iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59325

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Beyond the Mask: Advancements in Facial Recognition through Convolutional Neural Networks

Prof. Md. Mizan Chowdhury[1], Abhishek Kumar Jha[2], Soudip Panja[3], Radha Kumari[4], Oiendrila Banerjee[5], Abhraneel Khan[6], Paramjeet Mukherjee[7]

[1]*Assistant Professor, Department of IT, Guru Nanak Institute of Technology, Panihati, Kolkata, West Bengal*
[2, 3, 4. 5, 6, 7]*UG Student, Department of IT, Guru Nanak Institute of Technology, Panihati, Kolkata, West Bengal*

Abstract: ***Facial recognition technology has gained immense popularity in recent years, yet challenges persist with face alterations and the widespread use of masks. Uncooperative individuals in real-world scenarios, such as video surveillance, often resort to masking, leading to a degradation in current face recognition performance. While extensive research has addressed face recognition under various conditions, the impact of masks has often been overlooked. This research focuses specifically on improving the accuracy of recognizing faces obscured by different masks. Our proposed approach involves initial detection of facial regions, addressing the occluded face detection problem through the Multi-Task Cascaded Convolutional Neural Network (MTCNN). Subsequently, facial feature extraction utilizes the Google Face Net embedding model, followed by a classification task executed by Support Vector Machine (SVM). Experimental results demonstrate the effectiveness of this approach in masked face recognition, with noteworthy performance even under excessive facial masks. Additionally, a comparative study enhances our understanding of the proposed method's capabilities.***

Keywords: Facial detection, MTCNN, Face Net, Facial embedding, Masked face, SVM, Facial recognition.

I. INTRODUCTION

Face recognition stands as a promising domain within applied computer vision, enabling the automatic identification of individuals from given images. Widely utilized in various daily activities such as passport checks, smart doors, access control, voter verification, and criminal investigations, face recognition serves as a reliable biometric recognition technology, surpassing other methods like passwords, pins, and fingerprints. Governments worldwide show considerable interest in face recognition systems to enhance security in public places, including parks, airports, bus stations, and railway stations. Over the years, face recognition has been extensively studied, and notable progress has been made in technology.

The rapid advancement and proliferation of machine learning techniques have substantially addressed challenges in face recognition. In recent years, deep learning has achieved significant breakthroughs in various computer vision domains, including object detection, classification, segmentation, and notably, face detection and verification. Deep learning methods eliminate the need for manual feature design, allowing convolutional neural networks (CNNs) to automatically learn valuable features from training images. The popularity of CNNs has grown as they apply convolution filters successively, accompanied by non-linear activation functions. Unlike earlier methods that used a limited number of convolution filters with average or sum pooling, recent approaches leverage a more extensive set of pre-trained filters on large datasets, proving effective in detecting faces under various conditions and poses. Comparing recognition accuracy, traditional methods like the Eigenface algorithm achieve approximately 60% on the Labeled Faces in the Wild (LFW) dataset. In contrast, advanced deep learning algorithms, exemplified by FaceNet, boast an accuracy rate of 99.63%, surpassing the normal human eye acceptance rate of 99.25%. International projects have successfully applied deep learning methods, including FaceNet and DeepFace, to achieve remarkable accuracy in face recognition.

Despite these achievements, popular face recognition systems still face challenges in handling unrestrained circumstances characterized by image degradation, changes in facial pose, occlusions, and other constraints. Individuals can alter their identity through face modifications or the use of various physical attributes. Facial occlusions, whether deliberate or unintentional, present challenges, such as those posed by sunglasses, scarves, hoods, beards, and medical masks. The presence of obstacles in front of faces, like different objects, glasses, scarves, caps, or other occlusions, further complicates recognition. Real-world face occlusion problems are categorized into three types: occlusion of facial landmarks, occlusion by different objects, and occlusion by faces.

©IJRASET: All Rights are Reserved

Failure to specifically consider mask analysis can significantly impact the performance of sophisticated face recognition methods. The reduction in key features for identifying a person due to various masks or occlusions leads to challenges in recognition accuracy. Terrorists and criminals often use masks to disguise their identities, making masked faces a major concern in face recognition. Deep learning networks face additional challenges in this context, given the insufficient quantity of training data for direct application. This necessitates the use of transfer learning, which, while generally effective, may not always yield satisfactory results when fine-tuning pre-trained deep learning networks due to limited training data. This work focuses on enhancing face recognition accuracy across different types of masks.

II. CONVOLUTIONAL NEURAL NETWORK

The Convolutional Neural Network (CNN) comprises multiple convolutional layers, pooling layers (e.g., min, max, or average), non-linear layers (e.g., sigmoid, ReLU), and a classification layer (e.g., Softmax) with various units. Deep CNN networks are commonly trained on large labeled datasets like ImageNet, extracting general features applicable to diverse detection and recognition tasks, including image classification, verification, object detection, segmentation, and texture identification. Combining CNN architecture with multiple detectors allows the isolation of different object portions, achieving state-of-the-art results in fine-grained recognition, such as identifying specific dog breeds, bird species, or car models. CNN excels at capturing local essential features from data, selecting global training components, and demonstrating success across various pattern recognition applications. The deep CNN exhibits robustness in handling scaling, shifting, and transposing. To address the objectives of this work, diverse forms of CNN architecture are employed. Figure 1 illustrates a classic CNN architecture.

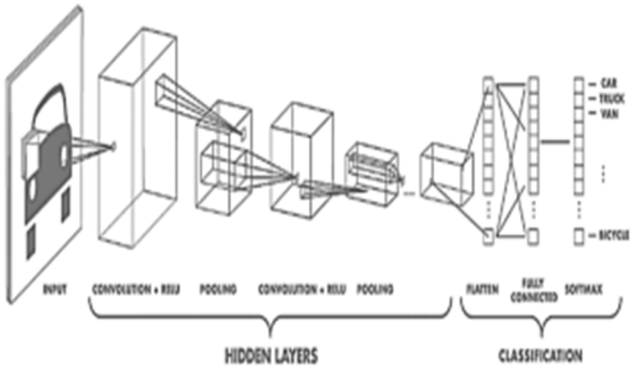


Fig. 1: Classic CNN architecture

III. METHODOLOGY

Our methodology comprises three core modules: Face detection in a given image, feature extraction, and recognition. Figure 2 illustrates our approach to the challenge of face recognition in masked conditions.

A. Facial Image Acquisition:

Image acquisition stands as the initial crucial stage in the face recognition process. We gathered masked and non-masked face images from the AR and IIIT-Delhi face databases. To address the limitation of a limited number of images, we implemented a data augmentation process on both masked and non-masked face images from the database, enhancing the dataset's size to ensure the reliability and efficiency of our work.

B. Masked Face Detection Using MTCNN:

Detecting faces from an input image is a crucial step in face recognition, as further progress relies on the presence of identifiable faces. To achieve this, we employ the Multi-task Cascaded Convolutional Neural Network (MTCNN), as proposed by K. Zhang, Z. Li, Z. Zhang, and Y. Qiao. MTCNN excels in face detection and alignment, surpassing many face-detection benchmarks while maintaining real-time performance. Utilizing a pre-trained MTCNN model, we detect candidate masked and non-masked face portions in the given image, interpreting them into high-dimensional facial descriptors. The MTCNN model consists of three cascading networks, each serving a specific purpose.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The first network, P-Net (Proposal Network), introduces candidate facial regions by rescaling the image using an image pyramid. The second network, R-Net (Refine Network), refines the bounding boxes, and the third network, O-Net (Output Network), determines facial landmarks. These networks perform face classification, bounding box regression, and facial landmark localization, collectively forming a multi-task network. The interconnection between these models involves feeding the outputs of one stage into the next, ensuring additional processing in different stages. Non-Maximum Suppression (NMS) is applied in the MTCNN model to refine candidate bounding boxes suggested by P-Net, R-Net, and O-Net before providing the final output. This face detection method excels under various lighting conditions, poses, and visual variations of the face. Figure 3 all the steps involved in the MTCNN method.

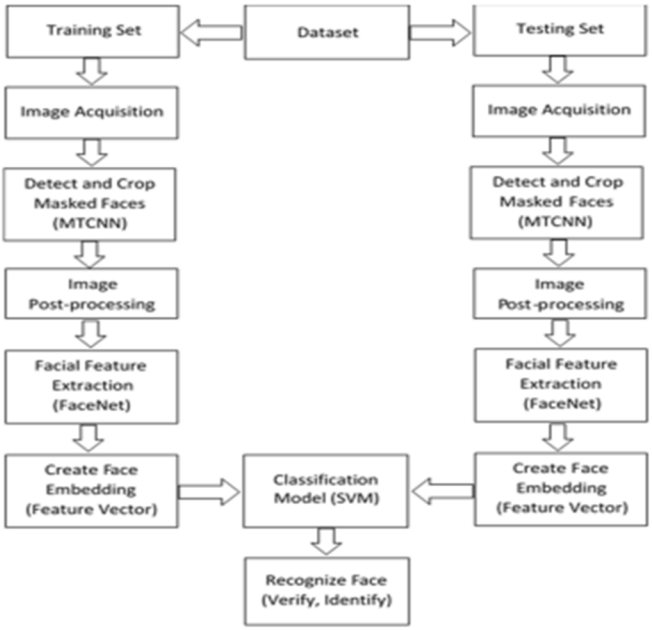


Fig. 2: Masked Face Recognition Flowchart

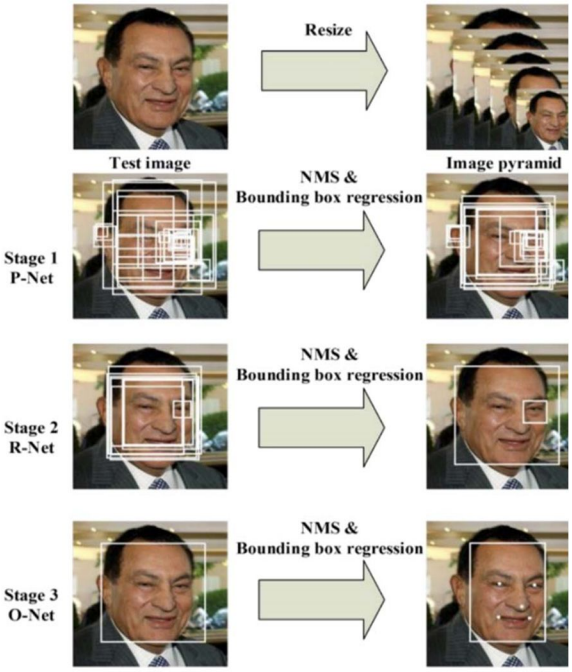


Fig. 3: Pipeline of Multi-Task Cascaded Convolutional Neural Network.

C. Image Post-processing

Following face detection, crop and resizing methods are applied to the input images using the bounding box generated during the face detection phase by the MTCNN model. This bounding box is utilized to crop the facial portion from the input images. Specific normalized parameters specified in the model architecture details guide the normalization process. In accordance with the FaceNet architecture details, all cropped images are resized to a standardized dimension of 160x160.

D. Feature Extraction using FaceNet

FaceNet, recognized as the state-of-the-art in face recognition, serves as the baseline for our deep neural network, contributing to identification, verification, and neural network clustering. In this methodology, the pre-trained FaceNet model is integrated with a batch layer and an extensive CNN network, supported by L2 normalization. The outcome of this normalization is the face embedding. During training, face embedding is optimized using triplet loss, where minimal distance occurs between an anchor and a positive when the identities match, while the distance is higher between an anchor and a negative when identities differ.



Fig. 4: The Pipeline of FaceNet model.

The FaceNet model is built upon 22 deep convolutional network layers, directly training its output on these layers to achieve a compact 128-dimensional embedding. After rectification, the fully connected layer serves as the face descriptor. These descriptors are transformed into a similarity-based descriptor using the embedding module. To create a unique feature vector from a template, the max operator is applied to the features. For the specific task of face recognition and verification, the network undergoes fine-tuning to expect a substantial performance boost. A substantial dataset of masked and non-masked face images is utilized for retraining the FaceNet model in this work. Figure 4 outlines the FaceNet model pipeline.

E. Face Verification Using SVM:

In the final step, the verification process is consolidated to recognize candidate faces by executing the classification task within a unified Support Vector Machine (SVM). SVM, a widely utilized algorithm in various classification-related problems since its introduction, determines a hyperplane to optimize classification tasks. It maximizes the margin between two classes within a given input-target pair, providing a classifier with a certain level of robustness against overfitting. The margin signifies the efficiency of class separation. In this phase, we compare a test face to other training faces using SVM. The classification result is deemed correct when the distance between the test image and the training image of the same person is minimized. The similarity of a masked face is measured by evaluating L2 normalization within the feature keypoints collected from the network structure, comparing masked and non-masked faces.

IV. DATASET CONSTRUCTION

Training deep learning networks necessitates a substantial number of images, and the scarcity of large masked face datasets poses a challenge for masked face recognition. To address this, we introduce our Masked Face Database (MFD), specifically designed to overcome the limitations of existing datasets. The MFD comprises diverse occlusions, orientations, and complexities, featuring 45 subjects with various disguises against different simple and complex backgrounds. The dataset includes 990 images of male and female subjects aged between 18 to 26 years. These images showcase various disguises, including (i) sunglass (ii) scarf (iii) medical mask (iv) beard (v) sunglass with scarf (vi) sunglass with a beard and (vii) sunglass with a medical mask.

In addition to our MFD, we leverage two existing databases, namely the AR Face Database and IIIT-Delhi Disguise Version 1 Face Database, for validation purposes. Figure 5 illustrates example images from each database.

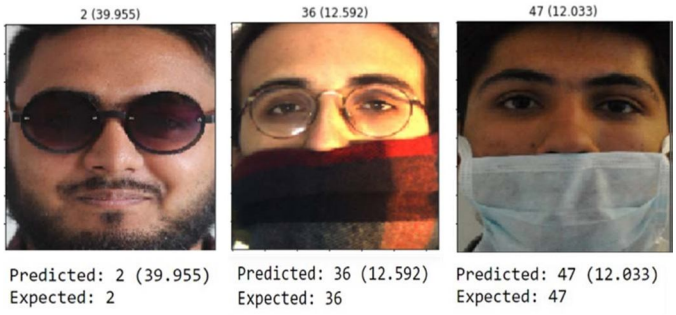


Fig. 5: Example images from (a) AR [17] (b) IIIT-Disguise V1 [18] (c) MFD face database.

The dataset images are partitioned into training and test sets, with each training set containing an average of 70% of total images and the test set comprising an average of 30% of total images for each dataset.

V. EXPERIMENTAL RESULTS

Our approach undergoes evaluation on the AR Face Database, IIIT Disguise Face Database, and our Masked Face Database (MFD). Testing outputs from different database images are presented in Figure 6.



(a) (b) (c)

Fig. 6: Test images from (a) MFD (b) AR (c) IIIT Disguise VI face database

The accuracy of masked face recognition is measured across various scenarios, Table-I outlines these scenarios, and Table-II presents the corresponding recognition accuracy.

Table I: Masked Face Recognition Training and Testing Scenario

Dataset	Scenario	Subjects			Train			Total Test Images (All Masked)	Total Images (Train + Test)
		Male	*Female*	*Total*	*Masked*	*Non-Masked*	*Total*		
AR	Scenario 1	76	59	135	0	1666	1666	760	2426
	Scenario 2	76	59	135	704	1666	2370	935	3305
IIIT Version 1	Scenario 3	58	17	75	0	1273	1273	533	1806
	Scenario 4	58	17	75	325	148	473	208	681
Proposed MFD	Scenario 5	31	14	45	0	900	900	233	1133
	Scenario 6	31	14	45	178	895	1073	315	1388
	Scenario 7	31	14	45	0	900	900	350	1250
	Scenario 8	31	14	45	253	899	1152	461	1613

From Table-I, it is observed that the accuracy of masked face image recognition using FaceNet is higher when mixing masked and non-masked images for training compared to using only non-masked images. Additionally, as the complexity of masked images increases, the recognition accuracy tends to decrease. A graphical representation of these recognition accuracies is provided in Figure 7 for enhanced data visualization.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

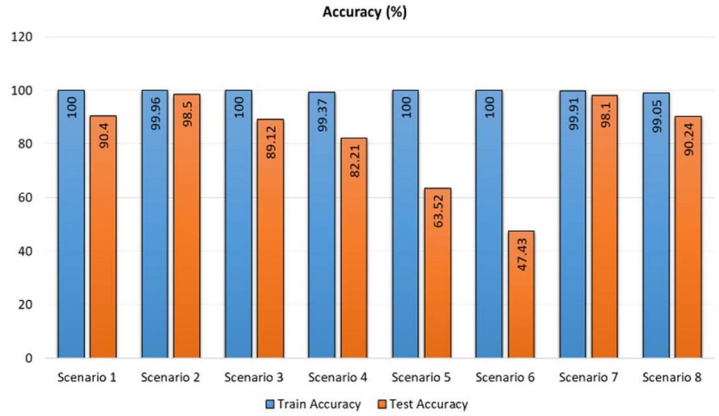


Fig. 7: Training and testing accuracy of masked face recognition.

Table II. Masked Face Recognition Accuracy

Scenario	Train Image	Test Image	Train Accuracy (%)	Test Accuracy (%)
Scenario 1	Non-masked Faces	Masked Faces	100.00	90.40
Scenario 2	Non-masked Faces + Masked Faces	Masked Faces	99.96	98.50
Scenario 3	Non-masked Faces	Masked Faces	100	89.49
Scenario 4	Non-masked Faces + Masked Faces	Masked Faces	99.37	82.21
Scenario 5	Non-masked Faces	Masked Faces	100.00	63.52
Scenario 6	Non-masked Faces + Masked Faces	Masked Faces	99.91	98.10
Scenario 7	Non-masked Faces	Masked Faces (Complex)	100.00	47.43
Scenario 8	Non-masked Faces + Masked Faces (Complex)	Masked Faces (Complex)	99.05	90.24

To further illustrate the classification algorithm's performance and correctness, we employ the Precision-Recall curve. This visualization technique demonstrates changes between precision and recall for various thresholds. A larger area under the curve indicates both a high precision rate and a high recall rate, signifying correct outputs from the classifier.

The Receiver Operating Characteristic (ROC) curve is another visualization method showcasing the actual performance of the classification model. It presents the correlation between recall and precision with varying threshold values for defining a positive in our model. In Figure 8, we present Precision-Recall and ROC curves for Scenario 2 (highest accuracy) and Scenario 7 (lowest accuracy) in this work.

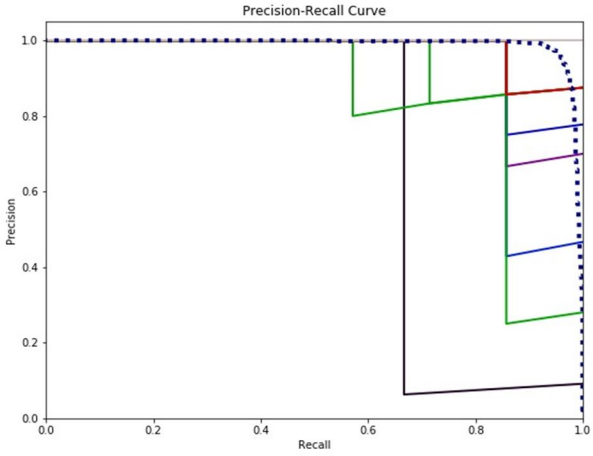


Fig. 8: Precision-Recall Curve (a)

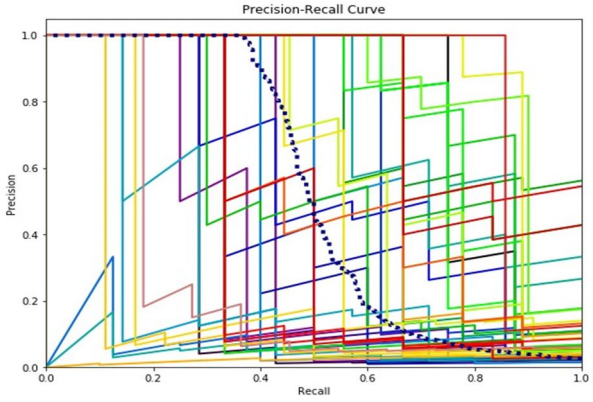


Fig. 9: Precision-Recall Curve (b)

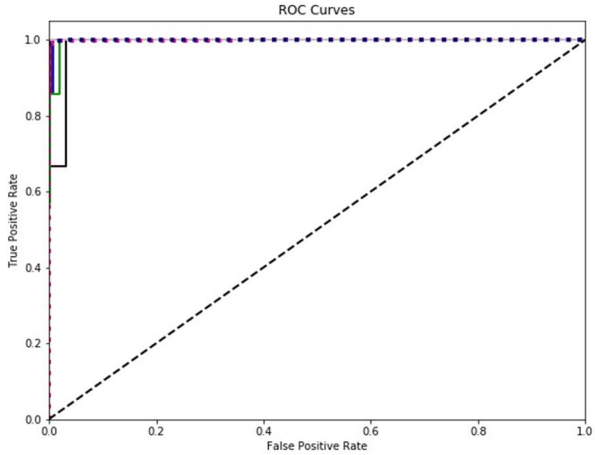


Fig. 10: ROC curves

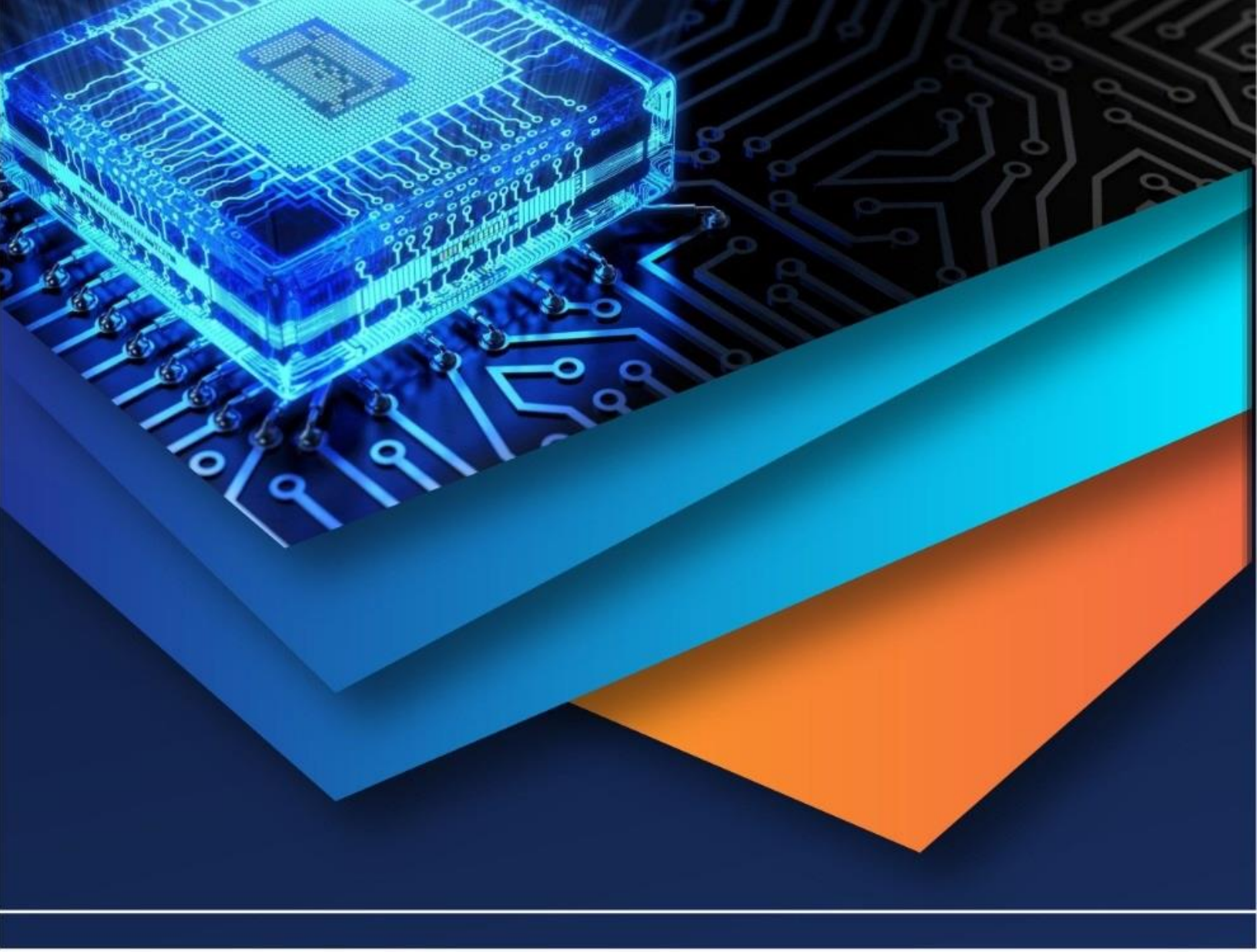
VI. CONCLUSION

In this study, the FaceNet pre-trained model has been employed to enhance masked face recognition. The proposed approach has been benchmarked against two established datasets and our created dataset, demonstrating improved recognition rates. Utilizing FaceNet trained on both masked and non-masked images has resulted in enhanced accuracy for simple masked face recognition. While our focus has been on masks induced by various factors like hats, sunglasses, beard, long hairs, moustache, and medical masks, the methodology can be extended to handle more complex occlusions and diverse sources. It is important to note that this method may not cover all types of masks, and further research may require more accurate and sophisticated approaches for comprehensive coverage. Future work aims to enhance and expand this approach to address different extreme mask conditions in face recognition. As industries move towards Industry 4.0 and Sustainable Technology, emphasizing computer adoption, mechanization, and autonomous intelligent systems fueled by data and machine learning, the importance of security in handling data becomes crucial. Our work contributes to making these smart and autonomous industries more self-governing, secure, accurate, and efficient, ultimately promoting increased production and reduced waste.

REFERENCES

[1] Roomi, Mansoor, Beham, M.Parisa, “A Review Of Face Recognition Methods,” in International Journal of Pattern Recognition and Artificial Intelligence, 2013, 27(04), p.1356005.
[2] A. S. Syed navaz, t. Dhevi sri, Pratap mazumder, “ Face recognition using principal component analysis and neural network,” in International Journal of Computer Networking, Wireless and Mobile Communications (IJCNWMC), vol. 3, pp. 245-256, Mar. 2013.
[3] Turk, Matthew A., and Alex P. Pentland. "Face recognition using eigenfaces," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991, pp. 586-591.
[4] H. Li, Z. Lin, X. Shen, J. Brandit, and G. Hua, “A convolutional neural network cascade for face detection,” in IEEE CVPR, 2015, pp.5325-5334.
[5] Wei Bu, Jiangjian Xiao, Chuanhong Zhou, Minmin Yang, Chengbin Peng, “A Cascade Framework for Masked Face Detection,” in IEEE International Conference on CIS & RAM, Ningbo, China, 2017, pp.458-462.
[6] Shiming Ge, Jia Li, Qiting Ye, Zhao Luo, “Detecting Masked Faces in the Wild with LLE-CNNs,” in IEEE Conference on Computer Vision and Pattern Recognition, China, 2017, pp. 2682--2690.
[7] M. Opitz, G. Waltner, G. Poier, and et al, “Grid Loss: Detecting Occluded Faces,” in ECCV, 2016, pp. 386-402.
[8] X. Zhu and D. Ramanan, “Face Detection, pose estimation and landmark localization in the wild,” in IEEE CVPR, 2012, pp.2879-2886.
[9] Florian Schroff, Dmitry Kalenichenko, James Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823
[10] Ankan Bansal Rajeev Ranjan Carlos D. Castillo Rama Chellappa, “Deep Features for Recognizing Disguised Faces in the Wild,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2018.
[11] K. Zhang, Z. Zhang., Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” in IEEE Signal Processing Letters 23, no.10, 2016, pp.1499-1503.
[12] Byun, Hyeran, and Seong-Whan Lee, “Applications of support vector machines for pattern recognition: A survey,” in International Workshop on Support Vector Machines, pp. 213-
[13] 236. Springer, Berlin, Heidelberg, 2002.
[14] Aleix M Martinez, “AR Face Database,” [Online]. Available: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html. [Accessed 1 September 2019].
[15] “Image Analysis and Biometrics Lab @ IIIT Delhi,” [Online]. Available: http://www.iab-rubric.org/resources/facedisguise. html. [Accessed 1 September 2019].
[16] N. Kohli, D. Yadav, and A. Noore, “Face Verification with Disguise Variations via Deep Disguise Recognizer,” in CVPR Workshop on Disguised Faces in the Wild, 2018.
[17] S. V. Peri and A. Dhall, “DisguiseNet : A Contrastive Approach for Disguised Face Verification in the Wild,” in CVPR Workshop on Disguised Faces in the Wild, 2018.
[18] K. Zhang, Y.-L. Chang, and W. Hsu, “Deep Disguised Faces Recognition,” in CVPR Workshop on Disguised Faces in the Wild, 2018.
[19] T. Y. Wang and A. Kumar, “Recognizing human faces under disguise and makeup,” in IEEE International Conference on Identity, Security and Behavior Analysis, 2016.
[20] Yingcheng Su, Zhenhua Guo, Yujiu Yang, Weiguo Yang, “Face Recognition with Occlusion,” in 3rd IAPR Asian Conference on Pattern Recognition, 2015.
[21] G. Righi, J. J. Peissig, and M. J. Tarr, “Recognizing disguised Faces,” in Visual Cognition, pp. 143–169, 2012.
[22] R. Singh, M. Vatsa, and A. Noore, “Face recognition with disguise and single gallery images,” in Image and Vision Computing, pp. 245–257, 2009.
[23] R. Min, A. Hadid and J. Dugelay, “Improving the recognition of faces occluded by facial accessories,” in Face and Gesture 2011(FG), Santa Barbara, CA USA, 2011, pp. 442-447.
[24] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa, “Disguised faces in the wild,” in IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018.
[25] G. Apostolopoulos, V. Tzitzilonis, V. Kappatos and E. Dermatas, "Disguised Face identification using multi-modal features in a quaternionic form," in 7th International Conference on Imaging for Crime Detection and Prevention (ICDP), Madrid, 2016, pp. 1-6.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)