ZANCO Journal of Pure and Applied Sciences

TYPE Original Research
DOI http://dx.doi.org/10.21271/ZJPAS.36.2.2

OPEN ACCESS

*Corresponding author
Omer T. Hamajan
omer.hamajan@su.edu.krd

RECEIVED :04 /04 2023
ACCEPTED :12/10/ 2023
PUBLISHED :15/ 04/ 2024

KEYWORDS:
Convolution Neural Network (CNN), Deep Learning, Masked Face Dataset, Masked Face Recognition, YOLO, Mask R-CNN.

Masked Face Recognition using deep learning models

Omer T. Hamajan and Abbas M. Ali
1Department of Software Engineering and Informatics, College of Engineering, Salahaddin University-Erbil, Iraq

ABSTRACT

Face recognition has become indispensable in our daily lives as a quick and painless technique of confirming our identities since in the era of wearing face masks the traditional face recognition system may not effectively recognize the concerned person as an important part of the face (mouth, nose, and chin) which makes a substantial contribution to the face recognition process are occluded and partially hidden. The objective of our research is to tackle the challenges of the partially occluded face with a mask by training our custom dataset using those powerful pre-trained deep learning models like YOLO and Mask R-CNN which have not been used before for this purpose and to compare which one is outperforming the better results. To this end, models like (YOLOv5, YOLOv7, YOLOv8, and Mask R-CNN) have been employed and trained on the created dataset to check the accuracy and robustness of the occluded face recognition process. In addition, an online dataset such as (mfr2) which contains celebrities, and politicians masked and unmasked faces after expansion with more images has been used. The experimental findings demonstrate that the proposed algorithms give an accurate result, we achieve an accuracy of 97.5% using YOLOv8s, an accuracy of 89.7% using YOLOv7, and an accuracy of 89% using YOLOv5x, while an accuracy of 94.5% using Mask R-CNN. The study concludes that YOLOv8s outperforms the other models in masked face recognition.

1.Introduction

Since the spread of (COVID-19) outbreak, traditional biometric systems which is based on password or fingerprint has become not safe anymore. Therefore, face recognition is safer for these situations there is no need to touch any devices. Recent studies showed that wearing a mask will reduce the probability of transmission of the virus among the population. However, wearing a mask leads to the following issues: Fraudsters and criminals make use of the mask to steal and conduct crimes while remaining anonymous. When a large portion of the face is occluded by a mask, community access control, and face authentication become extremely difficult. When wearing a mask that does not reveal the whole facial view for description, current face recognition technologies are ineffective. Because it is required for face normalization, exposing the nose region is critical in the process of face recognition [1], pose invariance [2], and face matching[3]. Due to those issues, the masked face has extremely challenged the existing face recognition process.

In recent years, face recognition technology has gained significant popularity in the study of pattern recognition or biometric technology, it has become a cutting-edge study topic. However, the recognition challenge of the occluded face such as mask, hairstyle, sunglasses, and hat occlusions, is often highlighted in the study of face recognition [4]. When a person wears a mask in which faces are occluded or partially hidden. Although there are many types of research that concentrate on authenticating persons wearing masks without the need to take mask off which is crucial to degrade the spread of the Coronavirus Disease (COVID-19) outbreak. However, this raises the challenge of face recognition since occluded parts of face like (mouth, chin, and nose) are essential for face detection and recognition [5].

Smartphone users prefer to use biometrics authentication method instead of traditional way of authentication which is password since it offers extra level of security [6]. In order to verify a person identity biometrics method are used which is based on physical like face, fingerprints, and iris or behavioral feature such as dynamic signature verification, and speaker verification, therefore, one of the most popular method of biometrics system is face recognition which allow more secure way of authentication [7].

Recently, deep learning technologies have made a significant advance in field of computer vision, most of the face recognition techniques have been stepped toward applying deep learning models since masked face recognition has become cutting-edge research in the field of partially occluded face [4],[8].

For instance, according to World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC) [9], the best approach to avoid spreading or becoming infected with the virus is to practice social distancing and wearing face masks [10]. The people need to remove their face masks to go through the gates at airports or access control systems at their working places.

It’s worth mentioning that when we talk about masked face recognition, our aims are to recognize a face that covered with mask partially (nose, mouth and jaw) and the other parts that not occluded is eyes, elbow and forehead regions which we could take an advantage from them and extract features that helps for training in deep learning methods.

Our contribution to the research community are as follows:

- We have tackled the challenges of partially occluded face with mask by training our custom dataset using most prominent deep learning models like YOLO and Mask R-CNN.
- We used transfer learning which already trained with something similar to our issue which has really good pre-trained weights to compare the performance and to get better-targeted results.

Related Works

In our study we have tried to characterize the existing researches by reviewing them comprehensively of the latest techniques and algorithms that developed in the scope of masked face recognition. below shows related researches in the filed of masked face recognition.

In the 1990s, holistic techniques ruled the face recognition community. Handcrafted local descriptors were popular in the early 2000s, and in the late 2000s, local feature learning algorithms were established. DeepFace [11] and DeepID [12] achieved a significant performance in 2014, and the focus of the research has evolved to deep-learning-based techniques. The LFW (Labeled Face in the Wild) dataset performance continuously improves from approximately 60% to around 90% as the representation process grows more complex, while deep learning enhances the performance to 99.80% in just three years. In figure (1) (holistic techniques), the beginning diagram of face recognition technology and how it progressed through time and how deep learning started to contribute to face recognition technology. are illustrated in four different period of time.

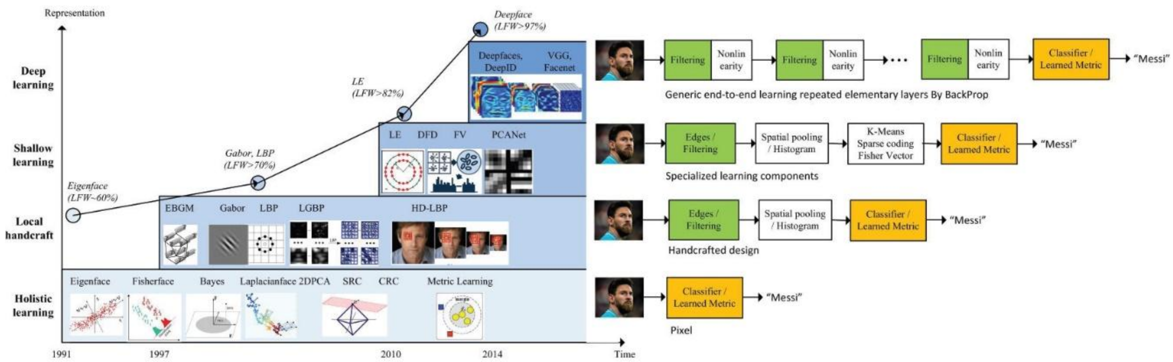

Figure 1: milestones of face recognition [13].

The study has been conducted a deep learning features combination for Masked Face Recognition (MFR). The methodology implemented feature level combination by applying two pretrained CNN architecture as a deep feature extractor, keeping in mind that CNN architecture that have yet to acquire a highest performance in ImageNet. Pretrained CNN architectures that used for masked face biometric system involved two strategies. The first one was pretrained GoogLeNet and VGGNet architecture were separately used to extract deep features, pursued by supervised learning models Support Vector Machine (SVM) classifier. The second one was that two feature vectors derived by pretrained GoogLeNet and VGGNet models were combined at the feature level, pursued by a multiclass SVM classifier. The experiment approach demonstrated the efficiency of algorithm by using three masked face recognition datasets. The proposed approach improved the performance and accuracy of CNN architecture in range of 94.62% to 95.33% on specified datasets [14].

Since the research proposed a reliable approach to remove occlusion by using deep learning-based features to resolve the issue of masked face recognition. Firstly, the masked face region is removed. Next, three pre-trained deep Convolutional Neural Network (CNN), which known as VGG-16, AlexNet, and ResNet-50 is applied and utilized to extract features from the acquired regions (generally eyes and forehead regions). The novel algorithm called The Bag-of-features (BoF) model is applied to the obtained feature of the last convolutional layer in order to quantize them to get a result to differentiate it from a fully connected layer of classical CNN. Finally, For the classification procedure, a Multilayer Perceptron (MLP) is used. In comparison to other state-of-the-art approaches, experimental findings on the Real-World-Masked-Face-Dataset indicate good recognition accuracy which is about %91.3 and for Simulated-Masked-Face-Dataset the accuracy of recognition is about %88.9 [15].

Since the publication of AlexNet architecture in 2012 Krizhevsky et al. [16], deep CNNs have been a prevalent technique in face recognition, It achieved a good result in face recognition under occlusion variation [16]. Although the idea of deep learning-based approach is inspired from the fact that human visual scene automatically neglects the occluded regions and only concentrates on the non-occluded parts. For example Song et al. [5] proposed a mask learning strategy for removing the masked region's feature parts from the recognition process.

A study reported the issue of masked face recognition by using deep learning-based algorithm that can effectively perform a face recognition process for those wearing face masks, The authors train a ResNet-50-based architecture that is effective in recognizing masked faces. The findings of this research might be smoothly incorporated into existing facial recognition algorithms that are used to recognize masked faces for security concerns [17].

A study has been conducted by using Gabor wavelet and machine learning technique which is deep transfer learning. The Gabor wavelet features are obtained from the unmasked part of the face and combined with deep learning CNN features to create a more reliable feature vector that may be used to improve real-time masked face recognition process. The experiment of proposed method used four benchmark datasets and achieved average accuracy of 97% [18].

A study has been presented a new artificial intelligence technique that utilizes hand-crafted and deep learning (YOLOv3 and CNNs) features and SVM classier. For training and testing the model 5 different face mask datasets are used in which testing phase demonstrated that the proposed algorithm performed comparable or better than traditional face mask recognition techniques. The algorithm can recognize a masked face with higher accuracy (99.4%) [19].

To remove occlusion, for instance [20] the image is divided into small local patches. Next, they applied Support Vector Machine (SVM) to get rid of occluded region. Finally, for face recognition on the non-occluded regions they applied a mean-based weight matrix. [21] the occlusion removal and restoration are applied. To eliminate the occluded regions, they applied the global masked projection. After that Principal Component Analysis (PCA) is then used to restore the data using eigenvectors.

A study has been evaluated K-Nearest Neighbors (KNN) classifier for face recognition and partial face recognition, it used images to train and test the model. The experimental results showed that (KNN) achieved significant accuracy of %100 in recognizing uncovered frontal faces. However, for covered partial face achieved a fair result about 74.7% [22].

A study has been proposed by combining deep learning and Local Binary Pattern (LBP) is a textual feature extraction to recognize masked face by using a well-known face detection module that can detect various scales of faces which is known as RetinaFace. In addition, to recognize a masked face local binary pattern features have been extracted from masked faces and combine their features that learned from RetinaFace. As an experimental result self-collected dataset called (COMASK20) and Essex dataset are utilized with the recognition results of 87% f1-score on the COMASK20 dataset and 98% f1-score on the Essex dataset. The proposed method outperforms Dlib and InsightFace, which has demonstrated the effectiveness and reliability of the proposed approach [23].

Datasets

We have generated our dataset which is called mfr2 or Masked Face Recognition Challenge (MFRC) (a small set of real masked face images. It contains 53 identities of celebrities, and politicians among 269 images, where each identity has an average of six images) [24]. And we have modified and expanded it with our own and other online images to more than 113 identities that contains 330 images with 100 classes each class contains about 6 labelled mask and unmask face images. After annotating this dataset, we generated a new version of dataset by applying preprocessing and augmentations to improve the quality and generalization ability of a models, eventually, we could have 790 images 87% as a training set, 8% as validation set and 4% as testing set.

Proposed Method

In this section we discuss about those methods that we have used in our study to recognize masked and unmasked faces in the image. The process of finding where an object is located in an image or video is called detection, and the process of finding what kind of object it is called recognition. The detection part of the process is where the bounding box appears, and the recognition part is where the class probability, which refers to the class name, appears [25]. As can be seen in the diagram (2) in order to make a comparison among those prominent models below we have used Mask R-CNN and three versions of YOLO to detect masked and unmasked faces as well as to recognize the face identity for masked and unmasked persons.

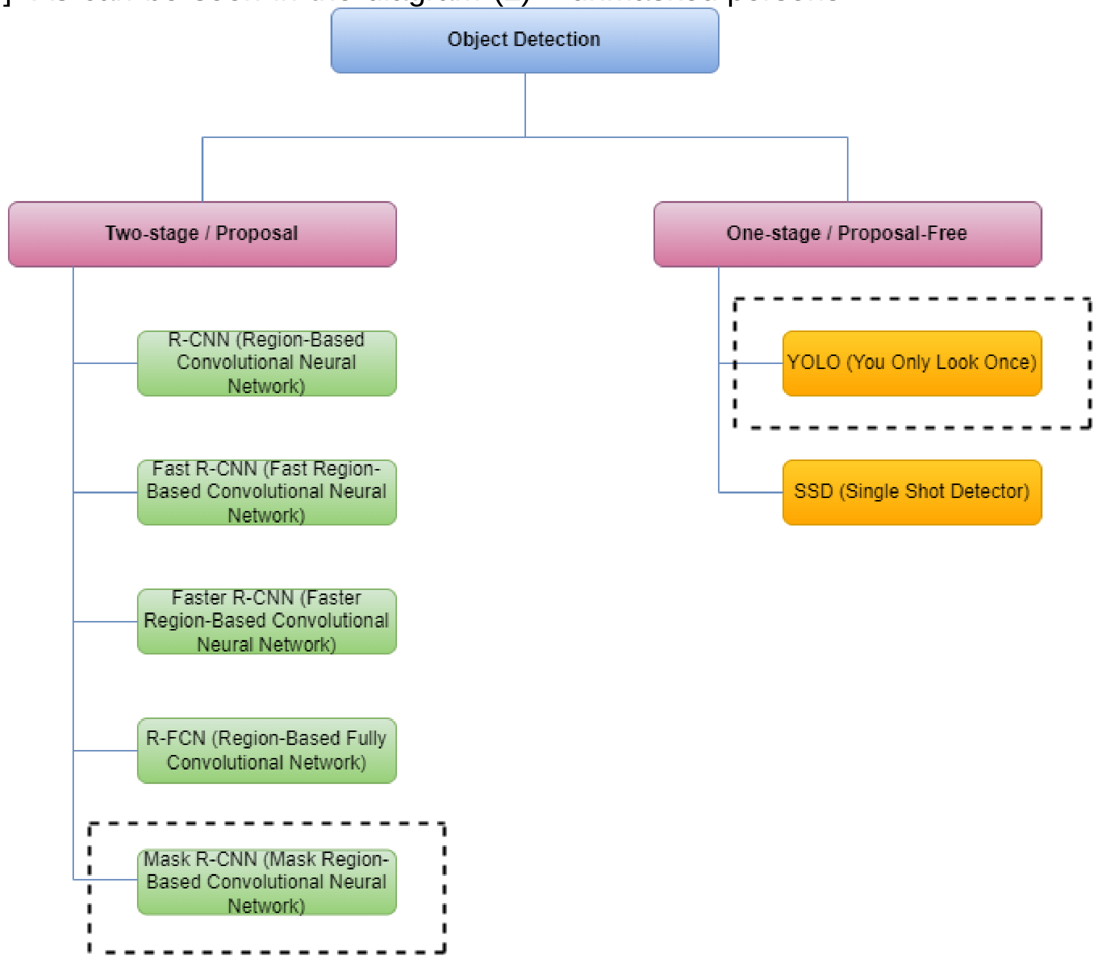

Figure 2: Proposed Object Detection methods for masked and unmasked face recognition

Mask R-CNN

It is a faster RCNN that has been modified to segment images at the pixel level. It's a combination of faster R-CNN plus FCN. It predicts the object mask using a new branch running in parallel with an existing branch for classification and bounding box regression and recognizes objects using pixels rather than just bounding boxes. The original implementation of the Mask RCNN architecture was built using the deep learning framework called TensorFlow which is an open source framework that developed by Google [26]. Mask RCNN is consists of two stages: the first scans the image and produces a proposal, and the second classifies the proposal and produces a bounding box and a mask [27].

The Mask R-CNN is an example of an **instance segmentation** algorithm. The collection of algorithms may be among the most efficient for object detection, delivering cutting-edge performances on benchmark datasets for computer vision. In comparison to other models like YOLO, which may be less accurate but are made for real-time prediction, the models, although being accurate, might be delayed when making a prediction [26].

The flow of Mask R-CNN model as follows figure (3) [26]:

- Given input image, either masked or unmasked to the network model using CNN (ResNet 101) architecture and the Feature Pyramid Network (FPN) that has been pre-trained to extract features and produce feature maps.
- This feature map uses Region Proposal Network (RPN) is used to extract Region of Interest (ROI) from images, and non-max suppression is utilized to choose the best suitable bounding boxes and to eliminate unnecessary ROI components.
- Region of Interest (ROI) is wrapped in defined dimensions by ROI Align.
- Fully Connected Layers – Consists of two parallel layers, one of which utilizes SoftMax for classification and the other regression for bounding box prediction.
- For each instance in an image, the Mask Classifier produces a binary mask.

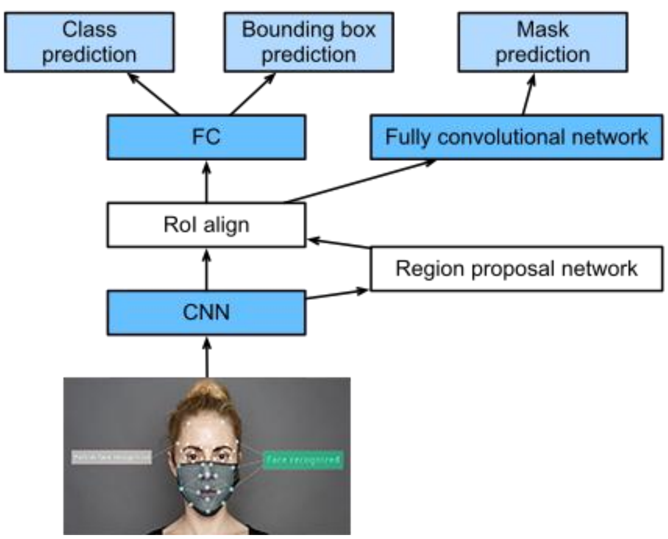

Figure 3: Mask R-CNN flow diagram

Although, to prepare our dataset for training in Mask R-CNN we have used VGG Image Annotator (VIA) which is an image annotation tool that used to identify areas in an image and provide textual descriptions of those areas by using polygon shape annotator to annotate our masked and unmasked face images.

YOLO The term "YOLO" (or "You Only Look Once") in 2016 publication invented by Joseph Redmon. On June 29, 2020, after a few days, the ultralytics firm launched YOLOv5. YOLOv5 was developed using PyTorch's implementation of YOLOv3 and was written in the Python programming language. PyTorch is a well-known deep learning framework that is driven by Meta [28]. YOLOv7 object detection model that outperforms all other object detectors in the speed and accuracy range of 5 FPS (Frame Per Second) to 160 FPS (Frame Per Second) with the accuracy of 56.8% mAP. For example, it could be used for image segmentation, posture estimation, image captioning, and tracking [35].
State-of-art and latest version of YOLO model, after 2 years of continuous research and development, finally YOLOv8 introduced by Ultralytics, they created models that are the best in the world at what they do: real-time object detection, image segmentation, classification, and key points detection. Faster, more accurate, and simpler, YOLOv8 places the power of AI in the hands of everyone. It can be trained on large datasets and is capable of running on a variety of hardware platforms, from CPUs to GPUs [36].

We have collected our dataset of masked and unmasked faces from online called mfr2 or Masked Face Recognition Challenge (MFRC), In order to train our custom model, we need to build up a dataset of representative images with bounding box annotations around the objects that we want to detect and labelled them appropriately by using Roboflow tool (which is a tool that used to convert raw images into custom trained computer vision model and deploy it for use in application). Roboflow supports object detection and classification models. Roboflow can be used for preprocessing images: resizing, grayscale, auto-orientation, contrast adjustments, Augmenting images: to increase your training data: flip, rotate, brighten / darken, crop, shear, blur, and add random noise, quickly assess your dataset quality [29]. After that we exported our dataset to train it in YOLOv5x, YOLOv7 and YOLOv8s, for training purpose we have used number of parameters (img) input image size of 416x416, (batch) batch size 16, (epochs) number of training epochs 100 we should be careful in choosing the number of epochs to prevent over-fitting and under-fitting, (data) our dataset, (weights) we choose the best pre-trained weight in order to get best results, (cache) cache images for faster training. Then we have evaluated our results to know the Mean average precision and loss function. Finally, test inference has been conducted to show the coefficient of recognizing faces with mask and without mask.

Results and Discussion

We
have provided many experiments in this section t o evaluate the suggested approaches.
The tests have been conducted using several sc enarios. The input datasets and the outcomes for each scenario will be provided and discussed in the sections that follows.

Experimental Setup

Here we have used most significant deep learning models like YOLOv5, YOLOv7, YOLOv8 and Mask R-CNN to detect and recognize masked faces. Then we have compared the performance and accuracy of each algorithm and we have prepared a dataset of representative images from online and trained our custom dataset which is called Masked Face Recognition Challenge (MFRC), we have used Google colab to implement YOLO the specification of free version of google colab was (12.7 GB of RAM, 78.2 GB of disk space, and 15.0 GB of GPU It's important to note that these resources are shared among all users, so the amount of available RAM and disk space may fluctuate depending on usage. Although, the free version of Colab has some limitations, such as 12 hours of continuous usage and maximum of 5 hours of GPU usage per day, we have used our local device to implement Mask R-CNN model and the specification of our device was (CPU corei7 11^{th} generation, 16 GB of RAM, and 4 GB of dedicated GPU NVIDIA GeForce GTX 1650) and the result showed that the latest and state-of-the-art YOLO model which is YOLOv8 outperforms among other algorithms. Implementation environment We have applied YOLOv5x, YOLOv7, YOLOv8s and Mask R-CNN models on our dataset that we annotated (labeled) and exported from Roboflow and VGG Image Annotator (VIA), to annotate our dataset in YOLO we have used Roboflow to draw a bounding-box around the face as shown in figure (4) below that used to recognize the masked and unmasked faces in image, while we have used VGG Image Annotator (VIA) for Mask R-CNN to draw a polygon line around the face in which Mask R-CNN is dealing with polygon as shown in figure (5) below. after that we evaluated our YOLOv5x, YOLOv7, YOLOv8s and Mask R-CNN model's performance in order to compare which model is outperforms the best result. Table 1 shows the mean average precision (mAP) is a measure metric to evaluate object detection models, the detected box is compared to the ground-truth bounding box to get the mAP's score. The model's detections get more accurate in case mAP score is high. Although metrics such as Precision, Recall and loss are shown in the table evaluate the performance of our models particularly in the context of classification tasks. They show how well the models could predict the correct class for a given input. Precision refers to correct positive predictions indicating how many positive predictions are made by the model is correct and it could be calculated as:

Precision=TruePositivePredictions/
(TruePositivePredictions+ FalsePositivePredictions)

On the other hand, recall refers to true positive predictions indicating how many of the actual positive instances were correctly identified by the model.

Recall=TruePositivePredictions/
(TruePositivePredictions+FalseNegativePredictions)

Loss is a measurement used to describe the difference between a machine learning model's estimated output and its actual output. The parameters of the model are changed to enhance performance using the loss function.

There are also estimated time of our proposed algorithms in order to train our dataset and in deep learning, the number of epochs is an important hyperparameter and can significantly impact the model's performance. Too few epochs can result in underfitting, where the model has not had enough time to learn the patterns in the data, while too many epochs can result in overfitting, where the model becomes too specialized to the training data and performs poorly on unseen data.

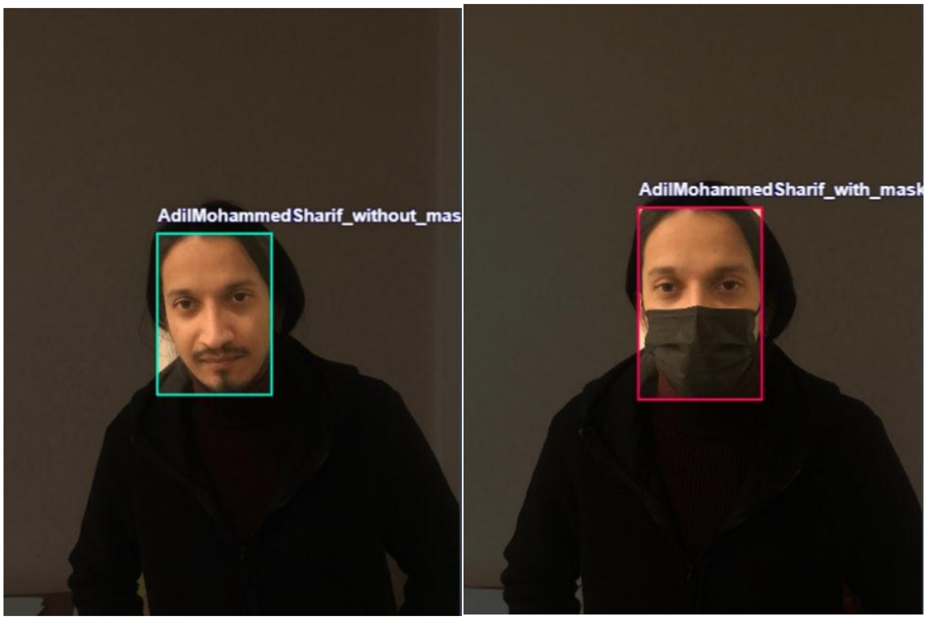

Figure 4: Using Roboflow to draw bounding-box in YOLO

Figure 5: Using VIA to draw polygon line in Mask R-CNN

It's worth mentioning that we have used older versions to make comparison among many versions of YOLO for detecting and recognizing the masked faces. YOLOv5x, YOLOv7, and YOLOv8s have been used. To show that YOLOv8s gives better accuracy than other versions of YOLO, however it takes more time to train in comparison with YOLOv5x.

Table 1: Comparison of performance and accuracy metrics among our proposed models

Algorithms	Performance and accuracy metric						
	mAP50	mAP50-95	P (Precision)	R (Recall)	Loss bounding box	Estimated time to train	Epochs
YOLOv5x	0.89	0.85	0.861	0.783	0.009729	47 minutes and 56 seconds	100

YOLOv7	0.897	0.848	0.692	0.829	0.01274	1 hour 15 minutes and 34 seconds	100
YOLOv8s	0.975	0.931	0.932	0.827	0.2086	1 hour 3 minutes and 35 seconds	100
Mask R-CNN	0.9453	-	-	-	0.01133	14 hours 7 minutes and 36 seconds	400

Discussion and Finding

Eventually, it could be observed that training our dataset on our proposed deep learning models, YOLOv8s gives better accuracy of 97.5% with an Intersection over Union (IoU) threshold of 50 and 93.1% with range of IoU thresholds between 50 and 95 and loss of 0.2086 when we use 100 epochs to train our model that takes about 1 hour, 3 minutes and 35 seconds. However, YOLOv7 gives accuracy of 89.7% with an Intersection over Union (IoU) threshold of 50 and 84.8% with range of IoU thresholds between 50 and 95 and loss of 0.01274 when we use 100 epochs to train our model that takes about 1 hour, 15 minutes and 34 seconds. Although YOLOv5x gives accuracy of 89% with an Intersection over Union (IoU) threshold of 50 and 85% with range of IoU thresholds between 50 and 95 and loss of 0.009729 when we use 100 epochs to train our model that takes about 47 minutes and 56 seconds. Finally, the most accurate result could be achieved by using Mask R-CNN which gives accuracy of 94.5% with an Intersection over Union (IoU) threshold of 50 and loss of 0.01133 but it takes more time to train our dataset with 14 hours, 7 minutes and 36 seconds when we use 400 epochs.

We have used the chart in order to visually represent the data and to provide comprehensive understanding of our findings. The chart in figure 6 below illustrates performance and accuracy comparison of our proposed deep learning models.

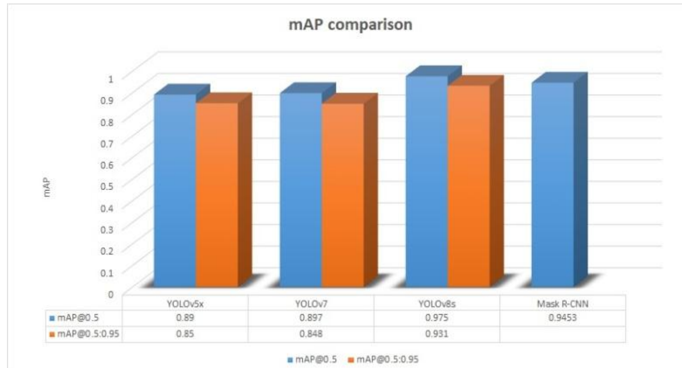

Figure 6: performance and accuracy comparison of our proposed models

To visualize our result, we have used TensorBoard tool which comes with TensorFlow. A tool for providing the measurements and visuals. It makes it possible to visualize the model graph, project embeddings into a lower dimensional space, track experiment metrics like loss and accuracy, when we train YOLO versions and Mask R-CNN on our dataset using 100 epochs the accuracy for YOLOv5x increases to 0.89 and the loss function decreases to 0.0097, the accuracy for YOLOv7 approximately the same as YOLOv5x which is 0.89 while the loss function decreases to 0.012, the accuracy for YOLOv8s increases dramatically to 0.97 while loss function decreases to 0.20, and the accuracy of Mask R-CNN increases to 0.94 while loss function decreases to 0.011 as illustrated below figure 7, figure 8, figure 9 , and figure 10 shows the accuracy and loss metrics for YOLOv5x, YOLOv7, YOLOv8s, and Mask R-CNN respectively. The significant comparison is that YOLOv8s outperforms other algorithms in term of accuracy while YOLOv5x gives lowest loss rate in comparison to other algorithms.

Figure 7: YOLOv5x Accuracy and Loss Metrics

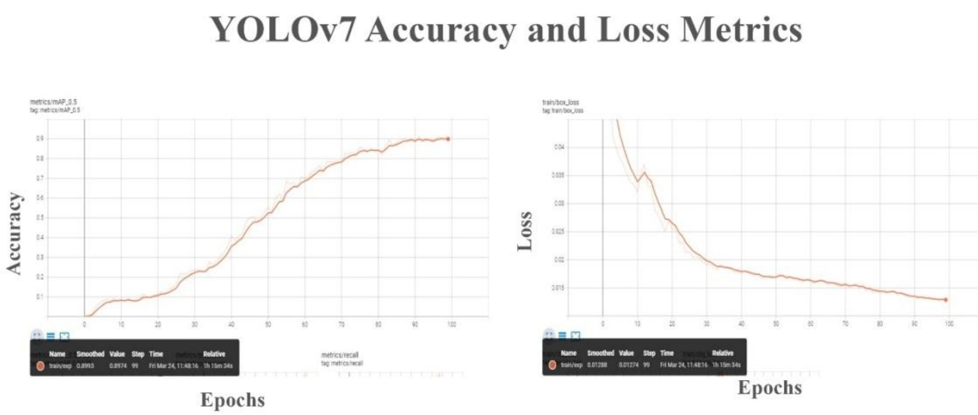

Figure 8: YOLOv7 Accuracy and Loss Metrics

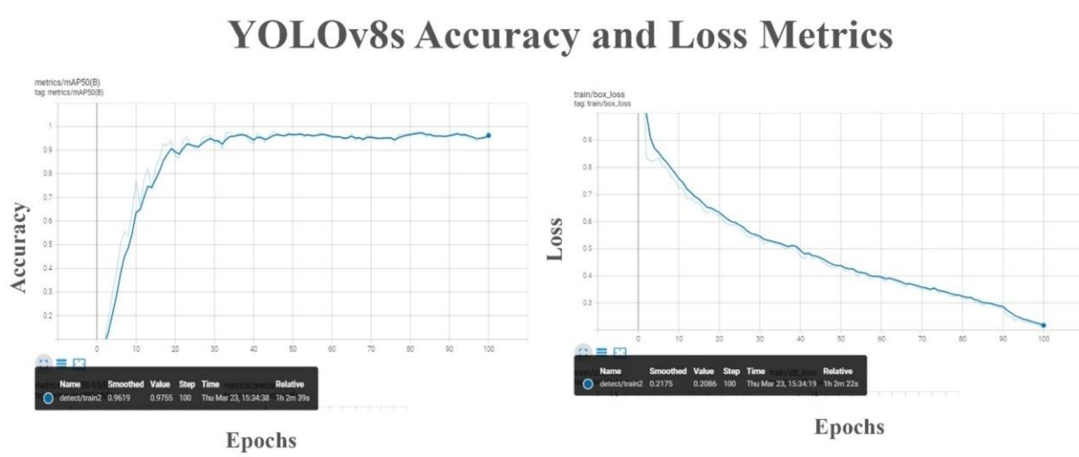

Figure 9: YOLOv8s Accuracy and Loss Metrics

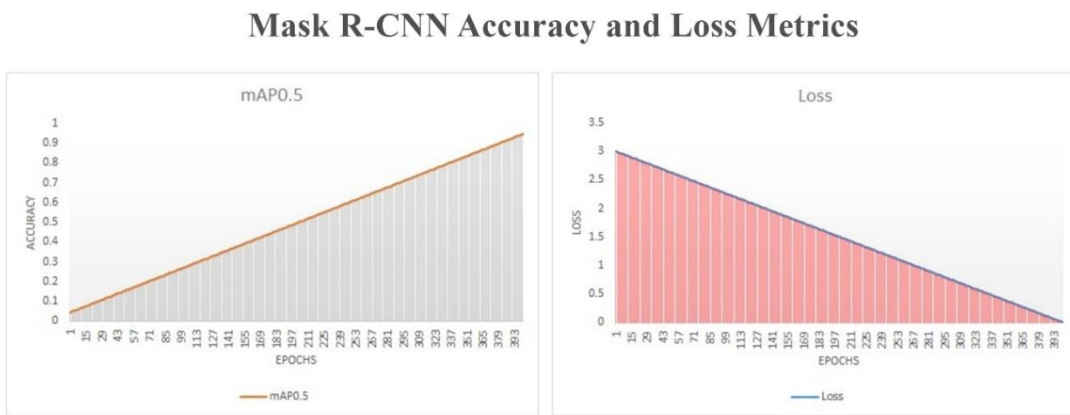

Figure 10: Mask R-CNN Accuracy and Loss Metrics

We have used the trained model to make inferences on recent images in order to find and identify masked faces the test result is shown in figure 11, figure 12, figure 13, figure 14 for YOLOv5x, YOLOv7, YOLOv8s, and Mask R-CNN respectively.

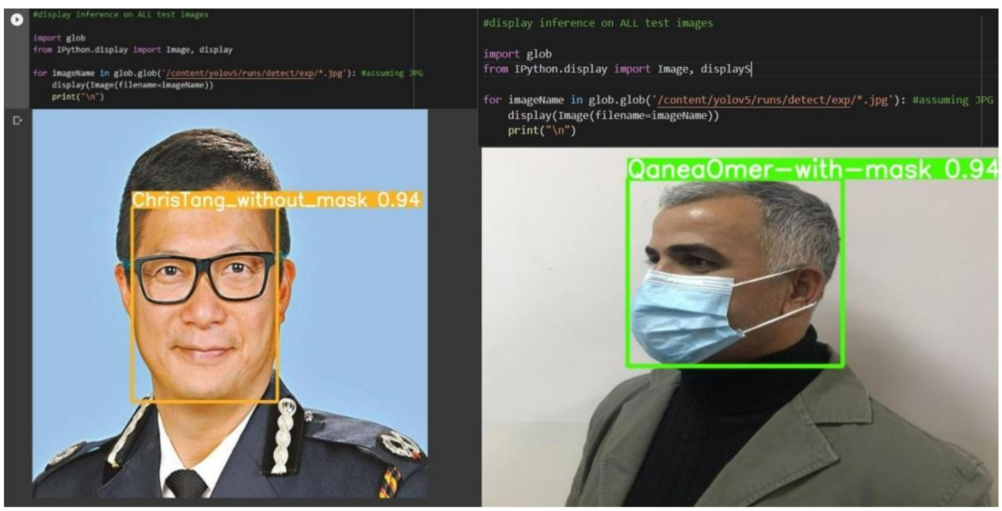

Figure 11: YOLOv5 inference of test images

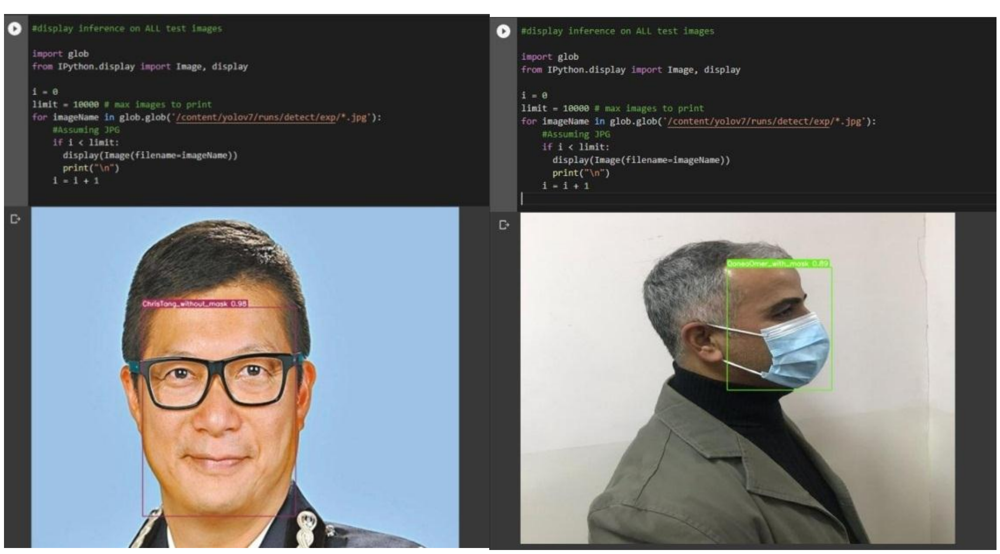

Figure 12: YOLOv7 inference of test images

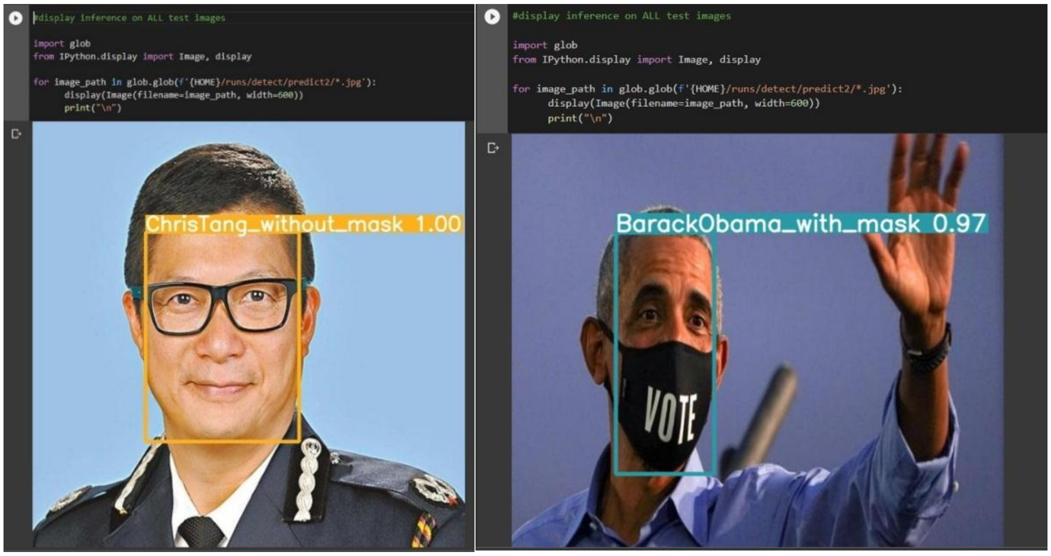

Figure 13: YOLOv8 inference of test images

Figure 14: Mask R-CNN inference of test images

Future Work

For future work, we want to improve the accuracy of masked face recognition model by using larger and more diverse datasets, and developing techniques that effectively detecting and extracting masks from images, and developing our model to detect and recognize masked faces in real time.

CONCLUSIONS

In our work, we have proposed state-of-art deep learning models to recognize mask covered faces. To develop our approach, we have used the concept of transfer learning by using pre-trained weights of Masked R-CNN, YOLOv5x, YOLOv7, YOLOv8s to detect the masked and unmasked faces from the image and to classify the identity of the person when the face is partially hidden with mask or fully visible no mask. We have generated a new dataset from the (mfr2) by gathering and manually labeling web images, since there is no publicly available dataset which is labelled properly. The system performance and accuracy are computed by extensive experiments conducted on benchmarked dataset and we have compared all four models. YOLOv8s gives a superiority over other models with an accuracy of 97.5%, Mask R-CNN gives an accuracy of 94.5%, YOLOv7 gives an accuracy of 89.7% and YOLOv5x gives an accuracy of 89%, The result showed that our proposed methods could improve the challenges of masked face recognition.

REFERENCES

[1] X. Peng, M. Bennamoun, and A. S. Mian, “A training-

free nose tip detection method from face range images," *Pattern Recognit.*, vol. 44, no. 3, pp. 544–558, 2011, doi: 10.1016/j.patcog.2010.09.015.

[2] X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 31–42, 2006, doi: 10.1109/TPAMI.2006.15.

[3] M. L. Koudelka, M. W. Koch, and T. D. Russ, "A prescreener for 3D face recognition using radial symmetry and the Hausdorff fraction," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2005-Septe, pp. 1–8, 2005, doi: 10.1109/CVPR.2005.566.

[4] H. T. Likassa, W. Xian, and X. Tang, "New Robust Regularized Shrinkage Regression for High-Dimensional Image Recovery and Alignment via Affine Transformation and Tikhonov Regularization," *Int. J. Math. Math. Sci.*, vol. 2020, 2020, doi: 10.1155/2020/1286909.

[5] L. Song, Di. Gong, Z. Li, C. Liu, and W. Liu, "Occlusion robust face recognition based on mask learning with pairwise differential siamese network," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 773–782, 2019, doi: 10.1109/ICCV.2019.00086.

[6] M. Haghighat, S. Zonouz, and M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 7905–7916, Jul. 2015, doi: 10.1016/j.eswa.2015.06.025.

[7] E. A. M. Al-Rahawe, E. A. M. Alrahawe, V. T. Humbe, and G. N. Shinde, "MULTI-BIOMETRIC TRAITS FUSION: A REVIEW Multimodal biometric system View project publication View project MULTI-BIOMETRIC TRAITS FUSION: A REVIEW," *Int. J. Res. Anal. Rev.*, 2020, [Online]. Available: www.ijrar.org.

[8] Y. Lecun, Y. Bengio, and R. 4g332, "Convolutional Networks for Images, Speech, and Time-Series."

[9] A. Alzu'bi, F. Albalas, T. Al-Hadhrami, L. B. Younis, and A. Bashayreh, "Masked face recognition using deep learning: A review," *Electron.*, vol. 10, no. 21, pp. 1–35, 2021, doi: 10.3390/electronics10212666.

[10] A. Anwar and A. Raychowdhury, "Masked Face Recognition for Secure Authentication," pp. 1–8, 2020, [Online]. Available: http://arxiv.org/abs/2008.11104.

[11] Y. Taigman, M. A. Ranzato, T. Aviv, and M. Park, "Taigman_DeepFace_Closing_the_2014_CVPR_paper," doi: 10.1109/CVPR.2014.220.

[12] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," *Adv. Neural Inf. Process. Syst.*, vol. 3, no. January, pp. 1988–1996, 2014.

[13] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021, doi: 10.1016/j.neucom.2020.10.081.

[14] J. M. Alrikabi and K. H. Alibraheemi, "A Combination Approach for Masked Face Recognition Based on Deep Learning," *Adv. Mech.*, vol. 9, no. 3, p. 2021, 2021.

[15] W. Hariri, "Efficient masked face recognition method during the COVID-19 pandemic," *Signal, Image Video Process.*, 2021, doi: 10.1007/s11760-021-02050-w.

[16] Y. Zhang, J. Gao, and H. Zhou, "Breeds Classification with Deep Convolutional Neural Network," *ACM Int. Conf. Proceeding Ser.*, pp. 145–151, 2020, doi: 10.1145/3383972.3383975.

[17] B. Mandal, A. Okeukwu, and Y. Theis, "Masked Face Recognition using ResNet-50," 2021, [Online]. Available: http://arxiv.org/abs/2104.08997.

[18] O. Mohammed Aarif K *et al.*, "Gabor-Deep CNN based Masked Face Recognition for Fraud Prevention," in *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, Apr. 2021, pp. 990–995, doi: 10.1109/ICCMC51019.2021.9418044.

[19] S. Liu and S. S. Agaian, "COVID-19 face mask detection in a crowd using multi-model based on YOLOv3 and hand-crafted features," no. April 2021, p. 20, 2021, doi: 10.1117/12.2586984.

[20] G. N. Priya and R. S. D. W. Banu, "Detection of Occluded Face Image using Mean Based Weight Matrix and Support Vector Machine," *J. Comput. Sci.*, vol. 8, no. 7, pp. 1184–1190, 2012, doi: 10.3844/jcssp.2012.1184.1190.

[21] N. Alyuz, B. Gokberk, and L. Akarun, "3-D face recognition under occlusion using masked projection," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 5, pp. 789–802, 2013, doi: 10.1109/TIFS.2013.2256130.

[22] X. Guo, "A KNN Classifier for Face Recognition," *2021 IEEE 3rd Int. Conf. Commun. Inf. Syst. Comput. Eng. CISCE 2021*, no. Cisce, pp. 292–297, 2021, doi: 10.1109/CISCE52179.2021.9445908.

[23] H. N. Vu, M. H. Nguyen, and C. Pham, "Masked face recognition with convolutional neural networks and local binary patterns," *Appl. Intell.*, 2021, doi: 10.1007/s10489-021-02728-1.

[24] Z. Zhu *et al.*, "Masked Face Recognition Challenge: The WebFace260M Track Report," 2021, [Online]. Available: http://arxiv.org/abs/2108.07189.

[25] G. Yang *et al.*, "Garbage Classification System with YOLOV5 Based on Image Recognition," *2021 6th Int. Conf. Signal Image Process. ICSIP 2021*, vol. 1, no. January 2020, pp. 11–18, 2021, doi: 10.1109/ICSIP52628.2021.9688725.

[26] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, 2020, doi: 10.1109/TPAMI.2018.2844175.

[27] Q. Zhang, X. Chang, and S. B. Bian, "Vehicle-Damage-Detection Segmentation Algorithm Based on Improved Mask RCNN," *IEEE Access*, vol. 8, pp. 6997–7004, 2020, doi: 10.1109/ACCESS.2020.2964055.

[28] D. Mallick Satya, "Custom Object Detection Training using YOLOv5." https://learnopencv.com/custom-object-detection-training-using-yolov5/ (accessed Jan. 30, 2023).

[29] "Roboflow." https://docs.roboflow.com/ (accessed Feb. 07, 2023).