IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 November 2022, accepted 24 November 2022, date of publication 2 December 2022, date of current version 28 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3226453

RESEARCH ARTICLE

Face Recognition via Multi-Level 3D-GAN Colorization

ZAKIR KHAN[1], ARIF IQBAL UMAR[1], SYED HAMAD SHIRAZI[1], MUHAMMAD SHAHZAD[1], MUHAMMAD ASSAM[2], MUHAMMAD TAREK I. M. EL-WAKAD[3], AND EL-AWADY ATTIA[4]

[1]Department of Information Technology, Hazara University Mansehra, Dhodial 21120, Pakistan
[2]Department of Software Engineering, University of Science & Technology, Bannu, Bannu 28100, Pakistan
[3]Faculty of Engineering, Future University in Egypt, Cairo 11835, Egypt
[4]Department of Industrial Engineering, College of Engineering, Prince Sattam bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia

Corresponding author: Syed Hamad Shirazi (syedhamad@hu.edu.pk)

ABSTRACT Rapid development in sketch-to-image translation methods boosts the investigation procedure in law enforcement agencies. But, the large modality gap between manually generated sketches makes this task challenging. Generative adversarial network (GAN) and encoder-decoder approach are usually incorporated to accomplish sketch-to-image generation with promising results. This paper targets the sketch-to-image translation with heterogeneous face angles and lighting effects using a multi-level conditional generative adversarial network. The proposed multi-level cGAN work in four different phases. Three independent cGANs' networks are incorporated separately into each stage, followed by a CNN classifier. The Adam stochastic gradient descent mechanism was used for training with a learning rate of 0.0002 and momentum estimates β and β as 0.5 and 0.999, respectively. The multi-level 3D-convolutional architecture help to preserve spatial facial attributes and pixel-level details. The 3D convolution and deconvolution guide the *G1*, *G2* and *G3* to use additional features and attributes for encoding and decoding. This helps to preserve the direction, postures of targeted image attributes and special relationships among the whole image's features. The proposed framework process the 3D-Convolution and 3D-Deconvolution using vectorization. This process takes the same time as 2D convolution but extracts more features and facial attributes. We used pre-trained ResNet-50, ResNet-101, and Mobile-Net to classify generated high-resolution images from sketches. We have also developed, and state-of-the-art Pakistani Politicians Face-sketch Dataset (PPFD) for experimental purposes. Result reveals that the proposed cGAN model's framework outperforms with respect to Accuracy, Structural similarity index measure (SSIM), Signal to noise ratio (SNR), and Peak signal-to-noise ratio (PSNR).

INDEX TERMS Convolutional neural network, generative adversarial network, sketch-to-Image translation, machine learning.

I. INTRODUCTION

Crime has taken a drastic revolution, so it demands enhancing the security of forensic files and records. There is an increased requirement to use technological measures in crime to identify, detect, and recognize suspects. For safety and security-related prompts, biometric recognition is necessary. One of the most common biometric techniques is face recognition. The face is the most convenient and reliable way of identification. Face-sketch recognition is a strong face identification domain when the photograph is not available. Face recognition systems have been evolving over the past few decades, particularly with the availability of large-scale databases and access to sophisticated hardware. Large-scale face recognition challenges such as MegaFace [1] and IARPA Janus Benchmark [2] provide further opportunities for bridging the gap between unconstrained and constrained face recognition. Sketch recognition is also an emerging trend in law enforcement agencies to identify suspects [3], [4]. Sketch recognition problems involve automated matching and

The associate editor coordinating the review of this manuscript and approving it for publication was Claudio Zunino.

generating coloured images from sketches [5]. There are two ways to reorganize the suspects by sketches: 1) Convert all the database images to sketches & compare the sketch with the sketch-database images. 2) Another way is to colourize the sketch and then find that colourized face image in the database. The first way is easy and less complex, but we lose too much information during the conversion process of images to sketching. So we are unable to find good & accurate results. On the other hand, if we convert the sketch into a coloured face image, this task is complex and challenging, but it is more effective to find out the suspect effectively.

Generative Adversarial Networks (GAN) have been used to colour the images and it may create sketches from coloured images. Due to the rapid development in GAN models [6], [7], [8], [9], the quality and efficiency of the sketch to coloured image translation have been improved significantly [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. Currently, translations from sketch-to-image or image-to-sketch have been extensively used in law enforcement agencies and digital image entertainment [20], [21], [22], [23], [24], [25], [26], [27], [28]. Zhang et al. [21] developed an architecture based on the dual transfer face sketch technique to improve the identification performance of sketched images. The Dual-transfer sketch approach comprises an intra-domain and inter-domain transfer process. It is used for identity-specific information loss and retrieval of common facial structures. Unlike a dictionary-based traditional approach, Zhang et al. [29] developed an end-to-end deep convolutional neural network (CNN) model for an image-to-image translation, while Isola et al. [30] work on conditional GAN by adding new condition y in traditional GANs. The condition y is used along with the input layer to handle the mapping between generated image and the input image. Zhang et al. [20] developed a generator for face sketching that addresses the problem of sketch generation using soothing properties. This work outperforms for the reduction of high-frequency loss with considerable performance. Zhang et al. [22] developed an automatic sketch generator comprising rough, fine, and finer face parts. The model colour the face sketch using mentioned parts with gentle and deep detail features. Probabilistic graphical models were used by Zhang et al. [23] to develop the face sketch architecture. They considered the generated sketch pixels and ground truth from training data to generate the face with fine details features. Zhang et al. [24] address heterogeneous lightning effect problems by developing cascaded face sketch synthesis models. This model comprises cascaded low-rank representation and numerous feature generators. The responsibility of the feature generator is to extract finely detailed features under different illumination. At the same time, the distance between the synthesized facial sketch and the corresponding ground truth was reduced by cascaded low-rank representation. To improve the efficiency of face sketching, Wang et al. [25] used new random sampling instead of an online KNN search method. Results show that this technique outperforms concerning quality and efficiency. Current face sketching techniques cannot select the neighbour feature during face synthesis. Bayesian techniques were used by Wang et al. [26] to consider the weight computation model and neighbor selection model to overcome it. This method competes with the existing techniques concerning subjective perceptions and objective evaluations. However, recent research [3], [4], [28], [31], [32], [33] [23], [34], [35] ignores the 3D-convolutional process for sketch to image colorization and controlling of pixel-level facial attributes during sketch to colored face translation. Existing sketch colorization and face sketching techniques are unable to outperform with heterogeneous lighting effects and fail to taking into account the neighboring features during face synthesis. The attention of all researchers was to achieve a balance between the target image and generated image to look more realistic and natural. But, due to minor facial feature selection, realistic and natural image generation is not achieved effectively. If the number of features increases directly, the desired results may be achieved. Still, it increases the complexity and depth of the model, which requires more computing power and other computational resources. The alternative way of increasing the feature is to increase the depth of convolutional layers and apply 3D convolutions instead of 2D. In addition, the existing research focuses on 2D-convolutions [36], [37] instead of 3D, which reduces the efficiency of the loss learning function to preserve spatial facial attributes of the input image. Existing research works does not provide any ground truth that may authenticate the performance using cross-match analysis.

Current research works, either GAN-based [23], [34], [35], [38] or CNN-based [3], [4], [28], [31], [32], [33], are insufficient to handle the facial attribute changing during sketch translation into RGB images. These techniques missed texture attributes and pixel-level details of facial attributes. However, the existing solution tends to overfit sketches because it outperforms training instances compared to test instances, thus requiring consistent professional sketches as inputs.

Most of the time, the photos of suspects obtained from surveillance cameras are poor in quality, so forensic experts draw face sketches of suspects and colour them to retrieve them from the database. To enhance retrieval performance and efficiency, we can synthesize face sketches from photos in the database and then match them with the suspect's sketch.

To overcome the problems mentioned above, a multi-level 3-dimensional conditional generative adversarial network (3D-cGAN) is proposed to translate and colourize sketches into realistic images. The proposed model translates and colours hand-drawn sketches into high-resolution RGB realistic images. It also controls spatial features and pixel-level details without affecting realistic attributes by imitating the condition. In addition to generating high-resolution RGB realistic images from sketches, the proposed model can also classify and recognize the input images. This architecture comprises four phases i.e., three cGANs followed by an image classifier. Each cGAN comprises Generator and Discriminator. The generator handles the 3D facial attributes during face sketch colourization and translation based on a conditional encoder-decoder network. It will be achieved by decoding optimum features extracted by the encoder, availing conditions. The framework converts the sketch into a high-resolution RGB image and classifies them. The whole process work in four different steps: In the first step, the input sketch is converted into a grayscale image. Secondly, the grayscale image is converted into an RGB image with the consideration of facial attributes. In the third step, the RGB image is converted to a high-resolution RGB image using a pixel modifier. The high-resolution RGB image is classified and labelled concerning the relevant class during the fourth step.

We have developed a face dataset for experimental purposes that consists of 1000 face images of 100 people (10 images per person). Each image is preprocessed and distributed into four versions: original RGB image, manually drawn sketches, Grayscale image, and high-resolution image for cross-match analysis. So, as a result, we have developed a fine-tuned state-of-the-art 4000 face image dataset. This dataset comprises 1000 original RGB images with different face positions for extracting spatial facial attributes, 1000 manually drawn sketched images, 1000 grayscale images, and 1000 high-resolution RGB images. These manually generated images and sketches work as training data and ground truth for cross-match analysis to authenticate the proposed model performance.

The key contribution of our research work is as follows:

1. First, we developed a multi-level 3-dimensional conditional generative adversarial network (3D-cGAN) that will colour and translate the sketch into realistic images and preserve spatial facial attributes and pixel-level details.
2. We process the 3D-Convolution and 3D-Deconvolution using vectorization, which trains more attributes and parameters without extra time consumption.
3. We also generate high-resolution RGB colour images from sketches that will be more realistic images.
4. The proposed technique also considers the spatial domain's heterogeneous lightning effect and neighbour feature selection.
5. We introduced a face dataset that consists of 1000 face images with four categories of 100 people (PPFD).
6. This work provides ground truth for each image at multiple stages that authenticate the performance of the proposed architecture using cross-match analysis.

II. RELATED WORK

Face recognition or person identification has been achieved by mutually using soft and hard biometric traits [39]. It is well-known that sketch information and facial attributes give more authentic results than sketch alone. It is due to the non-availability of complementary information in sketches such as skin, eye, hair colour, and ethnicity. Furthermore, other attributes like eyeglasses or wearing a hat would be considered secondary information to narrow down the results. In [40], Klare et al. proposed a direct approach for suspect identification using facial attributes without a sketch. Mittal et al. [41] try to increase the accuracy of their proposed algorithm by fusing multiple sketches and considering soft biometric traits like skin colour, ethnicity, and gender to reorder the ranked list of the suspects. Another framework has been developed by Ouyang et al. [42] to reduce the gap between photo and sketch by combining low-level features with facial attributes. The GANs have been widely used in image generation [7], [24], [29], [43], [44], [45], image translation [46], [47], and image synthesis [12], [13]. Recent literature regarding deep learning approaches [48], [49], [50], [51], [52], [53], [54] focuses more on face recognition and classification problems than classical methods [55], [56], [57]. These approaches can also be used for sketch-photo recognition problems. Face recognition through sketch is more complicated and challenging than classical face recognition problems. The main reason behind this is the heterogeneous nature of photo and sketch modalities and the non-availability of large datasets. For example, most datasets generate only a single sketch per face, making it challenging for a deep model to learn robust features [58]. Another CNN-based work with a new optimization objective function was introduced by Zhang and Lin [29] for end-to-end face image-to-sketch translation. They target the preservation of input image features during translation. Zhu. et al [45] are trying to solve the problem related to the non-availability of paired training data by introducing a new architecture cycle GANs. This GAN network tries translating the input images into target images without using paired training samples. Li et al. [59] proposed a deep CNN model named VGG-Face to overcome facial attribute preservation during translation. This model generates the expected output image based on desired facial attributes. Att-GAN was introduced by Zuo et al. [60] and worked as an attribute classifier and tried to guarantee the generation of correct faces based on desired facial attributes. Recently, conditional GANs networks [43] have greatly emerged in the image generation domain. These networks perform work based on conditions that are given as input. Based on cGAN, Karras et al. [61] introduce a substitute for the generator in GAN networks that can isolate stochastic variation and high-level facial attributes. This generator helps to generate high-quality facial images. In [30], Isola et al. developed a pix2pix architecture of GAN for image colourization, sketch-to-image creation, and semantic segmentation. An improved version of [56] has been proposed by Wang et al. [62], named pix2pixHD. This network demonstrates cGAN application concerning semantic label maps in the image generation domain. Hand-drawn sketches have been colourized by Sangkloy et al. [19] by taking user-centred sparse colour strokes as conditions. Researchers have explored component-based methods [16] for human face image generation by taking high-level features of human faces. Wu and Dai [63] introduced a three-step mechanism for sketch-photo-sketch conversion. They took sketches as input in the first step and then matched them with a face image dataset. The second step colored the sketch with the best-fit face image.

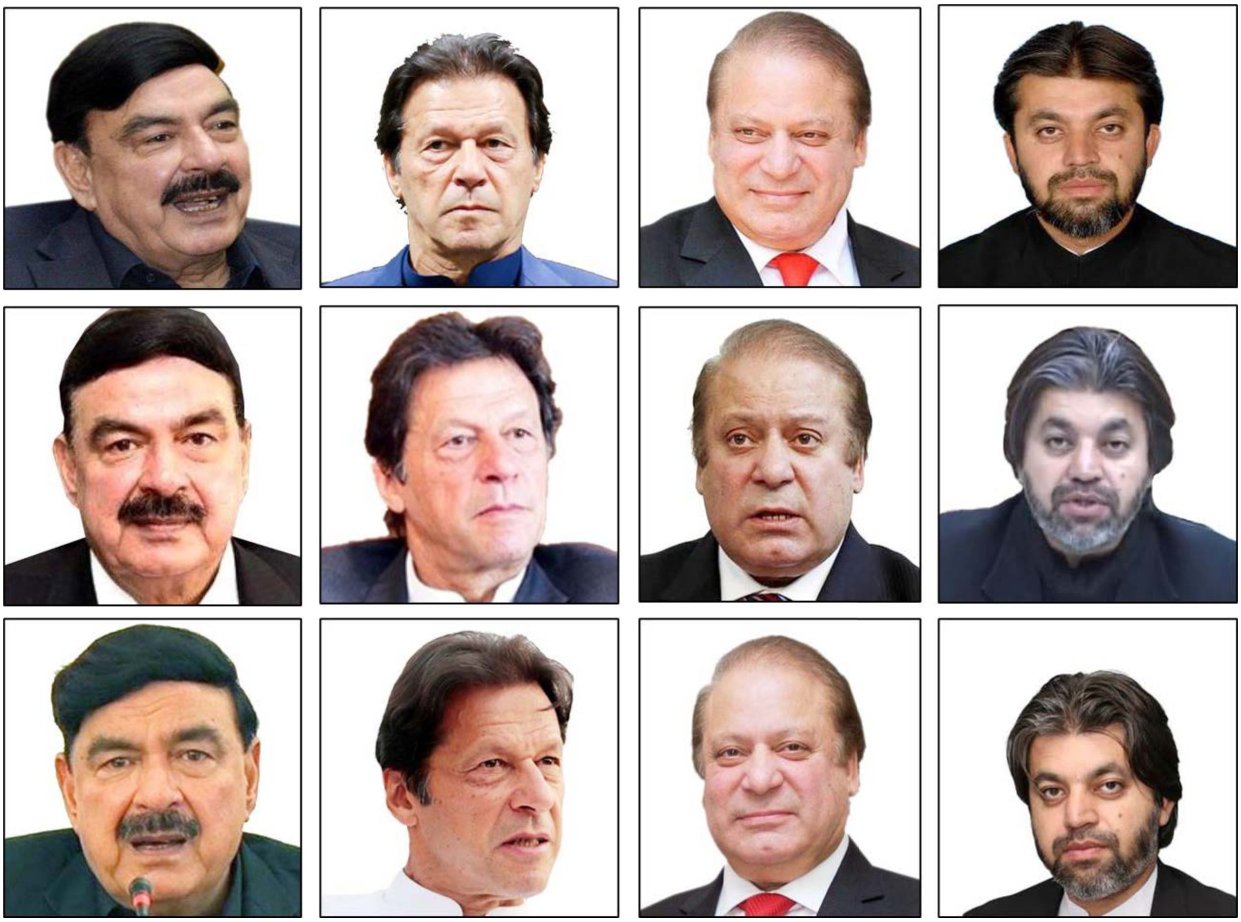


FIGURE 1. Facial images with different positions, angles, and heterogeneous lighting effects.

The sketch is re-drawn from the generated image during the last step to authenticate the output images. The problem with this technique was that it required a well-drawn sketch as input. Gu et al. [59] enabled component-level controllability of facial attributes using auto-encoders with the learning of embedding features from individual face components. They used mask-guided generative networks for the fusion of component feature tensors. cGAN networks have also been used to localize facial images using both facial sketches [8], [64] or semantic label masks [65], [66]. The semantic label mask-based editing is more flexible concerning style transfer and component transfer, while the former approaches give fine and direct control of facial components. To overcome the errors in manually generated sketches, [67] Portenier et al. proposed a conditional completion network that accepts the smooth edge semantic map and input sketch.The attention of recent approaches [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [44], [46], [68], [69] was to balance the targeted and generated images, making the generated image look more realistic and natural. But, less feature selection in the spatial domain reduces these models' effectiveness in heterogeneous lighting effects with different angles of the same face. Pixel-level details of facial attributes were not preserved accurately due to 2D convolutions of loss learning functions.

III. METHODOLOGY/RESEARCH PROCESS

A. DATASET PREPARATION

For experimental purposes, we have developed a face dataset that comprises a total of 100 participants. Each participant collected 10 facial images with different face positions and angles. So, a total of 1000 images were collected. These images with different facial positions and angles along with heterogeneous lighting effects are shown in Figure 1.

After that, a preprocessing phase was initiated, and four versions of each image were generated: original RGB image, manually sketched, grayscale image, and high-resolution image. Sketching and manual enhancement was performed under the supervision of expert artists and photographers. In this way, our fine-tuned facial image dataset is equipped with 4000 images with four categories: 1000 original RGB images, 1000 manual sketches, 1000 grayscale images, and 1000 with super-resolution, as shown in Figure 2.

The original image size is 256 × 256 × 3 with normal quality, high-resolution image 256 × 256 × 3 with high

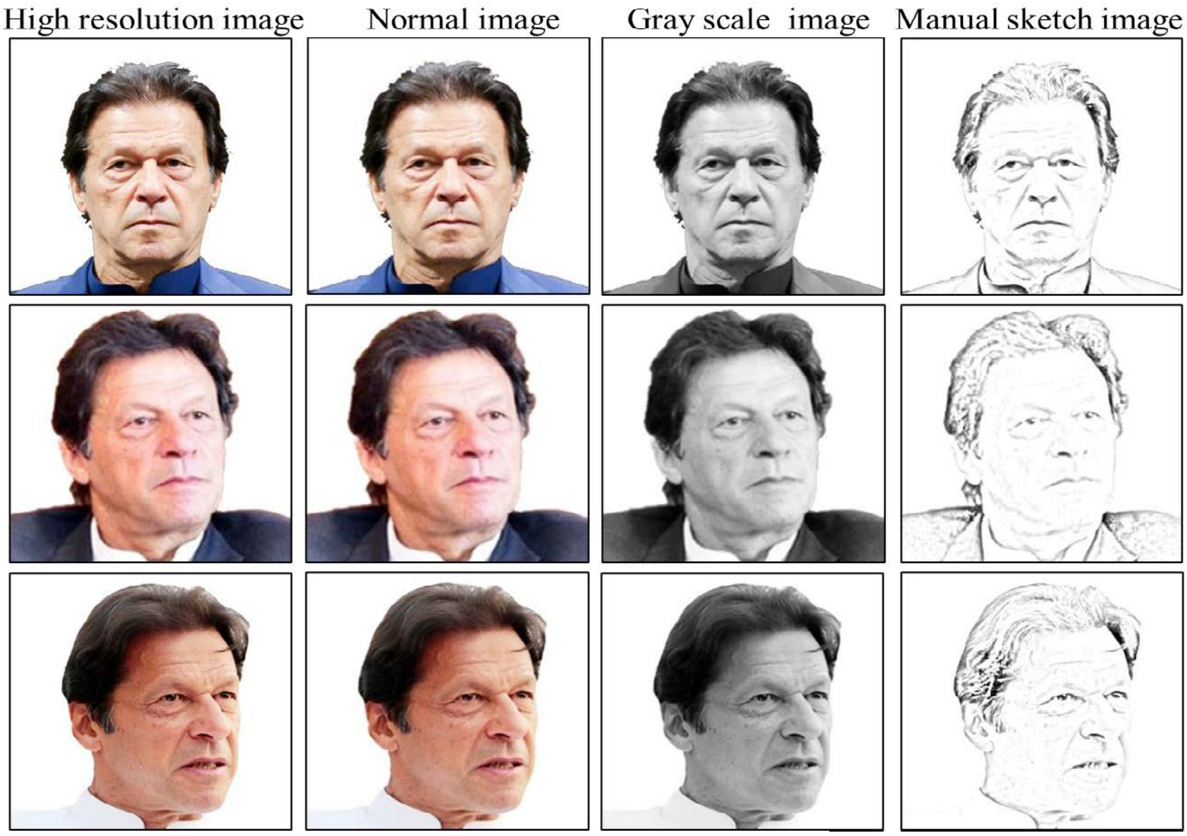


FIGURE 2. Four versions of dataset images. Original, High resolution, Grayscale, and manually sketched images.

quality, grayscale image 256 × 256, and manually sketched image 256 × 256.

B. PROPOSED FRAMEWORK ARCHITECTUR

The proposed model comprises four major phases. The first three phases are cGAN networks that generate images from sketch to high-resolution images step by step. In the proposed framework, every GAN network is a modified form of U-Net [70]. The final output of the first three phases acts as input to the fourth phase for classification and recognition. The fourth phase of the network contains the state-of-the-art CNN network. We re-trained three CNN networks for classification and recognition, i.e., ResNet-50, ResNet-101, and Mobile-Net. Based on the classification and recognition model selection, one of the above networks is selected for the classification and recognition of the input. Figure 3 shows the general framework of the proposed work. In this figure, the sketch image is an input of the first GAN *G1*. It encodes the sketch input and generates a grayscale image. The grayscale output of the *G1* will be the input of the second GAN*G*2. The*G*2 of the framework executed the grayscale image and generated an RGB image as output. The RGB image is given to the third GAN *G3* to generate high-resolution images. Image encoding and decoding processes are completed at this stage, and high-resolution images are passed to the CNN network for classification and recognition.

1) GAN ARCHITECTUR

The *G1*, *G*2 and *G3* have the same architecture for a sketch to grey, grey to RGB and RGB to high-resolution functionality, respectively. Each GAN Network of the proposed framework includes Generator and Discriminator. The internal architecture of the Generator and Discriminator is as follows

a: GENERATOR ARCHITECTURE

The generator of the GAN network consists of two blocks: Encoding and Decoding.

i) ENCODING

The encoding block extracts the features from the input image and encodes them into optimum features. The encoding block of the proposed GAN network's generator consists of eight 3D-encoding sub-blocks. Every sub-encoding block contains convolutional layers with a stride size of 2, Leaky ReLU, and batch normalization. During the preprocessing phase, the input image is resized into 256 × 256 × 3. The generator of the GAN network accepts images with the size of 256 × 256 × 3, and all encoding blocks convert them into 1 × 1 × 512.

ii) DECODING

The decoding block of the generator has seven sub-decoding blocks. The decoding block upsamples the input from

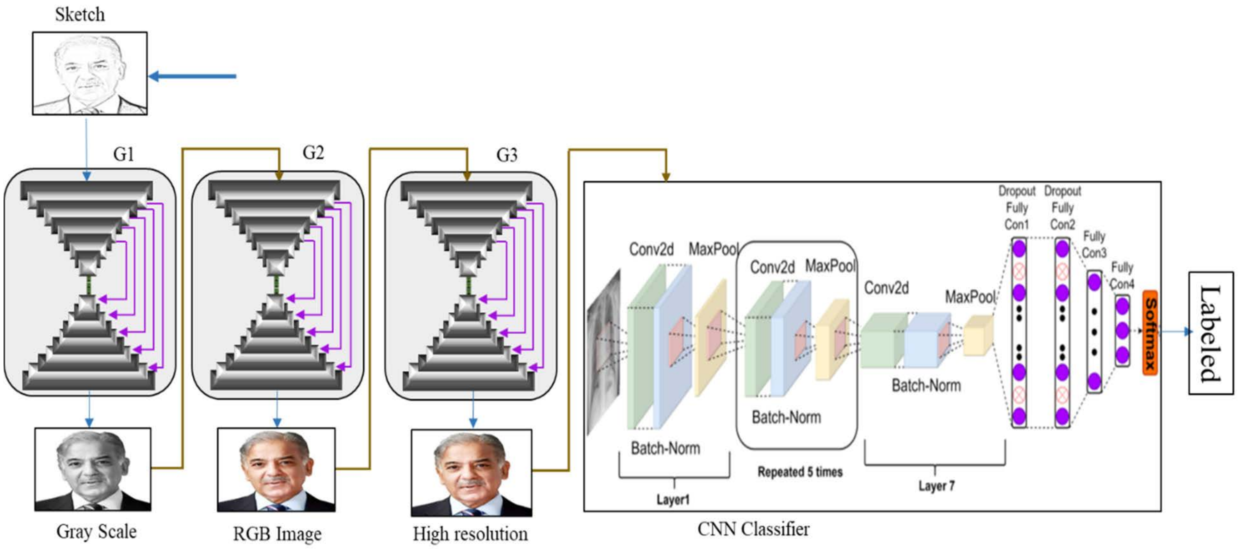


FIGURE 3. Flow chart of the proposed model with three different GAN and CNN classifiers. Here G1, G2 and G3 represent GAN networks.

1 × 1 × 512 to 256 × 256 × 3. The sub-decoding block applies transposed 3D convolution, batch normalization, and ReLU. The dropout of 0.5 was also applied only to the initial three sub-decoding blocks. This dropout was applied after convolution and batch normalization before ReLU to achieve suitable noise removal while maintaining the original face texture and features. Each sub-decoding block gets input from the previous block and the corresponding same-sized sub-encoding block. This approach helps to improve feature selection and texture preservation during encoding and decoding. The decoding part of the generator tries to generate a target image closer to the ground truth. The complete process of the encoding and decoding phase of the proposed cGAN is described inFigure 4.

b: DISCRIMINATOR ARCHITECTURE

The discriminator also comprised seven different sub-blocks. The initial three sub-blocks of the discriminator are similar to that of the sub-encoding blocks of the generator. After three sub-blocks, two separate convolutions are applied with a stride size of 1 for feature purification and preservation of the input. After that, batch normalization and Leaky ReLU are applied. The discriminator received two inputs: 1) an Image generated by the generator and 2) Target Images as ground truth. The primary function of a discriminator is to find discrimination between generated and ground truth images. It finds how much-generated images differ from the ground truth. Finally, the generated image of the discriminator of size 30 × 30 × 1 is used to decide generation quality, as shown inFigure 5.

2) CNN NETWORK

The final phase of the proposed network comprises three pre-trained CNN networks, as shown in Figure 3. 1) ResNet-50, 2) ResNet-101, and 3) Mobile-Net. These networks are used for the classification and recognition of colourized high-resolution images. We adopted the transfer learning technique for the training purpose of these CNN networks.

C. TRAINING DETAILS

1) PARAMETERS USE

The generator of the proposed model used 163,577,577 parameters in all GAN stages. These stages are sketched to grayscale conversion, grayscale to RGB, and RGB to the high-resolution image. The discriminator used a total of 8,311,299 parameters to find the originality and quality of the generated image. The CNN models, i.e., ResNet50, ResNet101 and mobileNet used 197525588, 216552832 and 175393895 parameters, respectively for classification purposes. The complete details of the total, trainable and non-trainable parameters of all stages are given in Table 1.

2) TRAINING PROCESS

The traditional Convolutional Neural Networks (CNNs) cannot explain the spatial relationship between features and the whole image. So, it will lose some of the targe's attribute information, such as direction and posture. To utilize the optimum attributes of the target image, the proposed multi-level 3D GAN applies 3D convolution to encode the input image into vectors as shown inFigure 6. The output vector of the encoder is given as input to the decoder to reconstruct the guided coloured-face image. The 3D convolution and deconvolution guide the *G1*, *G*2 and *G3* to use additional features and attributes for encoding and decoding. This helps to preserve the direction, postures of targeted image attributes and special relationships among the whole image's features. The process of 3D-Convolution and 3D-Deconvolution is

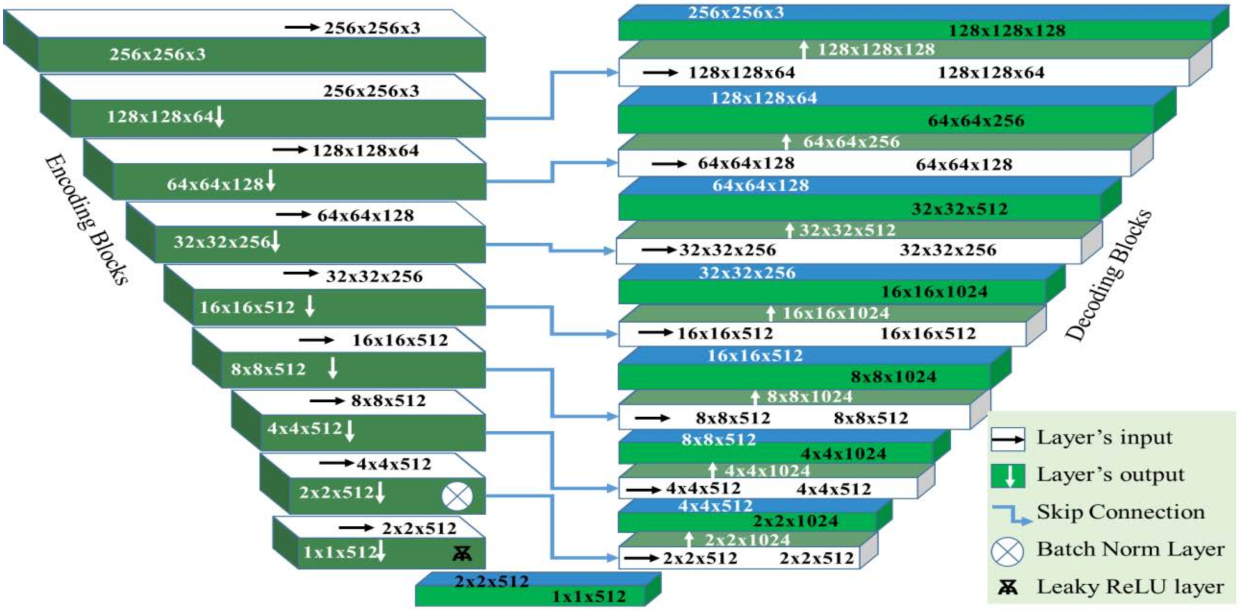


FIGURE 4. Internal structure of GAN-generator. The generator comprises two blocks, i.e., encoding and decoding. Encoding encodes the input and reduces it to minimum features, while decoding expands it and generates the output image.

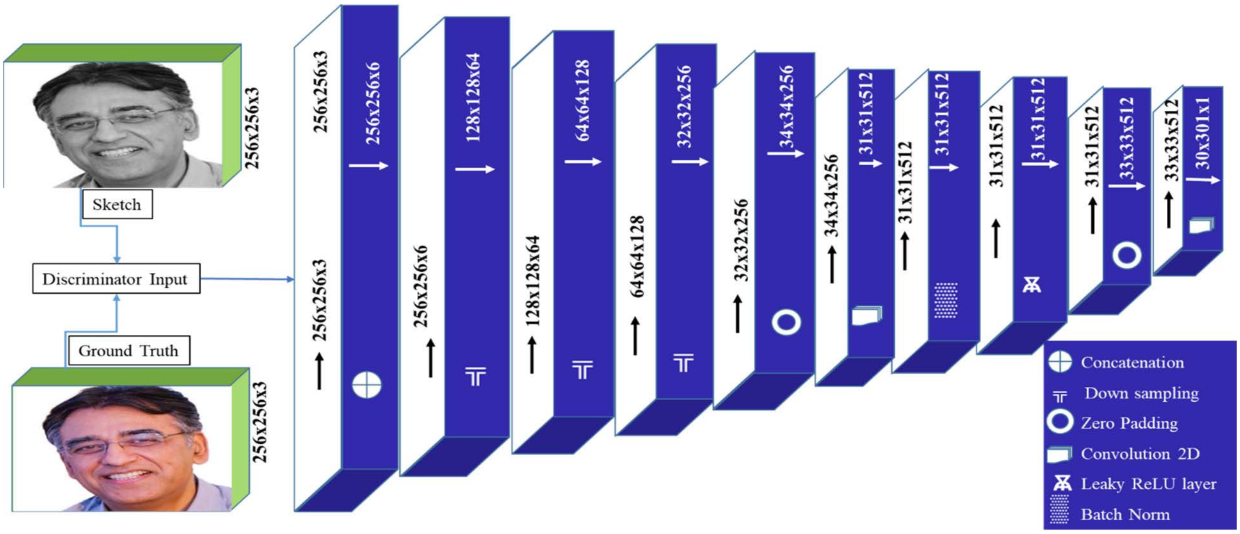


FIGURE 5. Discriminator received two images as input, i.e., original ground truth images and generated sketch images, and compared how much generated image differs from the ground truth.

handled by a vectorization process, that takes the same processing time as 2D-Convolution but extracts more features and texture information.

Generally, GAN networks generate the final image y from scratch, i.e., random noise vector z, $G : z \rightarrow y$ [43]. In the case of the proposed GAN, the network gets two inputs, i.e., random noise vector z and conditional vector x as sketch image, to construct the final output image y, $G : x, z \rightarrow y$.

The discriminator D trained adversarially to differentiate the generated vs. real images. The generator generates good-quality images indistinguishable from natural images as long as the generator is trained. The training mechanism of the proposed GAN Network is demonstrated inFigure 7. We used the Adam stochastic gradient descent [71] mechanism to train for the optimal learning rate of 0.0002 and momentum estimates as $\beta 1$ and $\beta 2$ as 0.5 and 0.999, respectively. The learning rate was reduced to 0.00001 after 150 epochs for fine-tuning model weights.

Before training, preprocessing phases were initiated to resize the image according to the underlying framework. Then, images are randomly cropped to the target size with horizontal flipping. For the training of the proposed GAN, the $G1$ trained on 250 epochs for Sketch to Gray transformation. The $G2$ trained on 300 epochs for grey to RGB image, while

TABLE 1. Trainable and non-trainable parameters of the proposed framework.

Parameters	Total		Trainable		Non-Trainable	
	G	D	G	D	G	D
Sketch to Gray (*G1*)	54525859	2770433	54414979	2768641	10880	1792
Gray to RGB (*G2*)	54525859	2770433	54414979	2768641	10880	1792
RGB to Sup-Res (*G2*)	54525859	2770433	54414979	2768641	10880	1792
ResNet-50	25636712		25583592		53120	
ResNet-101	44663956		44571411		92545	
Mobile-Net	3505019		3497757		7262	
Total Parameters GAN and Classification Network						
G1, G2, G3 and ResNet-50	197525588		197134452		91136	
G1, G2, G3 and ResNet-101	216552832		216122271		130561	
G1, G2, G3 and Mobile-Net	175393895		175048617		45278	

450 epochs were used to train *G3* to transform the RGB to a high-resolution image. *G3* was trained on 450 epochs due to texture enhancement and feature improvement.

The batch size was set to 1 for all three GANs networks. The proposed GAN took 120 minutes for *G*1 and 145 minutes for *G*2, and 230 minutes for *G*3 on P100s GPU for training. At the same time, every GAN needs approximately 0.35 sec to transform the input to output using the same GPU. So, the proposed GAN model needs only 1.25 sec to convert the sketch into a super-resolution coloured image. For classification, the input size of ResNet-50, ResNet-101, and Mobilenetv2 is 224 × 224 × 3. For the training of CNN networks, we used random crops of 224 × 224 × 3 from high-resolution images. Resnet-50 and Resnet-101 models were trained on 45 epochs with a batch size of 128 and a learning rate of 0.0001. While the batch size of 128 and the learning rate of 0.0001 was also set for Mobilenetv2 with 70 epochs.

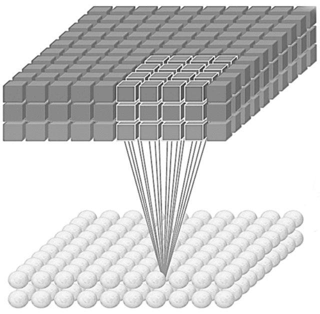


FIGURE 6. Three dimensional convolutional mechanism.

IV. RESULTS AND DISCUSSION

A. IMAGE GENERATION

The whole proposed multi-level GAN network generates sketches to high-resolution images in three phases. Three GAN networks are incorporated with each other to generate high-resolution images. During the training process, each GAN network was trained independently. The detail of each phase for generating images is given below.

Phase-I: The first GAN (*G1*) input is a sketch image, shown in Figure 3. For training purposes, we use the proposed PPFD dataset. The train-test ratio for *G1* was 7:3. A total of 500 epochs were carried out with a conditional training procedure on 1400 images.

These 1400 images include 700 sketches and 700 grayscale images. The output of *G1* was a grayscale image, as shown in Figure 8. As the training starts, *G1* generates a noisy and blurry image. But the noise is removed gradually as the number of epochs increases. At epoch no 210, the generated picture is more precise, and at epoch no 300, the generated image is more likely to ground truth, as shown in Figure 8.

Phase-II: Phase-II GAN network (*G2*) received 700 Grayscale images and 700 coloured images for training. The training of *G2* comprises 420 epochs with a learning rate of 0.0002. the output of *G2* was a coloured image, as shown in Figure 9. Initially, at epoch no 15, the generated image shows a blurry pattern, but at epoch no 110, the image looks more realistic.

As the execution goes ahead, from epoch no 210 to 300, the facial expression of the generated image shows a more realistic pattern than the ground truth.

Phase-III: The phase-II output is the normal coloured image. To generate a high-resolution image with a more realistic facial attribute, we incorporated *G3* with this network. *G3* network enhances regions with blurry and nosy patterns to convert normal images into high resolution using conditional attributes. The visual results of *G3* on epochs 7, 111, 285, and 390 is shown in Figure 10.

1) GAN EVALUATION

The generator generates output against every input, then the discriminator evaluates the Input image and generates the

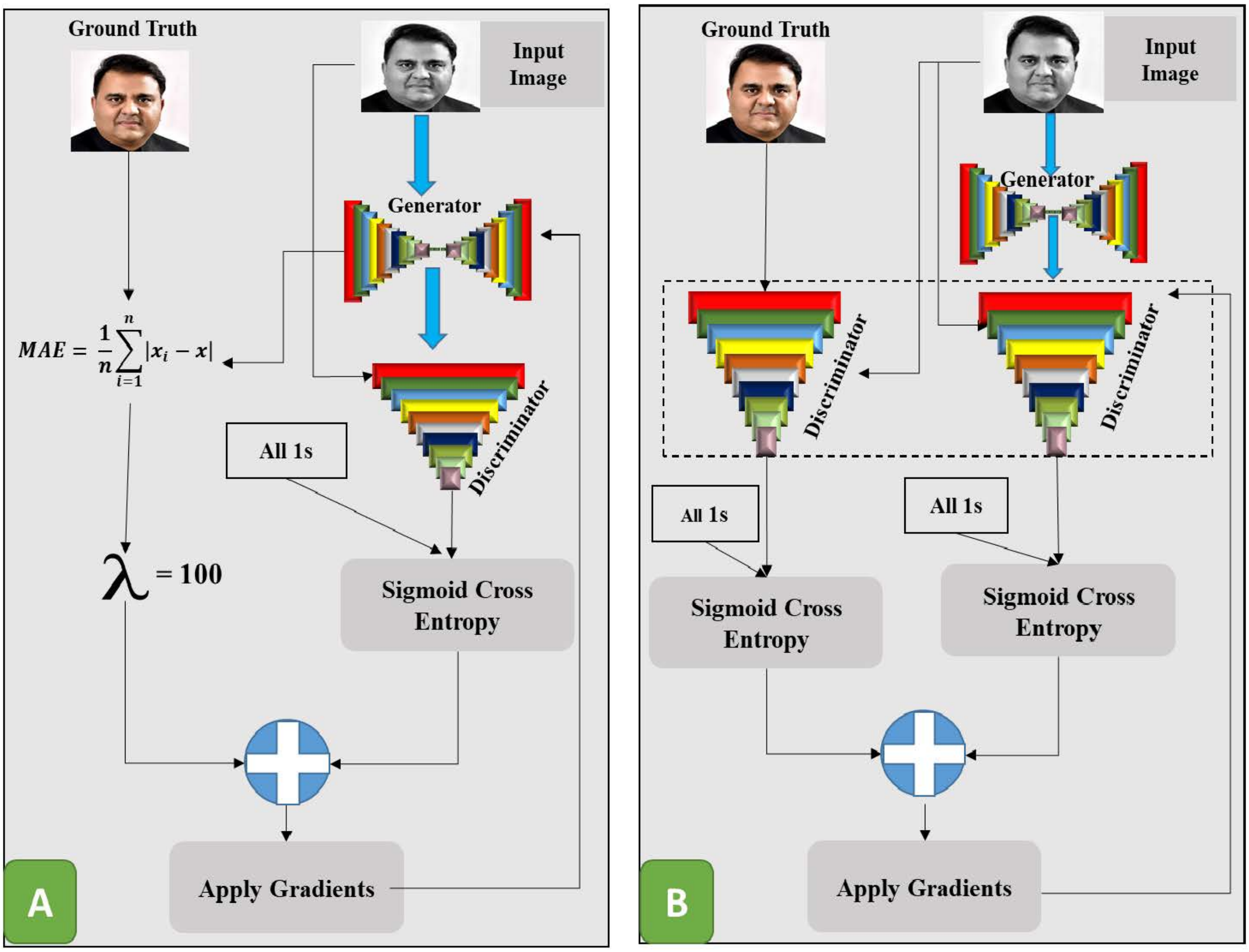


FIGURE 7. Training Mechanism of GAN Network A: Generator Training, B: Discriminator Training.

image in the first step. The discriminator evaluates the Input image and the targeted image during the second step. Finally, generator and discriminator losses were calculated with the gradient loss of the generator and discriminator's input. In this way, all the results were optimized.

a: GENERATOR LOSS

Generator loss is the sigmoid cross-entropy of generated images and an array of ones. Generator loss includes *L1*, which means absolute error *(MAE)* between the generated image and the target image. *L1* loss helps the generator to generate an image more realistic to the target image. Total generator loss is evaluated by equation 1.

$$Total\ Generator\ loss = gen_{loss} + \lambda * L1_{loss}$$
$$Here, \lambda = 100 \qquad (1)$$

Initially, the highest generator losses of the proposed model regarding $G1$, $G2$, and G3 were calculated as 3.31, 5.47, and 4.33, respectively, as shown in Figure 11. These losses approach zero by improving the accuracy with an increase in the number of epochs. To fool the discriminator, the loss function of the generator tries to improve the generated images near the ground truth. As the number of epochs increases, the learning proficiency increases, and generator loss decreases to 0.35, 0.06, and 0.02 for $G1$, $G2$, and $G3$, respectively, as shown inFigure 11. The proposed model achieved the highest training results at 300 epochs regarding $G1$ and $G2$. While for $G3$, the highest training results were achieved at epoch 400 due to the generation of texture details and high-resolution facial attributes.

b: DISCRIMINATOR LOSS

Two inputs were given to the discriminator loss function: 1) Real image and 2) Generated Image. A combination of sigmoid cross-entropy loss of real image and an array of ones were used for finding real loss. At the same time, the generated loss is the sum of the sigmoid cross-entropy loss of the generated images and an array of zeros. So, the total discriminator loss (L_{DT}) is calculated by the sum of real loss (LR) and generated loss (LG) as shown in equation 2.

$$L_{DT} = L_G + L_R \qquad (2)$$

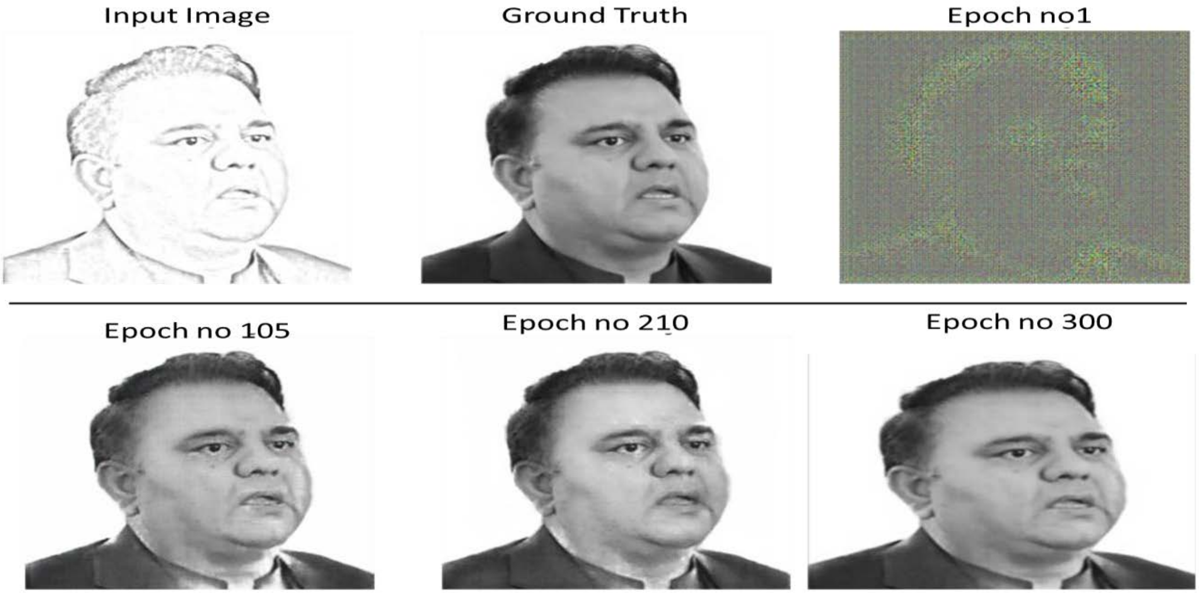


FIGURE 8. G1 Output results of training on different epochs.

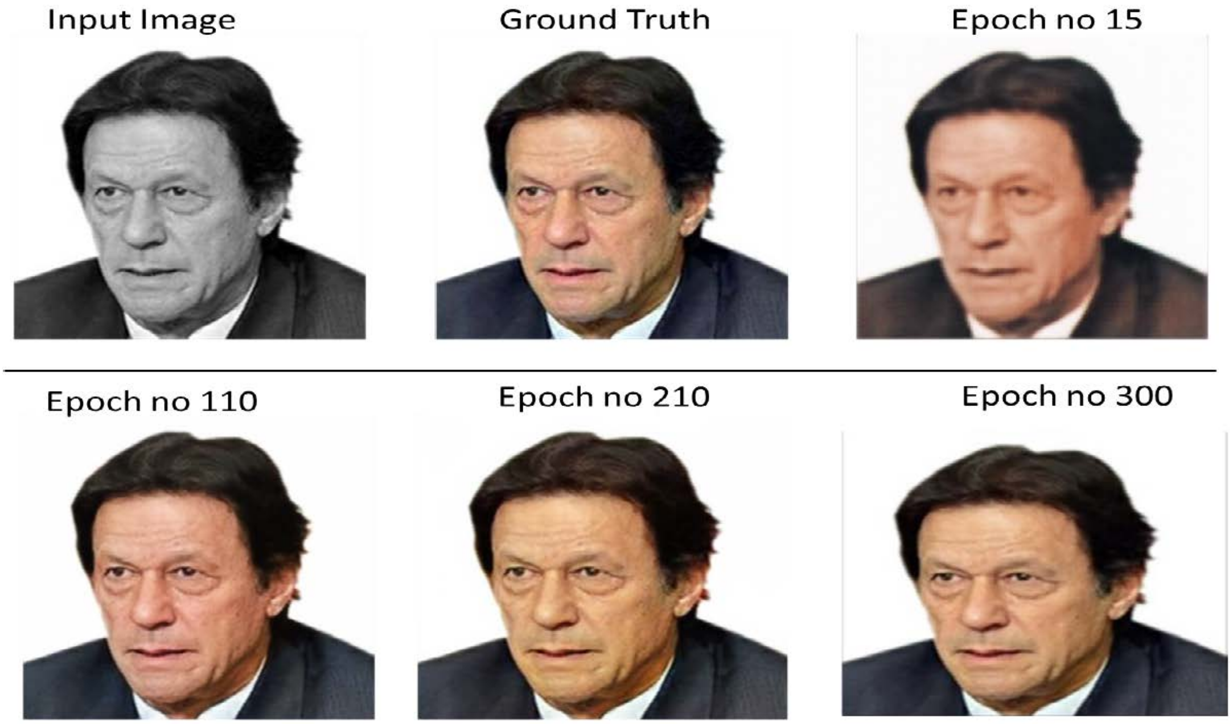
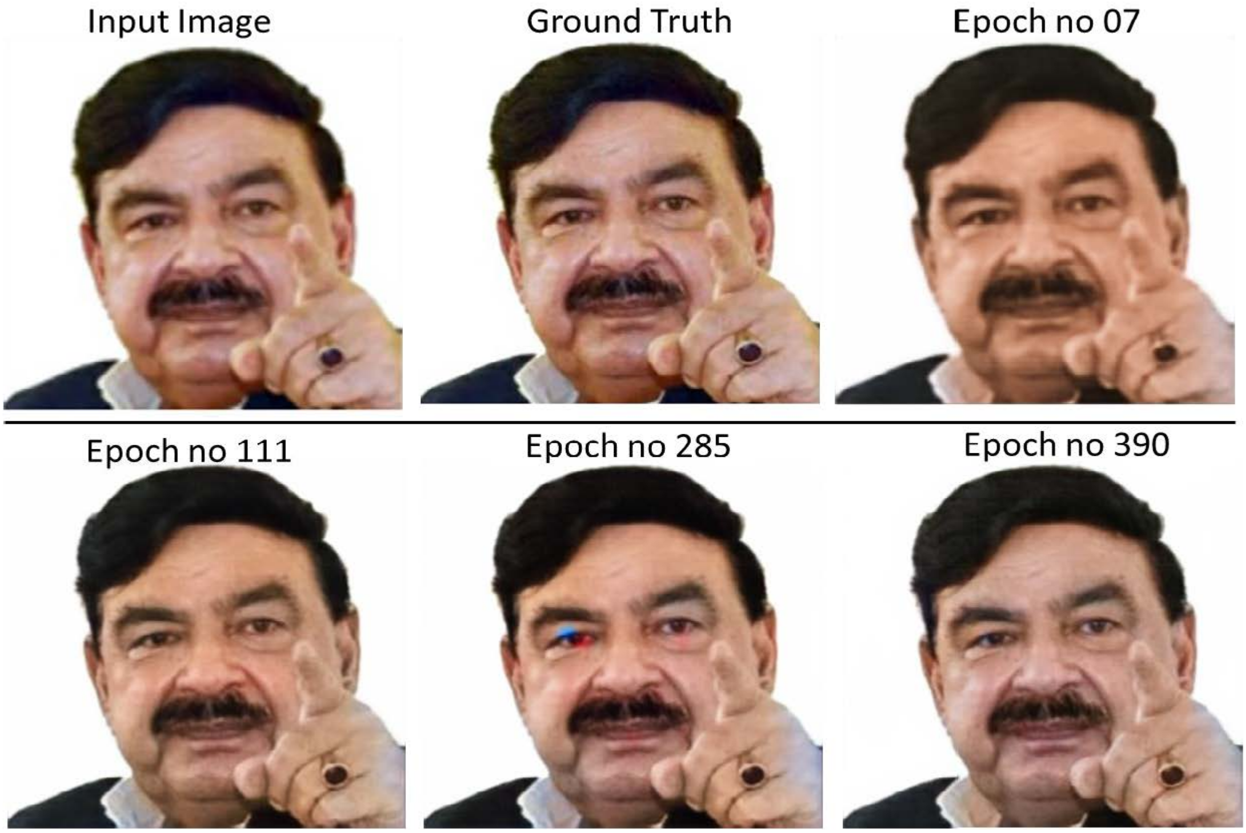


FIGURE 9. Output results of G2 at different epochs.

L_{DT} curves of *G1*, *G*2, and *G3* decrease from 1.42, 1.65, and 1.42, respectively, as shown inFigure 11. The trend line decreases as the number of epochs increases. The graph behaviour reveals that initially, the discriminator beat the generator and classified the generated image as fake. But as the learning of the generator increases up to 150 epochs, the generator tries to generate a realistic image. Finally, the losses decrease after 220 epochs, and the generated image looks more realistic and near the ground truth.

2) IMAGE QUALITY EVALUATION

To evaluate the generated image quality, we have used SNR, PSNR, and SSIM matrices.



Input Image

Ground Truth

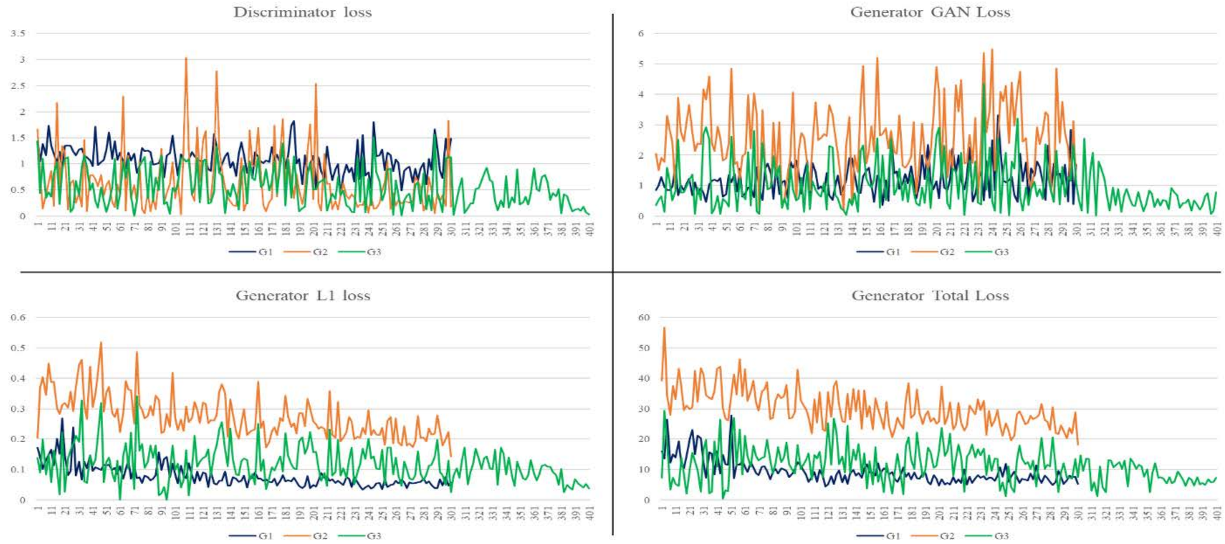
Epoch no 07

Epoch no 111

Epoch no 285

Epoch no 390

FIGURE 10. Visual output results of G3 at different epochs.



Discriminator loss

Generator GAN Loss

Generator L1 loss

Generator Total Loss

FIGURE 11. Discriminator, Generator GAN, Generator L1 and Generator Total losses of G1, G2 and G3.

a: SNR

Signal-to-noise ratio (SNR) is used in imaging to characterize image quality. The sensitivity of a (digital or film) imaging system is typically described as the signal level that yields a threshold level of SNR, as shown in equation 3.

$$SNR = 10log_{10}\left[\frac{\sum_{j=1}^{M}\sum_{k=1}^{N}\left(x_{j,k}\right)^{2}}{\sum_{j=1}^{M}\sum_{k=1}^{N}\left(x_{j,k}-x_{j,k}^{'}\right)^{2}}\right] \tag{3}$$

TABLE 2. Image quality Parameters of SNR, PSNR and SSIM.

Parameters	Images	SNR	PSNR	SSIM
G1	Included	26.13166	29.86521	0.925196
G2		38.17585	41.18612	0.989155
G3		39.80814	42.79287	0.994087
G1	Excluded	21.39811	26.11345	0.908497
G2		37.4459	40.2339	0.987543
G3		34.02675	37.85558	0.966104

We have calculated SNR on two different types of images. 1) Images that were included during training ($I_{include}$), and 2) Images not included during training ($I_{exclude}$). The SNR of $G1$, $G2$, and $G3$ on $I_{include}$ was calculated as 26.13, 38.17, and 39.80, respectively. While on $I_{exclude}$ images was 21.39, 37.44, and 34.02, respectively, shown in Table 2.

b: PSNR

SPNR is the ratio between the maximum Signal's power (Original target image) and the power of the noisy Signal (Generated image). To find the quality of the generated image based on pixels, peak Signal to noise ratio (PSNR) metrics were used. PSNR is formulated as in equation 4.

$$PSNR = 10 \log \frac{(2^n - 1)^2}{MSE}$$
$$PSNR = 10 \log \frac{(255)^2}{MSE} \quad (4)$$

Here

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (g_i - p_i)^2$$

MSE is the Mean square error between g and p. Here "p" is the predicted or newly generated image while "g" is the ground truth image. The $PSNR$ of $G1$, $G2$, and $G3$ on $I_{include}$ was 29.86, 41.18, and 42.79, respectively. On the other hand, PSNR for $I_{exclude}$ was calculated as 26.11, 40.23, and 37.85, as shown in Table 2.

c: SSIM

SSIM is abbreviated as structural similarity index measure. It measures the image statistics of the sliding window. The formula for $SSIM$ is derived as follows.

Let x = {xi|i=1, 2,……N} and y={yi|i=1,2,…….N}. The proposed quality index is defined as in equation 5.

$$Q = \frac{4\sigma_{xy} x^{/} y^{/}}{\left(\sigma_x^2 + \sigma_y^2\right) \left|\left(x^{/}\right)^2 + \left(y^{/}\right)^2\right|}$$

$$here, x^{/} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$y^{/} = \frac{1}{N} \sum_{i=1}^{N} y_i$$
$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(x_i - x^{/}\right)^2$$
$$\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(y_i - y^{/}\right)^2$$
$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} \left(x_i - x^{/}\right)\left(y_i - y^{/}\right) \quad (5)$$

The proposed multi-GAN network attains the $I_{include}$ SSIM value as 0.9251, 0.9891, and 0.9940 for $G1$, $G2$, and $G2$. However, $I_{exclude}$ SSIM was calculated as 0.9084, 0.9875, and 0.9661 for G1, G2, and G2 as shown in Table 2.

The comparative analysis of the proposed multi-level 3D-GAN with the existing GAN network is given in Table 3. The results reveal that the proposed model outperforms SSIM, SNR and $PSNR$. The proposed model outperforms due to the usage of multi-level GAN with 3-dimensional convolutions and deconvolution. 3D convolution extract optimal features and attributes with each pixel's direction and position. The relation among the attribute is also preserved, which helps decode the image with each object's actual position and relation.

B. IMAGE CLASSIFICATION

We proposed a multi-level GAN network with three phases to generate a high-resolution image from a sketch.

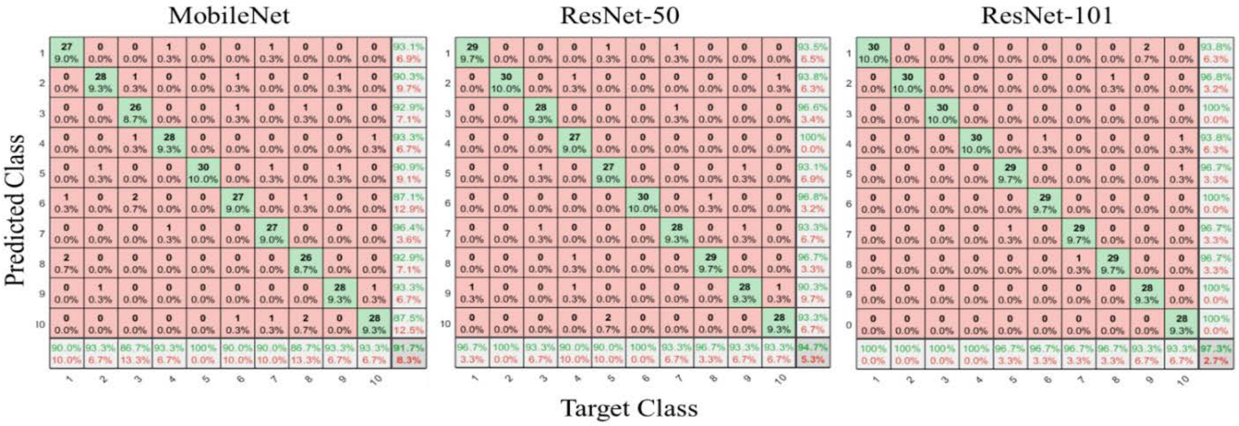
The CNN classifier is used to recognize the generated high-resolution image using the transfer learning technique with the help of three pre-trained models, i.e., ResNet50, ResNet101, and MobileNetV2.

The original high-resolution images were used for the training of these models. For testing purposes, we used the generated high-resolution images.

The confusion matrix is the most common and comprehensive way to represent classification evaluation. The confusion matrix includes four classes: 1) True Positive (TP), 2) True Negative (TN), 3) False Positive (FP), and

TABLE 3. Comparison of proposed 3D GAN with existing GAN networks.

Model	SNR	PSNR	SSIM
CAE-cGAN [72]	16.20	20.06	0.85
2D-FSRCNN [73]	26.38	31.55	0.88
3D-FSRCNN [74]	29.27	33.86	0.91
SRGAN [75]	28.63	33.48	0.90
FSCWRN [76]	26.47	30.96	0.90
CSN [77]	26.70	31.23	0.90
MI-SR-GAN [78]	35.04	38.83	0.95
Conditional GAN [8]	21.22	25.97	0.86
Cycle GAN [79]	19.86	24.89	0.84
SpA GAN [80]	25.89	29.30	0.93
Proposed ML-3D-GAN	**39.80**	**42.79**	**0.99**



MobileNet

	1	2	3	4	5	6	7	8	9	10	
1	27 9.0%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	93.1% 6.9%
2	0 0.0%	28 9.3%	1 0.3%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	90.3% 9.7%
3	0 0.0%	0 0.0%	26 8.7%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	92.9% 7.1%
4	0 0.0%	0 0.0%	1 0.3%	28 9.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	93.3% 6.7%
5	0 0.0%	1 0.3%	0 0.0%	0 0.0%	30 10.0%	0 0.0%	1 0.3%	0 0.0%	1 0.3%	0 0.0%	90.9% 9.1%
6	1 0.3%	0 0.0%	2 0.7%	0 0.0%	0 0.0%	27 9.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	87.1% 12.9%
7	0 0.0%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	27 9.0%	0 0.0%	0 0.0%	0 0.0%	96.4% 3.6%
8	2 0.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	26 8.7%	0 0.0%	0 0.0%	92.9% 7.1%
9	0 0.0%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	28 9.3%	1 0.3%	93.3% 6.7%
10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	1 0.3%	2 0.7%	0 0.0%	28 9.3%	87.5% 12.5%
	90.0% 10.0%	93.3% 6.7%	86.7% 13.3%	93.3% 6.7%	100% 0.0%	90.0% 10.0%	90.0% 10.0%	86.7% 13.3%	93.3% 6.7%	93.3% 6.7%	91.7% 8.3%

ResNet-50

	1	2	3	4	5	6	7	8	9	10	
1	29 9.7%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	93.5% 6.5%
2	0 0.0%	30 10.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	93.8% 6.3%
3	0 0.0%	0 0.0%	28 9.3%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	96.6% 3.4%
4	0 0.0%	0 0.0%	0 0.0%	27 9.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	1 0.3%	0 0.0%	27 9.0%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	93.1% 6.9%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	30 10.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	96.8% 3.2%
7	0 0.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	28 9.3%	0 0.0%	1 0.3%	0 0.0%	93.3% 6.7%
8	0 0.0%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	29 9.7%	0 0.0%	0 0.0%	96.7% 3.3%
9	1 0.3%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	28 9.3%	1 0.3%	90.3% 9.7%
10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	28 9.3%	93.3% 6.7%
	96.7% 3.3%	100% 0.0%	93.3% 6.7%	90.0% 10.0%	90.0% 10.0%	100% 0.0%	93.3% 6.7%	96.7% 3.3%	93.3% 6.7%	93.3% 6.7%	94.7% 5.3%

ResNet-101

	1	2	3	4	5	6	7	8	9	10	
1	30 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.7%	0 0.0%	93.8% 6.3%
2	0 0.0%	30 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	96.8% 3.2%
3	0 0.0%	0 0.0%	30 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	30 10.0%	0 0.0%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	93.8% 6.3%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	29 9.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	96.7% 3.3%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	29 9.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	29 9.7%	0 0.0%	0 0.0%	0 0.0%	96.7% 3.3%
8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	29 9.7%	0 0.0%	0 0.0%	96.7% 3.3%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	28 9.3%	0 0.0%	100% 0.0%
0	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	28 9.3%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	96.7% 3.3%	96.7% 3.3%	96.7% 3.3%	96.7% 3.3%	93.3% 6.7%	93.3% 6.7%	97.3% 2.7%

FIGURE 12. Confusion matrix of generated Image recognition.

TABLE 4. Classification results of MobileNet, ResNet-50 and ResNet-101.

Parameters	MobileNet	ResNet 50	ResNet 101
Accuracy	0.9167	0.9467	0.9733
Sensitivity/ Recall	0.9167	0.9467	0.9733
Specificity	0.9907	0.9941	0.997
Precision	0.9177	0.9474	0.9743
F1_score	0.9165	0.9465	0.9733

4) False Negative (FN). The confusion matrix of ResNet50, ResNet101, and MobileNetV2 is given in Figure 12.

The other classification evaluation matrices used to evaluate the proposed work includes accuracy, precision, recall and F1-score.

1) ACCURACY

Accuracy is used to find how much the proposed model produces accurate results. Accuracy is the ratio of correctly classified images and the total number of images evaluated. The accuracy of the proposed model is calculated by equation 6.

$$Accuracy = \frac{(TP + TN)}{(TN + TP + FP + FN)} \tag{6}$$

The ResNet-50, ResNet-101, and Mobile-Net achieved outstanding accuracies as 97.394.7% and 91.7% respectively.

2) PRECISION

Precision is the formulation of finding how many values are positive that are predicted as positive. It is beneficial when we have labelled data about our predictions. The formula for precision is given in equation 7.

$$Precision = \frac{TP}{(TP + FP)} \tag{7}$$

Due to deeper structural architecture, the proposed model got a high precision value in the case of ResNet-50, ResNet-101 than Mobile-Net. The precision of MobileNet, ResNet-50 and ResNet-101was 91.77%, 94.74% and 97.43%.

3) RECALL

Another way to evaluate the classification is recall. It helps us to find the ratio between correctly classified values as positives over the total values that are positives. The recall is formulated in equation 8.

$$Recall = \frac{TP}{(TP + FN)} \tag{8}$$

The global recall value of MobileNet, ResNet-50 and ResNet-101 calculated as 91.67%, 94.67% and 97.33%, respectively.

4) F1 SCORE

The overall picture of precision and recall can be calculated with the F1-Score. It gives the harmonic mean of recall and precision. The formula for F1-Score is given in equation 9.

$$F1_{score} = \frac{2 * (precision - recall)}{(precision + recall)} \tag{9}$$

ResNet-101 achieved the highest value for F1Score at 0.9733. However, ResNet-50 and Mobile-Net did not play better than ResNet-101 and got the F1Score of 0.9465 and 0.9165.

5) SPECIFICITY

It helps us to find the ratio between wrongly classified values as negative over the total values that are negative. The specificity is formulated in equation 10.

$$Specificity = \frac{TN}{(FP + TN)} \tag{10}$$

ResNet-101 outperformed in respect of specificity and got 99.70%. While ResNet-50 achieved 99.41MobileNet 99.07%. The values of evaluation matrices are also shown in Table 4.

C. ADVANTAGES AND LIMITATION

This approach has the following advantages over existing techniques.

- The proposed conditional GAN can perform work in 3-phases, i.e., sketch to colour and then high-resolution RGB image.
- The proposed framework used 3D-Convolution and 3D-Deconvolutional processes using vectorization.
- The proposed 3D-cGAN can translate sketches into more realistic images by preserving more spatial facial attributes and pixel-level information while using the same processing time as conventional 2D-Convolution.
- We have also developed a state-of-the-art facial PPFD dataset that contains 4000 images with four distinct categories along with a heterogeneous, multi-color, and different Luminus effect.
- Despite this, the proposed 3D-cGAN cannot generate full high-definition like 1024 × 1024 and more images due to the limited computational resources and complexity of convolutional neural networks.

V. CONCLUSION

This work proposed a framework with a multi-level 3D cGAN network to generate high-resolution images from sketches along with a classification network to recognize the image. We developed a state-of-the-art PPF dataset that comprises 4000 images collected from 100 people for experimental purposes. We have also generated the ground truth of each image to authenticate the proposed framework model results. The framework integrated three conditional cGAN networks for sketch-to-image generation, followed by pre-trained ResNet-50, ResNet-101, and Mobile-Net for classification. We use the 3D-Convolutional process for all GANs using vectorization, which extracts more features and texture information from images while using the same computational cost as 2D-Convolution. We used Adam's stochastic gradient descent mechanism to achieve the optimal results with a learning rate of 0.0002 and momentum estimates $\beta 1$ and $\beta 2$ as 0.5 and 0.999, respectively, during training. Multiple statistical measures were considered to authenticate the performance of the proposed framework. The framework got 97.33% accuracy with 99% image structure similarity index measure with high SNR and PSNR.

In the future, we will enhance the quality of the generated image using fewer parameters so that high-quality image generation may become possible with low-processing devices. We also try to generate images with the help of textual data.

VI. FUNDING

No funding is available for this study

VII. AUTHORSHIP CONTRIBUTION

Zakir Khan: Conceptualization; Methodology; Formal analysis; Data curation; Code Execution, Data Collection **Arif Iqbal Umar**: Supervision and reviewing. **Syed Hamad Shirazi:** Reviewing, Editing, Data Analysis. **Muhammad Shahzad**: Writing - Original Draft, Methodology, Writing, review & editing, Dataset Preparation, Data curation.

VIII. COMPETING INTEREST

All authors declare no conflict of interest.

IX. AUTHORS APPROVAL

All authors have read and approved the final manuscript.

REFERENCES

[1] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The MegaFace benchmark: 1 million faces for recognition at scale," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4873–4882, doi: 10.1109/CVPR.2016.527.

[2] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A. K. Jain, J. A. Duncan, K. Allen, J. Cheney, and P. Grother, "IARPA Janus benchmark-B face dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 592–600, doi: 10.1109/CVPRW.2017.87.

[3] M. Hu and J. Guo, "Facial attribute-controlled sketch-to-image translation with generative adversarial networks," Tech. Rep., 2020, vol. 5.

[4] Y. S. Ramya, S. Ghosh, M. Vatsa, and R. Singh, "Face sketch colorization via supervised GANs," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–6.

[5] Y. Zheng, H. Yao, X. Sun, S. Zhang, S. Zhao, and F. Porikli, "Sketch-specific data augmentation for freehand sketch recognition," *Neurocomputing*, vol. 456, pp. 528–539, Oct. 2021, doi: 10.1016/j.neucom.2020.05.124.

[6] A. Brock, T. Lim, J. M. Ritchie, and N. Weston, "Neural photo editing with introspective adversarial networks," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–15, doi: 10.5281/zenodo.807638.

[7] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," Tech. Rep., 2017, p. 20.

[8] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[9] R. Liu, X. Wang, H. Lu, Z. Wu, Q. Fan, S. Li, and X. Jin, "SCCGAN: Style and characters inpainting based on CGAN," *Mobile Netw. Appl.*, vol. 26, no. 1, pp. 3–12, Feb. 2021, doi: 10.1007/s11036-020-01717-x.

[10] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," in *Proc. ICML*, vol. 4, 2016, pp. 2341–2349.

[11] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez, "Invertible conditional GANs for image editing," 2016, *arXiv:1611.06355*.

[12] M. Li, W. Zuo, and D. Zhang, "Deep identity-aware transfer of facial attributes," 2016, *arXiv:1610.05586*.

[13] W. Shen and R. Liu, "Learning residual images for face attribute manipulation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1225–1233, doi: 10.1109/CVPR.2017.135.

[14] M. Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 701–709, 2017.

[15] S. Zhou, T. Xiao, Y. Yang, D. Feng, Q. He, and W. He, "GeneGAN: Learning object transfiguration and object subspace from unpaired data," in *Proc. Brit. Mach. Vis. Conf.*, 2017, pp. 1–13, doi: 10.5244/c.31.111.

[16] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, and M. Ranzato, "Fader networks: Manipulating images by sliding attributes," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 5968–5977, 2017.

[17] T. Kim, B. Kim, M. Cha, and J. Kim, "Unsupervised visual attribute transfer with reconfigurable generative adversarial networks," 2017, *arXiv:1707.09798*.

[18] T. Xiao, J. Hong, and J. Ma, "DNA-GAN: Learning disentangled representations from multi-attribute images," in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–14.

[19] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6836–6845, doi: 10.1109/CVPR.2017.723.

[20] M. Zhang, J. Li, W. Wang, and X. Gao, "Compositional model-based sketch generator in facial entertainment," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 904–915, Mar. 2018, doi: 10.1109/TCYB.2017.2664499.

[21] M. Zhang, R. Wang, X. Gao, J. Li, and D. Tao, "Dual-transfer face sketch-photo synthesis," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 642–657, Feb. 2019, doi: 10.1109/TIP.2018.2869688.

[22] M. Zhang, N. Wang, Y. Li, and X. Gao, "Bionic face sketch generator," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2701–2714, Jun. 2020, doi: 10.1109/TCYB.2019.2924589.

[23] M. Zhang, N. Wang, Y. Li, and X. Gao, "Neural probabilistic graphical model for face sketch synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 7, pp. 2623–2637, Jul. 2020, doi: 10.1109/TNNLS.2019.2933590.

[24] M. Zhang, Y. Li, N. Wang, Y. Chi, and X. Gao, "Cascaded face sketch synthesis under various illuminations," *IEEE Trans. Image Process.*, vol. 29, pp. 1507–1521, 2020.

[25] N. Wang, X. Gao, and J. Li, "Random sampling for fast face sketch synthesis," *Pattern Recognit.*, vol. 76, pp. 215–227, Apr. 2018, doi: 10.1016/j.patcog.2017.11.008.

[26] N. Wang, X. Gao, L. Sun, and J. Li, "Bayesian face sketch synthesis," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1264–1274, Mar. 2017, doi: 10.1109/TIP.2017.2651375.

[27] N. Wang, X. Gao, L. Sun, and J. Li, "Anchored neighborhood index for face sketch synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2154–2163, Sep. 2018, doi: 10.1109/TCSVT.2017.2709465.

[28] M. Zhu, J. Li, N. Wang, and X. Gao, "A deep collaborative framework for face photo-sketch synthesis," in *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 10, pp. 3096–3108, Oct. 2019, doi: 10.1109/TNNLS.2018.2890018.

[29] L. Zhang, L. Lin, X. Wu, S. Ding, and L. Zhang, "End-to-end photo-sketch generation via fully convolutional representation learning," in *Proc. 5th ACM Int. Conf. Multimedia Retr.*, Jun. 2015, pp. 627–634, doi: 10.1145/2671188.2749321.

[30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.

[31] B. Liu, J. Gan, B. Wen, Y. LiuFu, and W. Gao, "An automatic coloring method for ethnic costume sketches based on generative adversarial networks," *Appl. Soft Comput.*, vol. 98, Jan. 2021, Art. no. 106786, doi: 10.1016/j.asoc.2020.106786.

[32] N. Hamzah and F. H. K. Zaman, "Face aging on realistic photo in cross-dataset implementation," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 917, no. 1, Sep. 2020, Art. no. 012080, doi: 10.1088/1757-899X/917/1/012080.

[33] W. Chao, L. Chang, X. Wang, J. Cheng, X. Deng, and F. Duan, "High-fidelity face sketch-to-photo synthesis using generative adversarial network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*. Beijing, China: North China Electric Power Univ., Sep. 2019, pp. 4699–4703.

[34] L. Fang, J. Wang, G. Lu, D. Zhang, and J. Fu, "Hand-drawn grayscale image colorful colorization based on natural image," *Vis. Comput.*, vol. 35, no. 11, pp. 1667–1681, Nov. 2019, doi: 10.1007/s00371-018-1613-8.

[35] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9907, 2016, pp. 649–666, doi: 10.1007/978-3-319-46487-9_40.

[36] T. H. Chowdhury, K. N. Poudel, and Y. Hu, "Time-frequency analysis, denoising, compression, segmentation, and classification of PCG signals," *IEEE Access*, vol. 8, pp. 160882–160890, 2020, doi: 10.1109/ACCESS.2020.3020806.

[37] J. Karhade, S. Dash, S. K. Ghosh, D. K. Dash, and R. K. Tripathy, "Time–frequency-domain deep learning framework for the automated detection of heart valve disorders using PCG signals," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022, doi: 10.1109/TIM.2022.3163156.

[38] J. Li, K. Xu, S. Chaudhuri, E. Yumer, H. Zhang, and L. Guibas, "GRASS: Generative recursive autoencoders for shape structures," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, Jul. 2017, doi: 10.1145/3072959.3073637.

[39] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 3, pp. 441–467, Mar. 2016, doi: 10.1109/TIFS.2015.2480381.

[40] B. F. Klare, S. Klum, J. C. Klontz, E. Taborsky, T. Akgul, and A. K. Jain, "Suspect identification based on descriptive facial attributes," in *Proc. IEEE/IAPR Int. Joint Conf. Biometrics*, Sep. 2014, pp. 1–8, doi: 10.1109/BTAS.2014.6996255.

[41] P. Mittal, A. Jain, G. Goswami, M. Vatsa, and R. Singh, "Composite sketch recognition using saliency and attribute feedback," *Inf. Fusion*, vol. 33, pp. 86–99, Jan. 2017, doi: 10.1016/j.inffus.2016.04.003.

[42] S. Ouyang, T. Hospedales, Y. Z. Song, and X. Li, "Cross-modal face matching: Beyond viewed sketches," in *Proc. Asian Conf. Comput. Vis.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9004, 2015, pp. 210–225, doi: 10.1007/978-3-319-16808-1_15.

[43] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2014, doi: 10.1145/3422622.

[44] A. Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 658–666.

[45] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251, doi: 10.1109/ICCV.2017.244.

[46] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *Proc. 28th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, 2001, pp. 341–346, doi: 10.1145/383259.383296.

[47] M. Song, C. Chen, J. Bu, and T. Sha, "Image-based facial sketch-to-photo synthesis via online coupled dictionary learning," *Inf. Sci.*, vol. 193, pp. 233–246, Jun. 2012, doi: 10.1016/j.ins.2012.01.004.

[48] F. Schroff, D. Kalenichenko, and J. Philbin, ''FaceNet: A unified embedding for face recognition and clustering,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.

[49] S. Motiian, Q. Jones, S. M. Iranmanesh, and G. Doretto, ''Few-shot adversarial domain adaptation,'' in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Dec. 2017, pp. 6671–6681.

[50] A. Dabouei, H. Kazemi, S. M. Iranmanesh, J. Dawson, and N. M. Nasrabadi, ''Fingerprint distortion rectification using deep convolutional neural networks,'' in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 1–8, doi: 10.1109/ICB2018.2018.00012.

[51] S. Soleymani, A. Dabouei, H. Kazemi, J. Dawson, and N. M. Nasrabadi, ''Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification,'' in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 3469–3476, doi: 10.1109/ICPR.2018.8545061.

[52] S. Soleymani, A. Torfi, J. Dawson, and N. M. Nasrabadi, ''Generalized bilinear deep convolutional neural networks for multimodal biometric identification,'' in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 763–767, doi: 10.1109/ICIP.2018.8451532.

[53] S. M. Iranmanesh, A. Dabouei, H. Kazemi, and N. M. Nasrabadi, ''Deep cross polarimetric thermal-to-visible face recognition,'' in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 166–173, doi: 10.1109/ICB2018.2018.00034.

[54] A. Torfi, S. M. Iranmanesh, N. Nasrabadi, and J. Dawson, ''3D convolutional neural networks for cross audio-visual matching recognition,'' *IEEE Access*, vol. 5, pp. 22081–22091, 2017.

[55] A. Broumand, M. S. Esfahani, B.-J. Yoon, and E. R. Dougherty, ''Discrete optimal Bayesian classification with error-conditioned sequential sampling,'' *Pattern Recognit.*, vol. 48, no. 11, pp. 3766–3782, Nov. 2015, doi: 10.1016/j.patcog.2015.03.023.

[56] D. Alhelal, K. A. I. Aboalayon, M. Daneshzand, and M. Faezipour, ''FPGA-based denoising and beat detection of the ECG signal,'' in *Proc. IEEE Long Island Syst., Appl. Technol.*, May 2015, p. 4, doi: 10.1109/LISAT.2015.7160184.

[57] C.-Q. Huang, F. Jiang, Q.-H. Huang, X.-Z. Wang, Z.-M. Han, and W.-Y. Huang, ''Dual-graph attention convolution network for 3-D point cloud classification,'' *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 6, 2022, doi: 10.1109/TNNLS.2022.3162301.

[58] C. Galea and R. A. Farrugia, ''Forensic face photo-sketch recognition using a deep learning-based architecture,'' *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1586–1590, Nov. 2017, doi: 10.1109/LSP.2017.2749266.

[59] M. Li, W. Zuo, and D. Zhang, ''Convolutional network for attribute-driven and identity-preserving human face generation,'' 2016, *arXiv:1608.06434*.

[60] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, ''AttGAN: Facial attribute editing by only changing what you want,'' *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5464–5478, Nov. 2019, doi: 10.1109/TIP.2019.2916751.

[61] T. Karras, S. Laine, and T. Aila, ''A style-based generator architecture for generative adversarial networks,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4217–4228, Dec. 2021, doi: 10.1109/TPAMI.2020.2970919.

[62] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, ''High-resolution image synthesis and semantic manipulation with conditional GANs,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8798–8807.

[63] D. Wu and Q. Dai, ''Sketch realizing: Lifelike portrait synthesis from sketch,'' in *Proc. Comput. Graph. Int. (CGI)*, 2009, pp. 13–20, doi: 10.1145/1629739.1629741.

[64] Y. Jo and J. Park, ''SC-FEGAN: Face editing generative adversarial network with user's sketch and color,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1745–1753, doi: 10.1109/ICCV.2019.00183.

[65] S. Gu, J. Bao, H. Yang, D. Chen, F. Wen, and L. Yuan, ''Mask-guided portrait editing with conditional GANs,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3431–3440, doi: 10.1109/CVPR.2019.00355.

[66] A. Handa, P. Garg, and V. Khare, ''Masked neural style transfer using convolutional neural networks,'' in *Proc. Int. Conf. Recent Innov. Electr., Electron. Commun. Eng. (ICRIEECE)*, Jul. 2018, pp. 2099–2104, doi: 10.1109/ICRIEECE44171.2018.9008937.

[67] T. Portenier, Q. Hu, A. Szabó, S. A. Bigdeli, P. Favaro, and M. Zwicker, ''Faceshop: Deep sketch-based face image editing,'' *ACM Trans. Graph.*, vol. 37, no. 4, pp. 1–13, Aug. 2018, doi: 10.1145/3197517.3201393.

[68] C. Li and M. Wand, ''Combining Markov random fields and convolutional neural networks for image synthesis,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2479–2486. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Li_Combining_Markov_Random_CVPR_2016_paper.html

[69] C. Chen, X. Tan, and K.-Y.-K. Wong, ''Face sketch synthesis with style transfer using pyramid column feature,'' in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 485–493, doi: 10.1109/WACV.2018.00059.

[70] O. Ronneberger, P. Fischer, and T. Brox, ''U-Net: Convolutional networks for biomedical image segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9351, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[71] D. P. Kingma and J. L. Ba, ''Adam: A method for stochastic optimization,'' in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.

[72] Y. Ibrahim, A. Madani, and M. W. Fahkr, ''Coloring ancient Egyptian paintings with conditional generative adversarial networks,'' *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 26, no. 1, pp. 1–6, Jan. 2022, doi: 10.37934/ARASET.26.1.16.

[73] C. Dong, C. C. Loy, and X. Tang. *Accelerating the Super-Resolution Convolutional Neural Network*. [Online]. Available: http://mmlab.ie.cuhk.edu.hk/

[74] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, and D. Li, ''Brain MRI super resolution using 3D deep densely connected neural networks,'' in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 739–742.

[75] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, ''Photo-realistic single image super-resolution using a generative adversarial network,'' in *Proc. CVPR*, 2017, vol. 2, no. 3, pp. 4681–4690. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2017/papers/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.pdf

[76] J. Kim, J. K. Lee, and K. M. Lee, ''Accurate image super-resolution using very deep convolutional networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654, doi: 10.1109/CVPR.2016.182.

[77] X. Zhao, Y. Zhang, T. Zhang, and X. Zou, ''Channel splitting network for single MR image super-resolution,'' *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5649–5662, Nov. 2019, doi: 10.1109/TIP.2019.2921882.

[78] W. Ahmad, H. Ali, Z. Shah, and S. Azmat, ''A new generative adversarial network for medical images super resolution,'' *Sci. Rep.*, vol. 12, no. 1, pp. 1–20, Jun. 2022, doi: 10.1038/s41598-022-13658-4.

[79] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, ''Unpaired image-to-image translation using cycle-consistent adversarial networks,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 183–202. [Online]. Available: http://link.springer.com/10.1007/978-1-60327-005-2_13

[80] H. Pan, ''Cloud removal for remote sensing imagery via spatial attention generative adversarial network,'' 2020, *arXiv:2009.13015*.

ZAKIR KHAN received the M.S. degree in computer science from Hazara University Mansehra, Pakistan, where he is currently pursuing the Ph.D. degree with the Department of Computer Science and Information Technology. He is also a Lecturer in computer science and information technology with Hazara University Mansehra. His research interests include artificial intelligence, machine learning, deep learning, medical image processing, steganography, data hiding, and information security.

ARIF IQBAL UMAR is currently an Associate Professor with the Department of Computer Science and Information Technology, Hazara University Mansehra, Pakistan. His research interests include data mining, data encryption, neural networks, medical image processing, IT security, algorithms, and machine learning.

SYED HAMAD SHIRAZI is currently an Assistant Professor with the Department of Computer Science and Information Technology, Hazara University Mansehra, Pakistan. His research interests include computer vision, texture analysis, neural networks, object recognition, pattern recognition, digital image processing, machine learning, and wavelet transformation.

MUHAMMAD SHAHZAD received the M.S. degree in computer science from the Virtual University of Pakistan. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Information Technology, Hazara University Mansehra, Pakistan. His research interests include data mining, machine learning, deep learning, medical image processing, and natural language processing.

MUHAMMAD ASSAM received the B.Sc. degree in computer software engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2011, and the M.Sc. degree in software engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2018. He is currently pursuing the Ph.D. degree in computer science and technology with Zhejiang University, China. He has been working as a Lecturer (on study leave) with the Department of Software Engineering, University of Science & Technology, Bannu, Khyber Pakhtunkhwa, Pakistan, since November 2011. His research interests include brain–machine interface, medical image processing, machine/deep learning, the Internet of Things (IoT), and computer vision.

MUHAMMAD TAREK I. M. EL-WAKAD received the B.Sc. degree in mechanical engineering from Helwan University, Cairo, Egypt, the M.Sc. degree in medical engineering from George Washington University (GWU), Washington, DC, USA, and the Ph.D. degree in biomedical engineering from the Rensselaer Polytechnic Institute (RPI), Troy, NY, USA. He is currently a Professor and the Vice Dean Student's Affairs with the Faculty of Engineering and Technology, Future University in Egypt (FUE). He served as an Associate Dean for Society affairs, from 2011 to 2013, and the acting BME Department Head, Helwan University, from 2003 to 2004. He carried out research/teaching duties as a Staff Member in several international, regional, and national academic institutions in USA, Saudi Arabia, Oman, and Egypt. These institutions include GWU, RPI, The American University in Cairo (AUC), FUE, The British University in Egypt (BUE), Misr University for Science and Technology (MUST), and 10th of Ramadan Technological Institute. He established a strong research link initially with the Faculty of Oral and Dental Medicine, Cairo University, and later with similar faculties at Ain Shams University, Tanta, and Menia Universities.

EL-AWADY ATTIA received the Ph.D. degree in industrial systems engineering from the Department of Industrial Engineering, INPT/ENSIACET, Toulouse University, France. He was the responsible for the Department of Mechanical Technical, Al-Ameeria Integrated Technical Cluster, and Educational Development Fund (EDF) in Egypt. He is currently an Associate Professor with the Department of Industrial Engineering, College of Engineering, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia, and also the Faculty Member of the Department of Mechanical Engineering, Faculty of Engineering (Shoubra), Benha University, Egypt. He has published many papers in the international journals and conferences. Moreover, he reviewed many papers for many of industrial engineering top journals and conferences. His research interests include applicability of performance improvement tools in industry, risk management, lean manufacturing, simulation, fuzzy modeling, multi criteria decision making, data mining, maintenance planning, production planning, Kanban systems, optimization using heuristics techniques, and consideration of human factors in industry.