# Face Sketch Colorization via Supervised GANs

Ramya Y.S., Soumyadeep Ghosh, Mayank Vatsa, Richa Singh
IIIT Delhi, India
{yellapragada15117, soumyadeepg, mayank, rsingh}@iiitd.ac.in

## Abstract

*Face sketch recognition is one of the most challenging heterogeneous face recognition problems. The large domain difference of hand-drawn sketches and color photos along with the subjectivity/variations due to eye-witness descriptions and skill of sketch artists makes the problem demanding. Therefore, despite several research attempts, sketch to photo matching is still considered an arduous problem. In this research, we propose to transform a hand-drawn sketch to a color photo using an end to end two-stage generative adversarial model followed by learning a discriminative classifier for matching the transformed images with color photos. The proposed image to image transformation model reduces the modality gap of the sketch images and color photos resulting in higher identification accuracies and images with better visual quality than the ground truth sketch images.*

## 1. Introduction

Owing to recent instances of terrorism and public disorder, face recognition has become an indispensable tool for law enforcement. Unavailability of photo of the suspect often leads to law enforcement authorities preparing a sketch based on eye-witness descriptions. This sketch is then used to identify the subject by distribution in the media and/or matching with a mugshot database. The problem of face sketch recognition involves automatically matching sketch images to color photos [1, 12]. As shown in Figure 1, domain difference of the two input modalities, color photo and hand-drawn sketch, is particularly large with the abstraction, distortion and exaggeration of the sketch artists and the witnesses. In addition to that, a witness may have only fleetingly observed the suspect, augmented with the fact that they slowly lose memory over time [21], may result in sketches which might have incorrect/incomplete facial details of the suspect. The above points make it one of the toughest problems in heterogeneous face recognition (HFR).

978-1-7281-3640-0/19/$31.00 ©2019 IEEE

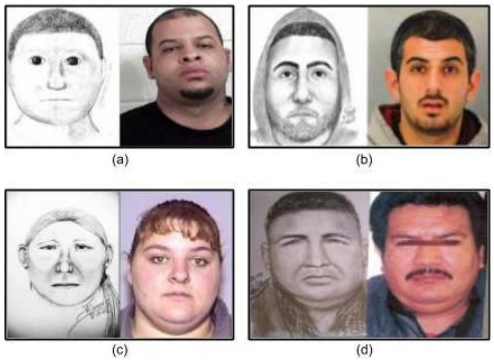



Figure 1: Sketch to photo matching is challenging due to their modality gap with color photos and inherent subjectivity of hand-drawn sketches. (a)-(d) are hand-drawn sketches (left) by a sketch artist along with their corresponding color photos (right).

Based on whether the person drawing the sketch has viewed the suspect themselves, sketches are classified into three categories: (1) Viewed sketches, (2) Semi-forensic (semi-viewed) sketches, and (3) Forensic (unseen) sketches. Existing literature in face sketch recognition generally involve projecting sketches and color photos into a common subspace where they can be matched efficiently [12, 20] or learning new descriptors [18] that remain invariant across the two modalities.

Memory is one of the core processes involved in the complex task of human cognition. Experiments reported in literature suggest that a person is more likely to recall a color version of a scene than its grayscale counterpart [2, 6]. Drawing inspiration from this observation, this paper aims to validate if this phenomena extends to automated algorithm for matching sketches. We propose a novel algorithm to colorize a hand-drawn sketch image to its colour counterpart and use them for recognition process. The contributions of this work are as follows:

- We propose a two step generative adversarial network (GAN) for image to image translation to convert grayscale sketches to color photos. In the first step, a grayscale sketch is converted into grayscale photo followed by its colorization. Thereafter, the color photo is matched to a gallery of color photos by learning a

discriminative deep CNN model.

- Extensive evaluation is carried out on the proposed method quantitatively in terms of recognition performance on viewed, semi-forensic, and forensic sketches and visual quality of the transformed sketch images using no-reference image quality metrics. The proposed method is shown to significantly improve sketch to photo recognition performance on the IIITD Sketch Database [1].

## 2. Previous Work

One of the earliest face sketch recognition papers by Uhl *et al.* [24] located facial features in both sets of images, photometrically standardized the sketch followed by geometric standardization of both and used eigenanalysis to compare the two images. Gao *et al.* [7] looked into ensemble of embedded Hidden Markov Models for transformation followed by eigenface based classification. Xiao *et al.* [28] used an HMM transformation model followed by kNN classification. Wang *et al.* [25] and Liu *et al.* [14] explored dictionary learning based methods for learning a mapping between the two domains. Forensic sketch recognition was addressed by Klare *et al.* [12] in which scale invariant feature transform (SIFT) features and multivariate local binary patterns (MLBP) were used to perform a local feature discriminant analysis. Subsequent works addressed recognition for software generated composite sketches by Han *et al.* [9] and Chugh *et al.* [3] where handcrafted features like MLBP, and HOG were used to extract features for matching. Bhatt *et al.*[1] proposed multiscale circular Weber local descriptor along with a memetic algorithm that improves the identification accuracies for sketch recognition. Mittal *et al.* [17] proposed a fusion based framework using HOG and DAISY features with attribute feedback to perform recognition on composite sketch images.

Limited work has been carried out in face sketch recognition using learning based algorithms. Mittal *et al.* [18] performed composite sketch matching using learned features by training on face photos using a combination of autoencoder and deep belief network. A framework that generates deep convolutional low level features that can be adapted to a domain, e.g. sketch, near-infrared or thermal was proposed by Pereira *et al.* [4]. Nagpal *et al.*[19] proposed a transform learning based approach to learn a transformation and mapping function between the features of two domains. Di and Patel [5] proposed a three stage pipeline for generating coloured faces from texture attribute information. Iranmanesh *et al.* [10] exploited facial attribute information and leveraged loss functions from facial attributes for identification and face verification tasks.

## 3. Proposed Method

In this section, the proposed algorithm for translating grayscale sketch to color photo is discussed. As shown in Figure 2, a two stage pipeline is proposed to generate a color photo from a sketch image with good structure and quality. The first stage transforms a sketch to grayscale photo and the second stage transforms the generated grayscale photo from the previous stage to a color photo. Before delving into the details of the proposed approach, a brief background on GANs is discussed.

### 3.1. Generative Adversarial Networks

Generative Adversarial Networksv (GANs) [8] learn to map a latent space to a given target distribution. Let $z$ be the random noise that is fed to the generator $G$, and $x$ be the real image. The discriminator $D$ takes in input data and outputs the probability of the data being real. The discriminator ensures that it maximizes the likelihood when the input, $x$, is real and minimizes the likelihood when, $x$, is synthetically generated by the generator $(G(z))$. The generator on the other hand minimizes the likelihood of the discriminator correctly predicting the generated image is fake. The final loss function, also called as the adversarial loss is of the form:

$$\min_G max_D V(D,G) = E_x[log(D(x))] + E_z[log(1-D(G(z)))] \quad (1)$$

### 3.2. Proposed Two Stage Image Translation

The sketch to grayscale photo stage presents a cyclicGAN style architecture which takes as input a grayscale sketch and outputs a grayscale photo. The blue and red GAN models (Figure 2) are part of the first stage that transform sketch to grayscale photo and grayscale photo to sketch respectively. The losses employed to train the models are annotated against each stage in Figure 2. In the second stage (grayscale photo to color photo), the output grayscale photo from stage 1 is sent into an image to image translation model (learned using grayscale and ground truth color photos), which takes in an additional noise vector and outputs a color photo. The output color photos are then utilized to learn a discriminative CNN model using triplet loss [23]. During testing, an input sketch image is given to the network which produces a color photo (through its two stage architecture), which is matched to color photos (gallery) with the discriminative CNN model learned. Next, each stage is illustrated in detail, along with the constituent losses that are utilized to learn the models.

#### 3.2.1 S2GSP: Grayscale Sketch to Grayscale Photo Translation

Let us consider $x$ as an image in domain $X$ and $y$, an image in domain $Y$, two generators, $G_{XY}$ and $G_{YX}$ that transform

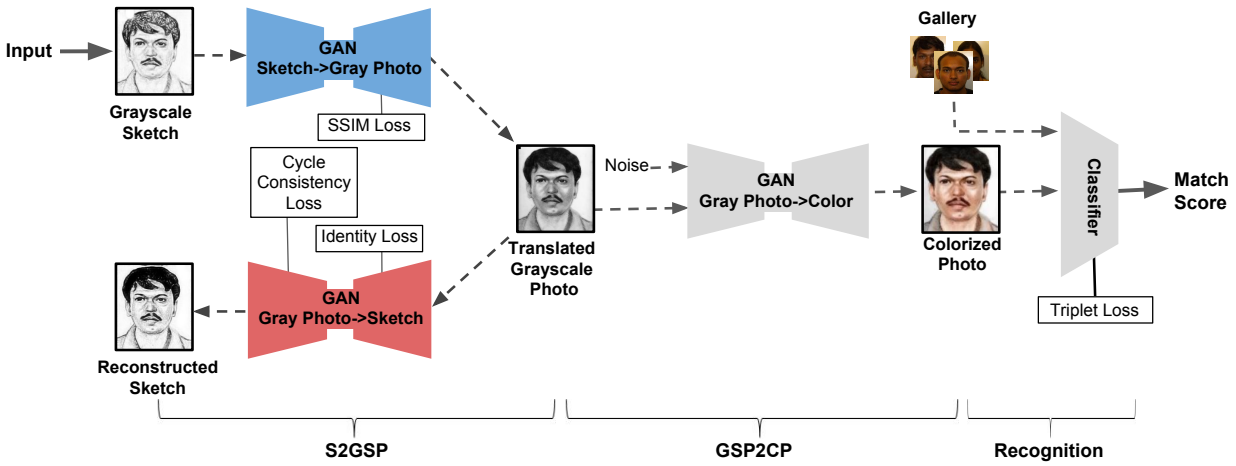



Figure 2: Illustrating the steps of the proposed method. The image to image translation network consists of two distinct stages. The S2GSP stage translates sketch images into grayscale photos (using a cyclic architecture of two GANs colored in blue and red), and the GSP2CP stage translates grayscale photos to color photos. The images produced by the GSP2CP stage are then used for matching with a gallery of color photos using a classifier trained by the triplet loss.

images from domain $X$ to $Y$ and $Y$ to $X$ respectively. Similarly, consider two discriminators, $D_Y$ and $D_X$ that distinguish real $Y$ from fake $Y$ and real $X$ from fake $X$ respectively. Losses employed in this stage are as follows.

**Cycle Consistency loss:** The idea behind cycle consistency is that the image in domain $Y$ formed after $G_{XY}(x)$ should give the original image, $x$, when it is passed through $G_{YX}$, that is, $x = G_{YX}(G_{XY}(x))$. Cycle consistency loss for the translation from $X$ to $Y$ is,

$$L_{cycXY} = E_x[G_{YX}(G_{XY}(x)) - x] \tag{2}$$

**Structural Similarity Loss:** Structural similarity (SSIM) [26], measures the image quality given a reference image. It compares the two windows, $a$, $b$ of same size from two images on the basis of three measurements:(i) luminance, $l(a,b)$, (ii) contrast, $c(a,b)$, and (iii) structure, $s(a,b)$ *.

A linear combination of luminance, contrast and structure is utilized to calculate the SSIM value of the two patches. The numerical range of SSIM is from $-1$ to 1, where the value is 1 when the image being compared is the same as the reference image. However, negative SSIM error can be misleading to the network, hence the following equation is used to calculate SSIM loss to ensure positive loss and direct proportionality of image quality and loss value,

$$L_{ssimX} = \frac{1 - SSIM(y, G_{XY}(x))}{2 + SSIM(y, G_{XY}(x))} \tag{4}$$

*The three measurement functions are calculated as:

$$l(a,b) = \frac{2\mu_a\mu_b + z1}{\mu_a^2 + \mu_b^2}, \quad c(a,b) = \frac{2\sigma_a\sigma_b + z2}{\sigma_a^2 + \sigma_b^2}, s(a,b) = \frac{\sigma_{ab} + \frac{z2}{2}}{\sigma_a\sigma_b + \frac{z2}{2}} \tag{3}$$

where, $\mu_a$ and $\mu_b$ be the averages, $\sigma_a^2$ and $\sigma_b^2$ be the variances of the windows $a$ and $b$, $\sigma_{ab}$ be the covariance of the two windows, R is the dynamic range of pixel values $2^{bitsperpixel} - 1$, $z_1 = (0.01 * R)^2$, and $z_2 = (0.03 * R)^2$.

**Supervised loss:** To ensure that the network does not corrupt important discriminative information in the process of image translation, a deep supervised loss is employed. Let $F(x)$ be the embedding of input $a$ from a model $F$. The $l_2$ loss of the embeddings of two images $a$ and $b$ is calculated as the supervised loss. The supervised $l_2$ loss between $x$ and $G_{YX}(y)$ can be written as:

$$L_{supX} = (F(x) - F(G_{YX}(G_{XY}(x))))^2 \tag{5}$$

**Identity Loss:** Ideally, if an image from the target domain, $y$, is passed through $G_{XY}$, the result should be $y$ itself. The identity loss is the $l_1$ difference between $y$ and $G_{XY}(y)$, expressed as,

$$L_{idenX} = ||G_{XY}(y) - y|| \tag{6}$$

When translating image from domain $X$ to domain $Y$, let $L_{ssimX}$ be the SSIM loss, $L_{supX}$ be the supervised Loss, $L_{cycXY}$ be the cycle consistency Loss, $L_{advY}$ be the adversarial loss, and $L_{idenX}$ be the identity loss.

Similarly, when translating image from domain B to domain A, let $L_{ssimY}$ be the SSIM loss, $L_{supY}$ be the supervised loss, $L_{cycYX}$ be the cycle consistency loss, $L_{advX}$ be the adversarial loss, and $L_{idenY}$ be the identity loss. The final loss function for the image translation process is,

$$\begin{aligned} loss = {} & L_{ssimX} + L_{supX} + L_{cycXY} + L_{advY} + L_{idenX} \\ & + L_{ssimY} + L_{supY} + L_{cycYX} + L_{advX} + L_{idenY} \end{aligned} \tag{7}$$

### 3.2.2 GSP2CP: Grayscale Photo to Color Photo Translation

The second stage involves the colorization of grayscale photos obtained from the previous stage. Using a conditional GAN [15], one can condition the generation of images in the target domain. Let $x$ be the input image, and $z$ be the noise vector added, then the objective function becomes,

$$\min_G max_D V(D,G) = E_x[log(D(x))] + E_z[log(1 - D(G(x,z)))] \tag{8}$$

In such a conditional GAN, the discriminator can be further conditioned so that the GAN can directly learn from the required target image *et al.* [11]. Let $y$ be the gallery image in the target domain to which we want $x$ to translate to. The objective function is expressed as:

$$min_G max_D V(D,G) = E_x[log(D(z))] + E_z[log(1-D(G(z)))] + \lambda E_{x,y,z}[\mid y - G(x,z) \mid] \quad (9)$$

#### 3.2.3 Face Recognition

Face sketch recognition is performed using a deep-CNN model. It is trained using triplet loss [23] on the training images of the evaluation dataset. Triplet Loss is an effective deep metric learning loss function for training a discriminative model. A triplet consists of an anchor (probe), a positive sample (correct galley image), and a negative sample (wrong gallery image). Consider one such triplet, an anchor image $A$, a positive image $P$, and a negative image $N$. The loss function can be expressed as,

$$triplet\_loss = \sum [|f(A) - f(P)|^2 - |f(A) - f(N)|^2 + \alpha]_+ \quad (10)$$

where $f(X)$ is the embedding of image $X$, $\alpha$ is the margin parameter, and $[k]_+ = max(0,k)$.

### 3.3. Network Architectures and Implementation

In this section the network architectures of the deep models used for image to image translation and other implementation details are outlined to ensure reproducibility of the work.

#### 3.3.1 S2GSP

S2GSP, similar to the cycleGAN [30], has two generators and two discriminators - one generator translates sketch to grayscale photo and the other translates from grayscale photo to sketch.

**Generator Architecture:** The generator, comprising an encoder, transformer, and decoder, transforms image from domain $A$ to domain $B$. The encoder network generates a feature vector of the image in domain $A$, the transformer consisting of ResNet blocks transforms the feature vector of the image in domain $A$ to a feature vector of image on domain $B$. The output of the transformer is then sent to the decoder which reconstructs the image from the feature vector.

**Discriminator Architecture:** To produce high quality output, patchGANs are proposed which have been previously explored in [11, 13] to preserve texture and/or style. In a patchGAN, the discriminator does not compare the image as a whole, it compares patches of say size $S \times S$ and gives an output as to whether the patch is a real, or fake.

#### 3.3.2 GSP2CP

This stage comprises of one each of generator and discriminator. Along with an input image in domain $A$ the network also takes in a Gaussian noise vector input $z$.

**Generator Architecture:** The generator uses a UNet-256 [22] architecture with skip connections between corresponding layers of encoder and decoder, that is, there exists a connection between layer $i$ and $n-i$ where $n$ is the total number of layers. A connection between two layers in a UNet implies that all channels from layer $i$ are concatenated and given as input to layer $n-i$.

**Discriminator Architecture:** The discriminator used is the same as that of S2GSP as illustrated in section 3.3.1.

#### 3.3.3 Training Parameters

Adam optimizer with learning rate $2 * 10^{-4}$ and $\beta_1$ value of 0.5 is slowly decayed to 0 over the course of 5 iterations in the S2GSP stage and 200 in the GSP2CP stage. Multiplier parameters were used to scale the loss functions in relation to each other in S2GSP stage. The best model has weights of 10, 0.5, 1, 1, 0.5 for cycle consistency loss, identity loss, adversarial loss, supervised loss and SSIM loss respectively.

#### 3.3.4 Training Face Recognition Classifier

A LightCNN [27] model, pretrained on MS-Celeb 1M dataset, is finetuned using the triplet loss. This finetuned model is used to match probe images (sketch images) and gallery (color photos) for face classification. The LightCNN model has 29 convolutional and 4 pooling layers. No hard mining was performed for generating the triplets. The Adam optimizer is used with a learning rate of $10^{-4}$ gradually decreasing to $10^{-6}$.

## 4. Results and Analysis

This section outlines the databases used, protocols followed, and the results and analysis obtained.

### 4.1. Databases

**IIITD Sketch Database [1]:** The proposed algorithm has been finetuned and evaluated on the IIITD Sketch database [1]. The database contains (1) 238 viewed sketches, (2) 140 semi forensic sketches, and (3) 190 forensic sketches.

**Augmented CelebA Database for Training:** Celeb Faces Attributes Dataset (CelebA) [29] is a large-scale face attributes dataset with more than $200,000$ celebrity images and $10,177$ identities. CelebA dataset is used to train the GSP2CP network. For the S2GSP network, there is a shortage of sketch-photo pairs to be used as training data. A new sketch-photo dataset is generated from CelebA using Adobe Photoshop. In order to generate variations in training samples, i.e., to simulate a forensic image setting, one grayscale sketch is mapped to an average of 10 different color photos of the same person. This generates a total of $92,720$ images.

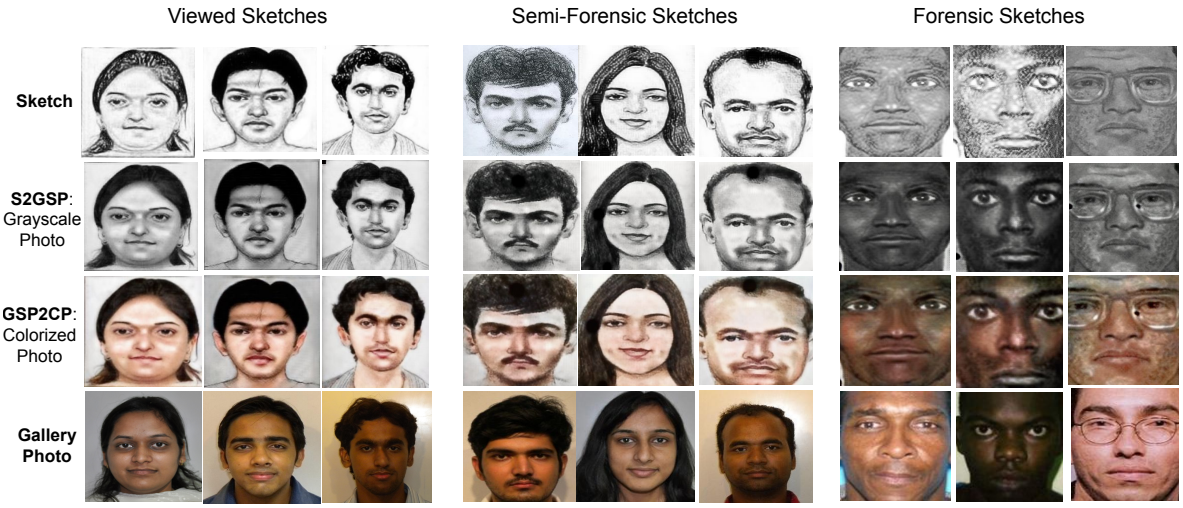
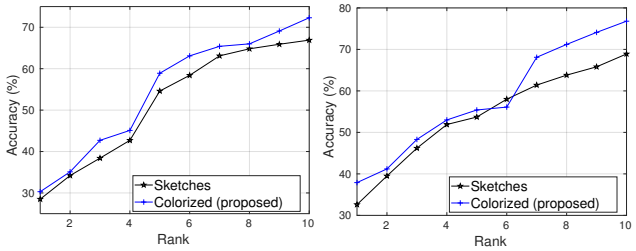



Figure 3: Results of image to image translation using the proposed two stage network. Visual inspection shows that the transformed images (third row) have much lesser domain difference with the gallery images, thus producing better identification accuracies when they are matched with them.

Table 1: Identification accuracies (%) on the IIITD sketch database.

| Source Images | Matched using | Rank 1 | Rank 10 |
|---|---|---|---|
| Viewed Sketches | Grayscale sketch | 32.63 | 68.94 |
| | Colorized Photos | **37.98** | **76.85** |
| Semi-Forensic Sketches | Grayscale sketch | 28.57 | 66.96 |
| | Colorized Photos | **30.35** | **72.32** |
| Forensic Sketches | Grayscale sketch | **5.26** | **33.55** |
| | Colorized Photos | 3.28 | 32.89 |




(a) Semi-Forensic sketches (b) Viewed sketches

Figure 4: CMC curves showing identification accuracies for viewed and semi-forensic sketch recognition.

### 4.2. Results and Analysis

Results are analyzed, both quantitatively (face sketch recognition and image quality metrics) and qualitatively (visual inspection of transformed images) as follows.

**Visual Quality:** On visual inspection (Figure 3) it is observed that the sketch images have significant domain difference with the ground truth color photos. However, the transformed sketches (third row in Figure 3) have significantly lower domain difference with the color photos. A popular no reference image quality metric namely BRISQUE [16] is utilized for image quality assessment. The BRISQUE value (lower score signifies better image quality) of the transformed semi-forensic sketches is 42.61, better than those of the ground truth sketches which is 43.19. Similarly, for the transformed viewed sketches, BRISQUE value is 40.17, and for the ground truth sketches it is 42.88. Thus, it can be established that the proposed method leads to improved quality images.

**Recognition Performance:** The colorized photos produced by the proposed method significantly outperforms (Table 1) the ground truth sketch images in terms of recognition performance. It is observed that the accuracies for the viewed sketches are the highest, as their domain difference with the gallery images are much lesser compared to semi-forensic and forensic sketches. For both viewed and semi-forensic sketches the images produced by the proposed framework yields higher identification accuracies than the ground truth sketch images. However, for forensic sketches the proposed method produces marginally lower identification accuracy (about 1.9% lower at rank 10). This may be due to the high domain difference between forensic sketches and color photos. Another reason could be the subjective nature and inherent variability of the forensic sketches. CMC curves for matching of semi-forensic and viewed sketches to color photos are outlined in Figure 4.

**Ablation Study:** An ablation study is performed by utilizing the images produced from stage 1 (S2GSP) of the image translation network. These images yield rank 10 identification accuracy of 71.43%, 70.74% and 30.18% for viewed, semi-forensic and forensic sketches respectively. These accuracies are lower that what we obtained (Table 1) with the colorized photos from stage 2 (GSP2CP), but higher than ground truth sketches, which shows that the second stage (GSP2CP) for colorization is essential and helps to reduce the domain gap with the gallery images (color photos).

## 5. Conclusion

A two stage framework is proposed for transforming grayscale sketches to color photos using the process of image to image translation by generative adversarial networks. Visual results show that the colorized photos have much lesser domain gap with the gallery images. Quantitative results show that the recognition accuracy of sketches improves after colorization. The proposed method may be utilized to colorize a sketch image for better matching performance (with color photos) or when improved quality images are required for visual inspection.

## 6. Acknowledgements

M. Vatsa and R. Singh are partly supported by the Infosys Center for AI, IIIT Delhi. S. Ghosh is partly supported by TCS Research Fellowship.

## References

[1] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. Memetically optimized mcwld for matching sketches with digital face images. *IEEE TIFS*, 7(5):1522–1535, 2012.

[2] S. Brédart, A. Cornet, and J.-M. Rakic. Recognition memory for colored and black-and-white scenes in normal and color deficient observers (dichromats). *PloS one*, 9(5):e98757, 2014.

[3] T. Chugh, H. S. Bhatt, R. Singh, and M. Vatsa. Matching age separated composite sketches and digital face images. In *IEEE BTAS*, pages 1–6, 2013.

[4] T. de Freitas Pereira, A. Anjos, and S. Marcel. Heterogeneous face recognition using domain specific units. *IEEE TIFS*, 14(7):1803–1816, 2018.

[5] X. Di and V. M. Patel. Face synthesis from visual attributes via sketch using conditional VAEs and GANs. *arXiv preprint arXiv:1801.00077*, 2017.

[6] M. A. Dzulkifli and M. F. Mustafar. The influence of colour on memory performance: A review. *MJMS*, 20(2):3, 2013.

[7] X. Gao, J. Zhong, D. Tao, and X. Li. Local face sketch synthesis learning. *Neurocomputing*, 71(10-12):1921–1930, 2008.

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014.

[9] H. Han, B. F. Klare, K. Bonnen, and A. K. Jain. Matching composite sketches to face photos: A component-based approach. *IEEE TIFS*, 8(1):191–204, 2013.

[10] S. M. Iranmanesh, H. Kazemi, S. Soleymani, A. Dabouei, and N. M. Nasrabadi. Deep sketch-photo face recognition assisted by facial attributes. *arXiv preprint arXiv:1808.00059*, 2018.

[11] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *IEEE ICCV*, pages 1125–1134, 2017.

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

[13] C. Li and M. Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *ECCV*, pages 702–716, 2016.

[14] J. Liu, S. Bae, H. Park, L. Li, S. Yoon, and J. Yi. Face photo-sketch recognition based on joint dictionary learning. In *IAPR MVA*, pages 77–80, 2015.

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

[16] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE TIP*, 21(12):4695–4708, 2012.

[17] P. Mittal, A. Jain, G. Goswami, M. Vatsa, and R. Singh. Composite sketch recognition using saliency and attribute feedback. *Information Fusion*, 33:86–99, 2017.

[18] P. Mittal, M. Vatsa, and R. Singh. Composite sketch recognition via deep network-a transfer learning approach. In *IAPR ICB*, pages 251–256, 2015.

[19] S. Nagpal, M. Singh, R. Singh, M. Vatsa, A. Noore, and A. Majumdar. Face sketch matching via coupled deep transform learning. In *IEEE ICCV*, pages 5419–5428, 2017.

[20] S. Nagpal, M. Vatsa, and R. Singh. Sketch recognition: What lies ahead? *Elsevier IVC*, 55:9–13, 2016.

[21] S. Ouyang, T. M. Hospedales, Y.-Z. Song, and X. Li. Forgetmenot: Memory-aware forensic facial sketch matching. In *IEEE CVPR*, pages 5571–5579, 2016.

[22] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241, 2015.

[23] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE CVPR*, pages 815–823, 2015.

[24] R. G. Uhl and N. da Vitoria Lobo. A framework for recognizing a facial image from a police sketch. In *IEEE CVPR*, pages 586–593, 1996.

[25] S. Wang, L. Zhang, Y. Liang, and Q. Pan. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *IEEE CVPR*, pages 2216–2223, 2012.

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.

[27] X. Wu, R. He, Z. Sun, and T. Tan. A light CNN for deep face representation with noisy labels. *IEEE TIFS*, 13(11):2884–2896, 2018.

[28] B. Xiao, X. Gao, D. Tao, Y. Yuan, and J. Li. Photo-sketch synthesis and recognition based on subspace learning. *Neurocomputing*, 73(4-6):840–852, 2010.

[29] S. Yang, P. Luo, C.-C. Loy, and X. Tang. From facial parts responses to face detection: A deep learning approach. In *IEEE ICCV*, pages 3676–3684, 2015.

[30] J. Zhu, T. Park, P. Isola, and A. Efros. Unpaired image-to-mage translation using cycle-consistent adversarial networks. In *IEEE ICCV*, pages 2849–2857, 2017.