Scrupulous SCGAN Framework for Recognition of Restored Images with Caffe based PCA Filtration

Khushboo Agarwal
Department of Computer Science and Engineering
Madhav Institute of Technology and Science, India
ka.agarwals@mitsgwalior.in

Manish Dixit
Department of Computer Science and Engineering
Madhav Institute of Technology and Science, India
dixitmits@mitsgwalior.in

Abstract: *Computer vision enables to detect many objects in any scenario which helps in various real time application but still face recognition and detection remains a tedious process due to the low resolution, blurriness, noise, diverse pose and expression and occlusions. This proposal develops a novel scrupulous Standardized Convolute Generative Adversarial Network (SCGAN) framework for performing accurate face recognition automatically by restoring the occluded region including blind face restoration. Initially, a scrupulous image refining technique is utilised to offer the appropriate input to the network in the subsequent process. Following the pre-processing stage, a Caffe based Principle Component Analysis (PCA) filtration is conducted which uses convolutional architecture for fast feature embedding that collects spatial information and significant differentiating characteristics to counteract the loss of information existing in pooling operations. Then a filtration method identifies the specific match of the face based on the extracted features, creating uncorrelated variables that optimise variance across time while minimising information loss. To handle all the diversification occurring in the image and accurately recognise the face with occlusion in any part of the face, a novel Standardized Convolute GAN network is used to restore the image and recognise the face using novel Generative Adversarial Network (GAN) networks are modelled. This GAN ensures the normal distribution along with parametric optimization contributing to the high performance with accuracy of 96.05% and Peak Signal to Noise Ratio (PSNR) of 18 and Structural Similarity Index Metric (SSIM) of 98% for restored face recognition. Thus, the performance of the framework based on properly recognizing the face from the generated images is evaluated and discussed.*

Keywords: *Computer vision, face recognition, image restoration, occlusions, pre-processing, spatial information, PCA, GAN networks.*

Received December 1, 2022; accepted May 24, 2023
https://doi.org/10.34028/iajit/21/1/10

1. Introduction

In the era of ubiquitous imaging, that comprises numerous basic products including surveillance systems with built-in imaging sensors, an increasing number of photographs are being taken in unrestricted settings Guo *et al*. [3]. Face photos are one of the most prevalent kinds of pictures we encounter every day, but they frequently become degraded due to a number of problems, such as low resolution, blur, noise, compression, etc., or a mixture of these factors Almabdy and Elrefaei [1]. On the other hand, accurate face photos are essential for tasks like facial recognition software and human perception. Face restoration and recognition is therefore a demanding although promising field of research in computer vision Loussaief and Abdelkrim [12].

Face Recognition is a biometric system application for identifying and verifying faces. A Face Recognition system must be able to detect or classify a non-cooperative face in an unmanaged scenario without the subject's knowledge Wang *et al.* [19]. Face recognition algorithms, on the other hand, were severely challenged by the diversity of the surroundings and scenarios. Although some amount of reliability has been reached by current automated identification techniques, many real-world applications place limitations on their performance Zeng *et al.* [21]. The biggest challenges in recognising faces stem from the wide range of characteristics imposed by disparities in acquisition conditions. The complexity of recognising the subjects in a multitude illumination environment and poses Wang *et al.* [18].

Since it seeks to reconstruct a precise and accurate face features from a damaged insight, face image restoration has attracted a lot of attention. Because of the unknown and complex degradation of Low Quality (LQ) face imagery in the wild Navabifar and Emadi [14], Blind Face Restoration (BFR) involving occlusion remains a challenging research problem despite substantial advancements Lee *et al*. [7]. A variety of BFR methods have been proposed employing spatial transformer networks Yu *et al*., [20], exemplar pictures Li *et al*. [10], 3D facial priors Hu *et al*., [5], including facial component dictionaries Li *et al*., [9] to recover a High-Quality (HQ) face image with photo-realistic textures from a Low-Quality (LQ) face image. Recent advancements in carefully constructed architecture as well as the application of relevant prior offenses in deep neural convolutional networks have increased the credibility and acceptability of the restoration findings. The blind restoration issue cannot be resolved because real LQ pictures frequently have complex and varied distributions which are challenging to synthesise despite great advances Karras *et al*. [6]. Reference-based techniques, which use a reference prior in an image restoration job to enhance network learning and decrease the channel's reliance on degraded input, have been presented as a solution to this issue Zhao and Qi [22].

Contrary to traditional image restoration, face restoration can restore features of face elements even if the images are significantly damaged by using prior information of the face. As a result, a number of recent studies on face super resolution use facial prior information to improve performance using techniques including parsing maps, face landmarks, or identification prior. The majority of these techniques use an encoder-decoder structure to acquire a direct black-box mapping from LQ to HQ images, following the standard image restoration approach Zhao and Qi [23].

Currently, the majority of this technology's related work focuses on identifying un-occluded facial expression images. However, in reality, partial obstruction of the image acquisition device's collection of facial expression image features by hands, glasses, masks, as well as other things is common. By preventing the retrieval of expression characteristics, these occlusions can decrease the precision of expression recognition Lin *et al*. [11]. The need of the present is for a system that can properly recognise expression under occluded conditions. Thus this paper works on developing a robust recognition approach which can solve the difficulty of face identification under occlusion with blind face restoration once the obstacles like illumination and noise are solved using following contributions,

- Scrupulous image refining technique is utilised that extracts the input, detects the face, and performs optimised processing of the requisite part of the image, enabling the complexity of processing unwanted regions to be deduced at an early stage.
- An optimal Caffe based Principal Component Analysis (PCA) filtration is conducted in which collects spatial information and significant differentiating characteristics to counteract the loss of information, provides the information of occlusion and identifies the specific match, creating uncorrelated variables that optimise variance across time.
- A novel standardized convolute GAN network is used to restore the image and recognise the face where the normal distribution along with parametric optimization contributing to the high performance for restored face recognition.
- Finally, this research work has been organised as follows: the conventional technique surveys are covered in section 2, the proposed method is covered in section 3, section 4 gives the performance and overall comparative results and section 5 provides the conclusion obtained from the research study.

2. Literature Survey

With the rapid growth of Convolution Neural Network (GAN) approaches, some methods for reconstructing faces from extremely low-resolution inputs have recently been developed. Lu *et al*. [13] proposed a WGAN-based occluded FER technique. This strategy consists of one generator and two discriminators. The generator network implements the occluded area under the twofold constraint of the weighted reconstructing with triplet loss parameters proposed in this study. Based on the initial un-occluded photos as well as the generated complementing pictures, an unnecessary area occluded image is additionally added as the discriminator's input. The Wasserstein distance was employed by the model to create an adversarial loss function among the produced complemented pictures, which improved the system's capacity to extract features. To finish the expression identification process, the classification loss function is introduced. However, for obtaining optimal recognition, the occlusion area was limited.

Sharma and Kumar [17], proposed a new 3D face reconstruction method as well as a sequential deep learning-based face identification method. It makes advantage of the voxels produced by the voxelization process. The mid-face plane is used to generate the reconstructed point in 3D using the reflection concept. A progressive deep learning architecture is developed using the reconstructed face to distinguish sexuality, feeling, exclusion, and individual. The deep features aid the Bidirectional LSTM in generating consistent embedding for better classification when combined with triplet loss training. The fuzzy c-means clustering method helps manage the sparsity of the voxels obtained after voxelization at a variety of positions. There is a huge loss of information while processing which reduces the accuracy.

Li *et al*. [8], studied by using Image Gradient Orientations (IGO) into robust programming coding. However, in the IGO domain, a weight-conditional Gaussian distribution as well as a uniform distribution can both construct the error distribution in the occluded section but also region, respectively, elegantly and simply. This research offers a joint probabilistic generative design for a novel IGO-embedded Structural Error Coding (IGO-SEC) model by including the two error distributions and a Markov random field for the prior distribution of the occlusion support. Two new methodologies-a novel reconstruction technique as well as a novel adaptive structural error metric—are introduced to enhance the performance of IGO-SEC. However, for improved performance, strong error coding techniques must be incorporated into deep learning algorithms.

An encoder network was used by Richardson *et al*. [16] to develop a general image-to-image translating system before being fed into a pre-trained generator. However, these techniques can only be developed to identify non-blind image super-resolution issues. Additionally, they did not make any changes to the pre-trained GAN during training to keep the uniformity & simplicity of face operations. This leads to inconsistent restoration quality when working with actual LQ face photographs with cluttered backgrounds since it is challenging to translate a face image with a little resolution to a necessary latent coding.

In this paper, Qiu *et al*. [15] present a novel occlusion-resistant end-to-end deep neural network-based facial recognition algorithm. Deep CNN may learn to recognise damaged characteristics, as well as a technique called Face Recognition with Occlusion Masks (FROM) cleans them up with dynamically learnt masks. Additionally, they produce sizable occluded facial images for effectively training FROM. In contrast to previous techniques that also employ shallow concepts which are less discriminatory than the suggested one or external sensors to identify occlusions. The network only has focus on occluded image for recognition and does not work on quality of the image.

Zhao *et al*. [24], recent 3D facial texture restoration experiments have mostly ignored occlusions like fingers, food that is about to enter the mouth, glasses, or other objects while focusing on improving resolutions. A deep learning-based strategy for precise 3D face restoration that doesn't need sizable 3D databases was presented in this paper. The weakly supervised design of our method is at its core, decoupling the problem of estimating a resilient fundamental shape from the process of estimating its mid-level details, which are represented here as bump maps. The 3D reconstruction framework based on weak supervision driven by the contour can build convincing 3D models, according to this research. There is a need to recreate the 3D face model using self-supervision for better applicability of the model.

Hariri [4] proposed a solid approach based on occlusion reduction with deep learning-based characteristics to handle the problem of masked facial recognition software. First, the area covering the face must be taken off. Then, they use three deep CNNs that are already trained to extract deep features from the retrieved areas. The feature maps are then quantized using the Bag-of-features approach to create a tiny approximation of a conventional CNN. Lastly, multilayer perceptron is used to carry out the categorization procedure. To enhance the efficacy, it is still necessary to merge deep classifier model with additional pre-trained systems.

However, in Lu *et al*. [13] the occlusion area was limited, in Sharma and Kumar [17] there is a huge loss of information while processing. Li *et al*. [10] still needs to include robust error coding models into deep learning systems for enhanced performance. When LQ face images with complex backgrounds are found, Richardson *et al*. [16] shows unsteady restoration quality, Qiu *et al*. [15] does not work on quality of the image, Zhao and Qi [23] needs to recreate the model using self-supervision, Hariri [4] need to combine deep ensemble models with additional pre-trained models to improve accuracy. Thus, there is a need to develop a framework which can overcome all the aforementioned issue in an optimized way.

3. Scrupulous CGAN with Caffe Framework

Computer vision enables to detect many objects in any scenario which helps in various real time application but still face recognition and detection remains a tedious process due to the low resolution, blurriness, noise, diverse pose and expression and occlusions. Thus, this proposal develops a novel scrupulous CGAN with Caffe framework as shown in Figure 1, for performing accurate face recognition automatically by restoring the occluded region including blind face restoration.

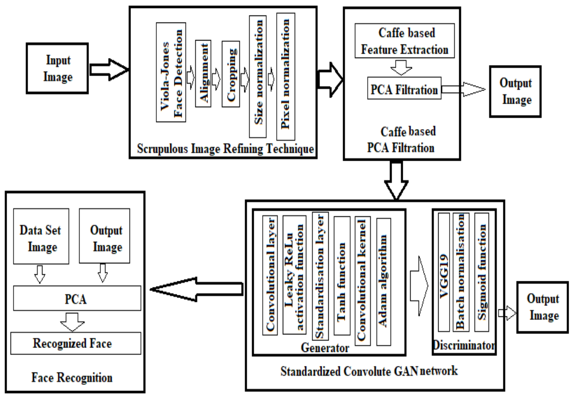

Figure 1. Proposed architecture diagram.

Initially, a scrupulous image refining technique is utilised that extracts the input, detects the face using the viola jones face detection algorithm, and performs optimised alignment and cropping of the requisite part of the image, enabling the complexity of processing unwanted regions to be deduced at an early stage. Then, for offering the appropriate input to the network in the subsequent process, a dual normalisation of size and pixels is performed. In the previous restoration method while extracting features there is a loss of information by the pooling layers which affects the accuracy of the recognition. The recognition system also must be automated to handle all the diversification occurring in the image and accurately recognise the face with occlusion in any part of the face. Hence, following the pre-processing stage, an optimal caffe based PCA filtration is conducted in which a caffe model is utilized that collects spatial information and significant differentiating characteristics to counteract the loss of information existing in pooling operations and provides the information of occlusion. Then the Principal Component Analysis (PCA) method identifies the specific match of the face based on the extracted features, creating uncorrelated variables that optimise variance across time while minimising information loss. If the correlation is greater than a threshold, the face is recognised and the operation is concluded; otherwise, if the image is not recognised, there is occlusion or low quality which should be rectified. Moreover, the previous conventional models in the literature can only handle one particular type of error. There is no method to handle all the artefacts and perform a perfect accurate recognition for occluded image. A novel Standardized Convolute GAN (SCGAN) network is used to restore the image and recognise the face where GAN network is modelled. SCGAN uses a CNN with a variational encoder structure that includes a leaky ReLu activation function and a tanh function in the final layer, also a standardisation layer is added after each convolutional layer to ensure normal distribution, and a convolutional kernel is utilised to convolute and de-convolute in an optimal way using the Adam algorithm for parametric optimization. Then, using VGG19, a discriminator is created, and batch normalisation is applied to the convolutional layers using the sigmoid function. Thus, the SCGAN contributes to the high performance for both occluded image with low quality. The face is then properly recognised using PCA from the generated images. As a result, the suggested framework can be used to accurately recognise faces in any type of image overcoming the issues faced in prior networks. The following section discuss the proposed framework in detail under various subsections.

3.1. Scrupulous Image Refining Technique

Initially, a Scrupulous image refining technique is utilised that extracts the input, detects the face using the viola jones face detection algorithm, and performs optimised alignment and cropping of the requisite part of the image, enabling the complexity of processing unwanted regions to be deduced at an early stage as shown in Figure 2. Then, to offer the appropriate input to the network in the subsequent process, a dual normalisation of size and pixels is performed.

Paul Viola and Michael Jones first presented the Viola-Jones face recognition system in 2001. It is the initial object recognition platform that provide good real-time recognition rate. The Viola-Jones system includes three frameworks for detecting face features:

- The rectangular Haar-like characteristics employed for feature extraction are determined by an entire picture.
- Ada boost is a Machine Learnig (ML) approach for recognising facial expressions. The classifiers that are complex in and of themselves at each level and are created from basic classifiers employing any of the four boosting procedures are referred to as “boosted” classifiers.
- A cascade classifier was utilised to efficiently merge several of the characteristics. The numerous filters on a subsequent classifier are determined by the classifier's “cascade.”

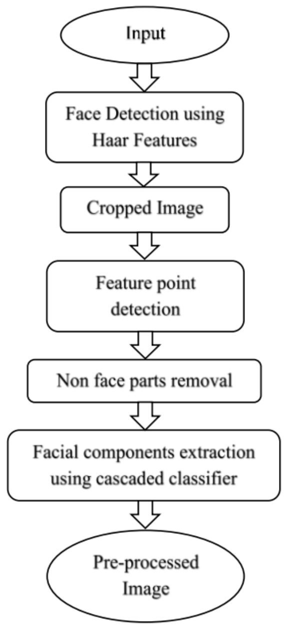

Figure 2. Flow chart of scrupulous image refining technique algorithm.

3.1.1. Haar Feature Extraction

Viola Jones' technique use rectangular characteristics rather than pixels. Ad hoc domain knowledge cannot be encoded using a limited amount of training data. This issue can be solved by using rectangular features.

The difference between the pixel summations within of two rectangular zones is known as the “two rectangle feature.” As a result, three rectangle characteristics denote the total of two peripheral rectangular regions less the addition of a centre rectangle. You can differentiate among diagonal pairs of rectangles using a four-rectangle function.

The integral images at positions *x*, *y* contains the total of the pixels above but to the left of *x*, *y*, inclusive:

$$ii(x,y) = \sum_{x' \leq x, y' \leq y} i(x', y') \tag{1}$$

The following pair of recurrences may be used to compute integral images in one pass over the original image:

$$s(x,y) = s(x,y-1) + i(x,y) \tag{2}$$

$$ii(x,y) = ii(x-1,y) + s(x,y) \tag{3}$$

Here, $s(x,y)$ means the cumulative row sum.

With four array references in rectangle D, it is possible to calculate the pixels' total number. These are: the integral image value at position 1 is allocated as the sum of the pixels in the rectangle designated as A, and position 2 is determined by the value at position A+B.

A+C is the value at position 3, while A+B+C+D is the value at position 4. The total within D is calculated as 4+1-(2+ 3). A two-rectangle feature is computed at any size using six array references.

3.1.2. Ada Boost for Feature Selection

For feature selection and classifier training, variant Ada Boost is utilized. It creates a classifier that is stronger by combining a number of weak classification algorithms. The weak learner needs to overcome a number of learning challenges. The examples are reweighted after the initial learning phase. The re-weighting method corrects for the less accurate classifications made by the prior wealth classifier. The final classifier uses a perceptron's structure and is powerful.

Given example images x_i and labels y={0, 1}, let us initialize weights according to the following equation:

$$w_{i,1} = \frac{1}{2m}, \frac{1}{2l} \tag{4}$$

Where *m* and *l* respectively mean the number of negative and positives. For *t*=1,…,*T*, we normalize the weights:

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}} \tag{5}$$

Afterward, it select the min-error classifier ht, $\epsilon_t = \min_{f,p,\theta} \sum_i w_i |h(x_i, f, p, \Theta) - y_i|$ the weight of xi is unaffected if it is erroneously categorized. If not, its weight must be changed in accordance with the following equaion:

$$w_{t+1} = w_{t,i} \frac{\epsilon_t}{1-\epsilon_t} \tag{6}$$

3.1.3. Cascaded Classifier

The final classifier is created by linearly combining weak classifiers. Each classifier's performance is considered while weighing it:

$$C(x) = sign\left[\sum_{t-1}^{T} \left(\log\frac{1-\epsilon_t}{\epsilon_t}\right)\left(h_t(x)\frac{1}{2}\right)\right] \tag{7}$$

Note that for $\epsilon_t = 0.5$, the classifier t fails to participate in the combination.

The first two features are easy to comprehend in terms of face recognition. Ada Boost chooses the area around the eyes as her initial characteristic. Typically, this area is darker than the area around the cheekbones and nose. The eyes, which are frequently darker than the nasal bridge, are the second distinguishing characteristic.

Cascaded classifiers are more effective since they are smaller. A classifier of this kind rejects the majority of negative sub windows and finds the majority of positive ones. Most of the sub windows are filtered out using simpler algorithms. Low false positive rates are achieved by using more complicated classifiers.

The rate of false positives for a trained classifier cascade

$$F = \prod_{i=1}^{K} f_i \tag{8}$$

Where *K* is the total amount of classifiers, *F* is the false positive rate across the board, and fi is the false positive rate for the i-th classifier. The rate of detection is

$$D = \prod_{i=1}^{K} d_i \tag{9}$$

Where, d_i is the detection rate of the ith classifier. Thus the image is ready to extract the features of occluded face efficiently.

3.2. Caffe based PCA Filtration

An optimal convolutional architecture for fast feature embedding based PCA filtration is conducted in which a convolutional architecture for fast feature embedding (Caffe) model is utilized that collects spatial information and significant differentiating characteristics to counteract the loss of information existing in pooling operations and provides the information of occlusion. Then the Principal Component Analysis (PCA) Cao and Liu [2] method identifies the specific match of the face based on the extracted features, creating uncorrelated variables that optimise variance across time while minimising information loss. If the correlation is greater than a threshold, the face is recognised and the operation is concluded; otherwise, if the image is not recognised, there is occlusion or low quality which should be rectified.

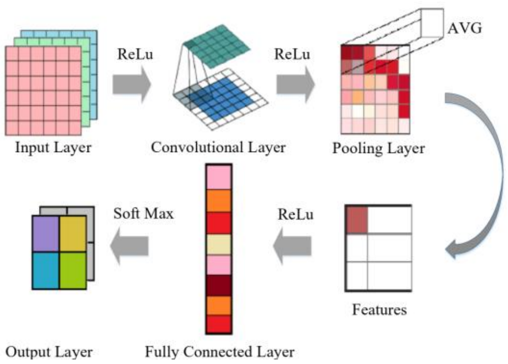

Figure 3. Basic caffe model.

Object detection is a goal shared by computer vision and ML. DL established techniques had been effectively used in current era to recognise specific items of interest in datasets comprising photos, videos, and live-feed recordings. We expand on prior research in the field of face recognition with the comparatively recent objective of facial-mask recognition. Berkeley AI Research (BAIR) created the deep learning framework Caffe as shown in Figure 3, as a quick and effective object identification technique. In the aim of finding faces in images and videos, we employed the Caffemodel. This jointly functions with the face mask method to categorise whether or not the provided faces are wearing masks.

OpenCV, an open-source computer vision library, offers the Caffe model. The Single Shot multi-box Detector (SSD) serves as the model's foundation, while the ResNet-10 design serves as the backbone network. Similar to the You Only Look Once (YOLO) approach, the Single Shot multi-box Detector uses Multibox to identify several items in a single shot. Next, the cv2.dnn.readNet programme was used to load the prototxt and Caffemodel files. The model's weights are provided by the Caffemodel, and other details about the model's architecture were provided via the prototxt file. After using the image recognition method, we get the amount of faces found, the locations of their individual bounding boxes, as well as the degree of confidence in those forecasts. We can extract every face from the image using this method, and it also enables quick and resource-efficient race identification in real time.

3.3. Standardized Convolute GAN Network

A novel Standardized Convolute GAN (SCGAN) network is used to restore the image and recognise the face where two distinct GAN networks are modelled. SCGAN as in Figure 4, uses a CNN with a variational encoder structure that includes a leaky ReLu activation function and a tanh function in the final layer, also a standardisation layer is added after each convolutional layer to ensure normal distribution, and a convolutional kernel is utilised to convolute and deconvolute in an optimal way using the Adam algorithm for parametric optimization. Then, using VGG19, a discriminator is created, and batch normalisation is applied to the convolutional layers using the sigmoid function. Thus, the SCGAN contributes to the high performance for both occluded image with low quality. The face is then properly recognised using PCA from the generated images.

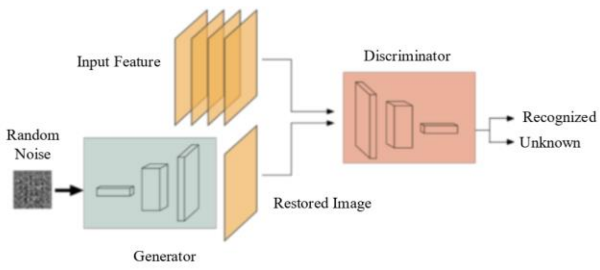

Figure 4. Process of proposed SCGAN model.

3.3.1. Generator

In this study, the generator uses a convolution network with a variation auto encoder structure. The encoder eploys 12 convolutional layers, one of which is fully connected, while the decoder employs 12 de-convolutional layers, one of which is fully connected. Simultaneously, LeakyRelu is employed as the activation function in the first 25 levels of the created network, while Tanh is utilised in the last layer. When the input image is down sampled and reaches the bottleneck layer, the image features are compressed to 1*1*4000 size before the corrected image is formed by up sampling. Due to the deep depth of the network in this article, the data distribution would progressively shift during the picture production process, affecting the model's convergence speed and resulting to the disappearance of some neural network layers during back propagation. As a result, a batch standardisation level is included after each convolution and deconvolution level to ensure that the input of each layer has the same normal distribution, hence accelerating network training convergence. To improve the generator's performance, a tiny convolution kernel is utilised to convolute or deconvolutions the input image. A composite layer is created for every two convolutions or deconvolutions. In a composite layer, the first convolution kernel size is set to 3*3, the step size is 1, and the 0 filling term is 1. The image in the coding stage is half reduced after passing through a composite layer, and the image in the decoding stage is twice enlarged after passing through a composite layer. The objective of neural network training is to optimise model parameters via back propagation to minimise the value of the loss function. The Adam method is used as the parameter optimization algorithm in this paper's generator model. The Adam algorithm, published in 2015, is a parameter optimization approach that combines the Momentum and AdaGrad algorithms to achieve an efficient search of parameter space. Figure 2 depicts the picture correction procedure of this generator.

3.3.2. Discriminator

The discriminator utilised in this work is based on VGG19 and contains of 13 convolution layers and 5 pooling layers. A batch standardisation layer is added to the middle 11 convolution layers to improve the discriminator's stability. The last layer employs the Sigmoid function to determine the likelihood that the input image is genuine. Thus, the proposed SCGAN performs perfect face recognition with image restoration for occluded images. The next section clearly explains the implementation details and results obtained from the proposed method.

4. Result and Discussion

The proposed Scrupulous SCGAN with Caffe framework is implemented in Python. The output of the implementation and the results obtained are discussed in this section.

4.1. System Configuration

The system configuration selected for implementing the proposed SCGAN model in Python 3.9 is provided in Table 1. The Intel Core i5 processor is a popular choice for general computing tasks and 16 GB of RAM is a good amount for running Python and handling data-intensive operations. Nvidia GPUs are commonly used for machine learning and deep learning tasks, and they can significantly accelerate computations for certain Python libraries like TensorFlow and PyTorch. A 1 TB Hard Disk Drive (HDD) provides ample storage capacity for storing your Python scripts, libraries, and datasets and it offers faster read/write speeds and can significantly improves overall system performance. Windows 10 is a widely used operating system that is compatible with Python and its various libraries.

Table 1. The configuration of the system used in implementation of proposed method.

Processor	**Intel Core i5, V generation**
RAM	16 GB
Graphics	Nvidia
HDD	1 TB
OS	Windows 10
Tool	Python 3.9

4.2. Dataset Description

A library of face images called Labelled Faces in the Wild (LFW) was utilized to examine the issue of occluded face identification. The dataset used in this research was built by the scholars of University of Massachusetts, Amherst. Then the faces in the images was detected using Viola Jones face detector, totally 5,749 persons in 13,233 images that were acquired from the internet. The datset contains two or more different images of 1,680 of the individuals. The database includes three distinct kinds of “aligned” images in addition to four separate sets of LFW images with deep-funnelled version. The dataset contains a variety of occlusions, such as sunglasses, scarves, and hats, making it suitable for evaluating occluded face recognition algorithms captured in challenging scenarios. It primarily consists of unconstrained, in-the-wild face images captured under various conditions, including variations in pose, lighting, expression, and background.

4.3. Implementation Results

In this research, we use Python 3.9 platform to build an improved generative adversarial networks model. During the training, we use Adam optimizer with a learning rate of 0.002 to optimize the model parameters, and carry out 100 cycle iterative training on 13,233 face image training set. Before the training set is loaded into the model, the image is pre-processed using Scrupulous image refining technique. There are 5, 749 identities represented in 13, 233 images in the LFW collection. To produce the test set, the images are pre-processed after the training dataset and a ratio of 3:1 is considered in splitting the dataset for training and testing. Then, each model receives the occluded face images as shown in Figure 5. To determine the efficacy of occluded face recognition, 6K pairings 3K positive pairs and 3K negative pairs are identified from among them using CAFFE based model. In this paper, the data loader function is used to obtain image data, in which the batch size is set to 4 and shuffle is set to true, so that the sequence of four image data of the same batch in different periods can be randomly scrambled to reduce the impact of input sequence on training using PCA.

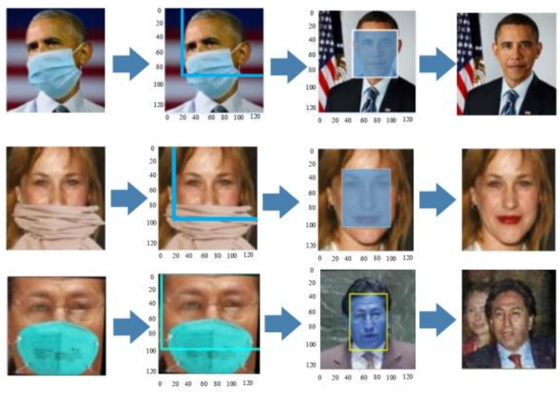

Figure 5. Processing of single image using the proposed framework.

4.4. Ablation Study

This research performed an ablation study on the effect of specific components of the SCGAN model. Starting from the original GAN (+Variational Encoder (CNN) + Standardisation Layer+Adam Optimization + activation layer), we gradually inject our modifications on the generator (adding Variational Encoder (CNN)), discriminator (VGG), and the loss (Standardisation Layer+Adam Optimization+activation layer). The results are summarized in Table 2. The results show that all our proposed components steadily improve both PSNR and SSIM. In particular, the Variational Encoder (CNN) + VGG module contributes most significantly. Also, adding PCA benefits both training stability and final results.

Table 2. Ablation study on the LFW dataset, based on DeblurGAN-v2.

Model	**PSNR**	**SSIM**
GAN	20.70	0.927
+ Variational Encoder (CNN)	20.26	0.931
+ Variational Encoder (CNN) + VGG	20.29	0.932
+ Variational Encoder + VGG + Standardisation Layer + Adam Optimization	19.37	0.943
GAN (+ Variational Encoder (CNN) + VGG + Standardisation Layer + Adam Optimization + activation layer)	19.55	0.954
GAN (+ Variational Encoder (CNN) + + VGG Standardisation Layer + Adam Optimization + activation layer) + PCA	18.01	98

4.5. Performance Metrics

The depicted Figure 6 shows the overall performance of the proposed model based on accuracy, Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Metric (SSIM). The proposed model has achieved accuracy of 96.05% and PSNR of 18 and SSIM of 98%.

Thus proving the better working of the proposed model. Also the following section proves the model with comparative analysis with conventional method. Convolutional GANs are capable of developing realistic faces with a variety of occlusion patterns. The discriminator network picks up crucial traits and patterns associated with occlusions as it learns to differentiate between real and fake faces. Face recognition models can increase their capacity to detect and recognise faces, even in the presence of occlusions, by utilising these acquired characteristics to establish a strong knowledge of occlusion changes. This upgraded information gives the model the ability to identify and match faces even when they are partially obscured, enhancing recall, accuracy, and precision in occluded instances.

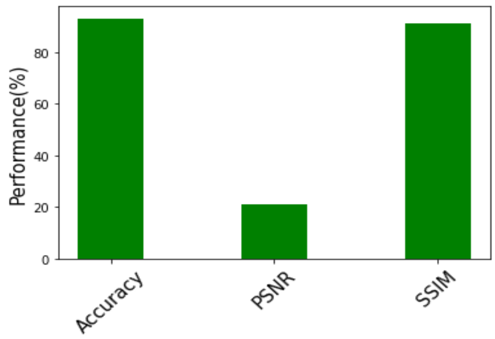

Figure 6. Overall performance of the proposed model.

The loss obtained from generator of the proposed GAN has been determined and depicted in Figure 7. As it is being trained, the discriminator separates the generator's actual information from its phoney data. It punishes it for mistakenly classifying a genuine example as phony or a false instance as real by maximising the following mechanisms.

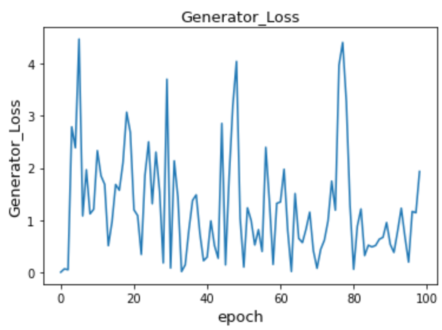

Figure 7. Loss obtained from generator of the proposed SCGAN.

$$\nabla_{\theta_i} \frac{1}{mn} \sum_{i=1}^{m} \left[log\ D\left(x^{(i)}\right) + log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \right] \quad (10)$$

- *log(D(x))* Refers the likelihood that the generator is correctly recognizing the real image.
- Maximizing *log(1-D(G(z)))* would improve in its ability to accurately classify the fake images produced by the generator.

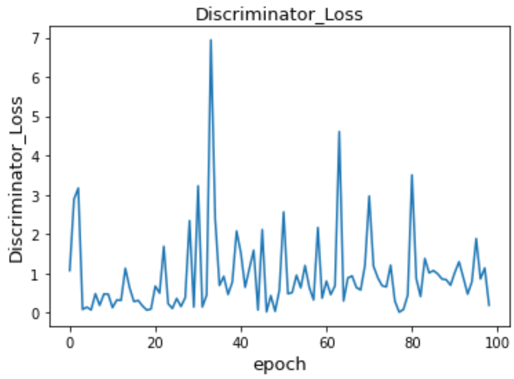

Figure 8. Loss obtained from generator of the proposed SCGAN.

As it learns, the generator recognises random noise and generates a result from it. The outcome is then sent to the discriminator, which determines whether it is “Real” or “Fake” dependent, the loss obtained by the discriminator is given in Figure 8 which shows how accurately it is able to classify them apart.

The discriminator's classification is then used to calculate the generator loss; if the generator is effective in deceiving the discriminator, it is awarded; if not, it is penalised.

Reduce the value of the given equations to train the generator:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \quad (11)$$

4.6. Comparison Metrics

To conduct a quantitative analysis of the restored and real face, they are utilized in this research. The basis for PSNR is the assessment of error-sensitive image quality. The restored face is more closely related to the original face the higher the value. With regard to image structure, brightness, and contrast, SSIM assesses the ability of image restoration.

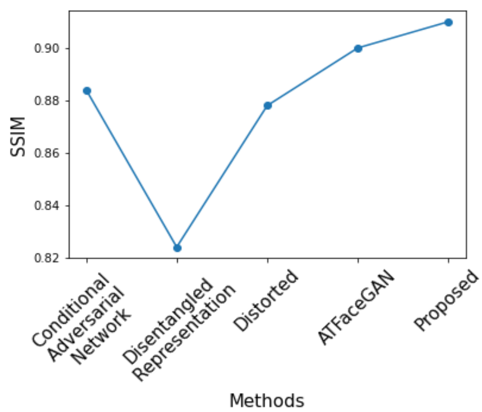

Figure 9. Comparison of proposed method based on SSIM obtained.

The impact of image restoration is improved by increasing the SSIM value, which ranges from 0 to 1. As shown in Figure 9, when the occluded area exceeds 50%, the proposed framework can repair the face, but the restoration impact is subpar, the repair trace is evident, and the occluded region and the non-occluded area have obviously different colours after repair.

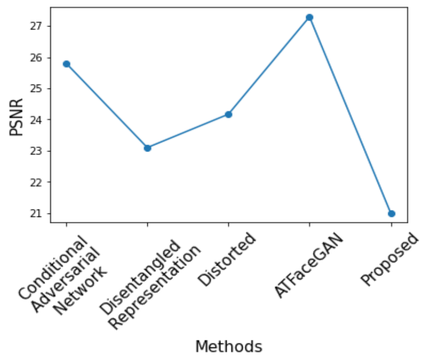

Figure 10. Comparison of proposed method based on PSNR obtained.

However, the facial expression recovered using this technique is complete, natural, and remarkably identical to the actual face as the PSNR value is lower than 20% as shown in Figure 10. When the face occlusion area is 50%, it can be shown that the restoration impact of the approach in this study is still much larger than that of the proposed framework, maintaining strong resilience.

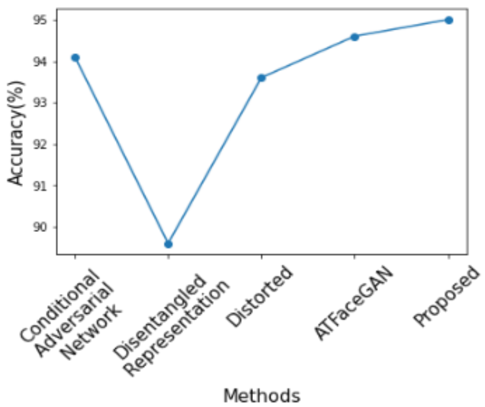

Figure 11. Comparison of proposed method based on accuracy obtained

Greater AUC indicates that the model does a superior job of separating the positive and negative groups. AUC values obtained are above 95% as shown in Figure 11, which suggests that the model successfully differentiates between the classes, while the values attained by the previous methods shows that the model has less discriminatory capacity than the proposed system.

5. Conclusions

This research presented a novel Scrupulous CGAN with Caffe framework for removing face image occlusion. Without knowing the nature or locations of the occlusions beforehand, the research has shown that the proposed framework can successfully denoise images that have been distorted by various forms of occlusions. The proposed framework was developed in Python 3.9 and evaluated using Labelled Faces in the Wild (LFW) benchmark database. The framework was validated with accuracy, the peak signal to noise ratio (PSNR) and Structural Similarity Index (SSIM) and proved to perform superior in all aspects. The proposed model has achieved accuracy of 96.05% and PSNR of 18 and SSIM of 98%. The loss of the discriminator and the generator of the proposed SCGAN was found to be very low. Also, the results obtained were compared with existing work and outperformed it all aspects. However, considering the performance of the computer, this experiment can further use cuda function to load all the processed data and models to GPU, which can speed up the training.

References

[1] Almabdy S. and Elrefaei L., "Deep Convolutional Neural Network-Based Approaches for Face Recognition," *Applied Sciences*, vol. 9, no. 20, pp. 4397, 2019. 10.3390/app9204397

[2] Cao L. and Liu D., "Face Image Super Resolution via Adaptive-Block PCA," *The International Arab Journal of Information Technology*, vol. 13, no. 6, pp. 699-706, 2016.

[3] Guo J., Zhu X., Yang Y., Yang F., Lei Z., and Li S., "Towards Fast, Accurate and Stable 3d Dense Face Alignment," *in Proceedings of the European Conference on Computer Vision*, Glasgow, pp. 152-168, 2020.

[4] Hariri W., "Efficient Masked Face Recognition Method During the Covid-19 Pandemic," *Signal, Image and Video Processing*, vol. 16, no. 3, pp. 605-612, 2022.

[5] Hu X., Ren W., LaMaster J., Cao X., Li X., Li Z., and Liu W., "Face Super-Resolution Guided by 3d Facial Priors," *in Proceedings of the European Conference on Computer Vision*, Glasgow, pp. 763-780, 2020.

[6] Karras T., Laine S., Aittala M., Hellsten J., Lehtinen J., and Aila T., "Analyzing and Improving the Image Quality of Stylegan," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Canada, pp. 8110-8119, 2020.

[7] Lee C., H., Liu Z., Wu L., and Luo P., "Maskgan: Towards Diverse and Interactive Facial Image Manipulation," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, pp. 5549-5558, 2020.

[8] Li X., Hao P., He L., and Feng Y., "Image Gradient Orientations Embedded Structural Error Coding for Face Recognition with Occlusion," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 6, pp. 2349-2367, 2020. https://doi.org/10.1007/s12652-019-01257-7

[9] Li X., Chen C., Zhou S., Lin X., Zuo W., and Zhang L., "Blind face Restoration Via Deep Multi-Scale Component Dictionaries," *in*

Proceedings of the European Conference on Computer Vision, Glasgow, pp. 399-415, 2020.

[10] Li X., Li W., Ren D., Zhang H., Wang M., and Zuo W., "Enhanced Blind Face Restoration with Multi-Exemplar Images and Adaptive Spatial Feature Fusion," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, pp. 2706-2715, 2020.

[11] Loussaief S. and Abdelkrim A., "Deep Learning Vs. Bag of Features in Machine Learning for Image Classification," *in Proceedings of the International Conference on Advanced Systems and Electric Technologies*, Hammamet, pp. 6-10, 2018. DOI: 10.1109/ASET.2018.8379825

[12] Lu Y., Wang S., Zhao W., and Zhao Y., "Wgan-Based Robust Occluded Facial Expression Recognition," *IEEE Access*, vol. 7, pp. 93594-93610, 2019. DOI:10.1109/ACCESS.2019.2928125

[13] Navabifar F. and Emadi M., "A Fusion Approach Based on HOG and Adaboost Algorithm for Face Detection under Low-Resolution Images," *The International Arab Journal of Information Technology*, vol. 19, no. 5, pp. 728-735, 2022.

[14] Qiu H., Gong D., Li Z., Liu W., and Tao D., "End2End Occluded Face Recognition by Masking Corrupted Features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6939-6952 2021. DOI:10.1109/TPAMI.2021.3098962

[15] Richardson E., Alaluf Y., Patashnik O., Nitzan Y., Azar Y., Shapiro S., and Cohen-Or D., "Encoding in Style: A Stylegan Encoder for Image-To-Image Translation," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, pp. 2287-2296, 2021.

[16] Sharma S. and Kumar V., "Voxel-based 3D Face Reconstruction and its Application to Face Recognition Using Sequential Deep Learning," *Multimedia Tools and Applications*, vol. 79, no. 25, pp. 17303-17330, 2020.

[17] Wang S., Cheng Z., Deng X., Chang L., Duan F., and Lu K., "Leveraging 3D Blendshape for Facial Expression Recognition Using CNN," *Science China Information Sciences*, vol. 63, no. 120114, pp. 1-120114, 2020.

[18] Wang Z., Wang G., Huang B., Xiong Z., Hong Q., Wu H., Liang J., and et al., "Masked Face Recognition Dataset and Application," *arXiv preprint arXiv:2003.09093*, 2020.

[19] Yu X. and Porikli F., "Hallucinating Very Low-Resolution Unaligned and Noisy Face Images by Transformative Discriminative Autoencoders," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, pp. 3760-3768, 2017.

[20] Zeng X., Peng X., and Qiao Y., "Df2net: A Dense-Fine-Finer Network for Detailed 3d Face Reconstruction," *in Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2315-2324, 2019.

[21] Zhao D. and Qi Y., "Generative Contour Guided Occlusions Removal 3D Face Reconstruction," *in Proceedings of the International Conference on Virtual Reality and Visualization*, Phu Quoc, pp. 74-79, 2021. https://doi.org/10.1007/978-3-030-98355-0_10

[22] Zhao D. and Qi Y., "Generative Landmarks Guided Eyeglasses Removal 3D Face Reconstruction," *in Proceedings of the International Conference on Multimedia Modeling*, Phu Quoc, pp. 109-120, 2022.

[23] Zhao D., Cai J., and Qi Y., "Convincing 3D Face Reconstruction from a Single Color Image under Occluded Scenes," *Electronics*, vol. 11, no. 4, pp. 543, 2022. DOI:10.3390/electronics11040543

Khushboo Agarwal presently working as an Assistant Professor in the Department of Computer Science and Engineering at Madhav Institute of Technology and Science, Gwalior, India. She is currently pursuing Doctoral Programme in RGPV, Bhopal, India. Her research interests include Computer Vision & Image Processing, Machine learning, Deep learning and Adhoc Networks.

Manish Dixit is currently working as a Professor and Head in the Department of Computer Science and Engineering at Madhav Institute of Technology & Science Gwalior. His areas of interest are Human-computer Interaction, Image Processing, Artificial Intelligence and Computer Graphics. He has published 11 SCI and more than a hundred research papers in several reputed journals and conferences. He received a Felicitation at the IETE Governing Council meeting for his efforts and achievement in working in IETE. He has also received an award for his research contribution in SCI, SCOPUS, and Conference publications. He is a Fellow of IETE, Senior Member of IEEE and Member of Computer Society of India.