International Journal of Computer Science Trends and Technology (IJCST) – Volume 10 Issue 5, Sep-Oct 2022

RESEARCH ARTICLE OPEN ACCESS

Real Time Face Mask Detection using Python

Mrs. B. Vijaya, MCA, M.Tech [1] , Sk Karimulla [2]
[1] Associate Professor, Department of Computer Applications
[2] Student, Department of Computer Applications
[1], [2] Chadalawada Ramanamma Engineering College (Autonomous)

ABSTRACT
With the rapid worldwide spread of Coronavirus (COVID-19 and COVID-20), wearing face masks in public becomes a necessity to mitigate the transmission of this or other pandemics. However, with the lack of on-ground automated prevention measures, depending on humans to enforce face mask-wearing policies in universities and other organizational buildings, is a very costly and time-consuming measure. Without addressing this challenge, mitigating highly airborne transmittable diseases will be impractical, and the time to react will continue to increase. Considering the high personnel traffic in buildings and the effectiveness of countermeasures, that is, detecting and offering unmasked personnel with surgical masks, our aim in this paper is to develop automated detection of unmasked personnel in public spaces in order to respond by providing a surgical mask to them to promptly remedy the situation. Our approach consists of three key components. The first component utilizes a deep learning architecture that integrates deep residual learning (ResNet-50) with Feature Pyramid Network (FPN) to detect the existence of human subjects in the videos (or video feed). The second component utilizes Multi-Task Convolutional Neural Networks (MT-CNN) to detect and extract human faces from these videos. For the third component, we construct and train a convolutional neural network classifier to detect masked and unmasked human subjects. Our techniques were implemented in a mobile robot, Thor, and evaluated using a dataset of videos collected by the robot from public spaces of an educational institute in the U.S. Our evaluation results show that Thor is very accurate achieving an F1 score of 87.7% with a recall of 99.2% in a variety of situations, a reasonable accuracy given the challenging datasetand the problem domain
Keywords--machine learning, convolutional neural networks, face detection, deep learning, mask detection, COVID-19

I. INTRODUCTION

The world is facing a health crisis due to the rapid spread of Coronavirus Disease 2019 (COVID-19). According to the World Health Organization (WHO) COVID-19 dashboard [1], more than more than 109 million people were infected by COVID-19 across 188 countries. The WHO published various reports that provide guidelines and mitigation measures to prevent the spread of the virus. According to the these reports and various research studies, wearing a face mask is highly effective in preventing the spread of respiratory viruses including COVID-19 [2]–[4]. For instance, Sim et al. [5] conducted a comprehensive study and reported that the effectiveness of wearing N95 mask in preventing SARS transmission is 91%. Since the outbreak of COVID-19, many organizations have updated their polices to require wearing face masks in public to protect their employees and community from the disease [6]. Therefore, a key role of the artificial intelligence and machine learning community is to propose new systems to automatically detect situations where people fail to wear face masks in public spaces to help mitigate the spread of COVID19 and other pandemics. For example, France integrated an AI-based system to the Paris Metro surveillance cameras [7] to provide statistics about the adherence to the face mask policy. Recent advances in deep learning techniques and their main component Deep Neural Networks (DNNs), have significantly improved the performance of image classification and object detection [8], [9]. Convolutional Neural Networks (CNNs or ConvNets) are a primary model of DNNs that have shown superior effectiveness in areas such asimage recognition and classification. CNNs have been very successful in detecting human subjects, faces, and other objects in images and videos because of their powerful feature extraction capabilities. In this paper, we ask the question: can we construct a deep learning-based classifier to detect unmasked faces from lowquality images? Our goal is to investigate the ability of deep learning to extract powerful features from low-quality images taken by a mobile robot (Thor) to construct a classifier that detects unmasked personnel with high accuracy. We describe low-quality images (and videos) not only as low-resolution images, but also other factors that significantly affect feature extraction from images. These factors are as follows: • The height difference between the camera and the face. Our mobile robot captures images with a camera that is 1-foot high from the ground, which provides partial facial images that are more challenging for feature extraction and classification than popular datasets that contain mostly images taken by cameras at the same height level of the face. • The angle between the camera and the face. Unlike most popular datasets, facial images are not always taken when human subjects are directly facing the camera. In practice, a dataset could contain images of human subjects that are walking away or with a 90 degree angle from the camera,which results in partial facial images that introduce more challenges to the image classification and mask detection tasks. • Quality of light. Unlike most popular datasets, using a mobile robot to capture videos or images results in images that are captured in spaces with a lighting quality that varies from low to intense. This variation in the dataset presents a new challenge to address. • Distance to human subjects. Capturing images at varying distances between the camera and human subjects makes the task of feature extraction and subsequently image classification more challenging because, at far distances, the areas of interest in the image (i.e., the human subject, face, and mask) are smaller which provides less powerful features to use for face and mask detection in such images.

II. RELATED WORKS

To help the organizations and the community defend against the rapid spread of Coronavirus Disease 2019 (COVID-19), there have been great efforts to spread awareness and sharecountermeasures with the public to mitigate the spread of COVID-19. Wearing a face mask in public is a key countermeasure to limit the spread of COVID-19 [4], and therefore, many educational and industry organizations have updated their policy to include having to wear face masks while on campus or inside buildings. In general, most research studies are focused on face construction and recognition for identify-based authentication. Loey et al. [10] presented a machine learning based framework for detecting face masks using a dataset of high quality conference-like face images. Their model achieved an accuracy of 99.64%-100% in detecting face masks. These images however, were taken when a face was looking towards a computer camera that is a few inches away. This is not appropriate to apply in realistic situations where people are walking around tens of steps away with varying angles that may only contain partial visibility of peoples faces and masks. Qin and Li [11] proposed a method that determines the correctness of mask wearing based on its placement. The approach classifies each situation into one of three categories: correct placement of the mask, incorrect placement, and no mask at all. The proposed approach achieved 98.7% accuracy in detecting face masks and mask positions. Ejaz et al. [12] analyzed and compared face recognition accuracy as an identity-based authentication using Principal Component Analysis (PCA) to recognize a person. They discovered that the accuracy of face recognition dropped to 73.75% when wearing masks.

Li et al. [13] proposed an approach for face detection using YOLOv3 algorithm. The approach constructed a classifier using more than 600,000 images of human faces provided by CelebA and WIDER FACE datasets. The approach achieved an accuracy of 92.9% in detecting faces. Nieto-Rodr´ıguez et al. in [14] and [15] proposed an approach for detecting surgical face masks in operation rooms. The main objective of this approach is to minimize the false positive face detection in order to only alert staff who are not wearing masks inside operating rooms. To achieve this, the approach takes advantage of the distinctive surgical masks color to reduce false positives. The approach achieved a recall above 95% with a false positive rate below 5% for the detection of faces and surgical masks. However, unlike operating rooms where the medical staff only wear the distinctively recognizable surgical masks, many other people wear masks with varying colors and styles. This variation will effect the the proposed approach in [14] and [15], and therefore, their reported accuracy might drop significantly.

III. PROPOSED METHOD

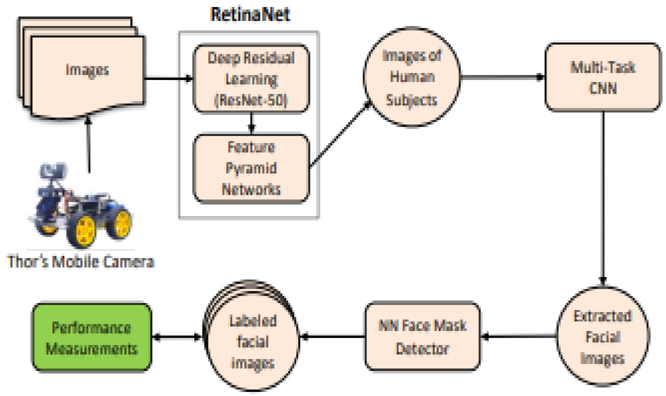

Fig. 1. The Architecture of Thor.

This research is conducted using a pipeline of three detection models that are constructed and tested on four publicly available datasets in addition to our own dataset that we collected to investigate our research objective. Table I summarizes these datasets. In this section, we describe each of these datasets. COCO Dataset. Microsoft Common Objects in Context (COCO) [19] is a large-scale object detection dataset that contains a total of 330,000 images of 91 object types (including human subject). These images have been labeled in this dataset with 2.5 million object labels. The COCO dataset was used to construct a pre-trained model that detects human subjects in captured images by our approach. We provide more details about this process in Section IV. CelebA. The CelebFaces Attributes Dataset (CelebA) [20] is a large-scale face dataset that contains 202,599 facial images of celebrities where each of these images have been annotated with 40 binary attributes. This dataset contains a largely diverse set of faces with many pose variations of human subjects which makes it a goldmine for training classifiers for face attribute recognition, face detection and extraction (of facial part) from the provided image. WIDER FACE. This is a face detection benchmark dataset [21] that contains 393,703 labeled faces with high variation in terms of pose and occlusion. theCelebA and WIDER FACE datasets were used to construct a pre-trained model that detects facial images (the facial area) in captured images by our approach. CMCD. The Custom Mask Community Dataset [22] contains 1,376 facial images that are well-balanced in terms of mask wearing. 50.15% (or 690) of these images contain masked faces and 49.85% (or 686) contain unmasked faces. Our approach uses this data to construct a convolutional neural network model for mask detection in the images it captures. Figure 1 illustrates the architecture of Thor including an Image Generator (IG), a human subject detector (HSD), a face detector and extractor (FD), a mask detector (MD). First, the IG continuously collects videos from various spaces and hallways in our organization. Then, it reduces the size of the video by sampling its images by keeping one image per second TABLE I SUMMARY OF DATASETS dataset images content number of images COCO [19] human subjects 330,000 WIDER FACE [21] CelebA [20] facial images 600,000 CMCD [22] masked & unmasked facial images 1,376 Standard Standard Images Deep Residual Learning (ResNet-50) Feature Pyramid Networks Multi-Task CNN Images of Human Subjects RetinaNet Extracted Facial Images Labeled facial images Performance Measurements NN Face Mask Detector Thor's Mobile Camera Fig. 1. The Architecture of Thor. and discarding the other images captured in that second. This sampling is important because the large number of images captured in each second significantly increases the size of the data and burdens the robot's resources. Therefore, we configured the robot to keep 1 image per second and discarded the other images captured in that second that are unlikely to provide additional information. Then, the HSD detects the presence of human subjects in these images and filters out the ones that do not have human subjects. The FD then detects and extracts human faces from these images and provides them for the MD. Then, the MD classifies the extracted faces into "Masked" or "Unmasked". The robot was equipped with two speakers to alert and provide unmasked individuals with a mask. A. Data Collection and Preprocessing As mentioned earlier, our robot (Thor) is equipped with a modified Donkey Car for mobility and a Raspberry Pi 1080p Camera that captures 20 images/second for data collection. We have used Thor to patrol a university campus and it collected over 150 videos from various hallways and spaces. These videos had varying lengths and most of their content was empty (e.g., no activity in images). We manually inspected these videos and discovered that our dataset contained 229 human subjects. 198 of the human subjects were facing the camera where the other 31 subjects were not facing the camera and therefore did not provide any facial footage. 133 of the subjects were wearing masks and 65 subjects were not wearing masks. To reduce the size of the data (i.e., number of images), our sampler selected only one image (frame) from the 20 frames captured in each second and discarded the other 19 images captured during that second. This process reduced the size of our data to 5% and boosted the performance of thefollowing detection modules by 95%. In the next part, we describe how we detect human subjects in these images.

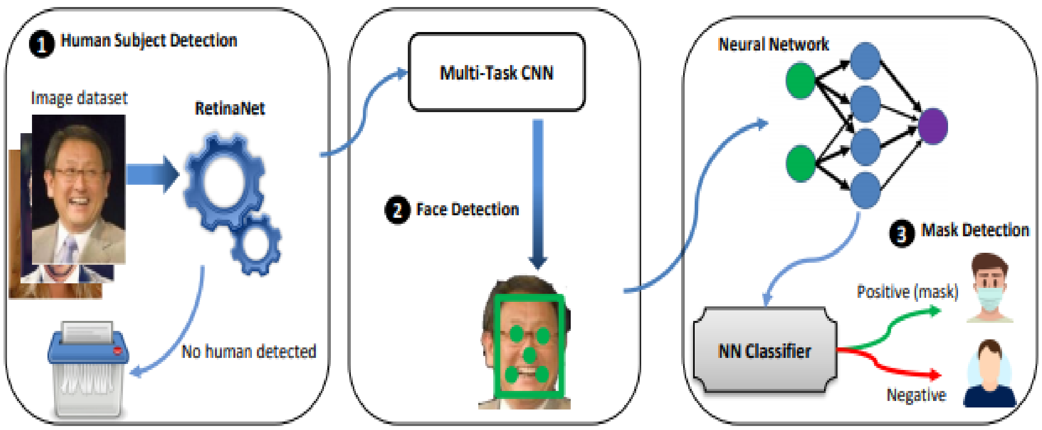

Fig. 2. Workflow of face mask detection using pipeline of deep learning techniques.

IV. RESULTS AND DISCUSSION

Our study shows that Thor takes a significant step to gather, detect, and mitigate situations of unmasked individuals inside indoor spaces. With the increased potential of spreading diseases, automatically detecting, alerting, and offering a mask for unmasked individuals can effectively counter current and emerging airborn diseases. However, our current implementation of Thor is still preliminary and in this section we discuss the limitations and potential future research of our approach. Error/misdetection analysis. Our evaluation shows that Thor has a high recall and precision given the nature of our dataset. However, Thor still mistakenly misses face masks and detects face masks that are not there. These problems mostly come from the limitations of existing datasets and subsequently the classifiers we use that are trained on these datasets. Specifically, the images provided by these datasets differ from images that are captured by a 1-foot-tall mobile robot in terms of the angle of capture that changes with distance and the available indoor lighting in the image. For example, using the Multi-Task cascaded Convolutional Neural Network (MTCNN) [8] on our dataset to detect human faces (facial areas) achieved an accuracy of 94.4%. The data responsible for the loss of 5.6% accuracy directly affects the outcome

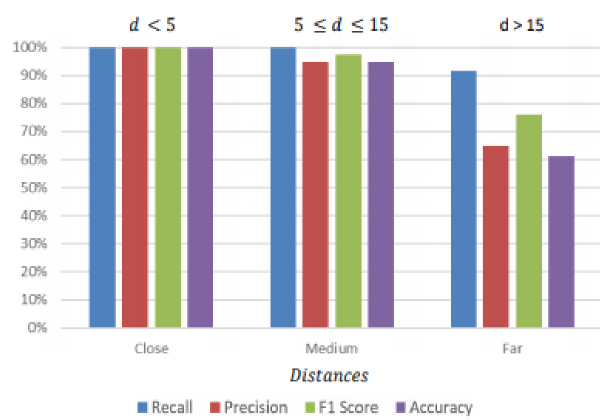

Fig. 3. The impact of the distance (d) between the human subject and the camera on the detection accuracy of face masks

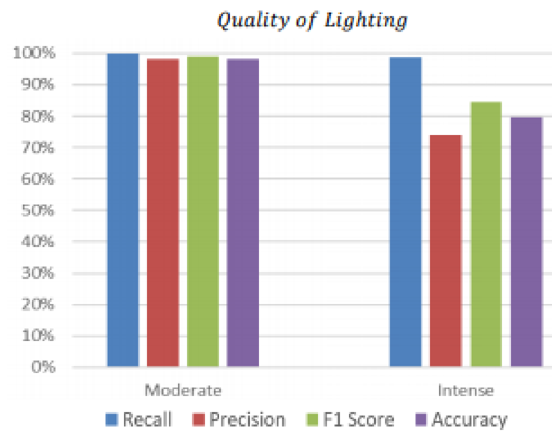

Fig. 4. The impact of the quality of lighting (q) in the space where image is captured on the detection accuracy of face masks.

V. FUTURE SCOPE AND CONCLUSION

In this paper, we present Thor, a system that implements deep learning-based techniques for automatic detection of unmasked personnel in public spaces. Thor developed an innovative approach that integrates different types of deep learning for face mask detection. Our prototype robot is comprised of three modules. The first module uses an integration of ResNet-50 and Feature Pyramid Network for feature extraction and human subject detection. The second module uses MultiTask Convolutional Neural Network (MT-CNN) to detect and extract faces from images containing human subjects. Then, the third module uses our constructed neural network model to classify the processed images to masked or unmasked. This classification enables identifying dangerous indoor situations of unmasked personnel. To mitigate such situations, Thor offers a surgical mask to the detected unmasked personnel. We evaluated our approach using a dataset of 229 human subjects collected by our mobile robot, Thor. The approach achieved a mask detection accuracy of 81.3% with a very high recall of 99.2%. To the best of our knowledge, this is the first effort that studies detecting face masks in images that are captured in various challenging settings such as space lighting and distance to camera.

REFERENCES

[1] “WHO Coronavirus Disease (COVID-19) Dashboard,” Accessed: January 10, 2021. [Online]. https://covid19.who.int, 2021.

[2] B. J. Cowling, K.-H. Chan, V. J. Fang, C. K. Cheng, R. O. Fung, W. Wai, J. Sin, W. H. Seto, R. Yung, D. W. Chu et al., “Facemasks and hand hygiene to prevent influenza transmission in households: a cluster randomized trial,” Annals of internal medicine, vol. 151, no. 7, pp. 437–446, 2009.

[3] S. M. Tracht, S. Y. Del Valle, and J. M. Hyman, “Mathematical modeling of the effectiveness of facemasks in reducing the spread of novel influenza a (h1n1),” PloS one, vol. 5, no. 2, p. e9018, 2010.

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

[5] S. W. Sim, K. S. P. Moey, and N. C. Tan, “The use of facemasks to prevent respiratory infection: a literature review in the context of the health belief model,” Singapore medical journal, vol. 55, no. 3, p. 160, 2014.

[6] H. Elachola, S. H. Ebrahim, and E. Gozzer, “Covid-19: Facemask use prevalence in international airports in asia, europe and the americas, march 2020,” Travel Medicine and Infectious Disease, 2020.

[7] “Paris Tests Face-Mask Recognition Software on Metro Riders,” Accessed: January 10, 2021. [Online]. https://Bloomberg.com, 2021.

[8] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[10] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic," Measurement,vol. 167, p. 108288, 2020.

[11] B. QIN and D. LI, "Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19," 2020.

[12] M. S. Ejaz, M. R. Islam, M. Sifatullah, and A. Sarker, "Implementation of principal component analysis on masked and non-masked face recognition," in 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1–5.

[13] C. Li, R. Wang, J. Li, and L. Fei, "Face detection based on yolov3," in Recent Trends in Intelligent Computing, Communication and Devices. Springer, 2020, pp. 277–284.

[14] A. Nieto-Rodr´ıguez, M. Mucientes, and V. M. Brea, "System for medical mask detection in the operating room through facial attributes," in Iberian Conference on Pattern Recognition and Image Analysis. Springer, 2015, pp. 138–145.

[15] A. Nieto-Rodr´ıguez, M. Mucientes, and V. M. Brea, "Mask and maskless face classification system to detect breach protocols in the operating room," in Proceedings of the 9th International Conference on Distributed Smart Cameras, ser. ICDSC '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 207–208. [Online]. Available: https://doi.org/10.1145/2789116.2802655

[16] J.-S. Park, Y. H. Oh, S. C. Ahn, and S.-W. Lee, "Glasses removal from facial image using recursive error compensation," IEEE transactions on pattern analysis and machine intelligence, vol. 27, no. 5, pp. 805–811, 2005.

[17] M. K. J. Khan, N. Ud Din, S. Bae, and J. Yi, "Interactive removal of microphone object in facial images," Electronics, vol. 8, no. 10, p. 1115, 2019.

[18] N. U. Din, K. Javed, S. Bae, and J. Yi, "A novel gan-based network for unmasking of masked face," IEEE Access, vol. 8, pp. 44 276–44 287, 2020.

[19] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollar, "Microsoft coco: ´ Common objects in context," 2015.

[20] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 3730–3738.

[21] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "Wider face: A face detection benchmark," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 5525–5533

[22] "Custom Mask Community Dataset (DMCD)," Accessed: January 10, 2021. [Online]. Available: https://github.com/prajnasb/observations, 2021.

[23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss ´ for dense object detection," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980–2988.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[25] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[26] R. Girshick, "Fast r-cnn," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), December 2015.

[27] V. Jain and E. Learned-Miller, "Fddb: A benchmark for face detection in unconstrained settings," UMass Amherst technical report, Tech. Rep., 2010.

[28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.

[29] GIMP, 2020, (Accessed: 12.07.2020). [Online]. Available: https://www.gimp.org/downloads/ SoutheastCon 2021 Authorized licensed use limited to: QIS College Of Engineering & Technology Autonomous. Downloaded on July 06,2021 at 04:23:26 UTC from IEEE Xplore