstructured dataset, such as one extracted from an image, on the other hand, contains so many features that this method becomes inefficient or impossible. Traditional machine learning algorithms cannot handle 2.4 million parts in a single 800 × 1000 pixel RGB color image. As the image passes through each neural network layer, DNN algorithms learn more about it. Initial layers learn how to detect low-level features such as edges, and later layers incorporate these features into a more comprehensive representation. For instance, a middle layer would detect edges to detect parts of an object in an image, such as a leg or a branch, while a deep layer might identify the entire object, such as a dog or a tree. You gather data from observations and integrate it into a single layer. (e layer produces an output, which becomes the input for the following layer, and so on. (is loops until the final output signal is received.

(ii) Types of Algorithms- (there are several different types of feature extraction algorithms, which can be classified into two categories.

(a) Algorithms that Rely on Classification- (the regions of interest are chosen in the first stage. After that, convolutional neural networks (CNN) are used to categorize specific areas. Since prediction must be run for each selected field, this solution may be prolonged. (is group includes algorithms such as the fast RCNN and faster RCNN which are enhanced variants of the region-based Convolutional Neural Network (RCNN).

(b) Algorithms that Rely on Regression- In contrast to the previous approach, algorithms in this category predict the class probability and define the bounding boxes surrounding the object of interest in a single run from the entire image point of view. (is group includes algorithms like You Just Look Once (YOLO) and Single Shot Multibox Detector (SSD). Deep learning and computer vision are used in various applications such as objection detection, medical image analysis, and action recognition. Recent re-search is focused on the use of mid-level features and deep learning models to build robust decision support systems and IoT applications.

(iii) Faster RCNN- In object classification and recognition, a deep learning technique known as area of interest polling is gaining much attention. Detecting objects from an image scene containing several things is one example. (e goal is to extract fixed-size feature maps using maximum pooling on the entire picture as reflected in Figure 1. (e object detection technique used by faster RCNN is divided into three stages.

(a) Region Proposal Network- Finding the spaces in the given input image where there is a possibility of finding an object is straightforward. (e position of an entity in an image can be determined. (e area where there is a possibility of finding an object is surrounded by the Region of Interest (ROI).

(b) Classification- (the next step is to assign corresponding classes to the regions of interest defined in the previous actions. Here, the CNN approach is used (Figure 2). (e proposed approach includes a detailed process for identifying all spaces of object location in an image. If no regions are placed in the first stage of the algorithm, there is no need to move on to the second step. In 2015, Girschik proposed the Region Proposal Network (RPN) and ROI pooling as a DLA-based object detection solution. ROI can achieve speed and usability for both training and research performance. (the ROI layers take a feature map as input, which is the output of a convolution neural network with multiple convolution layers and max-pooling layers.

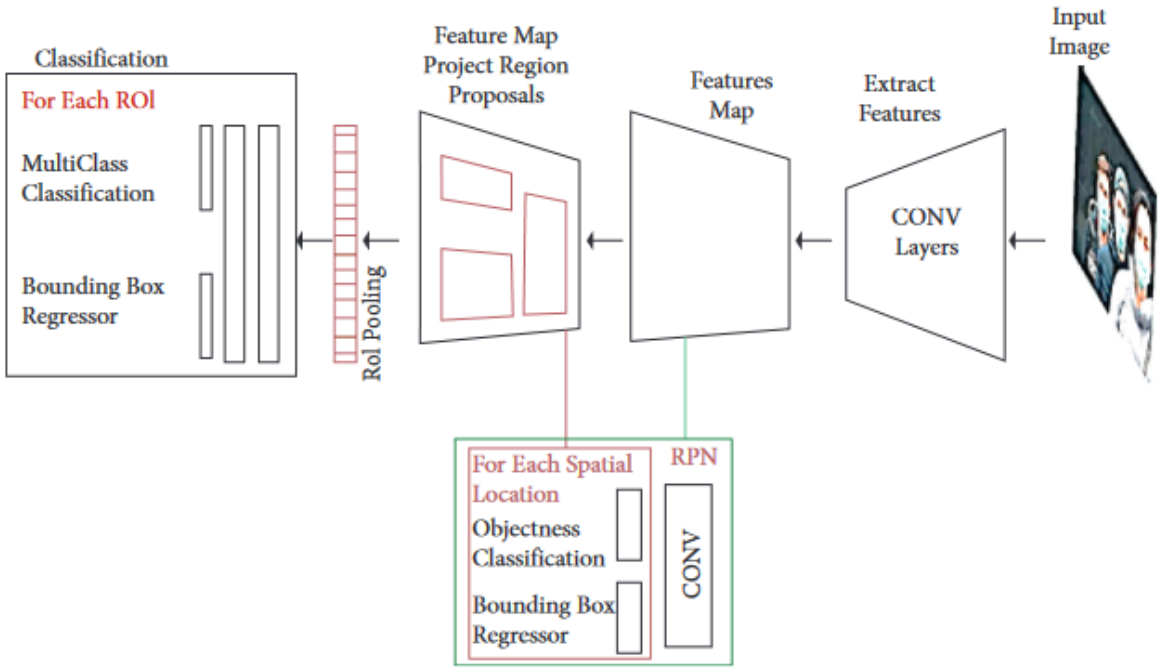

Figure 1- Faster RCNN

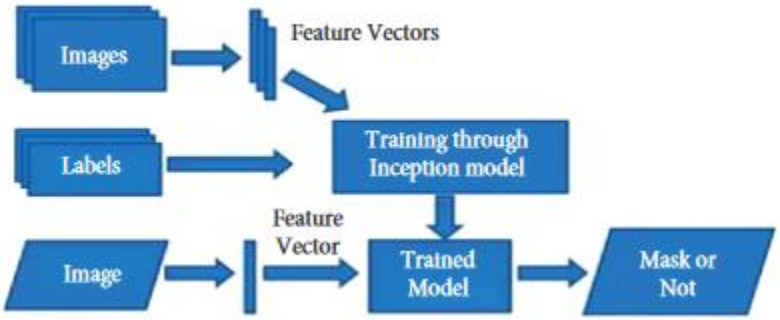

Figure 2- Model architecture

EXPERIMENTAL EVALUATION

We go over the dataset used for this study's analysis as well as the outcomes produced by the suggested methodology. On Google Colab, which has an Intel(R) Xeon(R) 2.00 GHz CPU, an NVIDIA Tesla T4 GPU, 16 GB of GDDR6 VRAM, and 13 GB of RAM, the experiments were

carried out. All of the programmes used OpenCV 4.2.0, Keras 2.3.0, and TensorFlow 2.2.0 and were written in Python 3.6.

(i) Dataset Used- The dataset used in this study is a compilation of YouTube films taken in public spaces across various geographic locations. The dataset contains a total of 15 video samples, each with an average runtime of one minute. The movies show how people move around in public spaces following the implementation of different safety guidelines and laws in the wake of the COVID-19 outbreak. The movies feature people of all races and also document various kinds of face masks worn by the general public. Our dataset includes videos that were recorded using several types of cameras and has a wide range of camera angles, noise, and frame rates (FPS) an average of 30. A few examples of the movies included in this collection are shown in Figure 3 and Figure 4.

(ii) Experimental Results and Statistics- The proposed method has been assessed by analysing the facial mask classifier's precision, recall, and accuracy metrics, accordingly.

Figure 4-Visualisation of the results obtained by the proposed approach

Figure 5- Some samples from the video dataset used in this work

CONCLUSION

We introduced a trustworthy DNN-based approach for mask identification in this study. This method used a quicker RCNN to train the data, which produced very high accuracy. In order to achieve a low computational cost, the model is trained on a GPU. We employed a multiphase detection model to accomplish our objectives. First, we labelled the face mask. Next, we detected the edge and computed edge projection for the selected face region within the face mask. According to the most recent research, quicker RCNN provided 97% accuracy while being effective and precise. After 200,000 epochs, the overall loss is 0.0503 and is trending downward. While the loss is declining, the results are becoming more precise. As a result, the training time was cut in half with improved accuracy using the quicker RCNN technique to determine whether or not someone is wearing a face mask. Future Deep Neural Network (DNN) implementations may train the data first, then compress the input dimensions to run on low-power devices at a lower computational cost.

REFERENCES

[1] T. Greenhalgh, M. B. Schmid, T. Czypionka, D. Bassler, and L. Gruer, "Face masks for the public during the covid-19 crisis," BMJ, vol. 369, p. 1435, 2020.

[2] S. Feng, C. Shen, N. Xia, W. Song, M. Fan, and B. J. Cowling, "Rational use of face masks in the COVID-19 pandemic," e Lancet Respiratory Medicine, vol. 8, no. 5, pp. 434–436, 2020.

[3] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: a review," IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212–3232, 2019.

[4] A. Kumar, Z. J. Zhang, and H. Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm," EURASIP Journal on Wireless Communications and Networking, vol. 2020, no. 1, 204 pages, 2020.

[5] L. Jiao, F. Zhang, F. Liu et al., "A survey of deep learning-based object detection," IEEE Access, vol. 7, Article ID 128837, 2019.
[6] Z. Li and Wu, "Efficient object detection framework and hardware architecture for remote sensing images," Remote Sensing, vol. 11, no. 20, p. 2376, 2019.
[7] H. Schneiderman and T. Kanade, "A statistical method for 3D object detection applied to faces and cars,"vol. 1, pp. 746–751, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662), vol. 1, pp. 746–751, IEEE, Hilton Head, SC, USA, June 2002.
[8] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in Proceedings of the International Conference on Image Processing, vol. 1, September 2003.
[9] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing features: efficient boosting procedures for multiclass object detection," in Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, June 2004.
[10] K. Hotta, "Robust face detection under partial occlusion," Systems and Computers in Japan, vol. 38, no. 13, pp. 39–48, 2007.
[11] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," International Journal of Computer Vision, vol. 61, no. 1, pp. 55–79, 2005.
[12] Y.-Y. Lin and T.-L. Liu, "Robust face detection with multiclass boosting,"vol. 1, pp. 680–687, in Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 680–687, IEEE, San Diego, CA, USA, June 2005.
[13] L. Goldmann, U. J. Monich, and T. Sikora, "Components and their topology for robust face detection in the presence of partial occlusions," IEEE Transactions on Information Forensics and Security, vol. 2, no. 3, pp. 559–569, 2007.
[14] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super- resolution via sparse representation," IEEE Transactions onImage Processing, vol. 19, no. 11, pp. 2861–2873, 2010.
[15] V. B. Subburaman and S. Marcel, Fast bounding box estimation based face detection,In ECCV, Workshop on Face Detection: Where We Are, and What Next? Lausanne, Switzerland, 2010.
[16] G. Mesnil, Y. Dauphin, G. Xavier et al., "Unsupervised and transfer learning challenge: a deep learning approach," in Proceedings of the ICML Workshop on Unsupervised and Transfer Learning, vol. 27, pp. 97–110, Washington, WA, USA, 2012.