International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Deep Learning Approach to Masked Face Recognition for Enhanced Attendance Systems

Jeena mary kurian

Department of Computer Engineering, Mangalam college of Engineering, Kottayam, Kerala, India

Abstract

Face recognition is the process of recognizing and identifying a face. Facial recognition uses specialized cameras to match faces in addition to unlocking phones. Face or picture recognition is useful for a variety of applications, including security phone unlocking, board control and airport operations, missing person searches, banking, retail crime reduction, marketing and advertising, healthcare, tracking employee or student attendance, and driver recognition. Biometric technology provides extremely attractive security option.

Face recognition systems are essential in practically every industry in our digital age. One biometric that is frequently utilized is face recognition. It has numerous more benefits in addition to being useful for security, identity, and authentication. Due to its non-invasive and contactless nature, fingerprint and iris recognition systems are still commonly employed despite their lower accuracy. Additionally, facial recognition systems can be utilized in companies, institutions, and schools to indicate attendance. The goal of this system is to create a facial recognition-based class attendance system because the current manual approach requires a lot of time and effort to maintain. Also, there might be opportunities for proxy attendance.

As a result, this mechanism becomes more necessary. The four stages of this system are face detection, face identification, database building, and attendance with a covered mask. Images of students in class are used to develop databases. The Haar-Cascade classifier and the Local Binary Pattern Histogram technique are used, respectively, for face detection and recognition. Faces are identified and detected from the classroom's live streaming footage. At the conclusion of the session, attendance will be kept on file in a server.

1. INTRODUCTION

Facial recognition technology can be used to verify or identify a subject based on a photo, video, or other audiovisual representation of their face. Among the uses for face recognition are voiceprints and security. Every face is different and possesses distinctive qualities. Systems, apps, or software for facial recognition compare additional uses for biometrics of the face and identification algorithms. The cornerstone for resolving contributions to deep learning is training with validation and testing. CNNs are particularly helpful for computer vision tasks like classifying and recognizing images because they are made to learn the spatial hierarchies of features by storing intricate patterns in deeper layers and critical features in earlier layers. This gets rid of the requirement for manual feature extraction, which was a difficult and labor-intensive operation in the past. Due to their numerous applications in people's daily lives—such as robotic intelligence, smart cameras, security surveillance, and even criminal identification—recognition systems have drawn a lot of interest.

2. LITERATURE REVIEW

2.1 New approach to similarity detection by combining technique three-patch local binary patterns (TP-LBP) with support vector machine

Face identification is the process of identifying the faces of the individuals whose photos are in a data set. Face recognition has always been a major area of research since it is non-intrusive and allows people to readily identify themselves, even in situations where other methods of identification might be more accurate. In an attempt to improve accuracy and reduce processing time, two facial recognition system approaches are analyzed here based on their detection rate and average authentication time. Two techniques are used in this method to extract features: PCA-SIFT (principle components analysis scale-invariant feature transform) and accelerated robust features (SURF). Afterwards, the random sample consensus (RANSAC) method is used to eliminate outliers. In the end, facial recognition is based on proximity determination. The support vector machine (SVM) classifier and the key point recovery techniques related are the foundation of the second strategy.

2.2 Face and liveness detection with criminal identification using machine learning and image processing techniques for security system.

Creating an anti-spoofing model with three main parts is the aim of this work: CNN classifier-based criminal identification, liveness detection, and face anti-spoofing detection. This kind has a very simple operating system. The face anti-spoofing module will recognize pictures, masks, posters, and smart phones after analyzing the data. When a face is found, the CNN classifier module gets the input and determines if the face is real or false. The liveness identification module will process the following data in order to identify lip movements and eye blinks. If both modules process the information, it is recognized as a real face.

2.3 Multiple face mask wearer detection based on YOLOv3 approach.

By merging the YOLOv3 object detection algorithm with many backbones, such as ResNet-50 and Darknet-53, a face and face mask detection model was developed. The datasets that were acquired from internet resources like Github and Kaggle were used to filter and classify the images. The models were assessed on recall, mean average precision, detection time, and precision following training on 4393 pictures. This study developed an automated face and face mask detection for numerous people that may be used in commercial cameras using YOLOv3 based deep learning object detection. This could be an effort to support law enforcement in keeping track of whether individuals are using face masks in public in an effort to stop the COVID-19 virus from spreading. During a pandemic, automated face mask detection must take the place of manual surveillance, which is ineffective at recognizing large groups of people and has a contactless monitoring approach. It needs to be precise and quick. For the purpose of producing a prototype, the model can be integrated into webcam software or a commercial camera. An assessment can be based on the live video's frames per second.

2.4 Face recognition in identifying genetic diseases: a progress review.

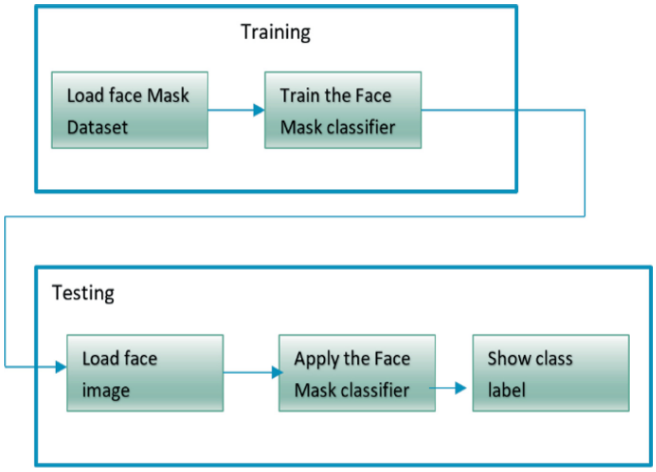
Genetic illnesses might differ greatly. It might be difficult for practitioners to diagnose hereditary illnesses. It is challenging to differentiate between hereditary diseases without conducting comprehensive testing on the patient, sometimes referred to as genetic testing. Nonetheless, some earlier research has shown that individuals with hereditary illnesses exhibit distinct physical traits. This makes it possible to identify variations in these physical traits to help medical professionals diagnose patients with hereditary illnesses. Research on facial recognition has been very busy in the last few years. It is still being developed by researchers using a variety of current techniques, databases, algorithms, and methodologies. Medical imagery is one of the sectors in which the application is used. Face recognition is one method of disease identification.

2.5 Attendance management system using face recognition.

Traditionally, attendance is tracked by teachers marking registers, which requires a lot of upkeep and human error. One important aspect of this method is time consumption. Our goal was to transform the digital tools that are currently in use, such as facial recognition. To address the shortcomings of the conventional system, this project has undergone a revolution. Our objective is to recognize faces and mark the present. A single folder has a database containing the faces of every student in the class. Attendance is noted if a student's face matches one of the faces that have been stored.

3. PROPOSED METHODOLGY

This study is devoted to a Face Mask Detection model. A mask detection model has been created to ascertain whether individuals are donning masks. In this method, the computer's primary camera was utilized, and the video was fed into the deployed model as input. Open-CV libraries and photos sent into the system during training are used in the construction of this system. To identify the system, the algorithms ResNET50V2 and Mobile-Net were combined. The goal is to create an automated model for face and face mask detection that goes through two phases: testing and training with validation. Figure 3.1 depicts the suggested two-phase training and testing approach for real-time face mask detection.



3.1 Two Stages of Proposed System

3.1 Training with Validation

The faces in the dataset are divided into three classes: those with the accurate mask, those with the incorrect mask, and those without one. The training process comprises of 11 phases, starting with dataset collection and ending with image classification. Utilizing Open-CV, the data frame is initially retrieved. To balance the unequal number of classes, random oversampling (ROS) utilizing ρ and imbalance computation are then carried out. The third phase involves applying face detection and picture augmentation by going through numerous neural layers that extract feature maps. In the following stage, the pre-trained model's final predicting layer is replaced with its predicting layers in order to accomplish fine-tuned transfer learning. Lastly, the final phase involved using ResNet50V2, a pre-trained classification model.

3.2 ALGORITHM 1: TRAINING PHASE

Input: MAFA dataset with pictures or videos in it.

Processes: Face identification, picture enhancement, random oversampling, frame extraction, and transfer learning with pre-trained categorization.
Frame Extraction using Open-CV:
Step 1: The video was split into frames using cv2.imwrite() and saved after being captured using cv2. VideoCapture() via the built-in camera.
Random over sampling:
Step 2: In order to balance the unequal number of classes, ROS and imbalance calculations are used.
Image augmentation and Face detection:
Step 3: The image is passed through a huge number of convolutional layers at different points.
Step 4: A 4*4 filter is applied to each of those feature maps in order to identify a small low default box and forecast the bounding box offset for each box.
Step 5: Each bounding box output contains the following five predictions: x, y, w, h, and confidence. In respect to the grid cell boundaries, x and y denote the centroid of the box.
Step 6: Additionally, each grid cell forecasts conditional class probabilities.
Step 7: Intersection over union (IOU), which is defined as intersection over union = area of overlap area of union, is used to match the truth boxes with the predicted boxes.
Transfer learning:
Step 8: To apply optimized transfer learning, swap out the pre-trained model's last prediction layer with one of its own.
Step 9: Early layers of the network obtain generic features from the obtained model; during training, their weights remain fixed.
Step 10: Higher layers are where task-specific features are learned; these layers can be pre-trained and refined.
Pre-trained classification:
Step 11: To categorize images, use the pre-trained ResNET50V2 classification model.
Output: Validation is performed on the masked faces.

3.2 Sample images that are selected in the final dataset

3.3 Testing the System

The 12-step testing process includes data collection (live video), the display of personal data (name and identity), and information about violations (location, timestamp, camera type, ID, and violation category, such as faces without masks and faces with the wrong masks). To determine if a picture is soft or hard, the initial step involves extracting the data frame using Open-CV and then predicting the image's complexity. We apply a semi-supervised technique to face categorization. To achieve fine-tuned transfer learning, the final predicting layer of the pre-trained model is swapped out with its predicting layers in the next step. The next stage in classifying photos is identity prediction using a pre-trained classification model, ResNet50V2.

3.4 ALGORITHM 2: TESTING PHASE

Input: Live video or images.

Processes: Frame extraction, image complexity predictor, transfer learning with pre-trained classification and identify prediction.

Frame extraction:

Step 1: Divide the video recorded by the built-in camera into frames using cv2.VideoCapture(), then store them using cv2.imwrite().

Step 2: Using training photos as a comparison, MobileNetV2 and ResNET50V2 perform face detection.

Step 3: Class-specific confidence scores are acquired during testing.

Step 4: The expected boxes and the truth boxes are matched using IOU, which stands for intersection over union = area of overlap area of function.

Image Complexity Prediction :

Step 5: Image classification using a semi-supervised approach is used to determine if the image is soft or hard.

Step 6: The MobileNet-SSD model, L=1/N(Lclass+Lbox), is used to predict the class of soft images. Lbox is the L1 smooth loss that indicates the error of matched boxes, Lclass is the softmax loss for classification, and N is the total number of matched boxes with the final set of matched boxes.

Step 7: Using a faster RCNN base on ResNet50V2, demanding picture prediction is accomplished

Transfer Learning:

Step 8: To apply optimized transfer learning, swap out the pre-trained model's last prediction layer with one of its own.

Step 9: The pre-trained Model is the source of generic features that the network's initial lower layers acquire, and during training, their weights remain fixed.

Step 10: Higher layers are where task-specific features are learned; these layers can be pre-trained and refined.

Pre-trained classification:

Step 11: Images are categorized on the masked face using the pre-trained classification model ResNet50V2.

Identify prediction:

Step 12: To determine if a face is wearing a mask or not, Open-Face is used.

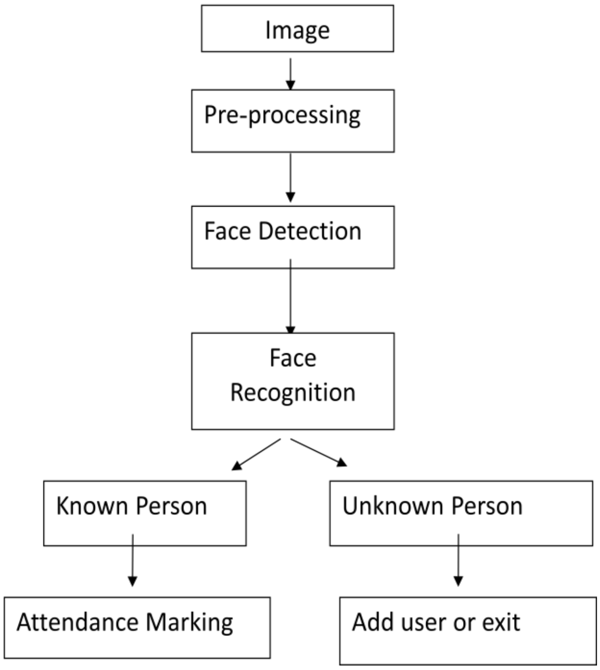
Output: Show personal data like name and identification as well as information about violations (such location, time stamp, type of camera, ID, and category of violations, including faces without masks and faces wearing the wrong kind of mask).

3.5 Mobile networks for the detection of face masks

Deep learning and computer vision disciplines have seen a significant increase in the use of face mask recognition techniques. The technique shown in was able to recognize face masks in real time using deep learning frameworks such as Tensor-Flow, Keras, and OpenCV libraries. A 99.9% F1 score and an accuracy score were obtained from the trained Mobile Net model. Mobile Net and other convolutional neural networks are made especially for embedded and mobile vision applications. Depth wise separable convolutions, a type of low latency, lightweight deep neural network designed for embedded and mobile devices, are used in their construction.

3.6 STEPS

This method's confined phases—unique ID enrollment, database creation, facial detection, facial training,
facial recognition, and graphical user interface—had to be completed.



3.4 Flowchart of working system

Phase 1: Registering using distinct IDs

In order to register with the database, an individual or student submits their general information.
This data will be retained for further processing at a later time. The person's picture will be taken with a digital digicam. After this phase, functions are extracted. As a result, specific functions may have their own ID and be kept in the database.

Phase 2: Creating databases

The presence tracking gadget needs to have some data entered into the system before it can be used. This data consists of the individual's face, ID, and basic details Using the camera to catch the subjects' faces is the most straightforward way to take pictures. In this way, the device will start by figuring out if a face is visible in the image. In order to obtain the hundred necessary images for each student and if no face is recognized, the system will prompt the user to snap another image . This process will continue until a certain number of photos are taken. After that, the photos will undergo several preparatory processing stages in order to generate a grayscale image.

Phase 3: Face detection

The facial recognition system operates on the.py format, which is also the primary format for storing learnt faces. The system first looks for faces in the photos or video streams that are provided to it using a system for facial recognition algorithm. It then proceeds through two crucial decision-making stages, which are explained in the face verification section. Face or no face determination and face verification, Whether the detected face matches any of the training faces is determined by the former, however the latter ascertains whether a face appears in the picture. To ensure that the system operates reliably and correctly, these distinct stages are crucial.

Phase 4: Training of faces

Once the images are taken with a camera, they are saved in grayscale. Since training establishes the resolution, the local binary pattern histogram (LBPH) recognizer is used to learn these faces.

The neighbors are thresholded against the center, which is one of the picture's elements. The center elem-

element's depth is indicated as 1 if it is more than or equal to its neighbor, and 0 otherwise. Binary patterns, also referred to as local binary pattern (LBP) code, may arise from this.

Phase 5: Face Recognition

The recognized faces are identified by matching them to the student IDs using the trained face data that has been saved. To ensure the correctness of the system, facial data is recorded in real time. This system is entirely dependent on the state of the camera.

Phase 6: Attendance Marking

This attendance system's most important feature is its automatic tracking of attendance, which starts as soon as a student's face is successfully recognized. The system will not record attendance for a pupil if its facial recognition algorithm is unable to recognize them. This method records attendance only when a known face is spotted, ensuring precise and efficient attendance monitoring.

Phase 7: Graphical User Interface

The Python tkinter module is used in the development of the graphical user interface (GUI). We designed two buttons labeled "Prediction" and "Exit," along with five Textboxes where we may add our biological information. The entered values are fed into the loaded model when we click on the predict button, and the labeled output appears at the textual content container that resembles an emotion.

RESULTS AND ANALASIS

4.1 Confusion Matrix

The prediction outcomes of face recognition models or classification algorithms are compiled into a confusion matrix. It offers details on the errors the classifier is committing, as well as—possibly more crucially—the kinds of errors it is making. The event row is labeled as "positive," the forecasts are classified as "true" or "false," and the non-event row is designated as negative. It was done on the two people's faces, which are never the same. As a result, the first person is used to evaluate true cases, whereas the second person is used to analyze false cases:

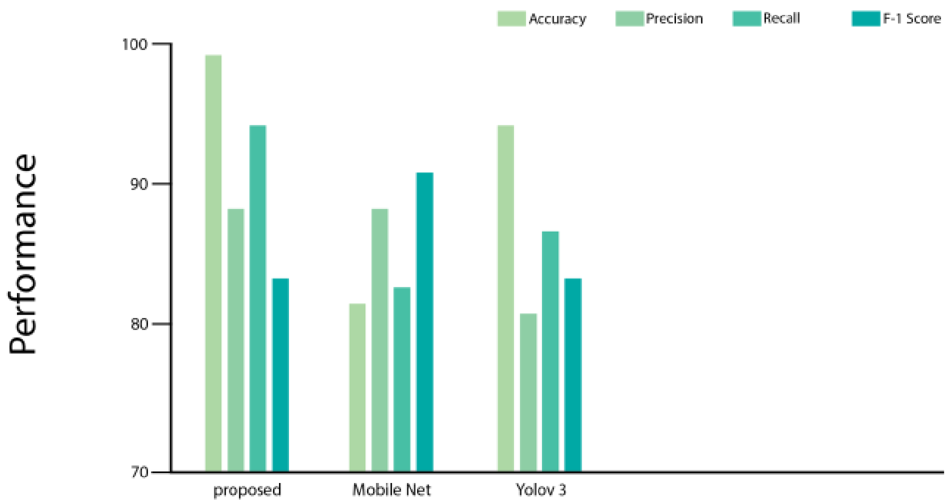
True positives (TP): These situations are recognized when a face pair in an image comparison shows differences between the predicted and real cases; in other words, the models have identified the two faces as distinct individuals.

True negative (TN): Models and identifiable faces have determined that the person is different but still the same.

False positives (FP): These are instances in which models have recognized distinct faces as belonging to distinct people despite the images' differing faces.

False negative (FN): Since the models identified them as the same individuals, the case is labeled as such.

4.1 Confusion Matrix



4.2 Evaluation Performance

Facial detection in crowd analysis can determine the number and composition of people in an open space. However, the detector may be affected by the facial obstructions' effectiveness in the facial frame image, similar to the presence of a mask. Few studies focused on mask-wearing facial recognition prior to COVID-19, mostly because there were insufficient datasets created specifically with mask-wearing facial recognition in mind. However, the strategic review reveals that most of the datasets utilized for face mask identification were artificial, meaning that more accuracy in reproducing real-world events is required. As a result, this disparity adversely affects the model's performance in practical scenarios.

Tools such as TensorFlow, Keras, OpenCV, a hybrid deep transfer learning model, public recognition databases, and others are used in deep learning. The range of its accuracy is 95% to 99.64%. Nevertheless, CNN outperformed DL when paired with deep learning or a multi-stage method, achieving 99.98% accuracy. Lightweight Region Proposal Networks (RPNs) had the lowest proportion (73%) when compared to other approaches.

The effectiveness of a training program is measured using the Intersection over Union (IoU) in image detection. It is used as a statistic to evaluate the degree to which the prediction's bounding box resembles the manually labeled ground truth. To determine how similar two regions are, one calculates the intersection area (IoU) of the predicted and ground-truth regions divided by the union of the two regions. The model will evaluate itself during the training phase based on the IoU and learn from it in order to improve its prediction as close to the ground truth as possible. In this report, the IoU threshold was set at 0.4.

Mean average precision (mAP), recall, and precision were computed to assess the model's performance. For the IoU, the experiment's true positive (TP), false negative (FN), and false positive (FP) criteria were set at 0.4. The forecast will be classified as TP, or the accurate prediction, if its IoU surpasses the cutoff. If a prediction's IoU score is less than a predefined threshold, it will be marked as FP. One of two conditions could result in the image being classified as FN: either there is no detection (no bounding box) or the classification is incorrect. Therefore, precision and recall can be computed using the number of TP, FP, and FN.

The precision of the model for categorizing samples as positive can be assessed by looking at its percentage of accurate positives (TP) among all predictions. Recall is defined as the proportion of accurate positive forecasts among all true positives, or ground truth. A high recall value means the model can accurately identify positive samples. Plotting a precision-recall graph with precision and recall allows one to calculate the average precision of the class by calculating the area under the graph. The average

precision between the masked and non-masked classes is found in this report as a mean, or mAP.

CONCLUSION

This work established an automated face and face mask detection for numerous users using YOLOv3 based deep learning image detection, which may be used in commercial cameras in the future. The dataset was gathered from online sources to start the dissertation's data collection phase. The pictures were then tagged and filtered. Two new models, ResNet50_YOLOv3 and DarkNet53_YOLOv3, were created by extracting and concatenating the extraction layers from pretrained ResNet-50 and Darknet-53 models using the YOLOv3 classifier. Using the darknet framework, both models were trained on over 4000 pictures on Google Colab Pro. The models were evaluated using precision, recall, mean average precision, and detection time. We presented algorithms in this dissertation report that can identify distinct faces when wearing masks, and the system's performance yields respectable, good outcomes.

The goal is to automate and create a system that an institution, for example, can use. The proper and efficient method of taking attendance in an office setting, which can replace antiquated manual procedures. An administrator can create a teaching account and add students and their data to the database using a facial recognition technology. After that, educators can access the system and verify the student's attendance. The student's attendance is recorded in the database, and their face is recognized by a camera. The student attendance report was visible to teachers and administrative staff.

REFERENCES

1. Salsabila Aurellia1, Siti Fauziyah Rahman1,' Face recognition in identifying genetic diseases: a progress review', *IAES International Journal of Artificial Intelligence*, Vol. 12, No. 3, September 2023.

2. A. Arjun Raj, M. Shoheb, K. Arvind, and K. S. Chethan, “Face recognition based smart attendance system,” *Proceedings of International Conference on Intelligent Engineering and Management, ICIEM 2020*
3. Ashish Jadhav1, Dhwaniket Kamble2, Santosh B Rathod1, Sumita Kumar2, Pratima Kadam3, Mohammad Dalwai1,’ Attendance management system using face recognition’,Vol. 13, No. 1, March 2024.
4. Ahmed Chater, Hicham Benradi, Abdelali Lasfar,’ New approach to similarity detection by combining technique three-patch local binary patterns (TP-LBP) with support vector machine’, *IAES International Journal of Artificial Intelligence (IJ-AI),* Vol. 12, No. 4, December 2023.
5. Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, “Face recognition systems: a survey,” *Sensors*, vol. 20, no. 2 Pratibha Shinde1,2, Ajay R. Raundale3,’ Face and liveness detection with criminal identification using machine learning and image processing techniques for security system’ *IAES International Journal of Artificial Intelligence (IJ-AI),* Vol. 13, No. 1, March 2024
6. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December.
7. Cheng Xiao Ge1, Muhammad Amir As’ari1,2, Nur Anis Jasmin Sufri1,’ Multiple face mask wearer detection based on YOLOv3 approach’, IAES *International Journal of Artificial Intelligence (IJ-AI,)* Vol. 12, No. 1, March 2023.
8. Yi Wang and Xinwei Duan, ‘Research on face recognition algorithm based on deep learning’, *In 2021 IEEE 21st International Conference on Communication Technology (ICCT).*