Advances in Artificial Intelligence and Machine Learning; Research 5 (2) 3768-3783 Received 01-03-2025; Accepted 21-05-2025; Published 28-05-2025

Harnessing Deep Learning Techniques for the Removal of Occlusion in Face Image

Jyothsna Cherapanamjeri jyothsnamtech@gmail.com
Research Scholar, Department of Computer Science and Engineering,
Jawaharlal Nehru Technological University,
Anantapur-515002, Andhra Pradesh, India

Narendra Kumar Rao Bangole narendrakumarraob@gmail.com
Professor and Program Head, Department of Artificial Intelligence and Machine Learning,
Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College),
Tirupati-517102, Andhra Pradesh,
India

Corresponding Author: Jyothsna Cherapanamjeri

Copyright © 2025 Jyothsna Cherapanamjeri and Narendra Kumar Rao Bangole. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Occlusion removal is one of the challenging tasks in face recognition system. Facial occlusion is nothing but face object is hiding the other object which is one important factor that influences how well facial recognition works. There is insufficient research regarding the issue of occlusion. Researchers proposed several methods for removal of occlusion to recognize faces. But are unable to produce results that are realistic for the removal of significant objects, particularly in images of faces. The performance test of face recognizer recognition system degrades when dealing with masked faces. The objective of this research work to remove mask object from face images to recognizes faces easily. We divide the problem into two modules: occlusion detection module and occlusion removal module. We proposed deep learning-based technique called GAN-based network. Two discriminators are used in a GAN-based network; one discriminator helps in learning the face's global structure, while the other learns the deep missing region. MaskedFace-CelebA dataset is used to evaluate our model that is utilised to compare masked images in pairs with ground truth in order to verify faces. The evaluation metric used in this research is SSIM. Using the proposed method can be useful task to generate new unmasked face samples and help criminal investigation agencies1to reveal the identity of criminals who committed the crimes behind the mask. Once the occlusion of face is removed, the system can easily recognize the face and the biometric authentication during COVID-19 pandemic no need to remove masks while taking biometric attendance, this will reduce the spread of COVID-19 disease.

Keywords: Computer vision, Deep learning, Face recognition, Facial occlusion.

Citation: Jyothsna Cherapanamjeri and Narendra Kumar Rao Bangole. Harnessing Deep Learning Techniques for the Removal of Occlusion in Face Image. Advances in Artificial Intelligence and Machine Learning. 2025;5(2):214.

1. INTRODUCTION

Face recognition systems have been used extensively to identify people by their facial features, and their effectiveness has been shown to be very high. However, these methods have faced a major challenge due to the COVID-19 epidemic and the widespread use of face masks. Since the unmasking1of masked faces is an interesting concept with many real-world uses, we concentrate on it. We initially detect the mask1region from an input image of a masked face. A GAN-based network is subsequently provided with the masked facial image as input along with a binary map indicating the location of the identified mask region. The network then generates an image free of the non-face item, in this case the mask object. There are two main reasons why individuals wear masks. To begin with, individuals take this action to shield themselves from contamination. Additionally, certain individuals feel insecure about how they look and choose to conceal their faces and feelings from others. The main goal of this challenge is to successfully remove the masked area from the human face and fill the resulting1hole with actual content. We synthesize images with an occluding item, such a mask, to provide face results without distraction. The mouth, nose, and eyes—also known as non-occluded faces—are the primary facial characteristics that are typically displayed to face recognition (FR) systems. People must wear masks in such situations include COVID-19 pandemic, laboratories, medical operations, escape from excessive pollution and finally escape from crimes. Normal face recognition system is very easy to recognize faces in real time. Unfortunately, occluded faces complicate to recognize faces accurately. When faces are masked, the majority of face recognition systems continue to perform poorly, even after the pandemic. Many applications for human-computer interaction—like face authentication-based mobile payments and face access control—rely on face recognition technology. Most of the face recognition system failed to recognize the masked faces. In order to tackle this challenge, early non-learning-based efforts remove undesired objects and use similar patches from the rest of the image to synthesize the missing content. From a library of millions of scene images [1], they identify recurring patterns and apply them to the damaged area. Some of the works use recursive error compensation and principal component analysis (PCA) reconstruction to remove eyeglasses from facial images. These non-learning algorithms, however, can only remove small objects from images.

New developments in learning-based techniques enable algorithms for image manipulation to learn from massive datasets. When it comes to removing objects with less structural diversity, learning-based image editing techniques perform well. However, because of the size of the objects—such as the mask object—these methods are not suitable for removing a masked face. We suggest a GAN-based network1that automatically eliminates the mask and completes the face's missing area in order to address this issue. We divide the problem into two modules which are occlusion detection module and occlusion removal module [2]. In the first module, we use an encoder and decoder network to recognize the mask object that is occluded and create the binary segmentation map of the masked object. We use a method for learning deep missing regions and global coherency in the second stage. First, we use one generator and discriminator to train our model. By considering the entire image, this discriminator helps to preserve global coherence. Even though this arrangement creates the structure of the face, the mask-covered chin and cheeks interact with the rest of the face in a way that makes it difficult to synthesize the deep area of the missing hole. The term "deep region of missing hole" refers to the area of the face that is distant from the boundary of occlusion formed by the masked object, such as the mouth area, particularly the lips and teeth. We enhance the model with a second discriminator that solely examines the missing region in order to concentrate more on producing the deep missing semantics. In order to complete the affected region1with fine details while preserving the overall structure1of the facial image, this technique requires the two discriminators to give the generator fair feedback. Our key contributions are summarized below:

1. This paper proposes a deep learning-based technique called GAN-based network. This network mainly consists of generator network, discriminator network and finally perceptual network which are helps to remove occlusion in face images.
2. Offering a thorough analysis of occlusion removal models' performance.
3. Performing experiments on synthetic dataset which is MaskedFace-CelebA. This dataset is created by CelebA dataset. This dataset is publicly available.
4. Performing a comparative analysis of two state-of-the-art segmentation models, namely Gated Convolution, GUMF.

2. RELATED WORK

In this section provides an in-depth review of important research on object removal and objects detection in images, image inpainting methods with their focus on face unmasking and their various applications in other fields, such as face recognition1and face reconstruction. We classify the methods into classical approaches, such as patch-based approach and diffusion-based approach, interpolation, and others that were commonly employed in the past. We discuss the developments in deep learning methods for image inpainting1and object removal.

In [3], the goal is to accomplish three types of facial recognition: occluded, de-occluded, and rebuilt. In this work present two distinct approaches based on CycleGANs and Laplacian pyramid mixing for face reconstruction. It uses two distinct feature extraction methods to validate of work: learnt features that take advantage of the last layers of a pre-trained1deep architecture model and hand-crafted features. Non-learning-based object removal techniques detect a similar structure in the input image or external data to fill in the blank region and remove undesired objects from an image. In order to find information that is most comparable to the input sample [4] search through hundreds of scene photos. After that, they duplicate and insert that data into the absent pixels in the input sample. But for images with different textures and intricate semantic frameworks, like a human face, they generate inconsistent content. By altering the patch-based function in the filling order calculation using a regularized component [5, 6] eliminate eyeglasses from facial images. Their work is only effective when it comes to taking out little objects like eyeglasses and fails to produce believable contents for facial image removal of huge objects. A statistical strategy based on the Kriging interpolation method was presented by [7, 8] to fill up damaged areas while maintaining spatial correlation. To take advantage of both texture synthesis and inpainting approaches [9–11] merged them. Their algorithm replicates both texture and structure by utilizing exemplar-based texture creation. They determined the values of the real color using the exemplar-based synthesis method. Wenqiu Zhu proposed Edge conditions and several recognizers are combined in the Internet of Things device to create generative adversarial networks [11]. Information restructuring is used to complete face recovery, which is predicated on satisfying the needs of users who are partially hiding. Nizam ud din suggested a GAN-based architecture that incorporates two discriminators [12]; one discriminator helps learn the overall structure of the face, while the other focuses on enhancing the learning in the deeper areas that are missing. A contextual feature-based restricted DCGAN with paired discriminator1method was introduced by for face completion in face recognition system [13–16]. They extracted features using a pre-trained VGG16 network and stabilized training by introducing a paired feature matching loss. Promising findings with better texture and semantic consistency were observed in the experimental data. Sinivasan proposed Viola Jones [17] for occlusion detection and face restoration. In [18], the suggested framework works better than earlier light field de occlusion (LF-DeOcc) techniques in both sparse and dense1LF images, reconstructing images with no occlusion, according to a number of experimental findings. In [19], a fully convolutional neural network (FCN) and K-means clustering algorithms were introduced for the purpose of occlusion removal utilizing segmentation and inpainting techniques. In [20], an innovative attention-based encoder-decoder framework was introduced for de-occlusion in head-mounted displays (HMD). In [21], a combined approach for eliminating occlusions is presented, utilizing Finite Element Based Bi-dimensional1Empirical Mode Decomposition1(FE-BEMD) alongside exemplar-based image1inpainting. In [22], There are three steps in this procedure. First, a curve evolution method is used to estimate the region of full occlusion. Second, each pixel in the partially obscured area has its alpha value assessed. Third, in areas of partial occlusion, the foreground occlusion intensity contribution is eliminated. In [23], a strong long short-term memory1(LSTM) Autoencoder approach was introduced to tackle the challenge of face de-occlusion in natural settings. In [24], this study presents a facial image de-occlusion technique guided by segmentation and 3D reconstruction that effectively eliminates various forms of occlusions. In [25], a method was proposed that utilizes a spatial attention module within a conditional generative1 adversarial network1to produce realistic images of faces with masks removed from the facial area. In [26], a conditional generative adversarial network called Pix2pix is employed for image restoration. This approach facilitates the transformation of one image into another by turning an obscured image into a clear one. In [27, 28], outline a method for eliminating occlusion in 3D multi-perspective imaging that employs the pixel depth mapping approach to enhance the reconstruction of images of obscured 3D objects. By reducing the statistical variation of the projection image1 pixels of 3D object points on various perspective images, depth mapping1is accomplished. Zhirong Gao proposed two stages [29], it consists of two stages: the first is intended to locate the test image's occlusion area, and the second is to categorize according to the test image with occlusion removed, in which we presume the test face has been identified and associated with the train samples. G. Rajeswari proposed a two-stage network for facial occlusion removal [30]. In first stage, to give the information required to recover from the lost parts of the face picture, a Face Similar Matrix1 (FSM) is created under the direction of the Structural Similarity Index Measure. Leveraging the connection between the occluded section and corresponding face images—which hold valuable information for reconstructing the occluded part—these corresponding face (CF) images are considered as relevant data in the subsequent phase and serve as input for creating eigenspaces through principal component analysis (PCA). Miao Zhang proposed the occlusions are extracted using a contour-based object extraction technique [31], that is resilient to posture fluctuation of the 3D occlusion objects. Tanvi B. Patel proposed the method The Viola-Jones [32], technique is utilized for face detection, while neural networks (NN) are employed for face recognition and rapid weighted PCA is used for occlusion detection and face reconstruction. Jiayuan Dong proposed the two-stage Occlusion-Aware1GAN (OA-GAN) [33], for de-occluded faces This will be used as the second GAN's extra input to synthesize the last set of de-occluded faces.

Summary of the Literature Survey:

3. METHODS

The proposed architecture of our research work is shown in FIGURE 1. Proposed system composed of two main modules such as occlusion detection module and occlusion removal module. The following provides an explanation of each module.

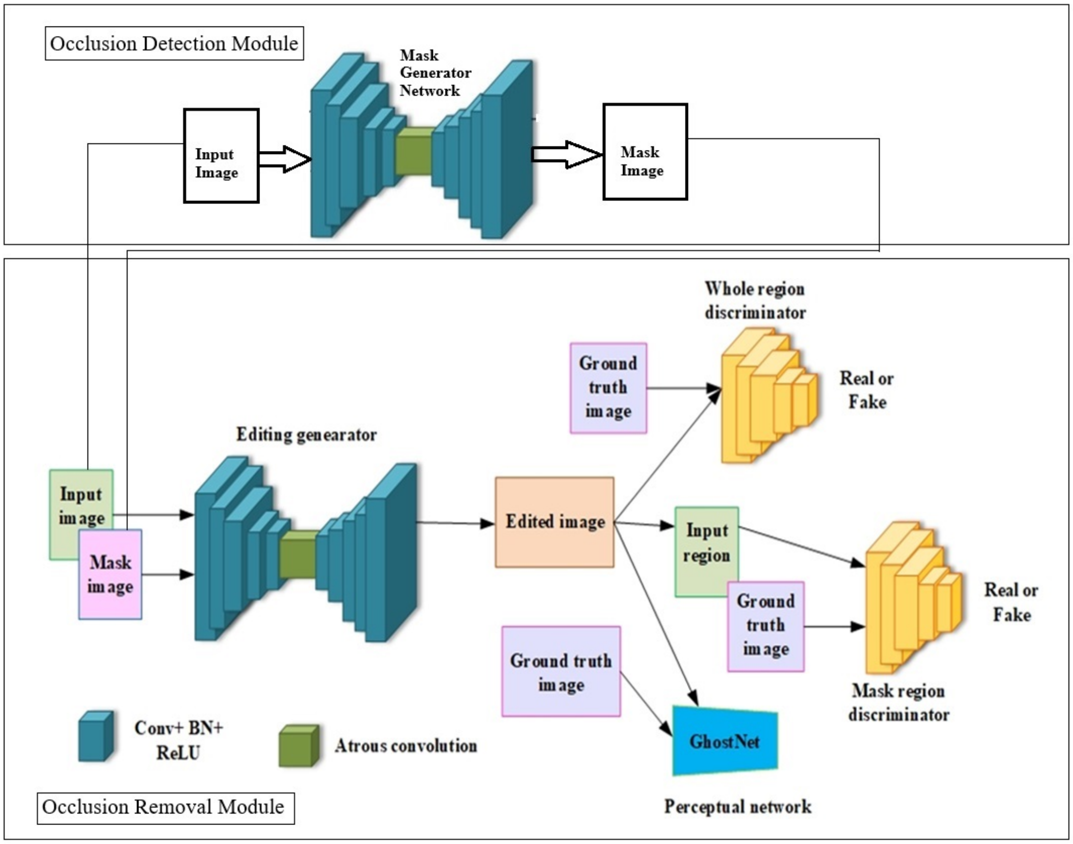


Figure 1: Proposed System Architecture

3.1 Occlusion Detection Module.

The purpose of occlusion detection1module is to detect an occluded area in face images. Here occluded area is considered as a masked area. The occlusion detection module produces a binary segmentation mask, where 1 indicates the face mask object and 0 corresponds to the other pixels in the image that we are analyzing. The first module called mask generator network which consists of U-Net architecture. U-Net is a fully convolution1neural network design that features an encoder-decoder framework with skip connections. The encoder path gradually down samples the input image, while the decoder path up samples the feature maps to produce the final segmentation map.

3.2 Occlusion Removal Module.

The aim of this module is to remove1 the mask and complete the remaining area in a way that corresponds both structurally and visually with the reference image. The primary components of this module include 1) Editing Generator 2) Discriminator 3) Perceptual Network

Step-1: Editing Generator

The architecture for the editing generator resembles that of the mask generator network; however, the distinction lies in the incorporation of a squeeze and excitation1 (SE) block1 at the output of the initial three blocks in the encoder. The purpose of squeeze and excitation block is that the Squeeze: Aggregates spatial information in each channel to produce a compact channel descriptor. Excitation: Learns to assign weights to channels, amplifying important1features and reducing the impact of irrelevant ones. Improved quality of the synthesized mask (e.g., realistic texture and blending), Better handling of complex facial variations (e.g., different skin tones or lighting). Moreover, we use an atrous convolution with four layers between the encoder and decoder paths which helps make the missing part. The generator receives the input image along with the output from the occlusion detection module and the mask image, and it creates a generated image, also referred to as an edited image.

$$\text{Edited Image} = \text{Editing Generator (Input Image, Mask Image)} \tag{1}$$

In order to compel the editing generator to create plausible missing content, we utilize reconstruction loss, which is a combination of L1 loss and structured similarity, known as SSIM loss.

$$\text{Reconstruction Loss (RC)} = \text{L1} + \text{SSIM} \tag{2}$$

L1 loss represents the variation in pixel values between the modified image and the ground truth image.

$$\text{L1} = \text{Edited Image} - \text{Ground Truth Image} \tag{3}$$

SSIM evaluates the structural likeness between the edited image and the ground truth1 image, and its associated loss function is expressed as:

$$\text{SSIM} = 1 - \text{SSIM (Edited Image, Ground Truth)} \tag{4}$$

Step-2: Discriminator

We employ two discriminators which are mask region discriminator and whole region discriminator as shown in FIGURE 1. The pix2pix discriminator architecture is used by both discriminators [34]. Unlike a standard GAN discriminator that outputs a single scalar to classify an entire image as real or fake, the discriminator1takes an input image (real or generated) and divides it into overlapping patches (e.g., 70x70 pixels). Each patch is evaluated independently, and the discriminator outputs a feature map where each value corresponds to the "realness" of a specific patch. The final loss is typically the average of all patch predictions. The size of the patches depends on the network's receptive field, controlled by the number of layers, kernel sizes, and strides. Smaller patches (e.g., 1x1, as in Pixel GAN) focus on pixel-level details, while larger patches (e.g., 70x70) capture more contextual information. Common patch sizes in practice: 1x1, 16x16, 70x70, or full image (effectively a standard discriminator). Both discriminators play a crucial role in compelling the editing generator to produce images that are not only aesthetically pleasing but also semantically coherent. Rather than training both discriminators simultaneously with the editing generator, we train the editing generator1together with the entire region discriminator for the first 40% of the total training iterations. This facilitates the enforcement of the output generated to align structurally with the initial input face image by reducing the following objective function:

$$\begin{aligned} \text{D (Whole Region Discriminator)} = &-E_{\text{ground truth}} \in \text{O log Whole Region Discriminator} \\ &\quad \text{(Edited Image, Ground Truth Image)} \\ &+ E_{\text{edited image}} \in \log(1 - \text{Whole Region Discriminator} \\ &\quad \text{(Editing Generator (Input Image, Mask Image)))} \end{aligned} \tag{5}$$

In this context, O represents the set of real images, while S signifies the set of synthesized images. We prioritize optimizing the mask area to generate meaningful semantics solely in the missing region. We incorporate a masked region in addition to the whole region for the editing generator. We continue to train them together for the remainder of the training iterations. In order to train the mask region, the subsequent objective function is optimized:

$$\begin{aligned} \text{D (Mask Region Discriminator)} = &-E_{\text{ground truth}} \in \text{O log Mask Region Discriminator} \\ &\quad \text{(Input Mask Region, Ground Truth Image)} \\ &+ E_{\text{edited image}} \in \text{S log}(1 - \text{Mask Region Discriminator} \\ &\quad \text{(Editing Generator (Input Image, Input Mask Region)))} \end{aligned} \tag{6}$$

Here
Input Mask Region = Input Image $\bigotimes$ (1 − Mask Image) + (Edited Image $\bigotimes$ Mask Image) and $\bigotimes$ represents multiplication performed on each corresponding element.

To train our model using a GAN framework, the generator deceives the discriminators by reducing the following loss1functions:

$$\begin{aligned} \text{Adversarial (Whole Region)} = &-E_{\text{edited image}} \in \text{S log (Whole Region Discriminator} \\ &\quad \text{(Editing Generator (Input Image, Mask Image)))} \end{aligned} \tag{7}$$

$$\begin{aligned} \text{Adversarial (Mask Region)} = &-E_{\text{edited image}} \in \text{S log (Mask Region Discriminator} \\ &\quad \text{(Editing Generator (Input Image, Input Mask Region)))} \end{aligned} \tag{8}$$

Step-3: Perceptual Network

Final building block is perceptual1 network. It is a pre-trained Ghost Net network. Existing works using VGG16 network. VGG16 classifier is utilized to differentiate between reconstructed faces and original images. In our study employing the Ghost Net module [35]. Convolutional neural networks designed with Ghost modules, which aim to produce more features with fewer parameters, are known as Ghost Nets. This network seeks to deliver modified images where the generator's results align more closely with the ground truth regarding feature representation. We employ a perceptual loss to penalize outputs that are not perceptually plausible by establishing a distance measure at the feature level between the intermediate feature maps of the edited image and the ground truth, utilizing a pretrained network. Let ϕi represent the activation map from the i-th layer of ϕ; the perceptual loss can be described as:

$$\text{Perceptual Loss (PC)} = \Sigma i \| \phi_i (\text{Edited Image} - \phi_i (\text{Ground Truth Image}) \| \tag{9}$$

The complete loss function used to train the editing1 module is characterized as:

$$\begin{aligned}\text{Total Loss} = {} & \lambda_{\text{RC}}(\text{RC} + \text{PC}) + \lambda_{\text{Whole Region Discriminator}}\text{D (Whole Region Discriminator)} \\ & + \lambda_{\text{Mask Region Discriminator}}\text{D (Mask Region Discriminator)} \\ & + \lambda_{\text{Adversarial Whole Region}}\text{Adversarial (Whole Region)} \\ & + \lambda_{\text{Adversarial Mask Region}}\text{Adversarial (Mask Region)} \qquad (10)\end{aligned}$$

We have established the weight parameters as $\lambda_{\text{RC}} = 100$, $\lambda_{\text{Whole Region Discriminator}} = 0.4$, $\lambda_{\text{Mask Region Discriminator}} = 0.6$, $\lambda_{\text{Adversarial Whole Region}} = 0.3$, $\lambda_{\text{Adversarial Mask Region}} = 0.7$. Total loss contributes to producing output that appears natural, maintains structural consistency, and is perceptually believable.

4. RESULTS AND DISCUSSION

The experimental setup and outcomes of our research are presented in this section. The training process is subsequently explained along with the details and configurations that were used. Our network performance is evaluated by using evaluation metrics that we have defined. Next, we compare the outcomes of our method with the most advanced deep learning-based techniques, and we show and discuss the findings.

4.1 Dataset

The Synthetic masked dataset repository contains the MaskedFace-CelebA synthetic masked dataset that is examined for this study. The MaskedFace-CelebA dataset1is available to the public. This dataset is constructed from CelebA dataset using the MaskTheFace tool. By using this tool to overcome the scarcity of masked face dataset. The collection of masked images also contained the original unmasked image. We have used two different kinds of masks such as a cloths and surgical-blue masks which accurately depict a range of masking situations. This was done in order to guarantee that the trained network works just as effectively with both masked and unmasked images. This dataset contains 21,844 masked face images unmasked face images and corresponding target or ground truth images on occluded area. Images of size are 256x256 pixels. These images are randomly divided into 17476 (80%) images for training, 2184 (10%) images for validation and 2184 (10%) images for testing from the MaskedFace-CelebA dataset.

4.2 Training Procedure Details

The specifics of our model's training are described as follows: The training process was split into two distinct phases. For the initial training phase1focused on the map module, we provided the network with a masked face image as input to produce a binary segmentation map. In training of second phase which is occlusion removal module the generated binary map in first module and masked face image as input to the editing network and generate the final result is depicted in FIGURE 2.

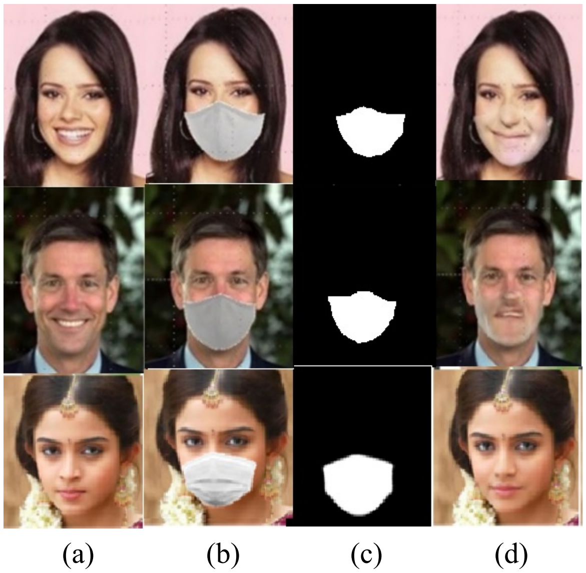


Figure 2: Results generated by our model: (a) Ground Truth Image (b) Masked face image (c) Binary segmentation map (d) Output of our model

For training the GAN-based network:

- **Input Pair**:
  - Masked image (face with synthetic or real mask).
  - Binary mask (indicating the mask region).
- **Ground Truth**: The original unmasked face image (used to compute reconstruction and perceptual losses).
- **Output**: The generator learns to produce images close to the ground truth. A complete image with the missing regions filled in a way that matches the surrounding context. That is face image without the mask.

We trained the first stage with 20 epochs and batch_size is 16. We used batch_size 64 and 100 epochs to train the second stage. In the second1step, we train the generator and discriminator with the entire region for nearly all training iterations in order to produce a suitable global structure of the face, rather than training the entire network at once. This helps in determining the facial region's accurate boundaries. In order to construct the deep missing semantics for the remaining training phases, we add a mask region after the face's reasonable global structure has been determined. The implementation has done in python. The training was performed with a batch_ size of 64 for image resolution1 of 256 and batch size of 8 for an image resolution2 of 512. Starting at 0.001, we used a variable leaning rate strategy.

4.3 Evaluation Metrics

4.3.1 Structural Similarity (SSIM)

The Structural Similarity Index1(SSIM) is a measure designed to assess the likeness between two images, emphasizing perceived quality over differences at the pixel level. Unlike metrics like Mean Squared Error (MSE), SSIM considers structural information, luminance, and contrast, aligning better with human visual perception. The formula for SSIM in equation (11).

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{11}$$

4.3.2 Peak Signal to Noise Ratio (PSNR)

The Peak Signal-to-Noise Ratio measures1the quality of a reconstructed or compressed image (or signal) in relation to the original, usually represented in decibels (dB). It quantifies the relationship between a signal's peak power and the power of the noise (errors) caused by distortion. Formula for PSNR as shown in equation (12).

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right) \tag{12}$$

where MAX_I The highest signal value present in the original image is referred to as the maximum signal value, and the mean squared error (MSE) is determined as follows:

$$\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[x(i,j) - y(i,j)]^2$$

Here, m and n are the image dimensions, and x(i,j), y(i,j) are pixel values at position (i,j).

4.3.3 Frechet Inception Distance (FID)

FID measures the similarity between the feature1distributions of generated images and actual images. Lower FID scores indicate that the produced visuals are more similar to the real ones, implying better generative model performance. FID is calculated as in equation (13).

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}) \tag{13}$$

4.4 Comparative Study

We use the following quantitative metrics to compare the outcomes of our method with other methods: 1) Structural Similarity (SSIM) 2) Peak Signal to Noise Ratio (PSNR) 3) Frechet Inception Distance (FID). The quantitative comparison is displayed in TABLE 1. The comparison is performed with GLCIC, GC, EC, GUMF. The best result is boldfaced. The results demonstrate that the best quantitative values are obtained by our model.

Table 1: Quantitative comparisons of our technique with structural similarity (SSIM, PSNR, FID)

Method	SSIM	PSNR	FID
Globally and Locally Consistent Image Completion (GLCIC) [36]	0.827	22.40dB	3.651
Gated Convolution (GC) [37]	0.829	18.65dB	4.012
Edge Connect (EC) [38]	0.864	20.86dB	3.555
GAN-based Unmasking the Masked Face (GUMF) [12]	0.864	26.19dB	3.548
Proposed	**0.875**	**27.13dB**	**3.435**

4.5 Ablation Study

4.5.1 Excluding occlusion detection module

We have eliminated the occlusion detection module and solely utilized the editing generator network to assess the impact of object1removal and image editing. This is solely provided with input that includes the mask image fed into the editing1network, while the remainder of the model remains unchanged. FIGURE 3 illustrates that omitting the mask generating network in the image editing system results in an irregular structure of the facial image.

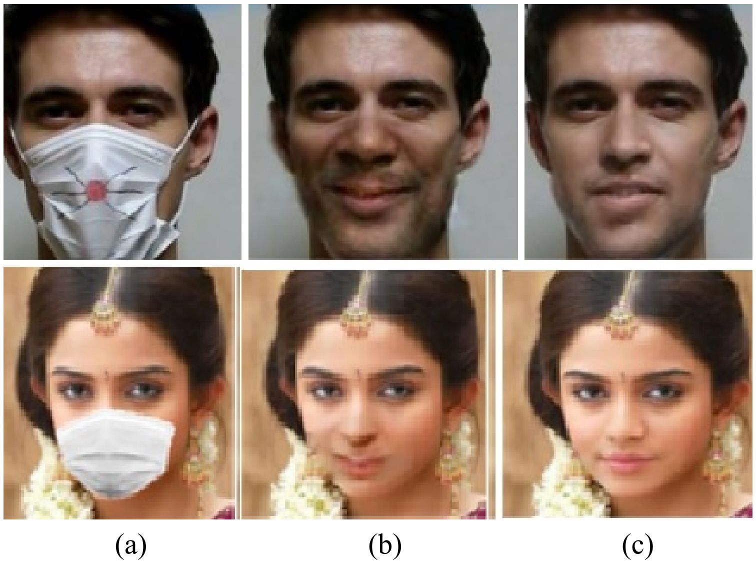
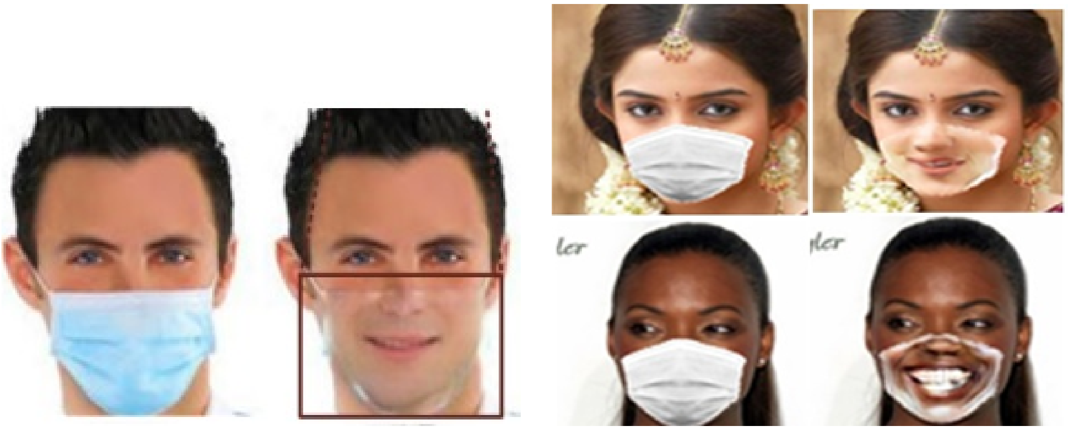


Figure 3: Excluding occlusion detection module. a) Input Image b) Results of our model without using occlusion detection module c) Results of our model using both occlusion detection module and occlusion removal module.

4.5.2 Purpose of using two discriminators

Outcome of our model utilizing a single discriminator as shown in the FIGURE 4. FIGURE 4(a) demonstrates the outcome of implementing a mask region discriminator while keeping all other settings identical to our model. It combines the chin alongside the neck by treating the neck region obscured by the mask object as a component of the chin. This configuration produces the least favorable outcomes when the mask color closely resembles the neck color. To address the issue, we opted to incorporate the whole region rather than just the mask region discriminator, keeping all other aspects consistent with our original model. We eliminate the mask region and utilize only the whole region together with the editing1generator. In FIGURE 4(b), it's evident that it creates a more uniform facial structure and keeps the chin separate from the neck; however, this setting struggle to generate realistic details in the deeper areas of the absent section, leading to teeth that appear neither symmetric nor natural.



(a) (b)

Figure 4: Results of using one discriminator. a) Result of using mask region discriminator b) Result of using mask region discriminator

To produce believable content in the areas that are incomplete and ensure it aligns along with the remaining part of the face, we employed both discriminators alongside the editing generator1by gradually incorporating each discriminator into the network as previously mentioned in the training section of the experiment.

4.6 Limitations

Our model encountered challenges in effectively retrieving the original content from images that were heavily occluded area or complex masking. The occlusion detection module does not accurately identify the mask map. This occurs when mask1objects significantly differ from those in our synthetic dataset concerning shape, size, color, and structure, as illustrated in FIGURE 5. The shape, color, and structure1of the mask objects are entirely distinct from the mask categories included in our synthetic1dataset for masked face, resulting in inadequate detection. The network cannot identify the entire masked area because of the intricate blend of colors. The network was unable to identify the area where its color resembled the texture of the face, as it was deemed part of the facial features. These are our challenges1and providing important perspectives for future research studies and improvements.

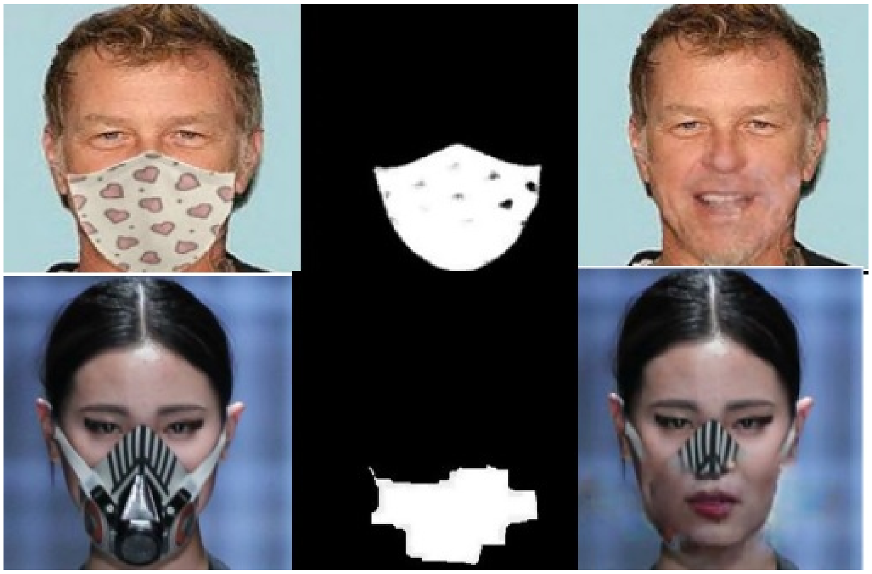


Figure 5: Failure results of complex masks, shapes and colors

5. CONCLUSION

In this study, we describe a novel method for removing occlusion from facial images; in this case, the mask region of the face is taken into consideration. For image completion purpose, we have used GAN based Ghost Net approach to produce plausible results. By utilizing modern deep learning techniques and integrating feature extraction modules, the GAN model successfully recovers required details and reliably reconstructs the basic information of masked images. Actually, our approach divided into two modules. First module is occlusion detection module. This module is trained by using U-Net module. This module produces binary segmentation map on occluded region. In second module we are using GAN with Ghost Net producing realistic and structurally consistent outputs. When compared to existing approaches image editing techniques show that our model can provide outcomes with good quality for huge missing holes in face images. This research work is helpful for some real-time applications such as biometric authentication during COVID-19 and help criminal investigation1agencies to reveal the identity of criminals who committed the crimes behind the mask.

References

[1] Herley C. Automatic Occlusion Removal From Minimum Number of Images. IEEE International Conference on Image Processing. IEEE. 2005:II-1046.

[2] Yang X, Gousseau Y, Maitre H, Tendero Y. A Fast Algorithm for Occlusion Detection and Removal. 25th IEEE International Conference on Image Processing (ICIP). 2018:2905-2909.

[3] Ouannes L, Ben Khalifa A, Ben Amara NE. Comparative Study Based on De-Occlusion and Reconstruction of Face Images in Degraded Conditions. Traitement du Signal. 2021;38:573-585.

[4] Kahatapitiya K, Tissera D. Context-Aware Automatic Occlusion Removal.2019 IEEE International Conference on Image Processing (ICIP). IEEE. 2019:1895-1899.

[5] Hays J, Efros AA. Scene Completion Using Millions of Photographs. ACM Trans Graph. 2007;26:4.

[6] Park JS, Oh YH, Ahn SC, Lee SW. Glasses Removal From Facial Image Using Recursive Error Compensation. IEEE Trans Pattern Anal Mach Intell. 2005;27:805-811.

[7] Bincy Antony M, Narayanankutty KA. Removing Occlusion in Images Using Sparse Processing and Texture Synthesis. Int J Comput Sci Eng Appl. 2012;2:117-124.

[8] Jassim FA. Image inpainting by Kriging interpolation technique. 2013. ArXiv preprint: https://arxiv.org/pdf/1306.0139.

[9] Kwiatkowski M, Hellwich O. Specularity, Shadow, and Occlusion Removal from Image Sequences using Deep Residual Sets. In VISIGRAPP (4: VISAPP). 2022:118-125.

[10] Criminisi A, Pérez P, Toyama K. Region Filling and Object Removal by Exemplar-Based Image Inpainting. IEEE Trans Image Process. 2004;13:1200-1212.

[11] Zhu W, Wang X, Wu Y, Zou G. A Face Occlusion Removal and Privacy Protection Method for Iot Devices Based on Generative Adversarial Networks. Wirel Commun Mob Comput. 2021;2021:6948293.

[12] Din NU, Javed K, Bae S, Yi J. A Novel Gan-Based Network for Unmasking of Masked Face. IEEE Access. 2020;8:44276-44287.

[13] Yang X, Xu P, Xue Y, Jin H. Contextual Feature Constrained Semantic Face Completion With Paired Discriminator. IEEE Access. 2021;9:42100-42110.

[14] Yu J, Lin Z, Yang J, Shen X, Lu X, et. al. Free-Form Image Inpainting With Gated Convolution. In: Proceedings of the IEEE/CVF international conference on computer vision. IEEE. 2019:4471-4480.

[15] Saleh K, Szénási S, Vámossy Z. Generative Adversarial Network for Overcoming Occlusion in Images: A Survey. Algorithms. 2023;16:175.

[16] Khamele SV, Mundada SG. Mundada, An Approach for Removal of Occlusion Using Examplar Based Image-Inpainting Technique: A Review. Int J Eng Res Technol (IJERT). 2015;4.

[17] Srinivasan A. Balamurugan V. Occlusion detection and image restoration in 3D face image, TENCON 2014 - 2014 IEEE Region 10 Conference. IEEE. 2014:1-6.

[18] Hur J, Lee JY, Choi J, Kim J. I see-through you: A framework for removing foreground occlusion in both sparse and dense light field images. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE. 2023:229-238.

[19] Jiang A, Xu H. Decoupling: Image Occlusion Removal Based on Image Segmentation and Image Inpainting. J Phys Conf Ser. 2023;2646:012003.

[20] Gupta S, Shetty A, Sharma A. Attention Based Occlusion Removal for Hybrid Telepresence Systems. 19th Conference on Robots and Vision (CRV). IEEE. 2022:167-174.

[21] Devasruthi D, Menon HP. Fe-Bemd and Exemplar Based Hybrid Image Inpainting for Occlusion Removal. Int J Comput Appl. 2011;28:0975–8887.

[22] McCloskey S, Langer M. Removal of Partial Occlusion From Single Images. IEEE Trans Pattern Anal Mach Intell. 2011;33:647-654.

[23] Zhao F, Feng J, Jian Zhao, Wenhan Yang, Shuicheng Yan. Robust Lstm-Autoencoders for Face De-occlusion in the Wild. IEEE Trans Image Process. 2018;27:778-790.

[24] Yin X, Huang D, Fu Z, Wang Y, Chen L. Segmentation-Reconstruction-Guided Facial Image De-occlusion. In2023 IEEE 17th international conference on automatic face and gesture recognition (FG). IEEE. 2023:1-8.

[25] Kumar A, Kaushal M. SAM C-GAN: A Method for Removal of Face Masks From Masked Faces. Signal Image Video Process. 2023;17:3749-3757.

[26] John S, Danti A. Removal of Occlusion in Face Images Using pix2pix Technique for Face Recognition. In Congress on Intelligent Systems: Proceedings of CIS. Springer. 2022;2:47-57.

[27] Xiao X, Daneshpanah M, Javidi B. Occlusion Removal Using Depth Mapping in Three-Dimensional Integral Imaging. J Disp Technol. 2012;8:483-490.

[28] Lee BG, Kang HH, Kim ES. Occlusion Removal Method of Partially Occluded Object Using Variance in Computational Integral Imaging. 3D Res. 2010;1:2.

[29] Gao Z, Li D, Xiong C, Hou J, Bo H. Face Recognition With Contiguous Occlusion Based on Image Segmentation. International Conference on Audio Language and Image Processing Shanghai China. 2014:156-159.

[30] Rajeswari G, Ithaya Rani P. Face Occlusion Removal for Face Recognition Using the Related Face by Structural Similarity Index Measure and Principal Component Analysis. J Intell Fuzzy Syst. 2022;42:5335-5350.

[31] Zhang M, Piao Y, Wei C, Si Z. Occlusion Removal Based on Epipolar Plane Images in Integral Imaging System. Opt Laser Technol. 2019;120:105680.

[32] Patel TB, Patel JT. Occlusion Detection and Recognizing Human Face Using Neural Network. International Conference on Intelligent Computing and Control (I2C2).IEEE. 2017:1-4.

[33] Dong J, Zhang L, Zhang H, Liu W. Occlusion-Aware Gan for Face De-occlusion in the Wild. IEEE International Conference on Multimedia and Expo (ICME).IEEE. 2020:1-6.

[34] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In 2017 Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2017:5967-5976.

[35] Han K, Wang Y, Tian Q, Guo J, Xu C, et. al. GhostNet: More Features From Cheap Operations. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2020:1577-1586.

[36] Iizuka S, Simo-Serra E, Ishikawa H. Globally and Locally Consistent Image Completion. ACM Trans Graph. 2017;36:107.

[37] Yu J, Lin ZL, Yang J, Shen X, Lu X, Huang TS. Free-Form Image Inpainting With Gated Convolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). IEEE. 2018:4470-4479

[38] Nazeri K, Ng E, Joseph T, Qureshi FZ, Ebrahimi M. Edgeconn Gener Image Inpainting Adversarial Edge Learn. 2019. ArXiv Preprint: https://arxiv.org/pdf/1901.00212