Computers, Materials & Continua
DOI:10.32604/cmc.2022.018871
Article

Tech Science Press

Efficient Facial Recognition Authentication Using Edge and Density Variant Sketch Generator

Summra Saleem[1,2], M. Usman Ghani Khan[1,2], Tanzila Saba[3], Ibrahim Abunadi[3], Amjad Rehman[3,*] and Saeed Ali Bahaj[4]

[1]Department of Computer Science, UET, Lahore, Pakistan
[2]Al-Khwarizmi Institute of Computer Science, UET, Lahore, Pakistan
[3]Artificial Intelligence & Data Analytics Lab CCIS Prince Sultan University, Riyadh, 11586, Saudi Arabia
[4]MIS Department College of Business Administration, Prince Sattam Bin Abdulaziz University, Alkharj, Saudi Arabia
[*]Corresponding Author: Amjad Rehman. Email: rkamjad@gmail.com
Received: 24 March 2021; Accepted: 15 May 2021

Abstract: Image translation plays a significant role in realistic image synthesis, entertainment tasks such as editing and colorization, and security including personal identification. In Edge GAN, the major contribution is attribute guided vector that enables high visual quality content generation. This research study proposes automatic face image realism from freehand sketches based on Edge GAN. We propose a density variant image synthesis model, allowing the input sketch to encompass face features with minute details. The density level is projected into non-latent space, having a linear controlled function parameter. This assists the user to appropriately devise the variant densities of facial sketches and image synthesis. Composite data set of Large Scale CelebFaces Attributes (ClebA), Labelled Faces in the Wild (LFWH), Chinese University of Hong Kong (CHUK), and self-generated Asian images are used to evaluate the proposed approach. The solution is validated to have the capability for generating realistic face images through quantitative and qualitative results and human evaluation.

Keywords: Edge generator; density variant sketch generator face translation; recognition; residual block

1 Introduction

PAKISTAN as an underdeveloped country is facing a lot of challenges such as exponential growth of population, ever-decreasing rate of economic growth and current waves of crimes, etc. Pakistan is facing 2.77% per annual growth of population, which looks quite demanding for developing countries. Metropolitan cities i.e., Lahore, Karachi, and Islamabad are concentrated with a major ratio of population. This further weakens the planning infrastructures of government agencies. These cities are concentrated with migrating people, who abruptly find a deficiency of social laws and feel curious to violate established laws for social control. This situation leads to an exponential increase in poverty and unemployment, which are prime factors for the increased rate of crimes.

In addition to theft activities, black markets run by criminologists put a severe burden on the economy of the country by not providing license fees and strict security measures should be applied. In Pakistan, security-sensitive areas such as less populated areas are mostly crime scenes. There has been a rapid increase in serious security threats i.e., motor-vehicle theft, robbery, theft, and burglary. In recent years, automatic face recognition systems [1] are extensively used by security agencies. Most conventional face identification systems [2] target photo-to photo matching. However, in law enforcement agencies manual sketching of the suspect is performed, as photos are generally unavailable. They usually sketch some narrative descriptions by the victim prone to human errors. Sketch recognition systems attain low efficiency due to considerable variation between drawn sketches and photos.

Some of the basic limitations of current face recognition systems from sketches are: 1) fewer details in sketches tends to decrease the accuracy of face recognition system 2) noise in sketches can also degrade the performance of face recognition systems 3) skin color-based detection and recognition systems can completely fail on sketches 4) sketch images have less detail of beard and mustaches that might lead to the wrong prediction. Conversion of sketches to photos is a solution to the above-mentioned problems, as photos of suspects are usually available in police records. Transforming sketches to images for criminal cases is a solution for increasing security measures especially in metropolitan cities of Pakistan.

Our proposed system will be efficient enough to extract information from the synthesized photo. This information can be used for security purposes in sensitive areas like less populated and less crowded areas. Moreover, information can be utilized for security authorized areas like hospitals, offices, judicial courts by security alerts. This research is an effort to develop a low power and low-cost efficient sketch to image synthesis system. In the proposed system manually drawn sketches are transformed into photo-realistic images. After extracting the visual depiction of criminals, information is processed to all central security authorities for further inquiry. This will make the security system more efficient for quick disciplinary actions. Our major target is Pakistan and as we see in Pakistan crime rate is increasing exponentially day by day to reduce criminals by achieving the following objectives: a low-cost automatic system to facilitate law enforcement systems for investigation.

There has been plenty of work for the image-to-image translation [3,4]. Image translation has gained remarkable improvement after the birth of generative adversarial networks (GAN) [5]. Providing training data from multiple representations enables GAN based network to transform an image from one representation to another. For example, the images of sketches can be one representation and other representations can be a realistic face. The efficiency of face recognition systems is dependent on the availability of facial features. Face identification from sketch images has to deal with a high rate of false predictions because of lesser details. This research is focused on a sketch to realistic image synthesis system that can assist security personals to increase safety in society.

Existing facial sketch to image translation data-sets is com-posed of faces from European and Chinese ethnicity. There is variation in facial features and skin tone among different ethnicity. For example, faces from European ethnicity might have light-colored eyes in general but in other communities such as Asia, people usually have dark-colored eyes. Therefore, there is a need for a benchmark data-set that is specific to the Asian community. There is no end-to-end system available that can recognize a person using the hand-made sketch. In traditional methods, human resources are employed to identify a person's sketch. In the proposed research, we automate the

process of identification by synthesizing photo-realistic facial images from sketch and high-level feature information. Major contributions of research work are mentioned below:

- Face translation system generates realistic faces of human-based on attribute guided approach using the contrastive learning features preserving low-level details such as color and ethnicity.
- The accuracy of the sketch-based face identification system has improved remarkably to assist law enforcement and security agencies.
- Asian data-set is generated for the learning of the network for a sketch to image translation with annotation.

A detailed survey for face recognition and face translation system has been given in Section 2. Section 3 discusses the proposed methodology. Sections 4 and 5 explain experiments and evaluations, respectively. Lastly, the conclusion and future work is presented in Section 6.

2 Related Work

In this section, research work related to face recognition and image translation techniques has been discussed.

2.1 Face Recognition System

Face recognition techniques have become a demand of the current era due to increased challenges in security issues of society. During past decades numerous algorithms have been devised for face recognition. This research work discusses methodologies based on computer vision and neural network-based. The most well-known techniques for face recognition are Eigenface [6] and Fisher face [7]. The Eigenface technique employs Principal Component Analysis (PCA); feature reduction technique. PCA is used to reduce the feature vector set along with maximum change. The Eigen-face is a low-level feature based-technique; preserving the texture features of the face. In contrast to the unsupervised Eigen-face approach, Fisher-face is a supervised approach. It finds unique face descriptors by using a linear discriminator. Both of the above-mentioned approaches rely on the Euclidean structure for extracted face features. The local binary pattern has also been used by researchers for face recognition [8,9].

Neural networks have brought remarkable improvements in terms of performance and it is highly dependent on big data-sets. Deep neural networks automatically extract feature vectors from data to learn face attributes. Lu et al. [10] proposed a methodology based on ResNet to identify faces. They employed face recognition architecture using a residual block comprising of two supporting networks. FaceNet architecture proposed by Schroff et al. [11] preserved the euclidean feature vector directly. The trained model tends to minimize the distance between similar faces and vice versa. The inception model was employed to extract feature-set and produced remarkable results on the LFW [12] data-set.

2.2 Image Transformation

Generating images by learning image collection is a fascinating task in the computer vision and graphics field. Successful approaches in the past years tend to use image fragments for non-parametric techniques [3,4,13,14]. In previous years, parametric deep learning networks achieved promising results [1,15,16]. GAN's [5] are the most promising techniques for image synthesis. A discriminator network is trained simultaneously for the classification of images as original or synthesized. GAN attempts to fool the discriminator network, while discriminator alarm generator

from synthesizing fake images. The trained generator network is capable of producing distinct images by learning low dimensional latent space.

Optimization in latent space representation can lead to the manifold of natural images in network visualization [17] and image editing [18]. Furthermore, latent space is not well formulated semantically, specific dimension does not correlate to semantics, aligning them to intermediate image structure can give more understanding. Rather than drawing hard constraints on the input sketch, the method proposed by [19] learns the joint distribution of sketch and corresponding real image. Hence keeping the constraint on the input sketch, weak. The output from their model depicts the freedom in the appearance of the generated image. Unsupervised image translation is proposed by Liu et al. [20] proposed an efficient probabilistic approach by learning the most likely output and rendering style. They change the image contents by updating [20] the statistics of the input image. By changing these, they changed image clarity, resolution, appearance, and realism. Wang et al. [21] proposed a novel technique for image translation using semi-coupled dictionary learning (SCDL). Hitoshi Yamauchi et al. worked on removing defects from input images [22]. Deep image synthesis is learning low-level latent representation to regenerate images with Generative Adversarial Network (GAN's) or Variation Auto-encoders (VAE's) networks [23]. Generally, the deep image synthesizing networks can be conditioned on different input vectors [23] like gray-scale images [24], object characteristics and 3-dimensional (3D) view parameters, attributes [25], or images and desired view [26].

Pattern Sangkloy et al. proposed a novel approach of image-to-image translation from a sketch of the image as input [26]. Conditional GANs synthesize images based on conditions that are generated on more relevant input from the rest of the data-set. Different techniques are formulated for relevant inputs such as low-resolution images [27], class labels [28], incomplete or partial images [29], or text [30] rather than generating images from latent vectors. Conditional GANs have also been implemented for specific applications such as diverse artistic styles [31], super-resolution [14], video prediction [32], texture synthesis [33], and product images. Image to image translation for general-purpose requires a huge number of paired labeled images as presented by Isola et al. [31].

Discriminator can be conditioned on specific inputs like input text embedding condition on generator and discriminator [34], which contributes to the powerful discriminator. The unsupervised approach for image translation proposed by Taigman et al. proposed a network that can learn image translation without labeled pair images [35]. Furthermore, this mechanism needs a pre-trained function for mapping images to an intermediary representation of cross-domain which relies on labeled images in other formats. Kazemi et al. proposed a framework [24] based on facial attributes and is a conditional version of Cycle-GAN has been presented in this paper. Rather than based on aligned face sketch pairs, the purposed framework only required facial attributes like skin and hair color for training purposes. The performance is evaluated on the FERET data-set and WVU Multi-modal data-set.

To reduce the requirement of labeled data, a dual learning approach was introduced by Xia et al. [36]. The main idea of the dual learning mechanism is to involve two learning agents. In CycleGAN [35] concept for unpaired image translation is introduced, for cyclic mapping dual relation in DualGAN is required. The predominant characteristic of CycleGAN is deter-mined by numerous problems where training data is hard to find like weather transfer and painting style transformation.

3 Proposed System and Methodology

This section provides a detailed methodology for trans-forming the sketches to realistic facial images for enhanced face recognition. We proposed a novel system for facial image generation based on the generative adversarial network. The generated face from the proposed system is feed as the input to the face identification system. We have employed contrastive learning using edge and density variant generative adversarial network for transforming the input image of the sketch to a realistic face image.

3.1 Edge Generator

Generally, speaking the Architecture of EdgeGAN. Straightforwardly demonstrating the planning between a solitary picture and its relating portrays, for example, SketchyGAN [9], is troublesome in light of the gigantic size of the planning space. We hence all things considered address the test in another plausible manner all things being equal: we gain proficiency with a typical portrayal for an article communicated by cross-space information. To this end, we plan ill-disposed engineering, which is appeared in Fig. 1, for EdgeGAN. Instead of straightforwardly construing pictures from draws, Edge-GAN moves the issue of the sketch-to-picture age to the issue of creating the picture from a quality vector that is encoding the articulation purpose of the freehand sketch. At the preparation stage, EdgeGAN learns a typical trait vector for an item picture and its edge maps by taking care of ill-disposed organizations with pictures and their various drawing-style edge maps. At the derivation stage Fig. 1, EdgeGAN catches the client's appearance goal with a quality vector and afterward creates the ideal picture from it. Structure of EdgeGAN. As appeared in Fig. 1, the proposed. EdgeGAN has two channels: one including generator GE and discriminator DE for edge map age, the other including generator GI and discriminator DI for picture age. Both GI and GE take a similar clamor vector along with a one-hot vector prosecuting a particular class as info. Discriminators DI and DE endeavor to recognize the produced pictures or edge maps from the genuine conveyance. Another discriminator DJ is utilized to energize the produced counterfeit picture and the edge map portraying a similar item by telling if the created counterfeit picture coordinates the phony edge map, which takes the yields of both GI and GE as info (the picture and edge map are connected along with the width measurement). The Edge Encoder is utilized to energize the encoded trait data of edge guides to be near the commotion vector took care of to GI and GE through an L1 misfortune. The classifier is utilized to gather the classification mark of the yield of GI, which is utilized to energize the produced counterfeit picture to be perceived as the ideal class using a central misfortune [20]. The itemized structures of every module of EdgeGAN are delineated in Fig. 1.

We actualize the Edge Encoder with the equivalent encoder module in bicycleGAN [37] since they assume a comparative job practically, i.e., our encoder encodes the "content" (e.g., the posture and shape data), while the encoder in bicycleGAN encodes properties into dormant vectors. For Classifier, we utilize a design like the discriminator of SketchyGAN while disregarding the antagonistic misfortune and as it was utilizing the central misfortune [20] as the arrangement misfortune. The design of all generators and discriminators depends on WGAP-GP [16]. Target capacity and all the more preparing subtleties can be found in the valuable materials.

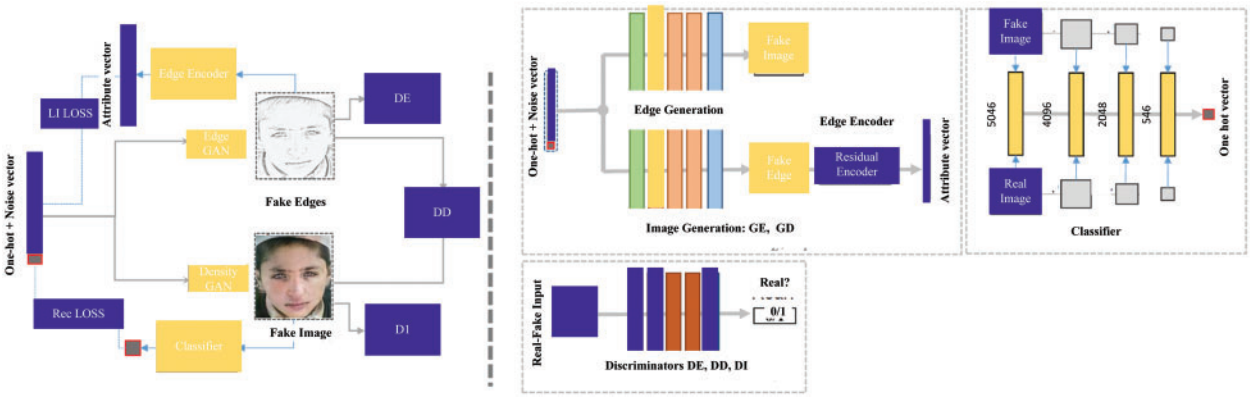

Figure 1: Architecture of proposed system for face transformation

3.2 Density Variant Sketch Generator

Our objective for DVSG is to create sketches with persistent solidity, which is difficult because of the accompanying reasons: clearly, it is difficult to acquire constant ground truth pictures for each scale factor, which implies the DVSG is a semi-administered model; moreover, we intend to create density variant sketches and utilize a scale factor to control its visual unpredictability, it is interesting to develop an exact association between high dimensional dissemination (picture) and a scalar. This can help for assessing the visual multifaceted nature in the sketch age stage, however non-straight planning between the scalar and sketch densities can likewise be embraced, it is troublesome to gauge the real thickness of yield with a given number. Even though there is no constant substance picture, nevertheless; all the same. Similar nevertheless, we can still utilize a few sketch pictures that depict the basic semantic data of the input picture as the key density pictures K $= K_1$; $K_2 \ldots K_{ng}$, K_2 Y, which are relating to a set of inspected thickness $s = s_1, s_2, \ldots, s_n$. At that point, the undertaking now is changed to produce semantically nonstop substance pictures between two key thickness pictures. We first stretch out the thickness factor to a thickness veil M_s by filling the veil with the thickness factor, joined with the reference picture, and send them into the content generator G_c, where $(Y_s) = G_c(M_s; X)$. Considering the circumstance that we have the ground truth picture, and afterward, a picture recreation misfortune can be utilized to recreate the key thickness pictures as shown in Eq. (1):

$$L_{recon_c} = \mathbb{E}_{x,s,G_s}\|K_i - G_c(M_{si}, X)\| \tag{1}$$

Since there are no such ground truth pictures among K_i and $K_i + 1$, it is a semi-directed issue requiring circuitous control. Roused by InfoGAN [3], we attempt to fabricate an association between 5 the created sketch pictures and its comparing thickness factors by embracing a thickness encoder Es, which intends to encode the produced portrays (^ Y (s)) back to a thickness scale that ($\hat{}s = E\_s((\hat{}\ Y\_(s)))$. Since we need to express authority over the generated pictures, and then it is improved by a scale reconstruction loss as shown in Eq. (2):

$$L_{recon_s} = \mathbb{E}_{x,s,\varepsilon_s,G_c}[\|s - E_s(M_s, X)\|] \tag{2}$$

where, K_i and $Y\_(si)$ show a similar sketch picture and ^ Y (s_i) is the reproduction of Y_{si}. In our underlying investigations, we found that the scale reproduction misfortune can just assist with assessing the sketch pictures around key thickness sketches. On the off chance that there is a huge

mathematical or semantic variety between two key thickness pictures, the thickness encoder just as the scale remaking misfortune is not, at this point ready to guarantee the linearity of pictures between the two key thickness outlines. Thusly, we plan a versatile component distance loss (AFD) to drive the organization to lean the connection between two key thickness sketches.

For a scale s0 coordinating the non-key thickness sketch Y (s_0). It has two neighbors key thickness outlines K_i and $K_i + 1$ that are relating to two thick factors si and si + 1, where I $= h(s_i)$, we characterized the AFD loss as follows in Eq. (3):

$$L_{AFD} = \left\| \Gamma(Y(s')) - \left(\frac{1}{s' - s_i} \Theta(K_i) + \frac{1}{s_{i+1} - s'} \Theta(K_{i+1}) \right) \right\| \tag{3}$$

where above parameter shows the feature extraction from the encoder of the content generator G_c and the scale encoder E_s. The AFD loss guarantees the linearity in latent space as the versatile loads are contrarily identified with the distance current content picture and its neighbor key thickness outlines. At the end of the day, the produced sketch is compelled to be of higher closeness to its nearest neighbor key thickness portrays. Ours explores likewise show that the AFD misfortune fundamentally improves the coherence of produced content pictures.

3.3 Discriminator

The point of the discriminator is to recognize the produced pictures and genuine pictures. The discriminator needs its production to be valid for genuine pictures. For produced pictures G_s, the discriminator needs its production to be bogus. The deficiency of the discriminator is composed as shown in Eqs. (4):

$$L_G = \frac{1}{2}\mathbb{E}_{(s,x)\sim pdata(S,x)}[(D(s, x) - 1]^2 + \frac{1}{2}\mathbb{E}_{s\sim pdata(S)}[(D(s, G(s))]^2 \tag{4}$$

Moreover, multi-scale discriminators D_1, D_2, and D_3 are used, which are regular in picture production. The coarse to the fine model can improve the nature of the produced picture. At the coarse scale, it can catch the worldwide data and improve the consistency of the produced picture with the enormous open field. At the fine scale, it catches the data at the nearby view and jellies the subtleties, for example, edges, and lines. Also, the multi-scale techniques can diminish the weight on the organization and make the organization simpler for preparing.

3.4 Losses Involved in GAN

Multiple loss functions need to be optimized for generating a perfect GAN-based system.

3.4.1 Adversarial Loss

Adversarial loss is utilized for creating a realistic image that looks like the original image. The adversarial loss can be described by Eq. (5).

$$L_a = E_x[logD_i(x)] + E_{x,c}[log(1 - D_i(G(x, c)))] \tag{5}$$

The above equation shows the generator G for generating images G(x; c) under the condition c. In the above equation, D represents a discriminator that differentiates between real and fake imagery. The generator tries to minimize the objective function of falseness and the discriminator tries otherwise. Domain Classification Loss: This part of the thesis describes the conversion of a single input image x to a transformed image y based on some condition variable c. For converting the sketch to multiple colored faces we utilized domain classification loss with

generator and discriminator. Domain classification loss of real images is used for optimizing the discriminator and similarly, domain classification loss on fake images to optimize the generator. Eq. (6) shows the basic loss of real images.

$$L^{t}_{class}eal = E_{x,\, c_{\gamma eal}}[-log\, Dclass(D(c_{\gamma eal},\, x))] \tag{6}$$

In the above equation, 6 D depicts the probability of real images by the discriminator. Discriminator minimizes this function to find the exact class type of the real image. The loss classification for fake images can be described by following Eq. (7).

$$L^{t}_{class} = E_{x,\, c_t}[-log\, Dclass(c \mid G(c_t,\, x))] \tag{7}$$

3.4.2 Reconstruction Loss

Reflection of the original sketch is necessary to enable sketch-based face identification of any person. Optimizing the above two losses ensures that generated images are real enough and belong to the specific color category but to make sure whether the transformed image reflects the original identity of the person, we utilized reconstruction loss. Following Eq. (8) shows the reconstruction loss.

$$\mathfrak{L}_{\gamma ec} = E_{x,\, c,\, c1}[x - Gen(C(x,\, c),\, c1)] \tag{8}$$

where Gen represents the generator that takes $C(x;\, c)$ image as the translated input image and reconstructs the original image from the translated image. All of the above losses contribute towards the full objective function for conditional GAN. The overall objective function of the system for generator and discriminator is shown in Eqs. (9) and (10), respectively.

$$\mathfrak{L}_{D}is = \mathfrak{L}_{a} + \lambda_{class}\, \mathfrak{L}^{\gamma}_{class} \tag{9}$$

3.5 Proposed Face Recognition System

The proposed face recognizer is residual, aiming to extract deep features. The extracted deep features are robust for generated faces from the sketch. Features extracted from fc7 layers of discriminator are stored in the database for making a comparison with the input probe image features. The system inherently compares stored features from the database and features of an input image for finding the identity of the person. The comparison of prob image features with the database image features is made by using Euclidean distance Eq. (10).

$$D(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} \tag{10}$$

4 Dataset and Experiments

In this section, we discuss the details of the evaluation for face recognition and sketch-face transformation system. We have provided an evaluation of both systems (combined and face generation systems). The major evaluation steps involved in our proposed system are elaborated in detail.

4.1 Training and Validation Dataset

Details are provided in the below section for traning and validation dataset used for experimentation.

4.1.1 CelebA Dataset

Celeb Faces Attributes Dataset (CelebA) is a large-scale face attributes dataset with more than 200 K celebrity images, each with 40 attribute annotations. CelebA has large diversities, large quantities, and rich annotations. There are 10,177 identities, 202,599 face images, 5 landmark locations, 40 binary attributes annotation per image. From 1–162770 images represent the training set, from 162771–182637 images are representing the validation set. and from 182638–202599 images are considered as the testing set. Sample images from the CelebA dataset are shown in Fig. 2a.

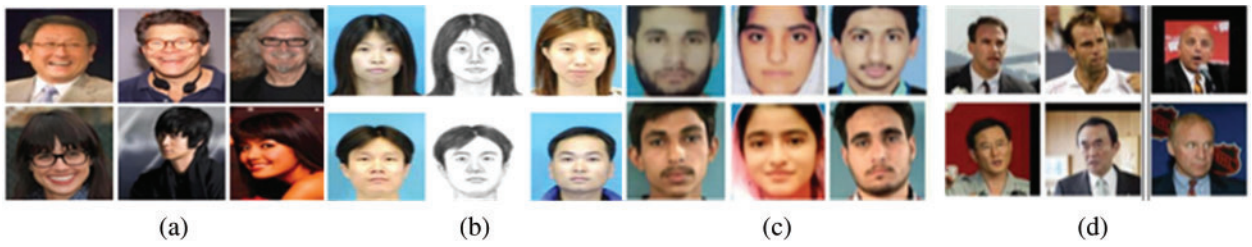

Figure 2: Sample images from the dataset (a) CelebA dataset (b) CHUK dataset (c) Self-generated dataset (d) Labeled faces in the wild (LFW)

4.1.2 CHUK Dataset

CHUK dataset comprises 188 sketches of students collected from the Chinese University of Hong Kong (CHUK). Out of 188 faces, 100 are selected for training while the rest of 88 faces are used for testing. Sketches and images of the dataset are of size 200 250. Some sample instances from the CHUK dataset are shown in Fig. 2b.

4.1.3 Self-generated Dataset

For Asian community face images are collected from the local community at the University of Engineering and Technology, Lahore. Total 10,764 out of which 4,567 sketches are of females and 6,197 sketches are of males. Sample images from the self-generated dataset are shown in Fig. 2c.

4.1.4 Labeled Faces in the Wild (LFW)

LFW dataset [12] is gathered to think about unconstrained face acknowledgment issues. This dataset comprises more than thirteen thousand facial pictures gathered from an assortment of sources. Complete 1680 subjects are envisioned and they have at least two discriminative pictures in the dataset. The confinement and size of appearances in unique LWF were dictated by utilizing computerized locater (Viola jones), new cut face pictures, in the LFW crop dataset portrays sensible situations, including mix arrangement, scale decent variety, the pivot of appearances in-plane and out of the arrangement. Few sample images from the LFW face dataset are shown in Fig. 2d.

4.2 Training and Implementation Details

In this section, we throw light on the training part of our proposed system for identifying faces. We employed a Soft-max classifier for learning the best possible features on the face. Optimization of the proposed system is achieved by using stochastic gradient descent. It is

the optimization function used for back-propagation we have employed 64 batch sizes with a momentum 0.9. For reducing the over-fitting of the proposed system, we introduced a dropout layer with a 0.35 removal rate. The second important parameter that is set externally is the learning rate during the training procedure. We employed a dynamic learning rate by changing the value of the learning rate to train the system optimistically. Initially, the learning rate was set to 0.001 which is dropped by a factor of ten when the loss stopped decreasing. The final learning rate e for the proposed identification system m was 0.00001. We have utilized Gaussian distribution with zero as a mean value and 0.01 as a standard deviation value. Initial biases are set to zero that are further updated to non-zero value based on back-propagation.

5 Evaluation

In the current section details of evaluation for face recognition and sketch-face transformation system have been discussed. We have provided an evaluation of both systems (combined and face generation systems). Extensive images from diverse ethnic groups are employed for training the sketch to photo translation system. We used the standard Celebrity Attribute dataset along with some additional images from the Asian community. CelebA dataset is a diverse dataset in terms of facial features, posed, ethnicity, and poses. This dataset is mostly for European and English actors as shown in Fig. 3. It contains quite fewer images from the South Asian community. Considering our local environment we have enhanced this data of 0.2 million im-ages with 12 thousand images from the local Pakistani community. The face transformation system is optimized based on adversarial loss involved during generator and discriminator competition. The loss is optimized based on Adam optimizer with two hyper-parameters of value $_1 = 0{:}5$ and $_2 = 0{:}99$. We have trained this system on Nvidia 1080 Ti GPU with 11GB of memory. The complete dataset is processed in the form of batches and we used a batch size of 16 for feeding the images. For optimizing the complete system we have set a learning rate of 0.001 initially with dynamic decay after fifty epochs. The network was trained for about eighteen hundred epochs. Five experts having profound knowlwdge of the domain and research task were elected for evaluation of results. The major evaluation steps involved in our proposed system are elaborated as follows in Tab. 1.

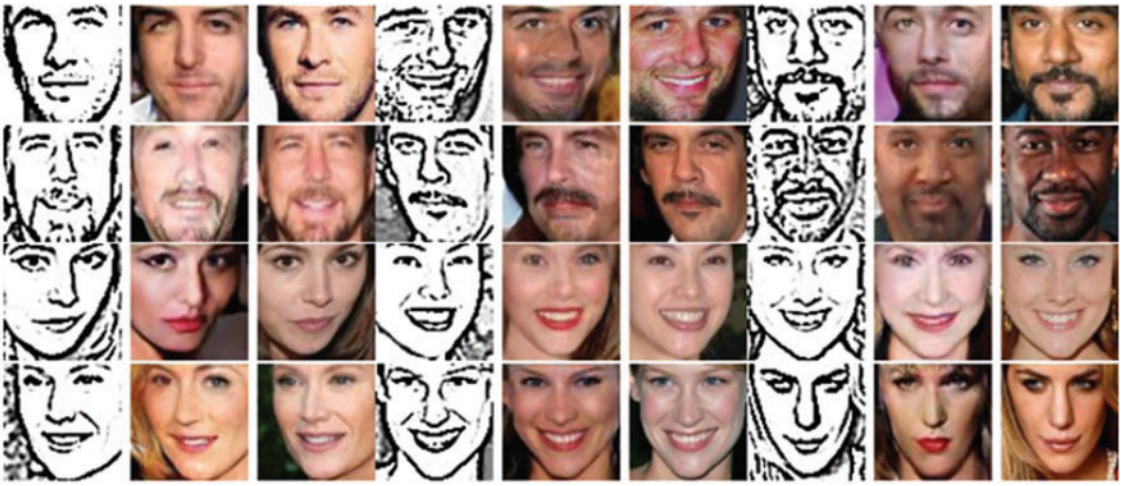

Figure 3: Sample sketches from ClebA dataset with translated realistic and ground truths

Table 1: Task-based evaluation of face realism system by five judges

2 * rating parameters	Judge				
	1	2	3	4	5
Image realism	8	7	8.5	8.5	8
Quality	8.5	9	9	9	8.5
Identity	10	9.5	10	10	9.5

5.1 Evaluation of Face Realism

For evaluating the quality of generated photos, three metrics are used namely structural similarity (SSIM), product-moment correlation coefficient R, and peak signal to noise ratio (PSNR) [37]. The proposed at1tributed guided network generates images preserving details of the overall structure. In comparison to this other methods generates an image with blurriness and low-frequency particulars. The comparison of results is shown in Tab. 2.

Table 2: Comparison with state of the art quantitative evaluation

Dataset	Method	PSNR	SSIM	R
CHUK	MrFSPS_SP [32]	–	0.6333	–
	Scribbler [37]	0.175	0.004	0.003
	Proposed	0.171	0.006	0.004

5.2 Qualitative Evaluation

Image realism.

- Quality of the generated image.
- Identity preserve.

Kappa calculates the inter-annotator agreement for the classification of the data into y target classes. The standard can be depicted as Eq. (11), The Face transformation system is evaluated by qualitative analysis and subjective evaluation. Fig. 3 shows the resultant visualization of output from the proposed sketch to photorealism. From the resultant image, it can be seen that our system outperformed the previous state-of-the-art systems in the past. From the image, it can be observed that the proposed system preserves the originality of the person by maintaining the identity of the individuals. The identity is preserved based on convolution features rather than low-level features. Results of the face recognition system are shown in Tab. 3.

$$\mathfrak{L}_G en = \mathfrak{L}_a + \lambda_{class}\mathfrak{L}^t_{class} + \lambda_{\gamma ec}\mathfrak{L}_{\gamma ec} \tag{11}$$

Other prominent systems do preserve the identity up-to some extent but they lose the detail of the original image. Systems like examples show less realistic results than our proposed systems in terms of realism, originality, and identity preservation. The basic contributing factor of the proposed system are listed below,

- Unique generator and discriminator for transforming the input sketch into different possible faces based on the color of the face.

- Features are selected from convolution layers.
- Data is augmented in the proposed system to enhance robustness.

Table 3: Confusion matrix for the proposed face recognition system

	Actual individual	Actual unknown
Predicted individual	99	3
Predicted	2	303

5.3 Inter-Annotator Agreement

For comprehensively checking the results of our proposed system, we have utilized an inter-annotator agreement to check the performance of the proposed system. We took advantage of the famously developed evaluation Cohen Kappa standard for subjective analysis. Cohen Kappa is the measure for inter-rater agreement for covering the realistic views of the developed system described in Eq. (12). It is considered more efficient because it takes the probability of chance rather than percentage agreement only. Provided an input photo to the system, five different judges were asked to rate the system based on questions about the performance of the system.

$$K = fracProb_a Prob_\gamma 1 - Prob_\gamma \tag{12}$$

where, above probabilities shows the observed and random agreement, respectively. The transformed photos from different baseline models and our system are shown to the judges. The transformed image can be generated based on two different conditions, as mentioned in Tab. 5. The complete results of the system are shown in Tab. 5. As depicted from Tab. 5 our system outperformed in terms of all conditional parameters. The results for white people for other systems are not that much worse because they are primarily trained based on the white people dataset. Five different judges ranked the 100 transformed images from the proposed system. Each image is transformed based on ethnicity and color condition. Following Tab. 3 provides the detailed ranks of the images.

5.4 Face Recognition System Evaluation

This section provides the detailed results of a combined sketch-based face recognition system. We have employed the Labeled Faces in the Wild dataset for training and evaluating the combined complete system. LFW dataset consists of 5,749 different individuals with two photographs on average. The dataset is split into two portions with 80% for training and 20% for testing the system. For evaluation of the complete system, accuracy has been used as an evaluation parameter. The system is checked based on the face verification task. In the verification task if we are given a pair of photographs of the same individual then the Euclidean distance decides that images are similar or not. The set of images belonging to the same person are represented as I_S and for different persons, the pair is represented as I_d. Similarity or dissimilarity is decided based on distance value is whether greater than a particular threshold or not. We set this threshold as 0.5. The correct identified person is decided based on the following Eq. (13).

$$TA(d) = (i, j) \in I_s \quad with\ D(x_i, x_j) \leq 0.5 \tag{13}$$

Similarly, false accepted are calculated as the following Eq. (14).

$$FA(d) = (i,j) \in I_d \quad with\ D(x_i, x_j) \leq 0.5 \tag{14}$$

Finally, the accuracy can be computed by Eq. (15).

$$Acc = \frac{TA(d)}{Is} \tag{15}$$

For training and testing the complete LFW is split into training and validation set. Tab. 5 provides the data-set division for LFW. For evaluating the complete system comprehensively, the validation set is further classified into 1000 known and 150 unknown individuals, means single image from 1000 individuals are placed in the gallery for matching purpose. Similarly, 2305 images are divided into 2000 and 305 images. Out of 2000 known images single image is placed in a gallery that makes overall 1305 testing images. Tab. 5 shows the detailed results of the proposed combined face recognition system. We have also compared the accuracy of the proposed hybrid system with previous systems. Tab. 5 shows the comparison of face recognition based on the sketch.

As depicted from Tab. 4 our system outperformed in terms of all conditional parameters. The results for white people for other systems are not that much worse because they are primarily trained based on the white people dataset. Five different judges ranked the 100 transformed images from the proposed system. Each image is transformed based on ethnicity and color condition. Tab. 5 shows the comparative analysis of face recognition based on sketch translation.

Table 4: Proposed edge and density variant GAN for varying attributes and noise

Attribute name	Conditional GAN (attribute + noise) (%)	Conditional noise vector (%)	Proposed system (%)
Ethnicity	15	14.5	70.5
Color	7.5	11	81.5

Table 5: Comparison table of sketch-based face recognition system

Method	Accuracy (%)
DeepFace [14]	77.3
FaceNet [11]	78.1
VGG face [5]	83.2
Center face [27]	79.2
Proposed network	95.5

6 Applications

Notwithstanding broad sketch-to-photograph interpretation, we illustrate a few applications requiring numerous thickness sketches as a contribution to our model.

6.1 Multi-Scale Face Editing

The coarse level compared to a little si permits the client to alter the enormous shapes while disregarding the subtleties which would be dealt with by the model. In the coarse level altering, the client can without much of a stretch change the overall trademark of a human face, including face shape, length of hair, facial articulation, and so forth, as appeared in Fig. 4. The fine Level comparing to a huge si underpins modern control on subtleties, for example, the hair surface (wavy to straight), and skin surfaces (adding or eliminating wrinkles), has appeared in Fig. 4. Contrasted with past face altering work dependent on division mask [24] or landmarks [10], our strategy has two preferences. The first drawing is a more instinctive and easy-to-use path for picture altering. Also, our strategy can control the altering cycle at various scales from significant article limits to explained miniature structures.

Figure 4: Face translation based on different attributes

Figure 5: Cartonification of sample face images

6.2 Anime Colorization

Our strategy is generally to various kinds of information, so it can be applied to Anime Colorization and Editing too. Not the same as the past sketch colorization strategy that our model can colorize pictures in both coarse and fine levels. In expansion, we additionally uphold post-altering after colorization. Such altering permits unpleasant adjustment as well as point-by-point changes like adding shadows or features, adjusting the minor surfaces, and so forth, results have appeared in Fig. 5.

7 Conclusion and Future Work

In this research study, we have developed a hybrid system for improving the robustness of face recognition technology. Face recognition tends to behave worst for sketch-based face recognition. Usually in crime scene photos of the culprit are not available but only eyewitnesses. Recognizing the person through sketches is a challenging task that may lead to false results. To make face recognition possible through sketches, we proposed a unified system by combining sketch to colored-photo transformation and face recognition. We have used contrastive learning using edge and multivariant generative adversarial network for the generating first part of the heterogeneous system. The generated colored image would be from the user-selected choice. Residual learning-based networks enabling the shortcut connections between lower layers and high-level layers are employed for face recognition. Evaluation results show that the proposed system performed satisfactorily for the recognition of sketches. Results also depicted that the transformation system is capable of generating original faces from sketches by minimum loss of identity. In the future, the proposed system can be extended to work for more noisy sketches. Furthermore, the addition of more high-level features for sketch transformation like pose, emotion, and hair color can be employed.

Acknowledgement: This research work was supported by Computer Science department, UET Lahore and KICS. This research is also supported by Artificial Intelligence & Data Analytics Lab (AIDA) CCIS Prince Sultan University, Riyadh, 11586, Saudi Arabia. The authors also would like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] X. Yan, J. Yang, K. Sohn and H. Lee, "Attribute2image: Conditional image generation from visual attributes," *European Conf. on Computer Vision*, vol. 99, no. 8, pp. 776–791, 2016.

[2] O. M. Parkhi, A. Vedaldi and A. Zisserman "Deep face recognition," *BMVC*, vol. 1, no. 1, pp. 6, 2015.

[3] C. Barnes, E. Shechtman, A. Finkelstein and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 24, 2009.

[4] T. Chen, M. Cheng, P. Tan, A. Shamir and S. Hu, "Sketch2photo: Internet image montage," *ACM Transactions on Graphics*, vol. 28, no. 5, pp. 1–10, 2009.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley *et al.*, "Generative adversarial nets," *Communications of the ACM*, vol. 63, no. 11, pp. 2672–2680, 2020.
[6] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Proc. CVPR'91*, Maui, HI, USA, IEEE, vol. 1, pp. 586–591, 1991.
[7] K. Kwak and W. Pedrycz, "Face recognition using a fuzzy fisherface classifier," *Pattern Recognition*, vol. 38, no. 10, pp. 1717–1732, 2005.
[8] T. Ahonen, A. Hadid and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
[9] A. Hadid, M. Pietikainen and T. Ahonen, "A discriminative feature space for detecting and recognizing faces," *Computer Vision and Pattern Recognition, IEEE Computer Society Conf.*, IEEE, vol. 2, pp. II–II, 2004.
[10] Z. Lu, X. Jiang and A. ChiChung Kot, "Deep coupled resnet for low-resolution face recognition," *IEEE Signal Processing Letters*, vol. 25, no. 4, pp. 526–530, 2018.
[11] F. Schroff, D. Kalenichenko and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 815–823, 2015.
[12] G. B. Huang, M. Mattar, T. Berg and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, vol. 1, 2008.
[13] J. Hays and A. A. Efros, "Scene completion using millions of photographs," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, pp. 4, 2007.
[14] J. Lalonde, D. Hoiem, A. A. Efros, C. Rother, J. Winn *et al.*, "Photo clip art," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, pp. 3-es, 2007.
[15] T. Qiao, J. Zhang, D. Xu and D. Tao, "Mirrorgan: Learning text-to-image generation by re-description," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 1505–1514, 2019.
[16] A. Radford, L. Metz and S. Chintala, "Un-supervised representation learning with deep convolutional generative adversarial networks," ArXiv Preprint, ArXiv: 1511.06434, 2015.
[17] A. Nguyen, A. Dosovitskiy, J. Yosinski, T. Brox and J. Clune, "Synthesizing the preferred inputs for neurons in neural networks via deep generator networks," *Advances in Neural Information Processing Systems*, pp. 3387–3395, 2016.
[18] Y. Zhao, P. Lai-Man, W. Xuehui, L. Kangcheng, Z. Yujia *et al.*, "Legacy photo editing with learned noise prior," in *IEEE/CVF Winter Conf. on Applications of Computer Vision*, Santa Rosa, CA, USA, pp. 2103–2112, 2017.
[19] Y. Ghosh, R. Zhang, P. Dokania, O. Wang, A. Efros *et al.*, "Interactive sketch and fill: Multiclass sketch-to-image translation," in *IEEE/CVF Int. Conf. on Computer Vision*, vol. 2019, pp. 1171–1180, 2019.
[20] M. Liu, T. Breuel and J. Kautz, "Un-supervised image-to-image translation networks," *Neural Information Processing Systems*, pp. 700–708, 2017.
[21] S. Wang, L. Zhang, Y. Liang and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," *Computer Vision and Pattern Recognition (IEEE)*, pp. 2216–2223, 2012.
[22] H. Yamauchi, J. Haber and H. -P. Seidel, "Image restoration using multiresolution texture synthesis and image in painting," in *IEEE Conf. on Computer Graphics Int.*, Tokyo, Japan, pp. 120–125, 2003.
[23] A. V. Oord, N. Kalchbrenner, L. Espe-holt, O. Vinyals, A. Graves *et al.*, "Conditional image generation with pixel cnn decoders," ArXiv Preprint ArXiv: 1606.05328, pp. 1–13, 2016.
[24] H. Kazemi, M. Iranmanesh, A. Dabouei, S. Soleymani and N. M. Nasrabadi, "Facial attributes guided deep sketch-to-photo synthesis," in *IEEE Conf. on Winter Applications of Computer Vision Workshops*, vol. 1, pp. 1–8, 2018.

[25] A. Dosovitskiy, J. T. Springenberg and T. Brox, "Learning to generate chairs with convolutional neural networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1538–1546, 2015.
[26] T. Zhou, S. Tulsiani, W. Sun, J. Malik and A. A. Efros, "View synthesis by appearance flow," in *European Conf. on Computer Vision*, vol. 99, no. 8, pp. 286–301, 2016.
[27] C. K. Sønderby, J. Caballero, L. Theis, W. Shi and F. Huszár, "Amortised map inference for image super-resolution," ArXiv Preprint ArXiv: 1610.04490, 2016.
[28] G. Antipov, M. Baccouche and J. Dugelay, "Face aging with conditional generative adversarial networks," in *IEEE Int. Conf. on Image Processing*, Beijing, China, pp. 2089–2093, 2017.
[29] A. Oord, N. Kalchbrenner and K. Kavukcuoglu, "Pixel recurrent neural networks," *Int. Conf. on Machine Learning*, vol. 48, pp. 1747–1756, 2016.
[30] S. E. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele *et al.*, "Learning what and where to draw," ArXiv Preprint ArXiv: 1606.05328, pp. 1–9, 2016.
[31] P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 1125–1134, 2017.
[32] A. Jamadandi, S. Kottur shettar and U. Mudenagudi, "PredGAN: A deep multi-scale video prediction framework for detecting anomalies in videos," in *11th Indian Conf. on Computer Vision, Graphics and Image Processing*, Hyderabad, India, pp. 1–8, 2018.
[33] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *European Conf. on Computer Vision*, Amsterdam, Netherland, vol. 99, pp. 702–716, Springer, 2016.
[34] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele *et al.*, "Generative adversarial text to image synthesis," in *Int. Conf. on Machine Learning*, New York City, NY, USA, pp. 1060–1069, 2016.
[35] Y. Taigman, A. Polyak and L. Wolf, "Unsuper-vised cross-domain image generation," in *Int. Conf. on Learning Representations*, Toulon, France, 2017.
[36] D. He, Y. Xia, T. Qin, L. Wang, N. Yu *et al.*, "Dual learning for machine translation," *Advances in Neural Information Processing Systems*, vol. 29, pp. 820–828, 2016.
[37] P. Sangkloy, J. Lu, C. Fang, F. Yu and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 5400–5409, 2017.