

## RESEARCH ARTICLE

# Skin Segmentation-Based Disguised Face Recognition Using Deep Learning

G. PADMASHREE<sup>1</sup> AND KARUNAKAR A. KOTEGAR<sup>2</sup>, (Senior Member, IEEE)

Department of Data Science and Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India

Corresponding author: Karunakar A. Kotegar (karunakar.ak@manipal.edu)

**ABSTRACT** Disguised face recognition refers to the ability of a computer system or algorithm to identify a person's face even when they are wearing some form of disguise, such as a mask, hat, sunglasses, or makeup. This is a challenging problem in computer vision and pattern recognition, as disguises can significantly alter a person's facial features and appearance, making it difficult to match the image with a known face. In this work, we propose a novel approach to disguised face recognition by focusing on the skin regions of the face, which are less likely to be covered by disguises. Skin regions are segmented from the faces by applying a region-based marker-controlled watershed algorithm and features are extracted from these skin regions using a convolutional neural network and classify the disguised faces using a deep learning-based recognition model. The results show that the proposed model achieves high accuracy on  $64 \times 64$  image size, with an overall accuracy of 94.92%. We also performed an ablation study to analyze the impact of different factors on the performance of the proposed approach, including the image size and the size of the kernel filter. Overall, our approach provides a promising solution for the challenging problem of disguised face recognition.

**INDEX TERMS** Deep learning, disguised faces, skin segmentation, Sejong face dataset, region-based marker-controlled watershed algorithm.

## I. INTRODUCTION

Face recognition is one of the most researched topics in computer vision and machine learning. With the advent of deep learning-based approaches, significant progress has been made in the field of face recognition [1], [2], [3]. However, the performance of face recognition systems can be severely impacted when faces are disguised. Disguises can range from simple items such as glasses or hats to more complex ones such as makeup, facial hair, or masks. The variability and unpredictability of disguises pose a significant challenge for face recognition algorithms [4], [5]. Disguised face recognition (DFR) is the process of identifying an individual wearing a disguise or altering their appearance. DFR is a critical task with several applications, including surveillance, law enforcement, security, and access control. Traditional face recognition algorithms typically rely on

the availability of high-quality face images for training and testing. However, the scarcity of disguised face datasets and difficulties in acquiring high-quality images in real-world scenarios makes DFR challenging. In recent years, there has been significant research interest in DFR, to develop robust algorithms that can perform well under different disguises. Various approaches have been proposed for DFR, including deep learning-based approaches, feature-based approaches, and template matching-based approaches. Deep learning-based approaches have shown promising results in DFR due to their ability to learn discriminative features from large amounts of data. Feature-based approaches, such as local binary patterns (LBP) and histogram of oriented gradients (HOG), have been widely used in DFR due to their simplicity and effectiveness. Template matching-based approaches, such as correlation-based matching, have also been used in DFR, although their performance is often limited by the variability of disguises. Some of the major problems in disguise face recognition are: (i) Changes in facial features:

The associate editor coordinating the review of this manuscript and approving it for publication was Shovan Barma<sup>1</sup>.

Disguises can alter the shape, color, and texture of a person's face, making it difficult for traditional facial recognition algorithms to accurately identify them. (ii) Lack of training data: Disguised face recognition requires a large amount of training data to develop effective algorithms. However, obtaining high-quality images of people wearing disguises can be difficult, leading to a lack of representative training data. (iii) Variability in disguises: Disguises can take many forms, such as masks, hats, sunglasses, makeup, or facial hair. Each type of disguise can alter the appearance of a person in different ways, making it difficult for algorithms to generalize and accurately identify individuals in different types of disguise. (iv) Intra-class variability: Even within the same type of disguise, there can be significant variability in the way it is worn and applied, making it challenging to accurately identify individuals. (v) Inter-class similarity: Disguises can make individuals with different faces look similar, leading to confusion for recognition algorithms. Some of the sample disguise images from the Sejong face dataset [6] are shown in Figure 1.

The key contribution of our proposed model is

- Skin segmentation has proven to be a valuable technique in achieving good results in recognizing disguised faces. Disguises often involve modifications or coverings of the facial features, and by concentrating on the skin region, we can effectively detect and analyze these changes which helps in reducing noise and irrelevant information in the image, thus improving the quality of the input data.
- By focusing on the skin region, we remove background clutter, non-facial objects, and occlusions, enabling the recognition system to focus on essential facial features and disguise-related patterns.
- By isolating the skin region, we can extract intrinsic skin-related features, including texture, color, and patterns, which exhibit a high degree of discriminative power for individual recognition.
- The proposed model incorporates an ensemble of filters and kernels with different sizes thus capturing a wide range of spatial information at different scales.
- By considering multiple filter sizes and kernel sizes, the model can effectively capture both fine-grained and coarse-grained details.

The paper is organized as follows. Section II thoroughly assesses previous studies on disguised face recognition. Section III goes over the proposed methodology in great depth. Section IV includes the results and discussions of the tests performed on the Sejong face datasets, along with visualizations and data analysis. Section V concludes with remarks on our findings.

## II. RELATED WORK

Over the recent decades, research in face recognition has seen magnificent development. Most of the current face recognition frameworks have achieved exceptionally high

accuracy for unconstrained face datasets starting from constrained environments. Although most contemporary face recognition frameworks continue to improve in accuracy, most systems are subject to failure under disguise, which remains a difficult challenge.

As we all know, a lot of research is going on in the field of disguise face recognition, which is considered one of the confounding tasks, [7] came up with the idea of developing a model to recognize disguised faces. Their proposed model consisted of a two-stage training approach. In the first stage, they used two Deep Convolution neural networks (DCNNs) for extracting identity features from aligned and unaligned images, which were fused later. Later in the second stage, the Principal Component Analysis (PCA) transformation matrix was computed to recognize disguised faces. From this, they achieved an accuracy of about 79%, which was one of the best results in the Disguise Faces in the Wild (DFW) competition phase-1.

Recognition of faces under variations of resolution of images, illumination, age, pose, etc. It becomes more challenging when they are disguised. Reference [8] developed an approach for identifying the disguised faces and rejecting the impersonators. During the training process, two different DCNNs (Inception ResNet-v2 and ResNet-101) are trained and L2-softmax loss is used to improve the Softmax loss. Features are fused from two different networks by taking the average of the scores. Once the training is complete, fused features are embedded into the discriminative subspace of the metric learning framework. When testing the given two faces, features are calculated from DCNN and embedded into the subspace. Finally, the similarity score is calculated between the two embedded features. The authors could achieve an accuracy of about 72.9% which was promising and provides direction for future work for better understanding.

To verify disguised faces, [9] created a transfer learning method based on deep learning that makes use of the Residual Inception network framework with center loss. In "Deep Disguise Recognizer(DDR)", the training phase works in two phases. In the first phase, the face representations are learned by training the deep inception ResNet network with a huge face database, and in the second phase, the pre-trained model is transferred to the DDR to encode the face representation of facial disguises. The authors observed that the verification accuracy varied among different databases and also among different DCNN models. Also, analysis proved that the DDR-MSCeleb framework had good performance for genuine male subjects as compared to genuine female subjects.

Reference [10] proposed a method for determining a person's identification among disguised and impostor photographs. Their strategy relies on a VGG-face design coupled with a Contrastive loss based on a cosine distance metric using the Disguised Faces in the Wild(DFW) dataset. In comparison to the DFW baseline, the suggested network increases accuracy by 27.13% by performing augmentation



**FIGURE 1.** Sample images were taken from the Sejong Face Dataset [6], a comprehensive collection of facial images displaying a wide range of disguises. This dataset incorporates numerous disguises, such as glasses, masks, hats, false beards, wigs, and combinations of these accessories, to simulate real-world settings. The deliberate use of disguises is designed to test facial recognition systems' robustness to various degrees of occlusion and modification. This dataset is used in our study to investigate the complexity of facial recognition in disguised situations, providing vital insights for developing biometric security systems in dynamic, real-world environments.

and also helped in increasing the generalization of the network.

The DFW data set, which includes more than 11000 photos of 1000 identities with changes in various types of disguise accessories in [3], is provided. Each identity has both imposter and real obfuscated face photos in this dataset. To demonstrate the complex nature of the problem, different levels of difficulty are assessed, such as easy, medium, and hard. In-depth descriptions of the DFW dataset are provided in this work, together with baseline results, evaluation methodologies, performance analyzes of the entries that were submitted as part of the First International Workshop and Competition on DFW, and the three difficulty levels of the DFW challenge dataset.

The research [11] introduces A2-LINK, an active learning system, to handle the challenge of face recognition in the presence of disguise. A2-LINK starts with a face recognition machine learning model, intelligently chooses training samples from the target domain, and

then uses hybrid noises like adversarial noise to fine-tune a network that performs well in both the presence and absence of disguise. Experimental findings on the DFW and DFW2019 datasets with cutting-edge deep learning feature models like DenseNet, ArcFace, and LCSSE and show the efficacy and generalizability of the proposed approach.

DED-Net an encoder-decoder network was proposed by [12] which learns the local and global features of both disguised and non-disguised images based on Cosine and Mahalanobis distance metrics along with their class variations. Disguise Resilient (D-Res) framework is the name given to the entire framework. Low-resolution images are also considered as part of this research among which  $32 \times 32$ ,  $24 \times 24$ , and  $16 \times 16$  image resolutions are considered from the benchmark datasets DFW2018 and DFW2019. An accuracy of 96.3% is achieved using this framework and an improvement of 3% was observed when compared to the state-of-the-art techniques.

To overcome the limitation of database availability, in the research domain, [6] presented a multi-modal disguised face dataset. In this dataset, there were 100 participants, each of which had 15 different disguised photos and 8 distinct face add-ons, as well as their combinations which were captured under different modalities which include infrared, visible, thermal spectra, and visible plus infrared.

Reference [13] suggested disguise invariant face recognition (DIFR) framework which detects faces using the Viola-Jones face detector by applying noise based augmentation technique. Features are learned by using a fine-tuned pre-trained Convolutional Neural Network (CNN) for identifying the disguised faces. Four different pre-trained models are used for identifying the disguised faces and Resnet-18 has outperformed with an accuracy of 98.19%.

Reference [14] proposed a joint segmentation and identification feature learning framework for occlusion face recognition which consists of an occlusion prediction (OP) module, channel refinement (CR) network, and feature purification (FP) module. The predicted occlusion mask is transformed into a channel-wise mask matrix, which is then utilized to reduce occlusion characteristics while emphasizing more discriminative visible features in spatial and channel dimensions. Instead of embedding the non-occlusion feature maps directly, an FP module has been designed specifically to enhance the viable candidate embeddings from the combined original and occlusion-free feature maps. Also, they proposed an upgraded occlusion face dataset Webface-OCC+ for evaluation to achieve generalization.

Reference [15] suggested a Deep Convolution Neural Network which used two biometric qualities, gait, and face for recognizing individuals. The feature vectors of gait energy images and facial images are merged and fed into the CNN model for feature extraction. This framework achieved an accuracy of 97.5% using ORL Face, FEI Face, and CASIA Gait datasets.

Owing to a scarcity of annotated datasets featuring low-resolution images in disguise, [16] proposed a dataset D-LORD which consists of HR mugshot images and Low-resolution surveillance videos of 2,100 subjects and 14,098 low-resolution surveillance videos with over 1.2M frames. Under varying lighting situations, the subjects appear to be wearing various disguise artifacts such as hats, monkey caps, wigs, sunglasses, and, face masks. The challenging issue of low-resolution face recognition with disguise variations is initially addressed by D-LORD.

Dosi et al. [17] introduced Seg-DGDNet, a model designed for disguised face recognition that adeptly identifies and removes occluded regions in facial images during feature extraction. The model's effectiveness is rigorously evaluated across various datasets and test scenarios, encompassing diverse disguise types and image resolutions. Experimental results highlight the remarkable performance of the Seg-DGDNet model, positioning it as superior to state-of-the-art models in both low-resolution face recognition and the demanding domain of disguised face recognition.

Cheema and Moon [18] presents a unique approach to Human Face Recognition (HFR) with the Deep Neighborhood Difference Relational (DNDR) network and Joint Discrimination Loss. Unlike using a handcrafted metric loss function, this method concentrates on learning relationships between images of the same identity. The network calculates relational distinctions among cross-modality images in the deep feature space. The fusion of embedding distance and match probability enhances classification accuracy, providing increased robustness in the recognition process.

Kumar et al. [19] introduce a novel thermal face recognition approach using the Radial Derivative Gaussian Feature (RDGF) descriptor. Their proposed cascaded framework combines BoCNN and the RDGF descriptor, assessing BoCNN performance before classification. It incorporates a dynamic classifier selector during runtime to choose between handcrafted features and the CNN framework, enhancing overall performance. Experimental assessments across diverse datasets highlight the efficacy of the RDGF descriptor, validated through comparative analyses against state-of-the-art descriptors.

Reference [20] delves into strategies for detecting disguises and analyzing faces within a dataset, focusing on facial key points for these processes. Challenges arise when props block, hide or disorient these key points. To overcome this, the proposed approach generates a face estimate using a limited set of key points. This estimated face is then compared with the available set to probabilistically assess the likelihood of the matched person's presence, even in the presence of a disguise. The solution aims to distinguish whether the obtained image represents an impostor or an intruder, incorporating obfuscation with a defined probability.

This field has received significant attention from researchers in recent years, owing to its potential applications in various domains. The studies reviewed in this section have highlighted the challenges associated with disguised face recognition, such as the quality of the image, illumination, and various occlusions. Several approaches have been proposed to address these challenges, including the use of deep learning-based methods. However, it is clear that there is still a significant gap in the performance of state-of-the-art methods compared to traditional face recognition systems, which suggests the need for further research and development in this area. Overall, the reviewed studies provide valuable insights and directions for future work in improving the accuracy and robustness of disguised face recognition systems.

### III. PROPOSED APPROACH

In this section, the description of the process of disguised face recognition in images is given. The steps involved are depicted in Figure 2 and the entire process is illustrated in Figure 3. A YOLO face detector is utilized for the initial detection of the faces. Then, the skin region is extracted from the detected faces through the use of a region-based watershed algorithm. The features are extracted from the

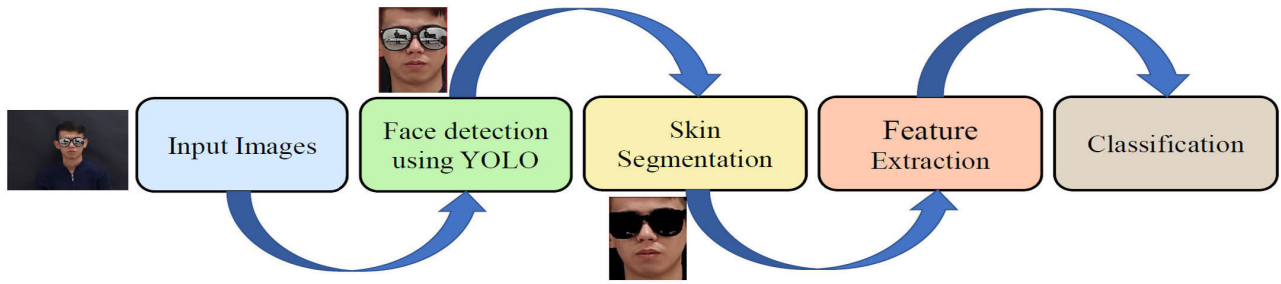


FIGURE 2. Steps involved in recognition of disguised faces.

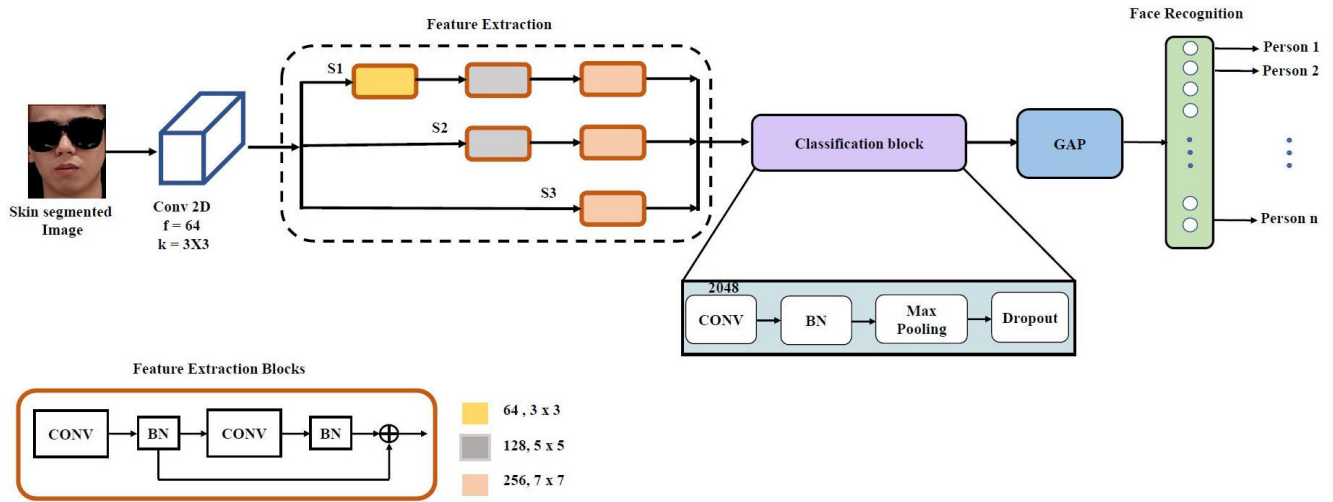


FIGURE 3. A Holistic Framework for Facial Recognition Under Disguise Challenges.

skin-segmented images and fed into the classification block for the recognition of disguised faces.

**A. PRE-PROCESSING AND FACE DETECTION**

In the pre-processing phase from the raw images of the Sejong dataset which are of  $1382 \times 1061$  resolution, faces are extracted using a YOLO face detector and are resized to various scales like  $32 \times 32$ ,  $64 \times 64$ ,  $128 \times 128$ , and  $224 \times 224$  for extracting the features and perform recognition. Table 1 exhibits some examples of subjects in various disguises, as well as their scaled images. Also, augmentation is applied to the images to increase the size of a training dataset by creating new, slightly modified versions of the original data to avoid over-fitting and increase the model’s accuracy and generalization ability. During the training process, different augmentation techniques have been applied such as horizontal shift, vertical shift, zooming, and horizontal flip and brightness has also been increased in the range of 0.2 to 1.2 to increase the size of the dataset.

**B. SKIN SEGMENTATION**

Here, we employ a marker-controlled watershed segmentation algorithm [21] for extracting the skin region from the disguised faces which are mainly used in an image of irregular

shapes. The algorithm consists of gradient transformation which transforms the image into a topographical representation, where the intensity values of the pixels are used to define the heights of the terrain, followed by marker placement to identify the regions that correspond to objects of interest, then flooding by filling the catchment basins until the basins merge or reach an edge of the image and finally segmentation. Here, we make use of HSV and YCbCr color spaces for extracting the skin region of the faces. By applying various morphological operations such as erosion and dilation, the images are pre-processed and then the region-based marker-controlled watershed algorithm is applied to extract the skin region of the detected faces.

**C. FEATURE EXTRACTION**

This model is a deep convolutional neural network (CNN) with three parallel branches ( $S_1, S_2, S_3$ ) that process the input image separately and then combine their outputs for classification. This parallel processing allows the network to extract diverse features from the image at different scales and levels of detail.

$S_1$  branch uses three convolutional blocks, each consisting of two convolutional layers followed by ReLU activation and batch normalization. The number of filters increases

**TABLE 1.** Sample images from sejong face dataset [6] of subjects in various disguises along with their scaled images.

| 32 x 32   | 64 x 64   | 128 x 128   | 224 x 224   |
|---|---|---|---|
|    |    |    |   |
|  |  |  |  |

progressively with each block (64, 128, 256), allowing it to capture increasingly complex features. Additionally, the outputs of the first and second blocks are concatenated before the subsequent layers, further enriching the feature representation. *S2* and *S3* branches both use two convolutional blocks, each with two convolutional layers followed by the same activation and normalization functions. However, they differ in their filter sizes and numbers. *S2* uses a  $5 \times 5$  filter size in the first block and a  $7 \times 7$  filter size in the second block, with 128 and 256 filters, respectively. *S3* uses only  $7 \times 7$  filter size with 256 filters in both blocks. The outputs of all branches are concatenated, creating a rich feature representation that combines diverse information about the

image which acts as the input to the classification block. This feature map contains a high-dimensional representation of the image features extracted through the network’s various convolutional and pooling layers. The first layer in the classification block is a fully connected layer with 2048 neurons. This layer takes each element in the feature map as an input and transforms it into a higher-level representation through linear transformation and bias addition. ReLU activation is applied to this layer’s output, introducing non-linearity, and helping the network learn complex relationships between features. After the fully connected layer, batch normalization is applied. This technique helps stabilize the learning process by normalizing input features across mini-

batches, leading to faster training, improved generalization, and reduced overfitting. Max pooling is employed to reduce the spatial dimensionality of the feature map, allowing the network to focus on the most relevant features. Dropout is further applied with a probability of 0.5, randomly dropping out neurons during training to prevent overfitting and improve the model's robustness. Then the feature maps are fed to the Global Average Pooling layer which replaces the spatial dimensions of the feature map with their average values, resulting in a single vector representation of the entire feature map. This effectively summarizes the global information captured by the convolutional layers, allowing the model to focus on overall image characteristics rather than specific spatial locations. Finally, the global average pooled vector is fed into a final classification layer. This layer typically uses a softmax function to compute the probability distribution over different output classes, enabling the model to classify the input image. The concept of Parallel Processing allows the network to extract features from different aspects of the image simultaneously, potentially leading to faster and more accurate results.

The initial convolutional branch,  $S_1$ , is specifically designed to capture fine-grained details present in the image. It utilizes smaller filters in its early layers to focus on intricate patterns and subtle features. On the other hand, branches  $S_2$  and  $S_3$  progressively prioritize larger structures within the image. As we delve deeper into these branches, the convolutional filters expand in size. This strategic approach enables the network to analyze features across a spectrum of scales, encompassing both fine details and more global structures. The concurrent utilization of parallel processing, along with the integration of progressively larger filters, is crafted to enhance the model's performance in tasks related to image understanding. This architectural design proves particularly advantageous in scenarios where the ability to capture information at diverse scales is paramount for achieving robust analysis.

#### D. CLASSIFICATION BLOCK

The classification block consists of a convolutional layer with 2048 filters, a size of  $3 \times 3$ , and ReLU activation, followed by batch normalization, max pooling, and drop out. The output of the dropout layer is then globally average pooled which reduces the spatial dimensions of the feature maps to create a one-dimensional vector, which is passed through a dense layer with 64 units and softmax activation to produce the final classification output.

## IV. EXPERIMENTS AND RESULTS

### A. DATASET

Subset-A and Subset-B are the two subsets of the Sejong database [6]. Subset-A comprises 30 individuals' facial images, 16 males and 14 females, recorded with one neutral and one add-on image in each modality. Frontal faces were used in all of the images. Subset-B comprises 70 individuals'

**TABLE 2.** Summary of images with various disguises available in Sejong face dataset. For example, breakdown of 'Natural Face' Category Images: The 'Natural Face' category in the table includes 15 images per subject, representing the maximum allowable quantity within the dataset. Subjects may be either male or female. This implies that the dataset consists of 15 distinct images for each subject, showcasing diverse orientations such as facing left, facing right, looking downwards, looking upwards, and more.

|                    | Accessories          | # of Images | Gender |        |  |
|--------------------|----------------------|-------------|--------|--------|--|
|                    |                      |             | Male   | Female |  |
| No Add-on          | Natural Face         | 15          | Yes    | Yes    |  |
|                    | Real Beard           | 10          | Yes    | No     |  |
| Accessory Add-on   | Cap                  | 5           | Yes    | Yes    |  |
|                    | Scarf                | 5           | Yes    | Yes    |  |
|                    | Glasses              | 5           | Yes    | Yes    |  |
|                    | Mask                 | 5           | Yes    | Yes    |  |
|                    | Makeup               | 5           | No     | Yes    |  |
|                    | Wig                  | 10          | Yes    | Yes    |  |
| Fake Add-on        | Fake Beard           | 5           | Yes    | No     |  |
|                    | Fake Mustache        | 5           | Yes    | No     |  |
|                    | Wig - Glasses        | 5           | No     | Yes    |  |
| Combination Add-on | Cap - Scarf          | 5           | No     | Yes    |  |
|                    | Glasses - Scarf      | 5           | Yes    | Yes    |  |
|                    | Glasses - Mask       | 5           | Yes    | Yes    |  |
|                    | Fake Beard - Cap     | 5           | Yes    | No     |  |
|                    | Fake Beard - Glasses | 5           | Yes    | No     |  |
|                    |                      |             |        |        |  |
|                    |                      |             |        |        |  |

facial images, 44 males and 26 females, with 15 neutral face images and 5 add-on images acquired in each modality for the remaining add-ons. In addition, 5 photos with actual beards for men and cosmetics for women were recorded. Subset-A contains 1, 500 images with 30 subjects, 4 modalities, 12 – 13 add-ons, and 1 pose, whereas Subset-B contains 23, 100 images with 70 subjects, 4 modalities, 12 – 13 add-ons, and 5 – 15 poses. The summary of various disguised images available in the Sejong dataset is provided in Table 2.

In our research, we exclusively utilized visible RGB images from the Sejong database. Our dataset comprised 64 subjects, with 23 females and 41 males. Each subject was represented by 15 neutral face images and 5 add-on images, which included 5 photos featuring actual beards for men and cosmetics for women. Furthermore, the dataset encompassed 12 – 13 add-ons and 5 – 15 poses for each subject. The number of images in each category varied from 5 to 15, demonstrating diversity across categories. To guarantee rigorous and unbiased evaluation, we used a strategic dataset split in our trials in the ratio 70 : 15 : 15. The dataset was divided into three parts: training(70%, validation(15%, and testing(15%. The training set, which included a significant fraction, allowed the model to learn from a variety of examples and patterns. Concurrently, the validation set was crucial in fine-tuning the model by acting as a reference point throughout training. A separate testing set was used to evaluate the model's performance on unseen data(refers to data from individuals who have not contributed images to the training set). This methodical approach to dataset splitting improves the dependability and impartiality of our experimental results, allowing us to gain a better grasp of the model's capabilities.

### B. EVALUATION METRICS

To provide a thorough picture of our model's performance, we evaluate it using a wide collection of metrics such as

precision, recall, F1-score, and accuracy. Accuracy, as a fundamental metric, offers an overall assessment of the model's correctness across all classes. It provides a high-level evaluation of overall performance. Precision is concerned with the model's capacity to detect positive cases among the expected positives. It is especially useful in situations when reducing false positives is crucial. Recall, also known as sensitivity, is the model's ability to capture all positive cases among the actual positives. When preventing false negatives is a priority, this metric is critical. The F1-score, which is a harmonic mean of precision and recall, provides a fair evaluation of the model's performance. It is useful in situations where striking a balance between precision and recall is critical.

The One-vs-Rest (OvR) AUC evaluation technique is a popular method for dealing with multi-class classification utilizing binary classification algorithms. The original multi-class problem is converted into numerous binary tasks, one for each class, by OvR. Each task considers a specific class to be the "positive" class while classifying all other classes as "negative." For each job, independent binary classifiers are trained and evaluated, with performance indicators such as AUC (Area Under the Curve) produced individually. These indicators can be combined using methods such as micro-averaging or macro-averaging to get an overall OvR score, which provides a comprehensive assessment of the model's capacity to handle many classes.

The utilization of this wide range of measures ensures a well-rounded review. While accuracy provides a broad overview, precision, recall, and F1-score dig deeper into specific elements like false positives and false negatives. This multi-metric method seeks to provide a sophisticated and detailed assessment that corresponds to the diverse needs of many applications and circumstances. We improve the accuracy and depth of our model evaluation by using a broad set of metrics. We believe that this diverse evaluation technique provides a robust and meaningful study of our model's performance.

### C. EXPERIMENTAL SETUP

A series of tests are carried out to illustrate the efficiency of the suggested model. These experimental results are thoroughly covered in this section. An HP Elite Desk 800G4 Workstation with a 48GB NVIDIA GeForce RTX A6000 GPU, 128GB RAM, and an i9 processor running at 3.7 GHz are used for all of the research. The Keras(2.3.1) Framework and Python(3.9) are used to implement the algorithm. All the experiments were carried out by employing categorical cross-entropy as the loss function and Adam as the optimizer. We utilize a batch size of 8 with an initial learning rate of  $10^{-4}$  and train our network for a maximum of 10 epochs.

### D. PERFORMANCE EVALUATION OF THE PROPOSED MODEL ACROSS DIFFERENT IMAGE SIZES

Security and law enforcement could benefit greatly from the development of disguised facial recognition, which is

a difficult problem. In this work, we suggest a novel deep learning model for recognizing disguised faces and assess its performance using a dataset of images of disguised faces from the Sejong face dataset. We evaluate the proposed model using numerous standard evaluation measures and undertake evaluations for varied image sizes. We test the proposed model on four distinct image sizes  $32 \times 32$ ,  $64 \times 64$ ,  $128 \times 128$ , and  $224 \times 224$  pixels. Table 3 shows the outcomes of the evaluations with the performance metrics, namely precision, recall, f1-score, accuracy, AUC, and training time per epoch for each image size.

As shown in the Table 3, the proposed model achieved the highest performance on the  $64 \times 64$  image size, with an overall accuracy of 94.92%. However, the performance decreased with increasing image size due to the decrease in image quality. This trend was observed in all the performance metrics, including precision, recall, and f1-score.

In terms of training time per epoch, the model took the longest time to train on the  $224 \times 224$  image size, which is expected due to the larger image size and more complex features. The  $32 \times 32$  image size had the shortest training time per epoch. Overall, the results demonstrate the effectiveness of the proposed model in disguised face recognition, with higher accuracy achieved on smaller image sizes.

To conduct further evaluations, we have selected  $64 \times 64$  images as they yielded the most optimal results.

### E. COMPARATIVE ANALYSIS OF PERFORMANCE USING HSV, YCBCR, AND RGB COLOR MODELS IN DISGUISED FACE RECOGNITION

In this section, we explore the utilization of different color models for accurate skin segmentation in the context of disguised face recognition. The primary objective is to extract precise skin regions from input images, which will serve as crucial input data for the proposed face recognition model. Table 4 shows some sample face images from which skin regions have been extracted considering different color models and Table 5 shows the results obtained for the recognition of disguised faces using the proposed model. From the results, it was observed that the segmented images obtained using by combining HSV and YCbCr color model images performed better in recognizing the disguised faces. Hence, all further experiments were performed by considering the segmented images which were obtained using a combination of HSV and YCbCr color models.

### F. ANALYZING THE IMPACT OF SKIN SEGMENTATION AND SKIP CONNECTIONS ON RECOGNIZING DISGUISED FACES IN THE PROPOSED MODEL

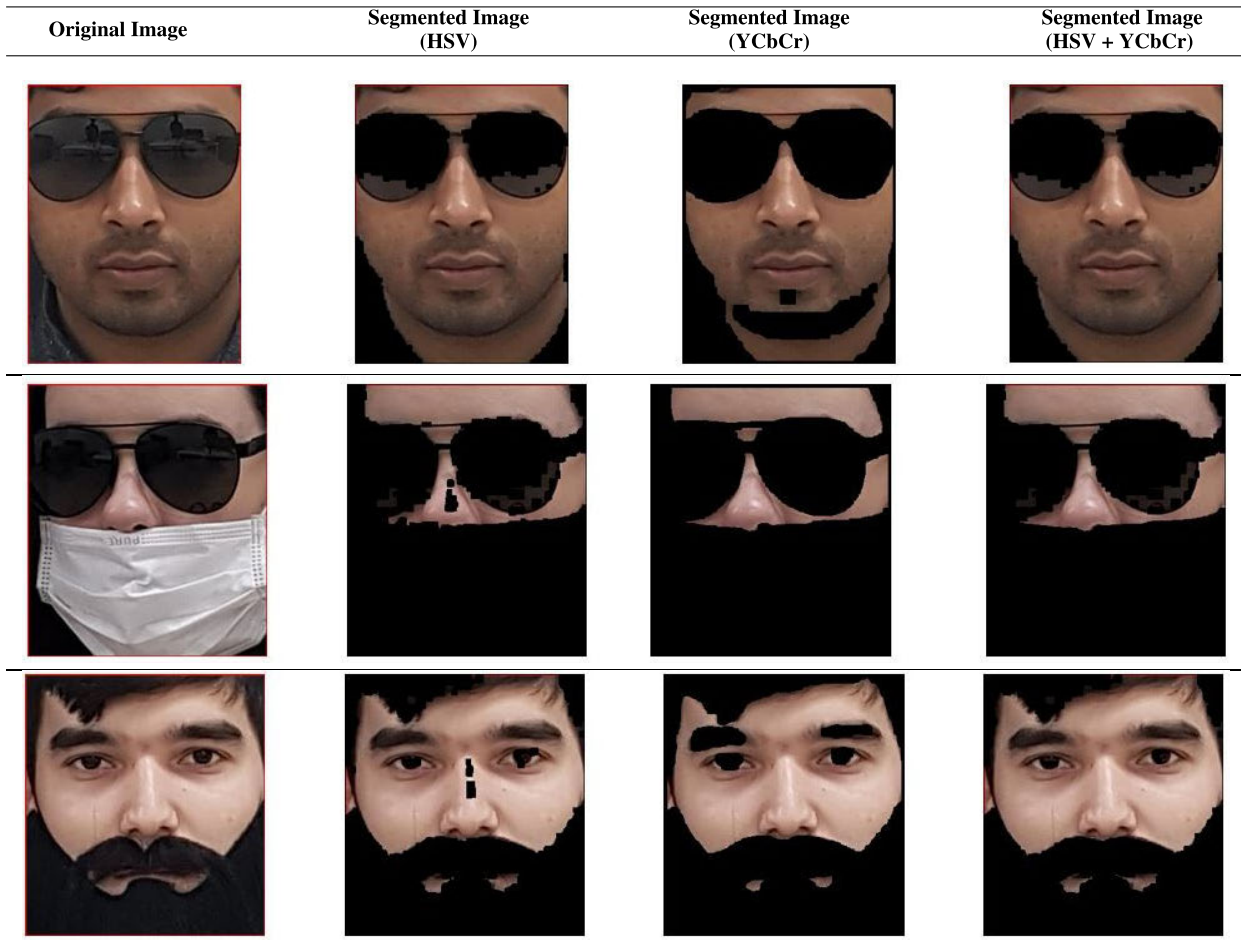
To investigate the impact of segmentation and skip connection in the proposed model, we conducted a comprehensive evaluation that encompassed both skin-segmented and non-segmented images. By segmenting the skin from the input images, we aimed to enhance the model's ability to focus on relevant facial features and improve recognition accuracy. The performance metrics, including precision, recall,



**TABLE 3.** Comparative performance evaluation of proposed model across different image sizes.

| Image Size | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) | AUC (%) | Training Time per Epoch (Sec) |
|------------|---------------|------------|--------------|--------------|---------|-------------------------------|
| 32 × 32    | 93.90         | 92.52      | 93.20        | 92.48        | 96.2    | 52                            |
| 64 × 64    | 95.57         | 94.98      | 95.27        | 94.92        | 97.45   | 140                           |
| 128 × 128  | 91.86         | 88.54      | 90.17        | 88.53        | 94.18   | 475                           |
| 224 × 224  | 88.35         | 80.52      | 84.25        | 80.07        | 90.10   | 1700                          |

**TABLE 4.** Skin Segmentation: Visual Comparison of different color schemes on sample images.



**TABLE 5.** Evaluation of proposed model performance using segmented images from different color models.

| Color Scheme | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) | AUC (%) |
|--------------|---------------|------------|--------------|--------------|---------|
| HSV          | 94.43         | 93.56      | 93.99        | 93.42        | 96.73   |
| YCbCr        | 95.63         | 94.35      | 94.99        | 94.17        | 97.13   |
| HSV + YCbCr  | 95.57         | 94.98      | 95.27        | 94.92        | 97.45   |

f1-score, and accuracy, of the proposed model are presented in Table 6. This analysis allowed us to examine the effectiveness of segmentation in improving the model’s performance.

Additionally, we investigated the impact of incorporating skip connections in the proposed model. Skip connections, also known as residual connections, are a technique commonly used in neural network architectures, particularly in

deep convolutional neural networks. The idea is to create shortcuts or connections that skip one or more layers. We have employed skip connections by concatenating the output of the batch normalization layer with the output of the batch normalization layer in the next sequential layer. Instead of simply passing this output to the next layer, skip connections facilitate the flow of information through the network providing an additional pathway for gradients during backpropagation. This helps in mitigating issues like vanishing gradients and accelerates the training of deep networks. Skip connections establish direct connections between layers, enabling the model to capture and utilize low-level features effectively. By evaluating the model’s performance with and without skip connections, we assessed the contribution of this architectural design choice in terms

**TABLE 6.** Performance assessment of proposed models with segmentation and skip connections: Comparing skin and non-skin segmented images.

| Method  | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) | AUC (%) |
|---------|---------------|------------|--------------|--------------|---------|
| Model 1 | 95.57         | 94.98      | 95.27        | 94.92        | 97.45   |
| Model 2 | 91.25         | 87.03      | 89.09        | 86.84        | 93.41   |
| Model 3 | 93.75         | 91.99      | 92.86        | 91.91        | 95.93   |
| Model 4 | 93.05         | 92.17      | 92.61        | 91.91        | 96.02   |

**TABLE 7.** Evaluating the performance of the proposed model in comparison to state-of-the-art models.

| Method         | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) | AUC (%) |
|----------------|---------------|------------|--------------|--------------|---------|
| Proposed Model | 95.57         | 94.98      | 92.57        | 94.92        | 97.45   |
| ResNet50       | 86.62         | 81.4       | 83.93        | 81.2         | 90.55   |
| ResNet101      | 85.17         | 80.48      | 82.76        | 80.26        | 90.08   |
| ResNet152      | 85.66         | 82.92      | 84.27        | 82.7         | 91.32   |

of recognition accuracy and the model's ability to handle variations in disguised faces.

The scenarios considered for the evaluation are as follows:

- **Model 1:** w/ Segmentation + w/ Skip connections
- **Model 2:** w/o Segmentation + w/ Skip connections
- **Model 3:** w/ Segmentation + w/o Skip connections
- **Model 4:** w/o Segmentation + w/o Skip connections

By conducting a thorough evaluation with both skin-segmented and non-segmented images, as well as assessing the impact of skip connections, we gained valuable insights into the effectiveness of these techniques in improving the performance of the proposed model for disguised face recognition. These findings contribute to our understanding of the importance of segmentation and skip connection strategies in addressing the challenges associated with recognizing disguised faces thus achieving an accuracy of 94.92%.

### G. COMPARATIVE ANALYSIS OF THE PROPOSED MODEL AGAINST STATE-OF-THE-ART DEEP LEARNING MODELS FOR FACE RECOGNITION

The performance of the proposed model is compared with state-of-the-art models like (ResNet50, ResNet101, and ResNet152 [22]) and the results demonstrate that the proposed model has achieved superior performance. Table 7 provides a detailed comparison of various evaluation metrics, including accuracy, precision, recall, and f1-score.

In terms of accuracy, the proposed model achieved a significantly higher accuracy of 94.92%, outperforming the best-performing state-of-the-art model by 12.22%. This indicates that the proposed model is more effective in correctly classifying disguised faces.

Furthermore, the precision of the proposed model was observed to be 95.57%, which is substantially higher than the precision achieved by the state-of-the-art models. This implies that the proposed model exhibits a greater ability to accurately identify disguised faces without misclassifying them.

Similarly, the recall and f1-score of the proposed model were notably higher compared to the state-of-the-art models.

This signifies that the proposed model excels in correctly detecting disguised faces, ensuring a comprehensive and reliable recognition system.

These comparative results strongly justify that the proposed model has outperformed existing state-of-the-art models in terms of accuracy, precision, recall, and f1-score. The superior performance of the proposed model showcases its effectiveness and potential for real-world applications in disguised face recognition.

### H. ANALYZING THE IMPACT OF FILTER SIZE VARIATIONS IN THE FEATURE EXTRACTOR BLOCKS OF THE PROPOSED MODEL

Using varying filter sizes in a deep learning model can have several advantages:

- **Capturing features at different scales:** By using filters of different sizes, a model can capture features at different scales. For example, smaller filters may capture fine-grained details, while larger filters may capture broader patterns. By combining filters of different sizes, a model can capture features at multiple scales, which may lead to better performance on tasks that require a diverse range of features.
- **Reducing overfitting:** Using filters of different sizes can also help to reduce overfitting, which occurs when a model learns to perform well on the training data but does not generalize well to new data. By using filters of different sizes, a model can learn to recognize features in different ways, which may help it to generalize better to new data.
- **Faster training:** Using larger filters can lead to a larger number of parameters in the model, which can slow down the training. By using filters of different sizes, a model can balance the number of parameters and the complexity of the model, which can lead to faster training times.
- **Improved performance:** By using filters of different sizes, a model can capture a wider range of features, which may lead to better performance on the task at hand. For example, in natural language processing tasks, using filters of different sizes in a convolutional neural

**TABLE 8.** Analyzing performance of various filter sizes in feature extraction phase on 32 × 32 images.

| Image Size 32 × 32 |          |          |               |            |              |              |         |
|--------------------|----------|----------|---------------|------------|--------------|--------------|---------|
| Filter_3           | Filter_5 | Filter_7 | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) | AUC (%) |
| ✓                  | ✓        | ✓        | 93.9          | 92.52      | 93.2         | 92.48        | 96.2    |
| ✓                  |          |          | 92.59         | 91.65      | 92.11        | 91.48        | 95.26   |
|                    | ✓        |          | 95.4          | 94.15      | 94.77        | 93.98        | 97.02   |
|                    |          | ✓        | 92.71         | 91.24      | 91.96        | 91.16        | 95.55   |

**TABLE 9.** Analyzing performance of various filter sizes in feature extraction phase on 64 × 64 images.

| Image Size 64 × 64 |          |          |               |            |              |              |         |
|--------------------|----------|----------|---------------|------------|--------------|--------------|---------|
| Filter_3           | Filter_5 | Filter_7 | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) | AUC (%) |
| ✓                  | ✓        | ✓        | 95.57         | 94.98      | 95.27        | 94.92        | 97.45   |
| ✓                  |          |          | 92.22         | 92.61      | 92.41        | 92.48        | 96.24   |
|                    | ✓        |          | 92.6          | 90.69      | 91.63        | 90.6         | 95.27   |
|                    |          | ✓        | 93.33         | 91.79      | 92.55        | 91.54        | 95.82   |

**TABLE 10.** Analyzing performance of various filter sizes in feature extraction phase on 128 × 128 images.

| Filter_3 | Filter_5 | Filter_7 | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) | AUC (%) |
|----------|----------|----------|---------------|------------|--------------|--------------|---------|
| ✓        | ✓        | ✓        | 91.86         | 88.54      | 90.16        | 88.53        | 94.18   |
| ✓        |          |          | 84.67         | 78.35      | 81.38        | 77.81        | 89.1    |
|          | ✓        |          | 88.01         | 83.41      | 85.64        | 83.08        | 91.57   |
|          |          | ✓        | 93.36         | 91.38      | 92.35        | 91.16        | 95.62   |

**TABLE 11.** Analyzing performance of various filter sizes in feature extraction phase on 224 × 224 images.

| Image Size 224 × 224 |          |          |               |            |              |              |         |
|----------------------|----------|----------|---------------|------------|--------------|--------------|---------|
| Filter_3             | Filter_5 | Filter_7 | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) | AUC (%) |
| ✓                    | ✓        | ✓        | 88.35         | 80.52      | 84.25        | 80.07        | 90.10   |
| ✓                    |          |          | 82.32         | 77.26      | 79.7         | 76.87        | 88.44   |
|                      | ✓        |          | 83.28         | 75.36      | 79.12        | 75.18        | 87.48   |
|                      |          | ✓        | 86.96         | 81.96      | 84.38        | 81.57        | 90.83   |

network can improve performance by capturing both local and global patterns in the text.

Overall, using varying filter sizes can be a useful technique for improving the performance and generalization of deep learning models. The proposed model was tested by varying the filter size in each feature extractor block on different image scales. The different filter sizes used for training the model are 3, 5, and 7. It was observed that the performance of the model drastically reduces when the filter size was increased due to Loss of spatial information, increased computational cost, and overfitting. Tables 8, 9, 10, and 11 depict the performance of the proposed model for varying image scales and filter sizes.

**I. EVALUATING CHANNEL VARIANTS IN FEATURE EXTRACTOR BLOCKS OF THE PROPOSED MODEL**

In this study, we attempt to evaluate the efficiency of several feature extraction blocks of differing sizes in recognizing disguised faces. The suggested model’s feature extraction block is divided into three branches (S1, S2, S3), each of which contains numerous feature extractors. Convolution

**TABLE 12.** Comparative analysis of model designs: Influence of feature extractor blocks on performance metrics.

| S1 | S2 | S3 | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) | AUC (%) |
|----|----|----|---------------|------------|--------------|--------------|---------|
| ✓  | ✓  | ✓  | 95.57         | 94.98      | 95.27        | 94.92        | 97.45   |
| ✓  |    |    | 93.8          | 92.9       | 93.34        | 92.85        | 96.39   |
|    | ✓  |    | 90.67         | 88.67      | 89.65        | 88.34        | 94.24   |
|    |    | ✓  | 90.34         | 87.25      | 88.76        | 86.84        | 93.52   |
| ✓  | ✓  |    | 89.26         | 84.62      | 86.87        | 84.58        | 92.19   |
| ✓  |    | ✓  | 91.71         | 90.28      | 90.98        | 90.22        | 95.06   |
|    | ✓  | ✓  | 91.91         | 89.12      | 90.49        | 89.09        | 94.47   |

layers with varied kernel sizes comprise the feature extractor blocks.

Here, analysis is carried out by taking into account both individual branches and combinations of different branches, and how traits extracted to aid in the identification of the disguised faces are carried out. Table 12 displays the results achieved by considering individual branches and combinations of branches that aid in the recognition of disguised faces.

We trained and analyzed versions of our baseline model to see how different filter sizes affected performance.

**TABLE 13. Comparison of disguised face verification accuracy: State-of-the-Art vs. proposed method.**

| Method                 | Accuracy(%)  |
|------------------------|--------------|
| [6]                    | 92.6         |
| <b>Proposed Method</b> | <b>94.92</b> |

We initially considered each branch separately before experimenting with combinations of branches. Our results suggest that varying the filter size can have a major impact on model performance as it captures different scales of information, reduces overfitting, increases model capacity (the model can capture more complex patterns and relationships in the data), and achieve better localization. Thus our results suggest that varying filter sizes is the most effective approach, but further research is needed to confirm this finding on other datasets and architectures.

### J. BENCHMARKING THE PROPOSED MODEL AGAINST STATE-OF-THE-ART TECHNIQUES: A COMPARATIVE ANALYSIS

In this section, we compare the performance of our proposed model with state-of-the-art techniques in the field of disguised face recognition. We evaluate the effectiveness and robustness of our model by considering accuracy as the key performance metric. Table 13 presents the performance of the proposed disguised face recognition model along with the baseline results of the Sejong face dataset [6]. The efficiency of the suggested system can be observed when it achieves state-of-the-art performance with a Precision of 95%, Recall of 93.92%, F1-score of 93.88%, and Accuracy of 93.79%. The comparison with state-of-the-art techniques not only showcases the advancements made in the field of disguised face recognition but also reinforces the significance and effectiveness of our proposed model. It establishes our model as a reliable and state-of-the-art solution for addressing the challenges of disguised face recognition, paving the way for enhanced security and improved authentication systems.

### K. VISUALIZATION AND DATA ANALYSIS

To better understand the behavior of our proposed model, we generated heatmap diagrams using three different visualization techniques: GRAD-CAM(Gradient-weighted Class Activation Mapping) [23], Guided Backpropagation [24], and Guided GRAD-CAM [25]. These techniques help us identify the regions of the input image that were most important for the model's prediction. GRAD-CAM highlights the most relevant regions of the image for the predicted class and guided backprop shows which parts of the image contribute positively to the prediction. Guided grad-CAM combines both techniques to produce a visualization that shows the positive contributions in the regions highlighted by GRAD-CAM. Table 14 depicts the heatmap diagrams of the disguised faces obtained using different visualization techniques. Table 15 depicts some of the predictions made

using the suggested model. Furthermore, t-SNE(t-distributed stochastic neighbor embedding) plots can be used to visualize the distribution of the features learned by the proposed model for different image sizes as shown in Figure 4. By comparing the t-SNE plots of different image sizes, we can analyze the effect of image size on the learned features and the model's overall performance. t-SNE allows us to examine how the learned features change as the image size increases.

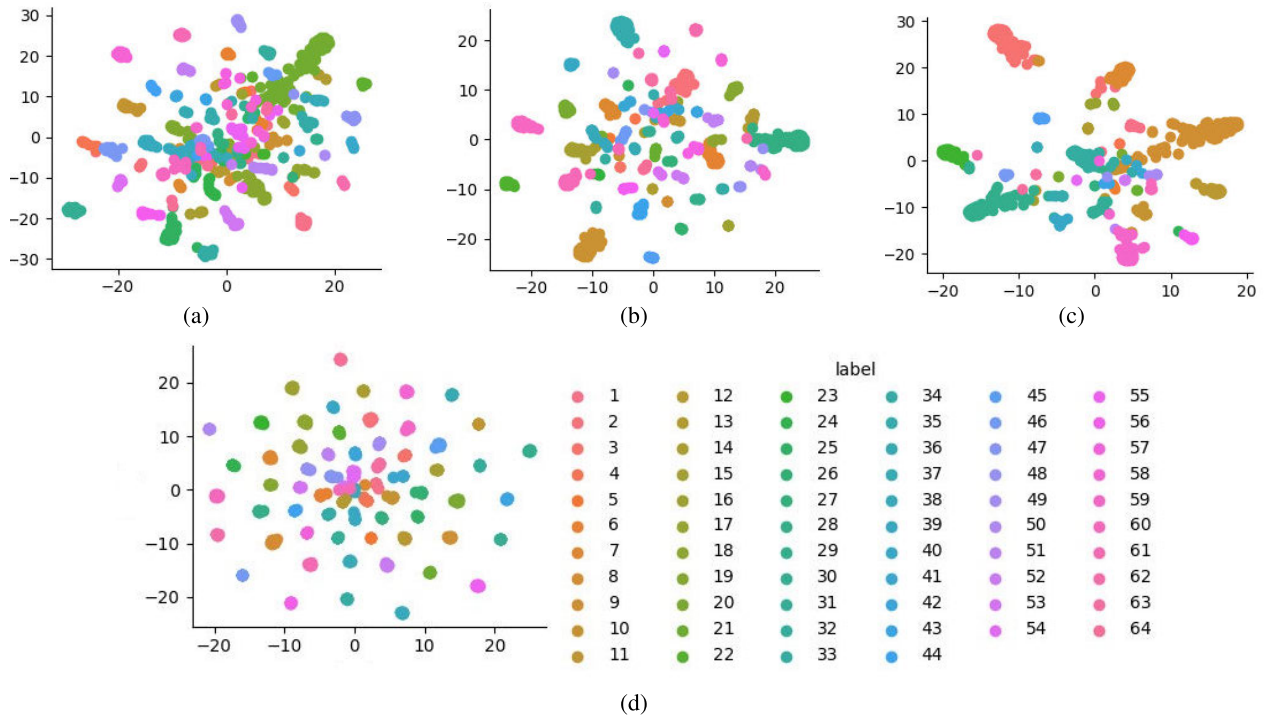
### L. DISCUSSIONS

We delved into instances where the model exhibited misclassifications, and a noteworthy factor contributing to these errors revolves around the resemblances in skin tones across certain images in the dataset. When individuals in the images share similar skin tones, the model encounters difficulty in making accurate distinctions. The model relies on a multitude of features and patterns for predictions, and when these features are akin due to similar skin tones, the model may falter in correctly identifying individuals. The subtleties in skin tones, especially within a diverse dataset, can be intricate. The model might lack the necessary information or distinctive features to make precise differentiations.

Consequently, misclassifications may arise when the model encounters images where the primary distinguishing features hinge on nuanced variations in skin tones. Several images from the test dataset are chosen at random for evaluation. It was also discovered that some of the classes in the Sejong face dataset were misclassified due to similarities in skin tone, as indicated in Table 16. Recognizing and comprehending these challenges provides valuable insights into the potential limitations of the model. Researchers can then explore strategies to enhance the model's capacity to navigate such complexities, be it through the incorporation of more diverse training data or targeted adjustments to specific layers to focus on pertinent features beyond just skin tone.

It's worth highlighting that the dataset was somewhat deficient in representing images with diverse illumination conditions and angles. The absence of varying lighting scenarios and viewing perspectives could have impacted the model's adaptability to real-world situations. A dataset with a broader range of illumination conditions and angles is pivotal for training models to handle the intricacies posed by different lighting setups and perspectives, ultimately enhancing their overall performance and robustness in diverse environments.

To tackle the constraints imposed by the dataset's lack of diversity in illumination and angles, we implemented augmentation strategies. Through techniques like rotation, scaling, and adjustments in brightness, we created variations of the existing images. This augmentation aimed to simulate a more extensive range of lighting scenarios and viewing angles. By exposing the model to this augmented dataset, we sought to improve its adaptability, enabling it to handle



**FIGURE 4.** Visualizing feature embedding distribution for disguised face recognition at various image sizes using t-SNE (a)  $32 \times 32$ , (b)  $128 \times 128$ , (c)  $224 \times 224$ , and (d)  $64 \times 64$ .

**TABLE 14.** Visualizations of disguised face recognition: Original image, GRAD-CAM, guided backpropagation, and guided GRAD-CAM.


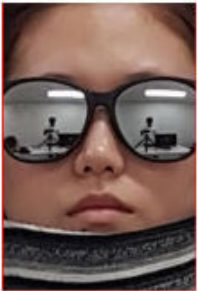


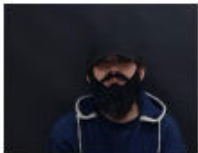






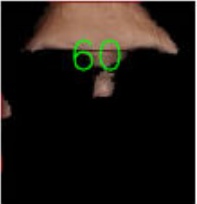




| Original image | GRAD-CAM | Guided backpropagation | Guided GRAD-CAM |
|----------------|----------|------------------------|-----------------|
|                |          |                        |                 |
|                |          |                        |                 |
|                |          |                        |                 |
|                |          |                        |                 |

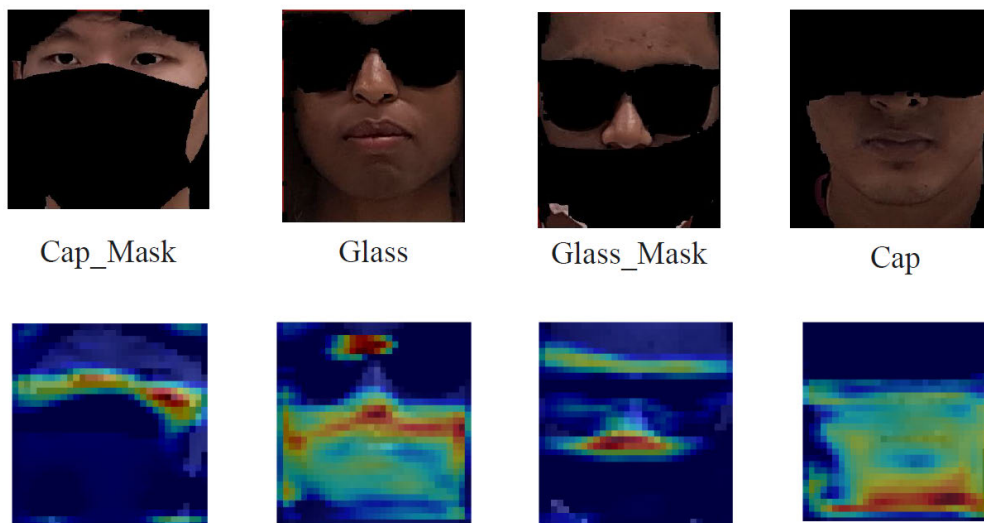
a wider array of real-world situations and minimizing the impact of the initial dataset’s limitations.

Also, the emphasis on expressly saying that no images from subjects utilized in the model’s training phase are included in the validation set is vital to maintaining our

evaluation’s integrity. This explicit distinction is critical to preventing data leaking, in which the model may mistakenly memorize features specific to certain participants during training, potentially recognizing them during validation and overestimating performance. Furthermore, this separation is

**TABLE 15.** Visualizing the disguised face recognition process: original image, face detection, skin segmentation, and prediction.

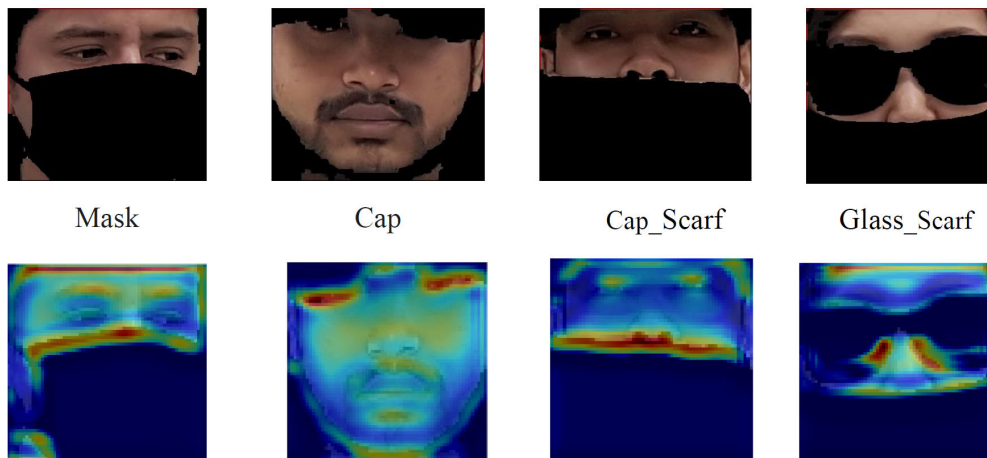
| Accessories                             | Original image  | Face detected Image   | Skin Segmented Image   | Prediction  | Confidence |
|---|---|---|--|---|------------|
| scarf and glasses<br>Actual Class : 38  |    |    |    |    | 0.9701     |
| Fake Beard and cap<br>Actual Class : 63 |    |    |    |    | 0.9963     |
| Glasses and mask<br>Actual Class : 60   |    |   |   |   | 0.8953     |
| Mushtache<br>Actual Class : 53          |  |  |  |  | 0.9998     |



**FIGURE 5.** Heatmap visualization on 32 × 32 disguised test images.

**TABLE 16.** Misclassification analysis of the proposed model: Accessory types, class IDs, and image comparisons.

| Accessories   | Original Face image |  |  | Misclassified Face image |  |  |
|---|---------------------|--|--|--------------------------|--|--|
| Glasses and Scarf<br>Actual Class : 4<br>Predicted Class : 56 |                     |  |  |                          |  |  |
| Glasses and mask<br>Actual Class : 13<br>Predicted Class : 58 |                     |  |  |                          |  |  |
| Mask<br>Actual Class : 22<br>Predicted Class : 5              |                     |  |  |                          |  |  |



**FIGURE 6.** Heatmap visualization on  $64 \times 64$  disguised test images.

critical to reducing the risk of overfitting, in which the model may overly adapt its predictions to the characteristics of the training subjects, limiting its capacity to generalize effectively to new, previously unseen data.

Also, it is observed that regarding the model's performance inconsistency across varying image resolutions. To provide a more comprehensive understanding, we have included heatmaps generated from the model for the low-resolution images, along with accuracy and loss graphs obtained during the training phases. The following figure shows the heatmaps generated for various test images with different accessories used as disguises on  $32 \times 32$  and  $64 \times 64$  image sizes as shown in Figure 5 and Figure 6 respectively. Furthermore,

the heatmaps generated from the model shed light on the features being learned at different resolutions. While we acknowledge that there may be challenges in generalizing patterns from low to high resolutions, our analysis suggests that the model adapts its feature representations accordingly. We observe consistent activation patterns across varying resolutions, indicating that the model learns relevant features for skin region detection.

Also, we have not observed evidence of overfitting during the training process. Our analysis of the accuracy and loss graphs indicates that the model maintains stable performance metrics throughout training, without significant signs of overfitting. The accuracy and loss graphs obtained for the

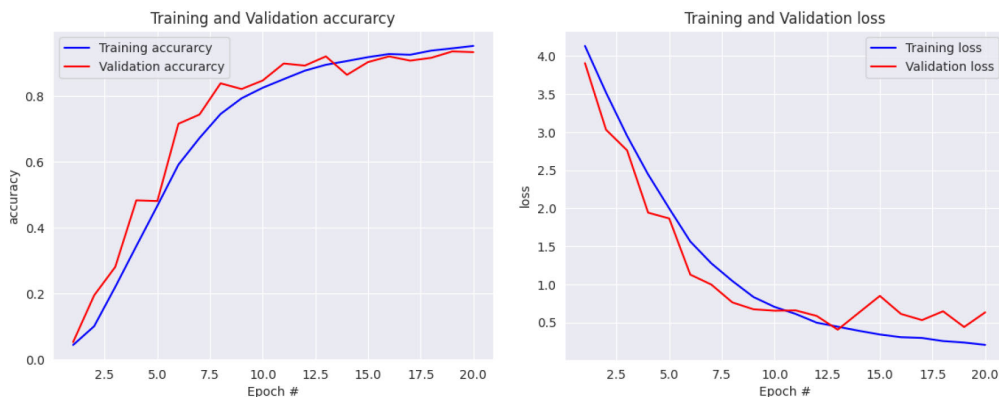


FIGURE 7. Accuracy and loss graphs obtained on 32 × 32 disguised images during the training phase of the proposed model.

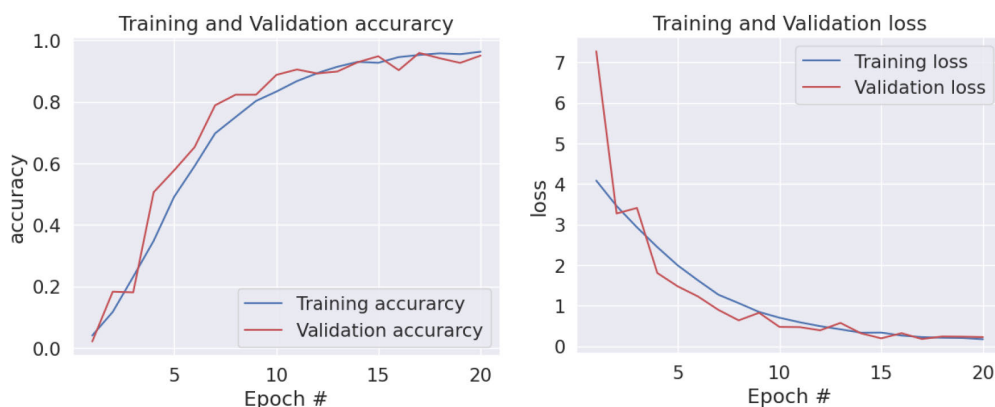


FIGURE 8. Accuracy and loss graphs obtained on 64 × 64 disguised images during the training phase of the proposed model.

proposed model utilizing 32 × 32 and 64 × 64 image sizes are shown in Figure 7 and Figure 8 respectively.

V. CONCLUSION AND FUTURE WORK

The proposed work introduces a CNN model that can effectively identify disguised faces. To address the challenges of variances and distortions present in disguised faces, we developed an end-to-end system that includes both segmentation and recognition modules. The experiments conducted using the Sejong Face dataset demonstrated that feature extraction blocks that incorporate skip connections yielded superior performance. The experimental results indicate that the method we propose outperforms state-of-the-art methods by a significant margin. Researchers have developed various algorithms and methods to improve the accuracy and robustness of disguised face recognition systems. However, there are still limitations and challenges that need to be addressed, such as the ability to recognize faces in real-time and under different lighting conditions. Improving scalability may become a focus in the future. This architecture is adaptable, allowing for seamless scalability by adding branches or expanding the layers inside each branch. This possibility for growth opens the prospect of

increased performance on larger datasets or more difficult tasks. Despite these challenges, the potential applications of disguised face recognition is significant, such as in security, surveillance, and law enforcement, and therefore it is an active area of research with much room for future advancements.

REFERENCES

- [1] T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar, "Recognizing disguised faces: Human and machine evaluation," *PLoS ONE*, vol. 9, no. 7, Jul. 2014, Art. no. e99212.
- [2] M. Singh, M. Chawla, R. Singh, M. Vatsa, and R. Chellappa, "Disguised faces in the wild 2019," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 542–550.
- [3] M. Singh, R. Singh, M. Vatsa, N. K. Ratha, and R. Chellappa, "Recognizing disguised faces in the wild," *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 1, no. 2, pp. 97–108, Apr. 2019.
- [4] G. Goswami, M. Vatsa, and R. Singh, "Face verification via learned representation on feature-rich video frames," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1686–1698, Jul. 2017.
- [5] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Context-aware local binary feature learning for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1139–1153, May 2018.
- [6] U. Cheema and S. Moon, "Sejong face database: A multi-modal disguise face database," *Comput. Vis. Image Understand.*, vols. 208–209, Jul. 2021, Art. no. 103218.



- [7] K. Zhang, Y.-L. Chang, and W. Hsu, "Deep disguised faces recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 32–324.
- [8] A. Bansal, R. Ranjan, C. D. Castillo, and R. Chellappa, "Deep features for recognizing disguised faces in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 10–106.
- [9] N. Kohli, D. Yadav, and A. Noore, "Face verification with disguise variations via deep disguise recognizer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 17–177.
- [10] S. V. Peri and A. Dhall, "DisguiseNet: A contrastive approach for disguised face verification in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 25–256.
- [11] A. Suri, M. Vatsa, and R. Singh, "A2-LINK: Recognizing disguised faces via active learning and adversarial noise based inter-domain knowledge," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 4, pp. 326–336, Oct. 2020.
- [12] M. Singh, S. Nagpal, R. Singh, and M. Vatsa, "Disguise resilient face verification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 6, pp. 3895–3905, Jun. 2022.
- [13] M. J. Khan, M. J. Khan, A. M. Siddiqui, and K. Khurshid, "An automated and efficient convolutional architecture for disguise-invariant face recognition using noise-based data augmentation and deep transfer learning," *Vis. Comput.*, vol. 38, no. 2, pp. 509–523, Feb. 2022.
- [14] M. Dosi et al., "Seg-DGDNet: Segmentation based disguise guided dropout network for low resolution face recognition," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 6, pp. 1264–1276, Nov. 2023, doi: [10.1109/JSTSP.2023.3288398](https://doi.org/10.1109/JSTSP.2023.3288398).
- [15] A. Sharma, N. Jindal, A. Thakur, P. S. Rana, B. Garg, and R. Mehta, "Multimodal biometric for person identification using deep learning approach," *Wireless Pers. Commun.*, vol. 125, no. 1, pp. 399–419, Jul. 2022.
- [16] S. Manchanda, K. Bhagwatkar, K. Balutia, S. Agarwal, J. Chaudhary, M. Dosi, C. Chiranjeev, M. Vatsa, and R. Singh, "D-LORD: DYSL-AI database for low-resolution disguised face recognition," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 6, no. 2, pp. 147–157, Apr. 2024.
- [17] B. Huang et al., "Joint segmentation and identification feature learning for occlusion face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 10875–10888, Dec. 2023, doi: [10.1109/TNNLS.2022.3171604](https://doi.org/10.1109/TNNLS.2022.3171604).
- [18] U. Cheema and S. Moon, "Disguised heterogeneous face recognition using deep neighborhood difference relational network," *Neurocomputing*, vol. 519, pp. 44–56, Jan. 2023.
- [19] S. Kumar, S. K. Singh, and P. Peer, "Occluded thermal face recognition using BoCNN and radial derivative Gaussian feature descriptor," *Image Vis. Comput.*, vol. 132, Apr. 2023, Art. no. 104646.
- [20] J. Mehta, S. Talati, S. Upadhyay, S. Valiveti, and G. Raval, "Regenerating vital facial keypoints for impostor identification from disguised images using CNN," *Expert Syst. Appl.*, vol. 219, Jun. 2023, Art. no. 119669.
- [21] F. Saxen and A. Al-Hamadi, "Color-based skin segmentation: An evaluation of the state of the art," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4467–4471.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [23] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
- [24] J. Tobias Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014, *arXiv:1412.6806*.
- [25] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?" 2016, *arXiv:1611.07450*.



**G. PADMASHREE** received the bachelor's and master's degrees from Visvesvaraya Technological University, Karnataka, in 2003 and 2012, respectively. She is currently pursuing the Ph.D. degree with the Department of Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. Her research interests include image processing, artificial intelligence, and machine learning.



**KARUNAKAR A. KOTEGAR** (Senior Member, IEEE) received the B.Sc. and M.C.A. degrees from Karnataka University, Karnataka, India, in 1995 and 1998, respectively, and the Ph.D. degree from Manipal Academy of Higher Education (MAHE), Manipal, Karnataka, in 2009. He is currently the Director of the International Collaborations of MAHE and a Professor with the Department of Data Science and Computer Applications, Manipal Institute of Technology, MAHE.

His research interests include image/video processing and communication, scalable video coding, media aware network elements, multi-view video coding, scalable video over peer-to-peer networks, error resilient and concealment for scalable video, stereo vision, and image and video forensics.

• • •