e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

ENHANCING REAL-TIME VIDEO SURVEILLANCE: FACE DETECTION AND TRACKING WITH CONVOLUTIONAL NEURAL NETWORKS

Monika*1, Prof. Dr. Rajesh Gargi*2

*1Deptt. Of C.S.E. PKG College Of Engineering & Technology, India.

*2Affiliated By Kurukshetra University, India.

ABSTRACT

Face detection and monitoring play a crucial role in a number of computer vision applications including surveillance, human-computer interaction, and augmented reality. The Convolutional Neural Networks to develop a method for face detection and tracking in real- time that is both efficient and accurate. Face alignment, occlusion, and shape illumination issues must be accurately interpreted for human emotion recognition from videos. Convolutional Neural Networks, a powerful class of deep learning models, have achieved outstanding results in image and video processing applications like entity recognition and tracing. CNNs are designed to autonomously learn and derive features from unprocessed input data, such as images and videos, via a sequence of convolutional and pooling operations. Face mask detection has numerous applications, including biometrics, real-time surveillance, etc. Automatic face mask detection and monitoring systems are a far superior option for managing public behavior and contributing to containing the COVID-19 outbreak than staff surveillance. Face mask detection is also advantageous for public surveillance in order to prevent the use of face masks in public areas. The RILFD dataset was created by capturing actual images with a camera and annotating them with two publicly accessible labels: with mask and without mask. In this study, machine learning models and pre-trained deep learning models YOLOv3 and Faster R-CNN are utilized to detect face veils. Face mask detection is proposed using four image processing stages and personalized CNN models. The method outperforms other face mask detection models and has demonstrated its robustness with a 98.5% accuracy score on the RILFD dataset and two publicly available datasets MAFA and MOXA.

Keywords: Face Detection, Face Monitoring, Computer Vision Applications, Convolutional Neural Networks, Human Emotion Recognition, Face Mask Detection, Yolov3, Faster R- CNN.

I. INTRODUCTION

The convolutional Neural Networks for real-time face detection and tracking is a ground breaking use of computer vision technology that has transformed numerous industries, from surveillance and security to entertainment and user experience. A wide range of applications, including video analysis, augmented reality, and biometric security, are made possible by this method, which harnesses the power of deep learning to automatically recognise and track human faces in real-time video streams. The core computer vision tasks of face detection and tracking have a wide range of practical applications. Traditional approaches struggled with differences in lighting, stances, and occlusions and relied on handcrafted features and intricate algorithms to recognize and follow faces. However, the accuracy and efficiency of these jobs have substantially improved with the introduction of CNNs and their capacity to automatically learn organizational features from raw data. Convolutional neural networks are networks that are known as of deep learning models designed specifically to process and evaluate grid-like data, such as images. They include several layers, including fully linked layers for classification and convolutional layers that extract features. CNNs' hierarchical structure makes it possible for them to recognize complex patterns and features in images, which makes them ideal for tasks like monitoring and recognition of faces. Real-time facedetection and tracking involves detecting faces in each frame of a video feed and preserving the identity of the detected faces across frames. The requires both precision and velocity, as video broadcasts are typically processed in real-time or close to real-time. Modern CNN- based architectures face the formidable challenge of achieving this equilibrium between precision and speed. Popular architectures such as YOLO and SSD Single Shot MultiBox Detector have become more popular as a result of their capacity to handle video frames rapidly and forecast bounding box coordinates and class probabilities in a single pass. Anchors or default boxes, which are predefined bounding boxes with varying sizes and aspect ratios, aid in the precise localization of features. The deployment of monitoring and recognition of faces system in real time has extensive ramifications. Its aides in the identification of persons of interest from live camera feeds in security

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

and surveillance. It improves user experiences in the entertainment industry by facilitating augmented reality applications that superimpose digital elements on real-world faces. Additionally, developments in this discipline contribute to the study of human-computer interaction, emotion recognition, and social robotics. Face recognition technology is commonly used to activate smartphones which are used by the majority of people. The technology provides an effective method for protecting personal information and ensuring that sensitive information remains inaccessible to criminals even if the phone is stolen.

There are numerous applications for face recognition technology, including safety, security, and payments. Face recognition refers to the challenge of accurately recognising or authenticating a person from a digital image or video frame using the biometric pattern of their face. To authenticate a person, the system collects a unique set of biometric data points associated with facial expressions. Biometrics is used by a system that recognizes facial features to map facial traits from a picture or video. Facial recognition is a technology-based method for human face recognition and matching identification in which the system compares information to a database of known features and can assist in establishing an individual's identity [1]. The faces that the camera captured are used as input before the Haar Cascade detection of faces algorithm is applied to the image. After that, feature values are extracted using the Facial Landmarks Detection approach and when compared to a feature database that has undergone SVM training earlier. The facial recognition is the consequence of software development [2]. Face recognition is accomplished using a multi-task convolutional neural network. The appearance, motion, and shape characteristics are utilized for tracking to compensate for tracking failures caused by object occlusion or rapid object movement. The relative weights of various characteristics for feature fusion are modified based on the tracking condition. The approach of modifying feature weights based on scenes is capable of addressing issues such as continuous tracking, interruptible tracking, and object interaction [3]. Biometric recognition, such as fingerprint recognition, palm recognition, voiceprint recognition, and retinal recognition, has been implemented in numerous attendance systems due to the accelerated development of artificial intelligence and machine learning. To accomplish the effect of access control, biological differences between individuals are used as the basis for discrimination [4]. Using CNNs, computer vision tasks like picture categorization and facial recognition have been successfully completed [5-6]. The image segmentation era of artificial intelligence deep architecture, GPU computation, and large training datasets are primarily responsible for the success of AI. As a result, advancements have been made in face recognition [7-8]. Currently, computers can already outperform humans in these areas [9]. CNNs are an artificial neural network type that are frequently employed for the recognition of tasks and classification of images. It is designed to identify patterns and attributes in images and is based on the structure of the human visual system. [10].

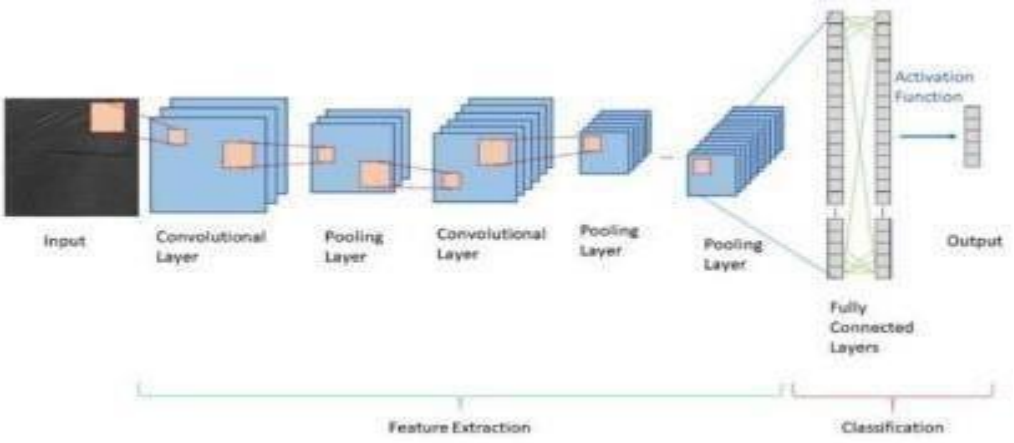

Figure 1: CNN Architecture sources: [Madhuri et al. (2023)]

1. Convolutional Layer:

The convolutional layer is a CNN's main structural component. In order to extract features, convolutions are applied to the input image. Each convolutional layer consists of multiple filters also called kernels, which are small-sized matrices. To calculate the product of dots over the filter and the surrounding region of the picture, the filters "convolve" or slide across the input vision. By doing this function, the network is able to recognize a variety of patterns and elements in the image, including edges, textures, and more intricate structures.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

2. Pooling Layer:

The feature maps' geographic extent is shrunk, but maintaining essential details by using pooling layers. Basic methods of pooling include maximum and average pooling. Pooling makes the network less computationally complex and more resistant to translations and distortions.

3. Activation Functions:

The network acquires non-linearity through activation functions, allowing it to recognise complex input patterns. Rectified Linear Activation (ReLU) is commonly used in CNNs, but there are other options like sigmoid and tanh.

4. Fully Connected Layers:

Fully linked layers are included to do advanced classification and reasoning after features are extracted using convolution and pooling. These layers take flattened feature maps from the previous layers and connect every node to every node in the subsequent layer, just like in traditional artificial neural networks.

5. Output Layer:

The network's estimations are generated by the final layer using the features that have been learned throughout the network. The task's specific requirements determine how the output layer is activated. As an illustration, categorization jobs frequently employ SoftMax.

Face mask detection has various applications in real-world settings, such as real-time authentication and virtual watching over people. Criminals frequently conceal the area of their faces around their mouths. The top of the head and shoulders of a person serve as a workaround for the problem of recognizing obstructed faces, according to researchers [11]. When a big number of people must be remotely checked for face mask usage, the task becomes more challenging. The novel coronavirus infection COVID-2019 epidemic also necessitated the usage of face masks and imposed many other restrictions. Due to its severe consequences, rapid spread, lack of proper medication, and lack of medical personnel, The WHO declared COVID-19 a pandemic and recommended several preventative actions, including the use of face masks [12, 13]. Wearing a mask is your best line of defence against the deadly conditions caused by COVID-19. The general public now accepts the idea that wearing a face mask can help to contain the propagation of COVID-19. The general people were under pressure all around the world to keep their distance and take precautions to halt the spread of this contagious sickness.

COVID-19 continues to infiltrate and spread because of its evolving forms despite intensive immunization campaigns in numerous countries. Therefore, constant use of face masks is necessary to stop its spread. Additionally, that will help to stop exposure and prevent people from coming into contact with the disease's germs. Face masks are a requirement in multiple nations, public buildings forbid admission without one. Due to the high volume of people that enter public buildings such that airports, train stations, and retail Centers, manual examination is almost impossible. Recently, there has been increased interest in the study of automatic face covering detection and identification. For applications in monitoring and surveillance, the COVID-19 has started to build automatic detection systems [14]. Since it is crucial to first recognise images before determining whether or not they are wearing masks, identification and classification are needed for the detection of face masks. As a result of the community's development of numerous face identification systems, the first objective has received considerable attention in the field of computer vision [15–17]. The detection of face coverings on multiple datasets required a significant amount of work. The complexity of face masks in different hues and styles, as well as the lack of a publicly accessible real-world image database, both impede existing research [18]. The difficulty posed by the large variety of face mask hues and ornamentation techniques, as well as the lack of a publicly accessible real- image dataset, limit the current research. If face mask simulations are utilised instead of non- facial mask pictures in simulated datasets, the models are no longer appropriate for use in real-world settings. Many various face masks and expressions make detection more difficult.

II. LITERATURE REVIEW

Wang et al. (2023) examined Almost everyone wears a respirator to prevent COVID-19 coronavirus. Security checks, community visit check-ins, and other face-based identification verification scenarios challenge the

traditional face recognition methodology. The most advanced face recognition algorithms use deep learning and many training instances. No masked identification of faces datasets, especially real ones, are accessible to the general public. Three masked face datasets MFDD, RMFRD, and SMFRD were used in this investigation.

Madhuri et al. (2023) described Recognition and tracking physical objects one of the difficult jobs regarding computer monitoring, used for object recognition, face recognition, and character recognition. Convolutional Neural Networks are powerful deep learning models that excel at entity detection and tracing in images and videos. CNNs employ convolutional and pooling techniques to automatically detect and identify features from raw input data, such as images and videos. CNNs can recognize and locate characteristics and objects in new photos and videos by training on massive datasets. The Raspberry pi runs the CNN model as an edge device and displays the recognized entities on the monitor. The system efficiently identifies and tracks entities with high precision.

Zarkasi et al. (2022) examined automatic facial mark recognition. Exterior nose corner landmarks are nosocomial. The data gathered landmarks are used to generate a triplet that consists of regions and geometric invariability. Later, triangular lines would be used to connect the nose's and eyes' outer corners. The Euclidean Distance formula will calculate the line length afterwards to get the triangle's area.

Reddy et al. (2022) described the introduces a CNN-based face expression recognition model called Deep Cross Feature Adaptive Network (DCFA-CNN). The form and texture feature blocks ShFeat and TexFeat make up the DCFA-CNN model. In order to separate expressive regions, the ShFeat block collects high-level responses. In contrast, the TexFeat block retains minute/micro changes in order to construct structural differences. Using a two-branch cross- relationship, DCFA-CNN additionally gathered the ShFeat and TexFeat blocks. Using four datasets are such that CK+, MUG, ISED, and OULU-CASIA the DCFA CNN is tested in single-domain and cross-domain experimental settings that are ethnicity-independent.

Khattak et al. (2022) described Facial expressions are variable, making emotion identification from photographs difficult. The findings harmed by poor layer selection in the convolutional neural network model, which was utilised to identify emotions from facial photos with a focus on emotion identification. The convolutional neural network model can be used to provide an efficient deep learning method for identifying emotions in facial photos and determining age and gender from expressions.

Li et al. (2021) examined in modern deep learning, face detection requires antecedent boxes and NMS post processing. Still, the working results of face detectors is highly impacted by anchor design and anchor matching technique, necessitating a large amount of anchor design work for a variety of business circumstances. A dual branch entirely convolutional framework is used in the two-branch centre face detector, a straightforward pure convolutional face identification that is efficient but NMS-free and does not need anchor design. Extensive experiments on four important face detection benchmarks AWW, PASCAL face, FDDB, and WIDER FACE method is faster while still being competitive with state-of-the-art methods.

Khaireddin et al. (2021) described Human-computer interaction, therapeutic practice, and behavioural description require facial emotion recognition. Computer models struggle to create accurate and robust FER due to human feature heterogeneity and image changes like facial position. Deep learning models, especially CNNs the most potential for FER due to their robust automatic efficiency of calculation and feature extraction.

Kamińska et al. (2021) examined Emotion recognition is complicated by face feature variation and ethnic and cultural differences. Faces also show people's complex emotions, which can be expressed through compound emotions. Compound facial emotion detection makes it harder because dominant and complementary emotions are sometimes hard to distinguish.

Feng et al. (2020) described China and other countries including South Korea and Japan have used face masks since the SARS-CoV-2 pandemic. The public and health care personnel in China are advised to use face masks based on risk., although certain provinces and municipalities have mandatory mask legislation.

Kumar et al. (2019) examined as video and image databases grow, intelligent systems need to automatically grasp and analyze information because manual processing is becoming difficult. Face conveys identity and feelings in social interaction. People can't distinguish faces as well as machines. Automatic face detection systems simplify face recognition, facial emotion identification, head-pose estimation, and interaction between people and computers. Digital image face detection uses computers to locate and size human faces. The

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

computer vision literature has highlighted face detection.

Iqbal et al. (2018) described Facial expression detection local feature descriptors are unstable, especially with feeble and Noise-distorted edges prevent them from working as well. In this study uses a local descriptor called the Neighbourhood-aware Edge Directional Pattern to get over these restrictions. The NEDP assesses the gradients at the target centre pixel and its neighbouring pixels to study a larger neighbourhood for feature consistency despite small distortion and noise in the local area., unlike existing local descriptors. To unambiguously depict local textures, provide template-orientations for neighbouring pixels that prioritize gradients in consistent edge directions and local edge neighbours. Due to effective featureless region management, NEDP never encodes a feature improperly. In person-independent recognition experiments using benchmark expression datasets, NEDP beats other descriptors, improving facial expression recognition.

III. METHODOLOGY

The system accurately recognizes the object using a convolutional neural network implemented on a Raspberry Pi. The object recognition and tracing using convolutional neural networks have gained popularity. YOLO is software that detects and recognizes objects using CNN. The YOLO software detects and localizes objects in images and videos by gridding the image and predicting bounding boxes and object class probabilities for each grid cell. Applying non-maximum suppression to eradicate overlapping bounding boxes refines the predictions. In this study provides an overview of the object recognition and tracing capabilities of the YOLO software. In the CNN architecture utilized in YOLO, the training procedure, and the outcome evaluation measures applied to gauge the software's effectiveness. The evaluation results indicate that YOLO performs well on multiple benchmark datasets, attaining high precision and quick processing times. Yolo is a promising object recognition and tracing software with potential applications in numerous domains, such as robotics, surveillance, and autonomous driving.

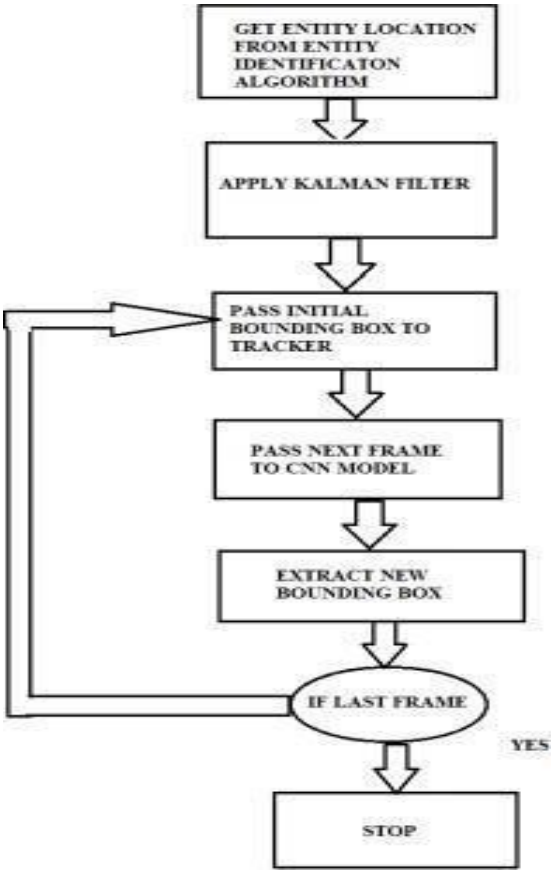

Figure 2: Object Tracing

Object tracing can be implemented once an object recognition CNN has been developed. The techniques such as region proposal algorithms, which locate areas in an image that might be object-containing regions, and then classify those areas using a trained CNN. Once an object has been identified, algorithms such as optical flow or

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

object tracking can be used to track its movement over time. As you trace objects that CNN performs less well on particular objects or in particular environments. The using additional data, you can fine-tune the CNN or modify the object tracing algorithms to enhance performance. The capture real-time images using cameras or sensors and perform object recognition and tracing in real-time. The data collection, and the framework's implementation processes. After gathering data from classes with masks and without masks, the images are pre-processed using four stages of picture processing. Facemask detection is complicated by changes in skin tone and illumination; hence picture processing stages are employed to normal the input photos. Next, A customised deep learning model evaluates the presence or absence of a face concealment. In addition, two commercially available, fine-tuned deep learning models are employed.

3.1 Image-based real Face Mask Labelled Dataset

It is a useful addition to the research community to have a hand-labelled and organized face mask dataset. Even though there are now some datasets that can be used for experiments in this way, each dataset has its own limits. For instance, some datasets have small pictures and others have images from the internet. the people in the datasets have been given fake face masks.

Stage 1: It requires applying filters to the source data. As shown in Figure 3c, filter size ([0, - 1,0], [-1,6, -1], [0, -1,0]) has been applied to images for this purpose.

Stage 2: At the third stage, blue, green, and red (BGT) images are converted to luma components, red and blue projections (YUV). This phase maintains the full resolution of the Y a luma component while decreasing the U and V resolutions.

So, luminance is more important than colour, a lessening of U and V also serves to reduce the training model's complexity. Figure 2d depicts the transformation from BGR to YUV.

Stage 3: In the last stage, pictures are normalized by returning to the BGR format. In this stage, applies histogram normalization and smoothes the images, as shown in Figure 3.

Stage 4: In this step, we adjust the contrast and resize the picture. The process entails making the input image smaller. Images in the dataset have the following dimensions: 11903x13096. Figure 3 depicts the image size after being decreased to 120 120 3. It aids in decreasing the computational time required for detection.

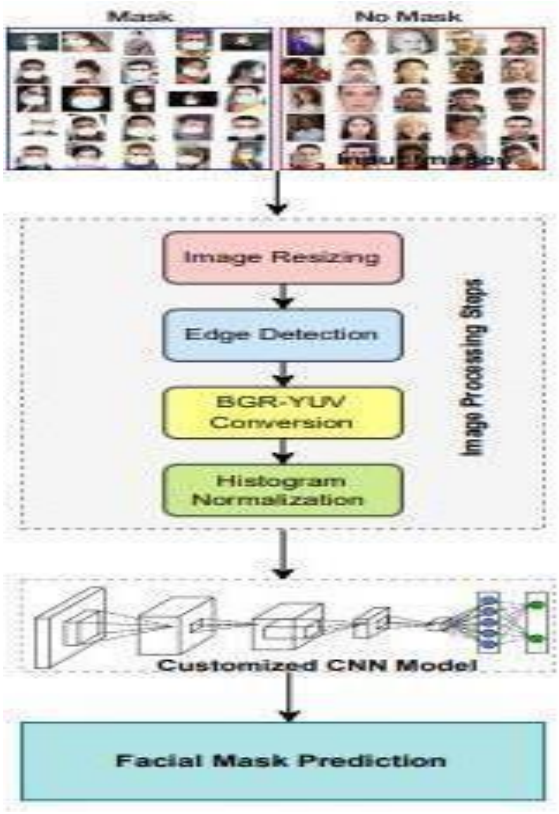

Figure 3: Workflow diagram of the proposed face mask detection framework.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

3.2 Face Mask Detection Model

The COVID-19 epidemic has been largely responsible for the rapid development of face mask detection algorithms. The development of deep learning models such as convolutional neural networks has facilitated advancements in face detection. Deep convolutional neural networks serve multiple purposes, including the processing of numerical data and text data.

CNN-based face detection methods show promising results in spite of a wide range of challenges, including pose, low-resolution pictures, and illumination, among others. Although research into face mask detection for various occlusions has been conducted, this field is far from mature. A common deep learning model used for object detection is the convolutional neural network.

The great feature extraction quality and affordable CPU costs of CNN make it essential for computer vision jobs. Its use in classifying photos into two groups is just one of its numerous uses. CNN convolutionally layers a number of kernels onto feature maps or images to extract abstract characteristics. But the most important problem to be solved is how to create a better CNN structural architecture. The inception model is another popular and well-known CNN. As a model of a more complex neural network, a residual network learns its mapping from the layer below it. In Mobile Net a network for mobile object detection that relies on a minimum computing cost. Deep and channel-wise compression are used in the model to reduce computing expenses.

The layers of a convolutional neural network (CNN) are convolutional, pooling, and entirely connected. Each stratum's functions are distinct. The CNN model employs a numeric matrix known as a kernel or filter, which convolves across the image and transforms it based on the filter's values. Images are efficiently processed when their size is decreased at the end of each convolutional stage. Given an input image I (x, y), and a convolution kernel f (x, y), we may describe the resulting output image y (i, j) as:

$$y(i,j) = (I, f)(x,y) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} I(x-u, y-v) f(u,v) \quad (1)$$

The effect of the kernel's motion on the image's center is quite minimal compared to its effect on the periphery. To compensate, a border can be created around the image. The following criteria should be met by any padding used:

$$P = \frac{(f-1)}{2}$$

where stands for padding and represents filter dimensions. The non-linear activation function of the Sigmoid type is employed.

$$sigmoid(x) = \frac{1}{1+e^{-x}}$$

The convolutional output is subsampled by the pooling layer, which may be a max or average depending on the task at hand. The formula for calculating pooling is:

$$X_{ij}^{[l]} = \frac{1}{MN}\sum_{m}^{M}\sum_{n}^{N} X_{(iM+m, jN+n}^{[l-1]}$$

where and display the sample sizes i, j and M and N present the output map positions.

3.3 Metrics for Evaluation

In this study, the accuracy, precision, recall, and F1-score were utilised to evaluate the efficiency of deep learning models. These measurements are predicated on the four terms true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Individuals depicted donning face masks but are not actually wearing them so by the model are denoted by the letter "FP," whereas "TP" denotes the individual actually wearing the mask. Similarly, TN denotes non-mask-wearing individuals anticipated to be NM, while FN denotes mask- wearers also predicted to be NM.

1. **Accuracy** = (Number of Correct Predictions) / (Total Number of Predictions) Mathematically, it can be represented as:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

2. **Precision** is the proportion of accurate positive forecasts to all positive forecasts

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

Precision = TP / (TP + FP)

3. Recall (also known as Sensitivity or True Positive Rate) is the ratio of genuine positive instances to true positive predictions.

Recall = TP / (TP + FN)

4. F1- Score is the harmonic mean of precision and recall, which provides a balanced evaluation of a model's performance:

F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

IV. RESULT

4.1 Results Obtained Without Image Preprocessing

Systematically evaluating the efficiency of each model using the four-step image processing pipeline, initial tests are conducted. When photos are fed directly into the models without any preprocessing such as colour modification, filter application, etc. Table 4 shows the resulting images. As can be observed, the models do not perform particularly well. Despite this, the customised CNN outperforms state-of-the-art models with a precision of 99.52 percent. In the terms of precision, recall, and F1-scores, it outperforms YOLO v3 as well as Faster R-CNN. Data analysis shows that facial masks are not easily distinguished from other objects. For instance, the colour of the mask may be similar to the wearer's skin tone or facial hair, making mask detection challenging. As a result, four distinct phases of picture preparation techniques have been implemented prior to classifier training.

Table 1: Deep learning models for face mask detection on the RILFD dataset using preprocessing methods.

Models	Accuracy	Precision	Recall	F1-score
YOLO v3	90.92	88.27	92.23	90.45
Faster R-CNN	93.66	92.82	95.24	93.24
Customized CNN	98.25	97.20	98.34	97.74

Utilising the previously discussed picture preparation techniques is the second set of experiments. Table 1's presentation of the experiment's findings shows how picture preparation improves performance when compared to doing nothing. Every model that uses deep learning has significantly improved performance. Image preprocessing effectiveness. 92.65% accuracy, 93.82% precision, 92.84% recall, and 95.24% F1-score are higher for the Rapid R-CNN than for other CNNs, demonstrating superior performance. Version 3 of YOLO. CNN had the best overall performance, scoring 97.24% on the F1 with 95.24% accuracy, 98.34% precision, 93.24% recall, and 977.44% F1-score. A straightforward, one- shot algorithm with a quick inference time is the YOLO V3 algorithm. Comparatively, faster R-CNN has produced better outcomes. Efficacy in general and speed are mutually exclusive.

The CNN model with the best performance's accuracy trajectory is shown in Figure 3. It shows how consistently the suggested framework improves the precision of training and testing, proving the practicality of the suggested strategy. CNN-based face mask identification performs better in this study than other deep learning models.

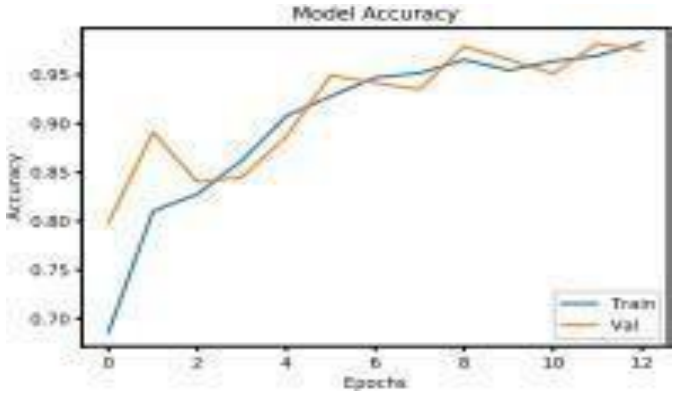

Figure 4. Curve of accuracy for the customized CNN model.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

4.2 Performance of Machine Learning Models

This work implements a number of machine learning models, such as RF, LR, ETC, SGD, and SVM, in addition to commercially available deep learning models and a distinctive CNN model. These models were selected based on their claimed effectiveness on tasks that were similar. The performance of the customised CNN model is compared with that of these machine learning models. These models are assessed using the RILFD dataset of face photos with and without face coverings. The performance of the models based on machine learning are shown in Table 2, and they show LR achieves the greatest accurate score of 83.45% with picture preprocessing steps. When picture preparation processes are used, machine learning model effectiveness is greatly increased. Despite having the best accuracy among machine learning models, LR's performance for face mask identification on the collected dataset is inferior to that of customised CNN models, which achieves a 98.25% accuracy rate.

Table 2: Comparative analysis of facial mask detection machine learning models

	Accuracy (%)	
Model	**With- preprocessing**	**Without -preprocessing**
SGD	79.80	74.48
RF	80.60	72.77
ETC	82.35	79.77
SVM	80.24	75.12
LR	84.46	78.34
Customized CNN	98.25	89.37

Figure 5: Accuracy comparison of customized CNN with other models.

4.3 Performance of Deep Learning Models Using MAFA and MOXA Datasets

The suggested customisable CNN model and other deep learning models have been further tested using three well-known datasets: MAFA, MOXA, and RMFRD. Due to their widespread use in face disguise detection, these datasets were chosen. On these datasets, the customised CNN outperformed YOLO v3 and Faster R-CNN, proving the usefulness of the suggested method. On the MAFA dataset, it achieves 95.74 percent accuracy and 94.29% recall, and on the MOXA dataset, it achieves 94.3 percent accuracy and 95.2 percent recall. On the RMFRD dataset, the adaptable CNN performs noticeably better, showcasing its remarkable accuracy with scores of 99.63% accuracy and 99.69% recall.

Table 3. Comparison of model performance utilizing MAFA, MOXA, and RMFRD datasets

Model	Dataset	Accuracy	Recall
YOLO v3	MAFA	91.47	89.24
Faster R-CNN	MAFA	92.65	91.87
Customized CNN	MAFA	96.74	95.29

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

YOLO v3	MOXA	87.34	89.85
Faster R-CNN	MOXA	88.35	88.31
Customized CNN	MOXA	95.37	96.28
YOLO v3	RMFRD	98.56	99.74
Faster R-CNN	RMFRD	98.97	98.31
Customized CNN	RMFRD	99.64	99.72

Additionally, the customised CNN performs well since real photographs from a publicly available dataset are used for comparison, despite the accuracy of simulated facial masks being normally rather good. As shown in Table 3, When compared to the selected studies and the proposed strategy, the novel method outperforms traditional models in terms of accuracy and precision. The high accuracy of face mask recognition is reported to have a 99.40% accuracy rate. Since this approach uses a face detection, region of interest, and face mask detection procedure and does not take the unpredictable nature of face detection into account, the accuracy stated is inflated.

V. CONCLUSION

Real-time face tracking and identification system design. This system tracks the human's face using a webcam-based system and compares the results with pre-recorded facial photos to perform its job. To prevent the spread of COVID-19, one of the most important restrictions imposed by governments is the requirement to wear a face mask. However, it becomes difficult and time-consuming to ensure that individuals in public areas are wearing masks on their faces. This paper presents a CNN-based solution to the problem of automatic face hidden form detection. Real-world, manually tagged high-definition image data is collected for this objective. Using a four-step image preparation procedure, a face mask with high efficacy can be identified. In addition to machine learning models such as LR, RF, SGD, ETC, and SVM, commercially available YOLO v3 and Faster RCNN are utilised in experiments. The results indicate that image preprocessing improves the performance of deep learning and machine learning models. Machine learning, YOLO v3, and Faster R-CNN cannot compete with the optimum accuracy CNN model, which has an accuracy rate of 98.25 percent. On the RILFD dataset, the proposed CNN model is more straightforward and less complex, and it trained quicker. To predict face coverings, YOLO v3 and Faster RCNN performed poorly and required additional time. Cross-validation experiments and trials on the two publicly available datasets MAFA and MOXA support the superior performance of the recommended CNN.

VI. REFERENCES

[1] Minaee, S., Luo, P., Lin, Z., & Bowyer, K. (2021). Going deeper into face detection: A survey. arXiv preprint arXiv:2103.14983.

[2] Ghoshal, A. M., Aspat, A., & Lemos, E. (2021). OpenCV Image Processing for AI Pet Robot. International Journal of Applied Sciences and Smart Technologies, 3(1), 65-82.

[3] Zarkasi, A., Abdau, F., Anda, A. J., Nurmaini, S., Stiawan, D., Suprapto, B. Y., ... & Kurniati, R. (2022). Implementation of Facial Landmarks Detection Method for Face Follower Mobile Robot. Generic, 14(1), 19-24.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

[5] Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 212-220).

[6] Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., ... & Liu, W. (2018). Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5265-5274).

[7] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[8] Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE international conference on computer vision (pp. 1520-1528).

[9] Sun, Y., Chen, Y., Wang, X., & Tang, X. (2014). Deep learning face representation by joint identification-verification. Advances in neural information processing systems, 27.

[10] Madhuri, G. M. G., Shaik, B., Tungala, S. S., Kumar, C. D. S., & Pilla, V. R. S. (2023). RECOGNITION AND TRACING OF OBJECT USING CNN. Journal Of Engineering Sciences, 14(03).

[11] Zhang, T., Li, J., Jia, W., Sun, J., & Yang, H. (2018). Fast and robust occluded face detection in ATM surveillance. Pattern Recognition Letters, 107, 33-40.

[12] Leung, N. H., Chu, D. K., Shiu, E. Y., Chan, K. H., McDevitt, J. J., Hau, B. J., ... & Cowling, B. J. (2020). Respiratory virus shedding in exhaled breath and efficacy of face masks. Nature medicine, 26(5), 676-680.

[13] Feng, S., Shen, C., Xia, N., Song, W., Fan, M., & Cowling, B. J. (2020). Rational use of face masks in the COVID-19 pandemic. The Lancet Respiratory Medicine, 8(5), 434-436.

[14] Wang, Z., Huang, B., Wang, G., Yi, P., & Jiang, K. (2023). Masked face recognition dataset and application. IEEE Transactions on Biometrics, Behaviour, and Identity Science.

[15] Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild: past, present and future. Computer Vision and Image Understanding, 138, 1-24.

[16] Li, X., Lai, S., & Qian, X. (2021). Dbcface: Towards pure convolutional neural network face detection. IEEE Transactions on Circuits and Systems for Video Technology, 32(4), 1792-1804.

[17] Kumar, A., Kaur, A., & Kumar, M. (2019). Face detection techniques: a review. Artificial Intelligence Review, 52, 927-948.

[18] Jignesh Chowdary, G., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020). Face mask detection using transfer learning of inceptionv3. In Big Data Analytics: 8th International Conference, BDA 2020, Sonepat, India, December 15–18, 2020, Proceedings 8 (pp. 81- 90). Springer International Publishing.

[19] Iqbal, M. T. B., Abdullah-Al-Wadud, M., Ryu, B., Makhmudkhujaev, F., & Chae, O. (2018). Facial expression recognition with neighbourhood-aware edge directional pattern (NEDP). IEEE Transactions on Affective Computing, 11(1), 125-137.

[20] Khaireddin, Y., & Chen, Z. (2021). Facial emotion recognition: State of the art performance on FER2013. arXiv preprint arXiv:2105.03588.

[21] Kamińska, D., Aktas, K., Rizhinashvili, D., Kuklyanov, D., Sham, A. H., Escalera, S., ... & Anbarjafari, G. (2021). Two-stage recognition and beyond for compound facial emotion recognition. Electronics, 10(22), 2847.

[22] Reddy, A. H., Kolli, K., & Kiran, Y. L. (2022). Deep cross feature adaptive network for facial emotion classification. Signal, Image and Video Processing, 16(2), 369-376.

[23] Khattak, A., Asghar, M. Z., Ali, M., & Batool, U. (2022). An efficient deep learning technique for facial emotion recognition. Multimedia Tools and Applications, 1-35.