**University of Sindh Journal of Information and Communication Technology (USJICT)**
Volume 5, Issue 2, July 2021
ISSN-E: 2523-1235, ISSN-P: 2521-5582
Website: http://sujo.usindh.edu.pk/index.php/USJICT/


# Deep Collaborative Model For Mix-Domain Applied To Photo-Sketch Synthesis

[1]Sheeraz Arif, [2]Rajesh Kumar, [3]Shazia Abbasi, [4]Khalid Hussain, [3]Kapeel Dev,

[1]Department Computer Science Salim Habib University
[2]Department of Computing Hamdard University
[3]Department of Computer science, University of Sindh
[4]Department of Computer Science, BUPT, China
sheeraz.arif@shu.edu.pk, rajesh.kumar@hamdard.edu.pk, shazia.abbasi@usindh.edu.pk, wangkhm@bupt.edu.cn, devkapeel22@gmail.com

***Abstract:*** The process of sketch synthesis from real face photos is of great importance in the area of face recognition and remains a challenging issue for law enforcement agencies. Due to the different characteristics between sketch and photo and limited training data, photo/sketch synthesis has become a topic of great concern for the research community. In addition, recent synthesis models are unable to generate a high-resolution realistic photo/sketch. To determine these issues, we propose a novel synthesis framework by employing contrastive loss and generative loss in the form of collaborative loss. This collaborative loss discovers the coherent features between the photos and sketches and recovers the underlying structure which can be helpful to generate high-resolution photo-sketch synthesis. We learn a multi-domain mapping relationship from the sketch-photo mix domain by transferring high-level quality from insufficient photo-sketch training data. The resultant identity-preserved face sketches can be treated as hidden data which can be combined with insufficient original data to recover the deficiency or underlying structure. We perform the qualitative and quantitative experiments on the challenging publically available photo-sketch datasets and yield better performance compared to the existing state-of-the-art framework.

**Keywords:** photo-sketch synthesis; face recognition; contrastive loss; collaborative loss; generative loss; multi-domain;

## I. INTRODUCTION

The face-sketch synthesizing process is a very complex and challenging computer vision task and has been the focus of interest for many researchers in the domain of face recognition which can be applied to criminal investigation or recognition of objects of interest in public areas. In addition, there are other numerous applications including video surveillance, remote system login, passport verification, entertainment and so on. For social media and Smartphone users, photo sketches are getting popularity where people using them as avatars and profile photos. Face photo-sketch process involves the computation of similarity or match between the corresponding photograph in law-enforcement agency databank and the sketch drawn by an artist according to the description of an eyewitness. Usually, face sketches are very useful for the identification of the suspects involved in criminal activities in the case of unavailability of suspect's facial photo or video recording at the crime spot. To handle this situation, law enforcement agencies are in need of alternate ways to catch criminals or suspects.

For the recognition or identification of a person, iris, face, and fingerprints are the three main biometric traits. Among them, iris and fingerprint are the most valuable and mature biometric technologies. However, the scope of these two biometric technologies is very limited in the case of surveillance applications. For the biometric face trait, facial images are captured in a convertible way, so it has significance in the surveillance framework. With the advancement in sensors and digital camera technology, robust and accurate face detection and recognition have become possible which play a crucial role to control, monitor and prevent crimes. However, in unconstrained conditions, accurate face recognition still has some limitations. So, significant research is needed to address these limitations and the development of robust and accurate feature extraction and matching models are required.

The sketch drawn by an artist according to the description of an eye-witness is one of the common methods

used by law enforcement departments to apprehend the suspects. In the context of synthesis models, the sketches are mainly classified into two categories i.e. viewed sketches and forensic sketches. The forensic sketches [Klare et al., 2011; Klare et al., 2013] are usually drawn by an expert according to the description given by an eye-witness. Viewed sketches [Wang& Tang, 2009; Zhou et al., 2012] are drawn on the basis of the available original photo of the subject and containing rich information so, we can obtain better accuracy. However, there so many issues associated with kinds of sketches. Due to the low-quality of hand-drawn sketches, the interpretation is a big problem. In addition, drawing a sketch by an artist is a time consuming process.

Recently, many law-enforcement departments are utilizing some software and synthesized application tools to generate facial sketches of criminals or suspects which is known as composite sketches [Han et al., 2013]. These generated sketches are more robust, flexible and affordable; and widely have been used as an alternative option to the forensic-sketches. Moreover, without any human intervention, some changes can be incorporated in these computer-generated sketches. Most of the models available for generating these sketches are based on mapping or matching between photos and sketches. However, the available photos and sketches may be different from each other in terms of appearance, size, resolution and style. So, the mapping and matching of these two modalities some times are not ideal and practical.

Face photo-synthesis technique is the most common practice to reduce the gap between photo and sketch and transmute photo images into sketch images by learning mapping functions. After this important process identification is performed by employing mutual recognition algorithms. Some researchers have introduced face photo-sketch synthesis models to map the images of both modalities in the same domain using generative adversarial networks (GAN) [Good fellow et al., 2014]. The architecture of GAN for the synthesis method mainly comprises of two steps. First is the generator (G) part which generates the corresponding sketch image for the input face photo image. Another part is known as a descriptor (D) which extracts important features and computes the difference between fake face photo image and available ground-truth face-sketch image.

Many manual and deep learning-based approaches have been proposed in recent years and achieved remarkable results. However, face recognition using photo-sketch images is a very challenging task because mapping or matching between a face-photo and drawn sketch is a very challenging and difficult process. More recently, GAN based models have secured remarkable results in image editing, image translation, image generation, and representation learning .However, these networks have limitations to produce high-resolution images/sketches and do not provide implicit refinement in the network. Moreover, the generated synthesized face-sketches may appear distorted and noisy so in result may be degradation of recognition performance. To resolve this problem, [Zhu et al. 9] proposed CycleGAN architecture by introducing a kind of loss known as a cycle-consistency loss to generate the high-resolution images. However, the cycle-consistency loss is not enough to address this complex problem. There may be consideration of other additional losses that behave as regularization during the learning process and can be beneficial for generating high-resolution useful patch additive sketches. Moreover, due to the unavailability of training data and identity-specific information, it is very challenging to overcome the noise and distortion problems.

In light of the above discussion, this research work proposes a novel framework based on generative modeling technique which introduces a collaborative loss. This loss is the mutual combination of two losses i.e. generative loss and contrastive loss. For the generative loss, we take advantage of U-net architecture as a generator model and learn non-linear multi-domain opposite mapping. Most of the existing models perform the mapping in a uni-directional feed forwarding way. In the proposed framework, we introduce bi-directional opposite mapping i.e. photo domain to sketch domain and vice-versa, and learn the middle domain as noise (loss) which contains the two models more similar, we can name this loss as a generative loss. For the contrastive loss, we input the output sketch image generated by U-net and ground-truth sketch to the Res-net which output is the fine estimation residual sketch image. The nonlinear mapping relationship is highly useful to designate the loss which can be termed as a contrastive loss. We combine these two losses in the form of collaborative loss which can be effective to compensate for the missing or lost underlying structure of the synthesized face-sketch image using the labeled training data.

In addition, we also explore a procedure to solve the problems of limited training labeled data. The introduced method concatenates the generated hidden data and available unique inadequate preparing information as adequate information. We move the elevated level quality data to the lacking preparing information to recuperate the fundamental structure of created shrouded information and acquire the last orchestrated sketch picture with high visual quality. We can sum up the critical commitment of this exploration fill in as follows:

1. We propose a collaborative loss which is the combination of two opposite mapping (bi-directional) and generative mapping thus tune the model more appropriate for photo-sketch synthesis.

2. We combine the generated identity-preserving sketch data with the original insufficient training data and compensate the underlying structure by transferring the needed high-level quality details from insufficient original training data.

3. We conduct quantitative and qualitative experiments to analyze and evaluate the performance and effectiveness of the proposed model validated on challenging benchmark photo-sketch databases and obtained superior results as compared to state-of-the-art models.

The rest of the article is coordinated as follows: Section 2 gives an outline of the connection works. In segment 3, we clarify our proposed system in detail. In Section 4, we exhibit exploratory assessment and conversation. At last, the end is attracted section5.

.

## II. Related Work

Wherever Face recognition using photo-sketch has been the topic of great importance and area of interest for many years and successfully applied to many applications such as criminal investigation, entertainment, and object of interest identification in videos. Many techniques have been proposed in this area which can be divided into two categories:

.

1) The shallow-learning based models

2) Deep-learning based models

The shallow learning-based models transfer manifold from a photo domain to a sketch domain. As our proposed framework is fully based on deep learning networks, so, this section mainly emphasizes the discussion of models related to deep learning-based networks.

### *A.* The shallow-learning based models

The shallow learning-based models transfer manifold from a photo domain to a sketch domain. As our proposed framework is fully based on deep learning networks, so, this section mainly emphasizes the discussion of models related to deep learning-based networks.

### *B.* Deep-learning based models

Profound neural organizations have acquired amazing advancement over shallow learning techniques in numerous applications and a wide scope of issues. The primary capacity of these ways to deal with figuring the planning capacity in the non-straight path between two modalities for example photograph and sketch area. In addition, information and character explicit data canalso be moved from the source space to the objective area. [Zhang et al., 2015] acquainted an end-with end completely convolutional network-based face photograph sketch blend technique. A few different models have likewise been produced for related errands, for example, photograph cartoon interpretation [Zhang et al., 2017],general sketch synthesis [Sangkloy et al., 2016], and the creation of parameterized avatars [Wolf et al., 2017].[Kazemi et al., 2018], applying the coupled deep convolution neural network and learn latent space between photo and sketch domain. In this research, an attribute-centered loss is introduced to train the network in order to match similar facial attributes between the two domains.

Recently, many investigators utilized generative modeling approaches which are highly successful for various image-to-image generation and translation tasks. Gatys et al., 2017] discovered a generator that has the tendency to preserve the detailed pixel information during the transfer of identity-specific information. [Zhang et al., 2016] proposed a fully convolution network (FCN) based model to transfer training samples using generative loss in a photo-sketch mixed domain.

Recently, Generative Adversarial Networks GANs [Good fellow et al., 2014] is employed to synthesize face photo-sketch images by learning the distribution of sample training images. GANs have achieved remarkable performance in image editing, image generation, and representation learning. GANs exploited the concept of game theory having two different competing networks: generator (G) and discriminator (D). The architecture of GAN utilized the generator (G) to generate the sample images from available training samples and discriminator D distinguishes the synthesized sample from the actual distribution. Recently, various models based on GAN have been developed for image-to-image generation and translation tasks. The conditional GANs has been proposed by [Isola et al., 2016]for different tasks such aslabels to facades, labels to street scenes, image colorization. [Zhang et al., 2019a] introduced the dual-transfer method by utilizing the GAN to discover the nonlinear relationship between the face-sketches and high-frequencies of the original photos. [Zhang et al., 2018] utilized the idea of the probabilistic graphic model to strengthen the GAN-based method and introduced a coarse to-fine method. This model has the ability to add sensitive information on the coarse sketches. However, some noises cannot be eliminated from the coarse sketches.

More recently, [Zhu et al., 2017 from sketch, respectively. However, CycleGAN uses only cycle-consistency loss and lacks the regularization process.

Keeping in view the above-presented discussion and highlighted issues, we propose a framework to generate high visualized sketch images with the help of collaborative loss. We transfer the high-level quality information from training data to recover and compensate the underlying structure.

## III. PROPOSED FRAMEWORK

Before This section provides a detailed description of our framework for face-sketch synthesis. The overall framework of the proposed method is demonstrated in Figure 1. Firstly, we briefly define the description of deep learning networks i.e. generative network (U-net) and convolution neural network (Res-net) and their architecture and then we explain the step-by-step procedure for the computation of loss function.

### *A. Deep learning-based networks*

D Recently, deep convolution neural networks provide an overall solution in the domain of photo-sketch synthesis by learning a direct mapping relationship between two modalities. In addition, these networks capture classified information in the feature maps in an end-to-end manner using a sequence of convolution layers as features extractors. In this manner, these networks transform photo-sketch images from source to target channel. In our framework, we utilize two networks i.e. U-Net generator for sketch generation from face photos and Res-Net to get the fine residual images.

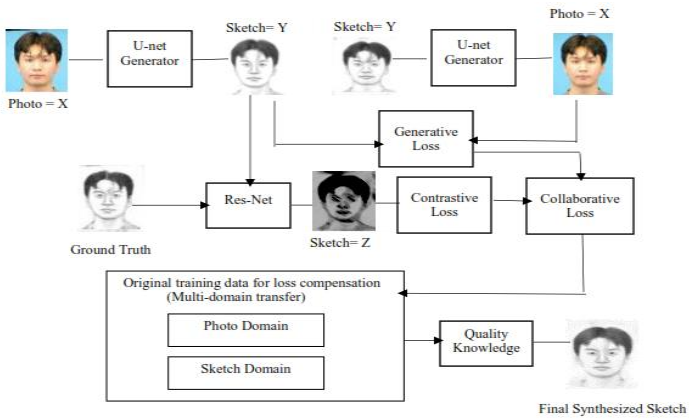


**Figure 1.** Illustration of the overall flowchart of the proposed multi-domain deep collaborative method.

### *B. U-Net Network*

In our framework, we utilize U-Net as a generator network which synthesizes realistic face sketch by computing nonlinear mapping relationship between training dataset and face photo-sketch. We perform a modification in U-Net by introducing the RDB module with the help of dense connected convolution layers to extract more related local features. In order to improve the quality of sketch images, we use instance normalization like [Ulyanov et al., 2016] after each convolution layer in RDB. The generator U-network architecture comprises convolution layers with a size of the kernel 3, ReLu layer and instance normalization. The three convolution layers having the number of filters as 32, 64, and 128, respectively. The stride of all layers is 2 except the first layer with stride 1. The yield is viewed as the contribution of the decoder organizations, which is the backwards cycle of the encoder. The decoder includes three convolution layers with 128, 64, and 32 channels separately. At last, the convolution layer with measurement 3×3 with a solitary channel is utilized to get the objective face sketch picture.

### *C. Residual Network*

We use Res-Net as our secondary network which comprises stacked layers and generates fine residual images using the residual mapping between the ground-label and synthesized sketch. The specialty of this network is that the input image/frame does not need to connect with the next layer, there is the ability of this network to skip forward and move backward to the input layer. Moreover, skip connections allow the signal to propagate from first to the last layer. After individual convolution layer normalization on a batch can be accomplished. In our framework, we utilize Res-Net as our secondary network to generate fine residual images and also we can estimate the contrastive loss using the training samples and obtain synthesized images generated by U-Net. Res-Nets are like VGG networks having small 3 x 3 spatial filters with stride 1. Convolution step is followed by global average pooling and then the fully connected layer is used with soft max for the classification process.

### *D. Loss Functions*

Generative Loss: Most of the existing models for photo-sketch synthesis only consider unidirectional mapping such as the photo to sketch or sketch to photo and resultant image may be of less visualized quality with noise and distortion. We can attain the benefits and generate images with better quality by simultaneously employing two opposite mapping and utilize mutual information. Let consider S is the mapping between photos (X) to sketch (Y) domain and can be represented as S: $Y \rightarrow Z$ . Similarly, G can be denoted as a mapping between sketch (Y) to photo (X) domain and can be defined as G: $Y \rightarrow X$ . Therefore, we can introduce the generative loss to regularize the intermediate representation between the opposite mapping and providing them the same distribution. The process of computation of loss function can be given as:

If S is represented as a mapping between the face-photo channel (X) and face-sketch channel (Y), the generated image can be represented as S(x). Similarly, we can denote the G as a mapping between sketch (Y) and photo (X) domain and generated sketch cab represented by G(y). Thus, the objective functions $\Gamma$ for both domains can be written as follows:

$$\Gamma(F) = \mathrm{E}_{y,z}[\| F(z) - y \|_1]$$

$$\Gamma(G) = \mathrm{E}_{x,y}[\| G(y) - x \|_1]$$

We utilize the L1 distance to avoid the distribution of the representation generated by S and G, so the generated images by both domains look similar. We can utilize the mutual interaction of both domain and represent the generative loss as follows:

$$\Gamma(S,G,F) = \Gamma_g(S,G) + \Gamma(F)$$

Contrastive Loss:For the contrastive loss, we consider the generated sketch S(x) by U-net to the input of our Res-net. The underlying reason for introducing this secondary network is to acquire the fine estimation residual image. There is a high possibility for the occurrence of noise and distortion in the sketch S(x). In order to reduce the noise in S(x), we introduce the Res-net to capture further fine sketch. We can introduce our third domain F: $Y \to Z$ to represent the mapping between S(x) and new domain Z. The output of the Res-net is again in the form of the sketch so we can represent it as a sketch domain (Z). If the generated sketch image is F(z) then the objective function can be computed as:

$$\Gamma(F) = \mathrm{E}_{y,z}[\| F(z) - y \|_1]$$

By combining the generative loss and contrastive loss, we can achieve the full objective as:

$$\Gamma(S,G,F) = \Gamma_g(S,G) + \Gamma(F)$$

In the testing phase, there is always a noise vector n which is produced as an intermediate domain, so intermediate products of domain S, G and F have the same distribution. In the presence of n, we can apply the conventional L1 distance to make the network generates synthesized images closer to the target image, the collaborative loss can be written as:

$$\Gamma_{col}(S,G,F) = \mathrm{E}_{x,y,n}[\| S(x,n) - y \|_1] + \mathrm{E}_{x,y,n}[\| G(y,n) - x \|_1] + \mathrm{E}_{x,y,n}[\| F(z,n) - y \|_1]$$

There is a term $\lambda$ which is the actually the weight of feature loss in each domain and also controls relative importance so if we consider the weight so our final objective can be illustrated as:

$$\Gamma(S,G,F) = \lambda\Gamma_g(S,G) + \lambda\Gamma(F)$$

The different values $\lambda$ may affect the quality of the final synthesized sketch-images. In our experiments and results section, we explore these phenomena furthermore.

### *E. Knowledge Transfer*

The existing training photo-sketch databases are limited and insufficient, only ideal for those face photo-sketches which are captured with simple view position and under the normal lighting. However, in the case of multi-view and abnormal lighting conditions, they are unable to recover the underlying structure. During the synthesizing process distortion or noise may appear and there is a chance of losing some critical information so if we have large and sufficient training data we can generate realistic sketch images.

To address this issue, we utilize the collaborative loss to learn a non-linear multi-domain (photo and sketch) mapping relationship using U-Net and Res-Net. From the training data mapping function can be transferred to the testing data and the hidden data which preserving the characteristics of the test data. The generated hidden data can be combined with original limited training data to form sufficient training data. This sufficient training data can be very handy to recover the lost/bad structure and learn a multi-domain transfer of high-level of useful quality knowledge to generate the final realistic synthesized sketch. To learn the high-level useful knowledge we adopt latent low-rank representation (LLRR) as proposed in [Zhang et al., 2019b].

## IV. EXPERIMENTS AND ANALYSIS

For the verification of the effeteness of our proposed model, the series of extensive experiments have been carried out in the context of the photo-sketch synthesis framework. Two well-known benchmark publically available databases: the Chinese University of Hong Kong face sketch (CUFS) database and CUHK student database have been used. First, we will give the descriptions of each dataset following by experiment setting/implementation detail as well as the experimental results and discussion.

### *A. Datasets and evaluation matrices*

We verify the performance of the proposed framework by conducting extensive experiments on two datasets i.e. The Chinese University of Hong Kong face sketch (CUFS) database [Wang& Tang, 2009] and CUHK student database [Tang & Wang, 2002]. There are 606 well-aligned photo-sketch pairs in the CUFS database which are divided into 268 pairs for training and the remaining for testing. In the CUHK student database, there are 188 pairs with 88 pairs for training and the rest of the subjects are for testing. All the images in both databases are geometrically aligned, however, some images are affected by lighting variations and shape exaggeration.

To assess the presentation of got results, we consider two assessment measurements for example top sign to-clamor proportion (PSNR) and basic comparability file metric(SSIM). PSNR works at pixel-level and identifies the pixel-level degradation between two images. SSIM captures the structural distortion between ground truth images and synthesized images and identifies how much structural information is preserved during the synthesis process. High values of SSIM and PSNR indicate the less noise and high preservation of structural information.

### *B. Implementation Details*

We use 12G Nvidia RTX 2080Ti for training and testing of our framework and model is implemented by PyTorch and Tensor Flow. We utilize CUFS and CUHK student databases for training the network and testing of the proposed model. We crop both images and sketches to 224 × 224 under constrained conditions and perform the normalization to the [-1, 1during both the preparation and testing stages. For the GAN model, we embrace Adam

[Kingma& Ba, 2015] analyzer utilizing back propagation. The preparation ages is 50, and the cluster size is 8. The organization is prepared without any preparation, like the organization instatement arrangement in [Wang et al., 2004], the learning rate is set to 0.0002 for the first 100 ages, and straightly rotting down to 0 for next 100 ages.
For the CNN model, we use Adam with back propagation to optimize our network and the momentum parameter is set to 0.8 and the weight decay parameter is set as $2 \times 10{-}4$. The rate of initial learning is set as $3 \times 10{-}4$ and being divided a half every 1 epochs.

### *C. Results and discussion*

We complete a few trials to asses the presentation of our proposed strategy. In this part, we introduced applicable test results and execution examination.

### *D. Exploration Results*

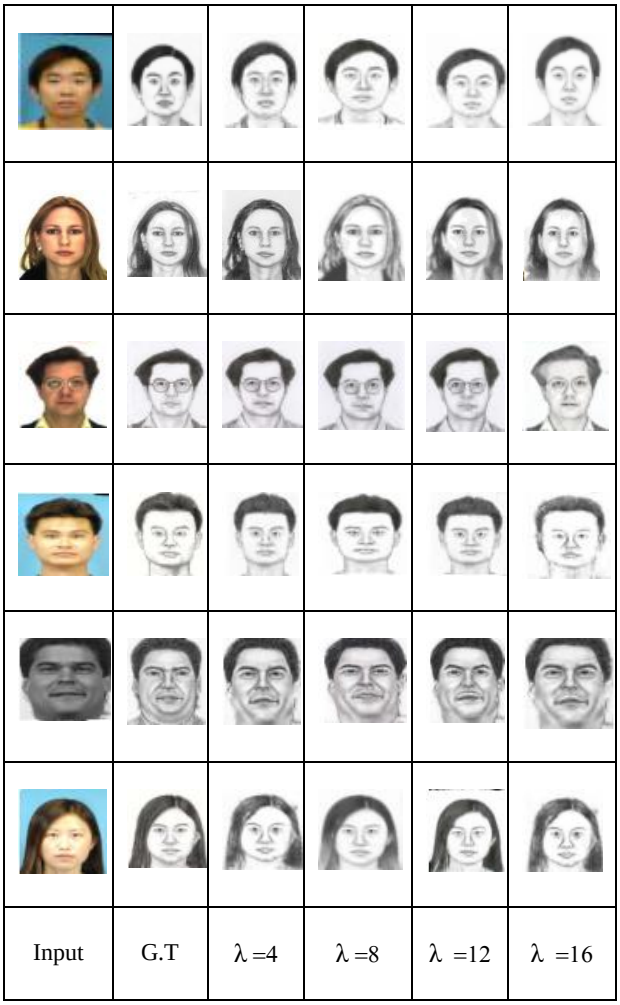


**Figure 2**. Visual results obtain under different values of *L*. Each row from top to bottom: Input, Ground truth and different values of *L*. First three rows are from CUFS and last three rows from the CUHK dataset.

We test our model by employing different exploration aspects. We conducted our experiments on CUFS and CUHK student database. To train the classifier, we randomly select 150 synthesized photos and corresponding ground-truth photos for the CUFS database and rest comprise of the gallery. For CUHK, we randomly select 100 synthesized photos and corresponding ground-truth photos for the CUFS database and rest comprise of the gallery. Firstly, we explore the performance of our proposed method by using different values of λ in a qualitative assessment way. The obtained results are demonstrated in Figure-2. We can notice from the qualitative results that the texture of the results appears clearer with the increase in λ. However, some irrelevant texture distribution decreases the visual effects of achieved results. We found that the results listed in column-4 in Figure 2 are much better and minimize the artifacts.

Furthermore, we also explore the effect of losses such as contrastive loss, generative loss, and collaborative loss and results are illustrated in Table 1. We consider two values of λ i.e. 8 and 12 for each of the loss and compute the average SSIM for all photo-sketch pairs from the CUFS database. We obtain better scores at λ=8 for all losses. This is the reason, we keep λ=8 for the rest of the experiments

| **Loss** | **Contrastive** | | **Generative** | | **Collaborative** | |
|---|---|---|---|---|---|---|
| | λ =8 | λ =12 | λ =8 | λ =12 | λ =8 | λ =12 |
| **SSIM** | 0.482 | 0.471 | 0.592 | 0.583 | 0.643 | 0.631 |

**Table 1**. Average SSIM Score for the three losses under selected values of *L* on CUFS database

Moreover, we use SSIM to compute the quality of synthesized images by considering three losses for our model i.e. contrastive loss, generative loss, and collaborative loss. We perform the evaluation process on all photo-sketch image pairs from the CUFS database. Figure 3 shows the obtained results. SSIM score can be represented by horizontal-axis the vertical axis shows the percentage of synthesized images used for the experiment. From the results, it is obvious that in the presence of collaborative loss, we achieve the best results while other losses are not able to keep the critical details.

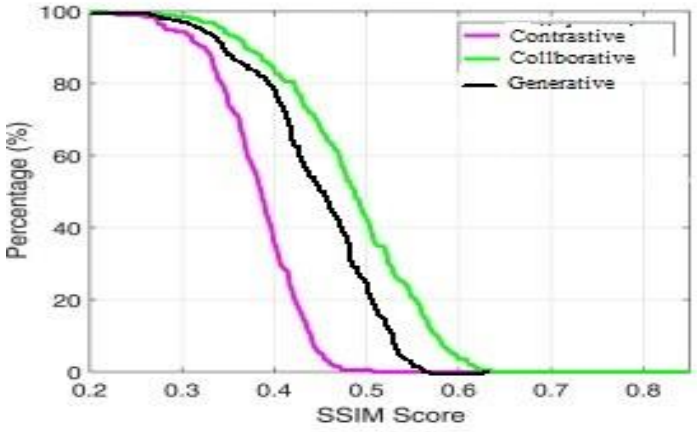


**Figure 3**. SSIM scores by considering three losses on the CUFS database

*E. Face Recognition Accuracy Results*

We also conduct face recognition experiments that can be used as the substitution of a quantitative evaluation metric. It is difficult to investigate the precision and viability of various models by just quantitative measurements, particularly when all the tried models are not all that great as far as target measurements. This the explanation, face acknowledgment is generally utilized as an elective metric that bodes well and considers as a very effective index. High-visualized synthesized images would have high recognition accuracy. For our experiment, we select 189 synthesized sketch images along with their respective ground truth from the CUFS database to learn the classifier. The rest of the images are treated as gallery images. To evaluate the quality of synthesized images, we adopt a null-space discriminative analysis (NLDA) to reduce the number of dimension feature vectors. For the face similarity measurement extracted feature vectors are very important and their variations with reduced dimension may affect the visual quality.

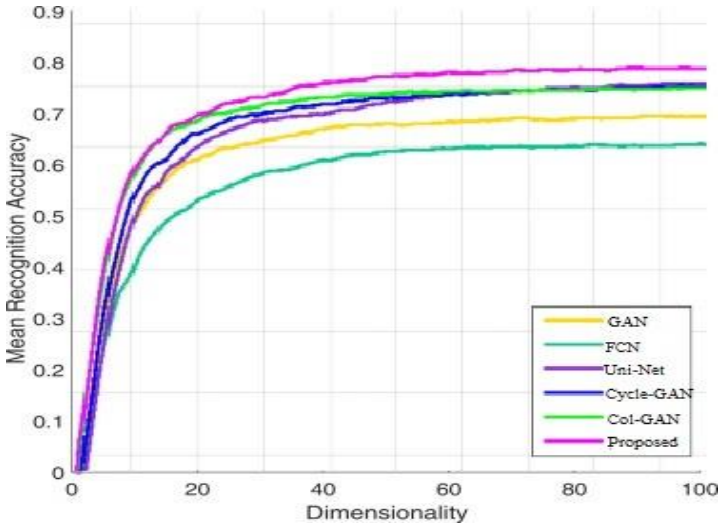


Figure4. The recognition accuracy on CUFS dataset by considering the different reduced number of dimensions

For the comparison, we consider existing state-of-the-art models regression-based and generative models such as Generative adversarial networks (GAN) [Goodfellow et al., 2014], fully convolutional networks (FCN) [Zhang et al., 2015], unidirectional-net (Uni-Net), CycleGAN [Ledig, 2016], Collaborative GAN Col-GAN [Zhu et al., 2019]. We consider extracted features in the form of descriptors generated by these above models. We demonstrate the results in Figure 4 which shows the performance comparison of synthesized photo-sketch recognition using the extracted features by our model and other prominent models. The x-axis represents the variations of the number of reduced dimension of feature vectors and y-axis represent the face recognition accuracy. It can be seen that our model obtains the highest recognition accuracy against state-of-the-art models. Col-GAN, uni-net, and cycle-GAN achieve almost the same recognition accuracy. The obtained results indicate that the extracted face features by the proposed model are much robust and discriminative.

Furthermore, we also represent the average recognition accuracy of the proposed model and other existing prominent models in Table 2. We have given the best average recognition accuracy of all models by considering the reduced dimension at which methods obtain the highest performance. As we can observe that our model achieved the best recognition accuracy of 80.90% at dimension 98 as compared to other methods.

| Models | GAN | FCN | Uni-Net | Cycle-GAN | Col-GAN | Proposed |
|---|---|---|---|---|---|---|
| **Recognition (%)** | 75.12 | 76.12 | 76.45 | 75.89 | 77.90 | 80.90 |
| **Dimensions** | 77 | 93 | 88 | 97 | 96 | 98 |

Table 2. Recognition accuracy based on NLDA on the CUFS database.

*F. Comparison with state-of-the-art models*

This section further verifies the performance and effectiveness of the proposed model, we compare our approach to different successful existing state-of-the-art photo-sketch synthesis approaches on both CUFS and CUHK student databases for all photo-sketch pairs. The obtained results are reported in Tables 3 and 4. For the comparison, we consider state-of-the-art models include: exemplar-based such as local linear embedding (LLE) [Liu et al., 2005] and Markov weight field (MWF) [Zhou et al., 2012]; regression-based model such as Uni-net and FC

N; generative models such as GAN, Cycle-GAN, and Col-GAN. We use the peak signal-to-noise ratio (PSNR) and structural similarity index metric (SSIM) to compute the quality and effectiveness of the photo-sketches synthesized by different methods. The higher values of both metrics indicate that the model preserves the more structural value with less noise. However, in table-4, the average PSNR value of col-GAN is slightly higher. The possible reason is that the quality knowledge transferred by training data overlaps the underlying patches of synthesized images in two or three cases and degrades the performance. Overall, it is crystal clear from the results that the proposed method with the help of collaborative results obtains the best performance. In addition, the mechanism of knowledge transfer from insufficient training data plays a vital role to improve the visual quality.

| Method | CUFS | CUHK |
|---|---|---|
| MWF | 0.539 | 0.619 |
| LLE | 0.525 | 0.601 |
| Uni-net | 0.547 | 0.622 |
| FCN | 0.521 | 0.521 |
| GAN | 0.493 | 0.499 |
| Cycle-GAN | 0.496 | 0.506 |
| Col-GAN | 0.550 | 0.629 |
| Proposed | 0.581 | 0.633 |

**Table 3**. Average SSIM results using the proposed model on CUFS and CUHK student databases

| Method | CUFS | CUHK |
|---|---|---|
| MWF | 16.65 | 17.23 |
| LLE | 16.37 | 16.99 |
| Uni-net | 16.73 | 17.66 |
| FCN | 16.58 | 17.69 |
| GAN | 15.77 | 16.21 |
| Cycle-GAN | 15.90 | 16.86 |
| Col-GAN | **16.77** | 17.72 |

| | | |
|---|---|---|
| Proposed | 16.71 | **17.88** |

**Table 4**. Average PSNR results using the proposed model on CUFS and CUHK student databases

## V. CONCLUSION

This research study introduces a deep learning-based approach for face photo-sketch synthesis. We compute the collaborative loss with the help of generative and contrastive loss by learning a multi-domain cross-mapping relationship. We used two network models i.e. Uni-Net and Res-net and compute two different losses such as generative loss and contrastive loss. The combination of two losses in the form of collaborative loss helps to guide the insufficient training data (photo-sketch) to transfer the high-level quality information to generate more visually pleasing sketch images. In addition, we generate hidden data which we offer a compliment to sufficient training data. We analyze the ability of the introduced framework on two challenging face photo-sketch databases and obtain at par results

## REFERENCES


[1] Chen L.-F.,Liao H.-Y. M., KoM.-T.,LinJ.-C. &Yu G.-J. (2000). A new LDA-based face recognition system which can solve the small sample size problem. *Pattern*Recognit. 33 (10):1713–1726

[2] Gatys L. A., Ecker A. S., &Bethge M. (2017). A neural algorithm of artistic style., arXiv:1508.06576, [Online]. Available: https://arxiv.org/ abs/1508.06576.

[3] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., &Bengio Y. (2014). Generative adversarial nets. In *Advances in NIPS*, pp. 2672–2680.

[4] Han, H., Klare B. F., Bonnen K.; Jain A. K. (2013). Matching composite sketches to face photos: A component-based approach. *IEEE Transactions on Information Forensics and Security (IFS)* 8(1), pp. 191–204.

[5] Isola P., Zhu J.-Y., Zhou T., &Efros A. A. (2016). Image to-image translation with conditional adversarial networks. arXiv:1611.07004.

[6] Kazemi H., Soleymani S., Dabouei A., Iranmanesh M., &Nasrabadi N.M. (2018). Attribute-centered loss for soft-biometrics guided face sketch-photo recognition. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA , pp. 499–507.

[7] Kingma D. P. & Ba J. (2015). Adam: A method for stochastic optimization. In *Proc. 3rd Int. Conf. Learn. Represent*., pp. 1–15.

[8] Klare A., Li Z., & Jain A. K. (2011). Matching forensic sketches to mug shot photos. *IEEE PAMI* 33(3): 639–646

[9] KlareB. F. & Jain, A. K. (2013). Heterogeneous face recognition using kernel prototype similarities. *IEEE PAMI*, 35(6): 410–1422.

[10] Ledig (2016). Photo-realistic single image super-resolution using a generative adversarial network." [Online]. Available: https:// arxiv.org/abs/1609.04802.

[11] Liu Q., Tang X., Jin H., Lu, H. & Ma S. (2005). A nonlinear approach for face sketch synthesis and recognition.IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, USA, pp. 1005–1010.

[12] Sangkloy P., Lu, J., Fang, C., Yu F., & Hays J. (2016). Scribbler: Controlling deep image synthesis with sketch and color. arXiv:1612.00835.

[13] Tang X. & Wang X. (2002). Face photo recognition using sketch. *Proceedings International Conference on Image Processing*,Rochester, NY, USA, USA, pp. 257–260.

[14] stylization," arXiv preprint arXiv:1607.08022.

[15] Wang X. & Tang, X. (2009). Face photo-sketch synthesis and recognition. *TPAMI* 31(11): pp.1955–1967.

[16] Wang Z.,BovikA. C.,SheikhH. R. &SimoncelliE. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process*., 13 (4): 600–612.

[17] Wolf L., Taigman Y., &Polyak A. (2017). Unsupervised creation of parameterized avatars. IEEE International Conference on Computer Vision (ICCV), Venice, Italy.

[18]

[19] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[20] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[21] Zhang M., Wang R., Gao X., Li J., & Tao D. Dual-transfer face sketch–photo synthesis. (2019a). *IEEETrans.ImageProcess* 28 (2): 642–657.

[22] Zhang M., Zhang J., Chi Y., LI Y., Wang N. &Gao X. (2019b). Cross-Domain Face Sketch Synthesis. *IEEE Access* 7: 98866 – 98874.

[23] Zhang S., X. Gao X., Wang N., Li, J. (2016). Robust face sketch style synthesis,'' *IEEE Trans. Image Process* 25 (1): 220–232.

[24] Zhou H., Kuang Z., & Wong, K.-Y. K. (2012). Markov weight fields for face sketch synthesis. *In IEEE Conference on Computer Vision and Pattern Recognition*(*CVPR)*, Providence, RI, USA, pp. 1091–1097.

[25] Zheng Z., Zheng H., Yu Z., Gu Z. &Zheng B. (2017). Photo-to-caricature translation on faces in the wild. arXiv:1711.10735.

[26] Zhu J.-Y., Park T., Isola P., &. Efros A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.*IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy.

[27] Zhu M., Li J., & Wang N. (2019). A Deep Collaborative Framework for Face Photo–Sketch Synthesis." IEEE transactions on neural networks and learning systems 30 (10): pp. 3096-3108.