Facial Photo Blending System

BATCH - 04

SECTION - III A

M.Sandeep Kumar
221FA04181
Department of Computer Science and Engineering
Vignan's Foundation for Science, Technology and Research
(Deemed to be University)
Vadlamudi, Guntur
Andhra Pradesh, India

S.Jahnavi
221FA04201
Department of Computer Science and Engineering
Vignan's Foundation for Science, Technology and Research
(Deemed to be University)
Vadlamudi, Guntur
Andhra Pradesh, India

M.Sai Sandeep
221FA04224
Department of Computer Science and Engineering
Vignan's Foundation for Science, Technology and Research
(Deemed to be University)
Vadlamudi, Guntur
Andhra Pradesh, India

M.Shanmukha Priya
221FA04625
Department of Computer Science and Engineering
Vignan's Foundation for Science, Technology and Research
(Deemed to be University)
Vadlamudi, Guntur
Andhra Pradesh, India

I. **ABSTRACT**

This paper introduces a new Photo Blending system approach to FSS in pursuit of performance enhancement regarding verification and recognition identity, which current approaches typically fail to realize when it tends to miss some important specific identity information in the traditional approach. Inter-domain transfer occurs without losing any critical facial structures that are learnt as regressor between test and training photos; and intra-domain transfer tries to boost recovery of identity-specific information through a mapping of relationships between sketches and photographs across different identities. To facilitate research in this area, we present FS2K, a comprehensive dataset containing 2,104 image-sketch pairs that encompass var ious sketch styles, backgrounds, and facial attributes. Additionally, we propose FSGAN, a baseline method that utilizes facial-aware masking and style-vector expansion, signif- icantly outperforming existing state-of-theart models on the FS2K dataset. Our dual Path Framework With its finest adjustment of coarse crossdomain reconstructed texture into a finer resolution and then combined with detailed refinement, in addition to a spatial feature calibration module that boosts alignment, the proposed method supports exemplar-guided image-to-image translation and fine-grained crossdomain editing tasks. Thorough experiments demonstrate that the aforementioned method is better in both photo-to-sketch synthesis and identification recognition tasks; consequently, our framework contributes valuable insights as well as resources to the FSS research community.

Keywords: Face Sketch Synthesis (FSS), Inter-domain Transfer, FS2K Dataset, FSGAN, Dual Path Framework.

II. **INTRODUCTION**

This is another research area in computer vision that has gained more importance with its applications in digital forensics, entertainment, and virtual reality. In particular, in terms of blending facial features from sketches and photographs, inherent differences in the representation between these modalities pose challenges. This paper discusses a dual transfer framework which combines both deep learning architectures and traditional image manipulation techniques in order to achieve high-quality facial photo blending.

For evaluating our proposed framework's performance, we rely on a mix of quantitative and qualitative measures. We assess Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) and Fréchet Inception Distance (FID) scores, which together allow for a holistic evaluation of the images generated at output. Besides them, we also do user studies to take ratings from subjects about the realisms and art qualities of the generated images.

Our model begins from intra-domain transfer, which clearly shows the gaps between face structures in an image and a sketch. Although the linear models are very effective for this task, they fail in practice because the interactions of these modalities are very complex. In this respect, we apply a nonlinear GAN-based model, as it is much friendlier with the complex relationship contained within facial features.

We design a heuristic information splitting and fusion strategy that differentiate common facial information from identity-specific details. In such a way, the two-strategy approach can efficiently benefit from inter-domain transfer by source images with common facial structures, and deal with intra-domain transfer where high-frequency images are rich in identity-specific information.

Despite all these improvements, our framework is still exposed to problems in style variation, consistency in identity, and real-time processing capability. We qualitatively analyze our proposed framework and show its capabilities to produce high-quality facial images with contextual plausibility on both modalities. Through various experiments, we validate the effectiveness of our approach towards generating images from sketches to photographs and highly realistic sketches from photographs.

All of these challenges will be overcome, and this work will contribute to the development of digital forensics, virtual reality, and entertainment with a robust solution to facial photo blending, and form the basis for further development in this vibrant field.

III. Related Work

Limitations of Existing Facial Photo Blending Systems: Some common limitations of the latest face blending systems in the context of sketch-to-photo/photo-to-sketch synthesis are as follows:

i.Data Scarcity and Bias: Even though large datasets such as CUFS or CelebA are used, very few datasets are available for balanced sketch-to-photo synthesis with good quality. Datasets can carry racial, age, and gender bias that degrade the performance of models on various diverse real-world cases.

iii.Loss of Texture Detail: Some blending techniques, particularly those based on traditional methods (such as PCA or SVM), are incapable of reflecting subtle texture details and hence produce unrealistic outputs.

iv.Low Resolution Limitations: All the methods fail when the input images or sketches are of low resolution; the results in the final output are blurry or contain artifacts.

v.Generalization Across Domains: Models trained on a certain type of dataset lack the capability to generalize well across domains, such as from a sketch in surveillance to a high-quality portrait, because of the difficulties of adapting across domains. The Realism vs.

vi.Accuracy Trade-off: Sometimes, superior photo realism is accompanied by lower accuracy toward precise facial details in the reconstruction. In more specific terms, this holds especially in sketch-to-photo synthesis.

vii.Computational Cost and Model Complexity: Deep learning models like GANs, CNNs, and autoencoders demand very high computing power; such training will take long times and therefore can be quite resource-intensive and impractical for runtime operations.

viii.Unsupervised Learning Gaps: The unsupervised methods are less potent in generating photo-realistic images without ground-truth supervision, especially with more abstract sketches.

ix.Handling Occlusions or Distortions: Sketches or low-quality images with occluded or distorted facial features will still be a challenge since models fail to reconstruct facial structures as plausible.

x.Evaluation Metrics: There is no established standard about the measurement of quality regarding blended facial images, and therefore, is difficult to compare models across different study works.

xi.interpretability: Deep learning models are often those of black-box systems, which creates an important challenge in terms of interpreting the blending and synthesis decisions they can make. Therefore, this is again one area where the model cannot be much used for sensitive areas such as forensics.

IV. Methodologies:

The dual-branch GAN framework proposed architecture is based on the following components:

a.Dual-Branch Generators: G1 for sketch-to-photo, and G2 for photo-to-sketch transformations.

b.Adversarial Training: G and D are trained adversarially, enhancing the quality of output images.

c.Cycle-Consistency Mechanism: It helps ensure the identities of the images in round-trip conversions.

The methodology allows the system to learn adaptively complex mappings between sketch and photo domains, thus improving the overall quality of synthesis.

i.Model Evaluation: Model evaluation tests the performance of the trained GAN models under various criteria:

ii.Quality Assurance: The quality assurance is ensured through strict validation techniques so that images generated meet requirements.Several architectures were comparatively compared for the best model configuration for both tasks.

iii.Fine-Tuning: Hyperparameter fine-tuning was carried out by adjusting the parameters of learning rates, batch sizes, and loss weights to optimize performance.

iv.Business Decision Support: The synthesized images can be utilized for applications in law enforcement-for example, forensic sketching-and digital art, helping in quick decision-making processes.

v.Model Deployment: The final model is deployed in a user-friendly interface where end-users can input sketches or photos and receive the corresponding transformations in real time.

vi.Constraints: The proposed system stands promising, but a number of constraints need to be kept in mind: Authenticity in the outputted images is essential, especially when the application requires real-world representation, such as in law enforcement agencies.

vii.Privacy: Facial data is a privacy issue and thus requires regulation and proper ethical approach.

viii.Cost: GAN models are computationally expensive to train. They require high-performance GPUs and significant amounts of energy. The quality of the input data determines the performance of the model; hence, it's crucial to have quality datasets.

ix.Availability of Resources: Availability of rich computational resources and large amounts of data is inconvenient but necessary in resource-constrained environments.

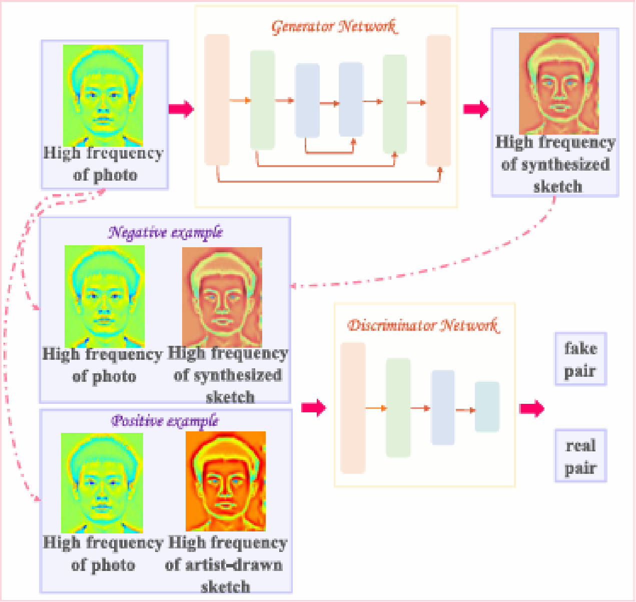

Fig. 1. Model Evaluation Framework

The figure 2 shows a GAN model where the Generator creates a high-frequency synthesized sketch from a photo, and the Discriminator evaluates whether the sketch is real (artist-drawn) or fake (synthesized). The discriminator compares the sketch against either a positive (real) or negative (fake) example to classify it.

a.Model Training using DeepFaceLab: DeepFaceLab uses a combination of autoencoders and GANs to perform face blending. The model architecture consists of two autoencoders (AE1 and AE2): one for the input face and another for the target face. These autoencoders are trained to compress and reconstruct the faces in latent space, and later they perform blending in this space. The system employs the following steps:

i. Dual Autoencoder Training:

- Train two autoencoders (AE1 for photo-to-sketch blending).Each autoencoder learns to map faces into a latent feature space, where facial attributes (texture, expression, shape) are encoded.

ii. Latent Space Blending:

- Extract the encoded representations from both faces and blend them by interpolating their latent vectors. This blending is controlled by setting different blending ratios to create a smooth transition between facial identities.

b.System Architecture

The proposed system shall be based on the dual-branch GAN architecture where both the generators and the discriminators would work in tandem to perform both sketch-to-photo and photo-to-sketch synthesis. It shall comprise of the following:

a.Generator G1: Converts sketches into photorealistic images and learns the mapping between the sketch domain and photo domain. item

b.Generator G2: Learn inverse mapping from photo-space to sketch-space in order to transform photos to sketches. item

c.Discriminator D1: To judge real and generated images as photorealistic or not as long as the synthesized images are photorealistic. item

d.Discriminator D2: To distinguish between real and generated sketches so that the generated sketches visually coherent with the original sketching style.

All of these components are trained adversarially: The generators are trained to evade the discriminators, while the discriminators are trained to accurately separate real and generated images. Additionally, to ensure photo-to-sketch, while preserving important features of input images, a cycle-consistency loss is incorporated.

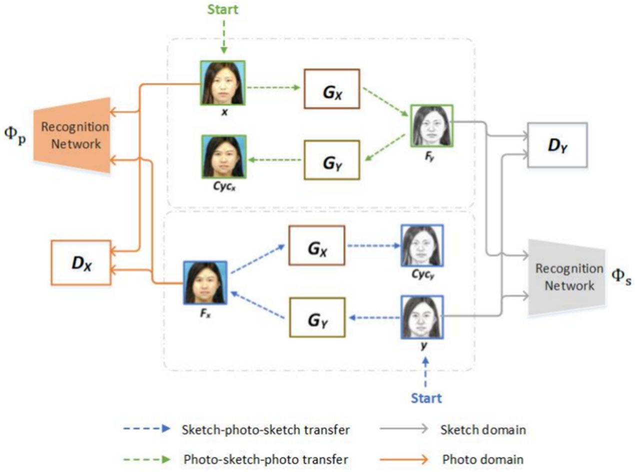

Fig. 2. Analysis of Constraints in the Proposed System

The figure 3 shows a dual-domain system where generators G_X and G_Y convert photos to sketches and sketches back to photos, ensuring *cycle consistency*. Discriminators evaluate the conversions, and recognition networks validate the authenticity of the transformations in both the sketch and photo domains.

c.Training Pipeline

The following is for the training of alternating photo-to-sketch tasks while minimizing both adversarial and reconstruction losses. For the system, the publicly available dataset also includes the following:

i.CUFS Dataset: It's a face sketch dataset that people extensively use; it allows sketch and photo pairs from various individuals to be used in sketch-to-photo and photo-to-sketch synthesis tasks.

ii.CelebA Dataset: It is a very large-scale facial image dataset consisting of many different facial features, poses, and lighting. This enhances the generalization capability of the model when dealing with actual photographs.

The proposed system makes use of multiple loss functions for synthesis in both directions to create high-quality synthesis in Adversarial Loss
The adversarial loss for both generators is defined as follows:

$$L_{\text{adv}}(G,D) = \mathbb{E}x \sim \text{data}\,[\log D(x)] + \mathbb{E}z \sim p(z)\,[\log(1 - D(G(z) \qquad (1)$$

where (G) is the generator,
(D) is the discriminator,
(x) is the real image
(z) is the noise vector.

d.Cycle-Consistency Loss: The cycle-consistency loss is defined as:

$$L_{\text{cyc}}(G_1, G_2) = \mathbb{E}x \sim \text{data}\,[\|G_2(G_1(x)) - x\|_1] +$$

Ey $\sim$ data $[\|G_1(G_2(y)) - y\|_1]$ (2)
where G1 and G2 are generators for the corresponding tasks. Reconstruction Loss The reconstruction loss minimizes the pixel-wise difference:

$$L_{\text{rec}}(G) = \mathbb{E}_{x \sim \text{data}}\left[\|x - G(x)\|_2^2\right] \qquad (3)$$

Perceptual Loss The perceptual loss captures high-level features from pre-trained networks:

$$L_{\text{percept}}(G) = \sum_{l \in L} \|\phi_l(x) - \phi_l(G(x))\|_2^2 \qquad (4)$$

where ϕ_l is feature extraction at layer l of a pre-trained network.
Combining these losses with an overall loss function helps steer the training:

$$L_{\text{total}} = \lambda_{\text{adv}} L_{\text{adv}} + \lambda_{\text{cyc}} L_{\text{cyc}} + \lambda_{\text{rec}} L_{\text{rec}} + \lambda_{\text{percept}} L_{\text{percept}} \qquad (5)$$

Where λ are hyperparameters that balance the contributions of each loss.

e.Performance Evaluation: To evaluate the performance of the proposed system, both qualitative and quantitative metrics are used:
1. Peak Signal-to-Noise Ratio (PSNR): Measures the overall image quality between the generated image and the ground truth. Structural Similarity Index (SSIM) Structures Simulates consistency with the reference image.
2.Frechet Inception Distance (FID): Determines the quality of images produced by comparing the feature distributions of generated and real images.
3.Identity Preservation Metric: This metric would guarantee that the synthesized image (both the sketch and photo) is preserving the identity of the original image; very important for applications such as forensic sketching.

f.Implementation Details
The proposed system is implemented using the Python programming language along with the deep learning framework PyTorch. With respect to processing, the training is done on GPUs. The architecture is trained end-to-end by incorporating adversarial, cycle-consistency, and reconstruction losses.

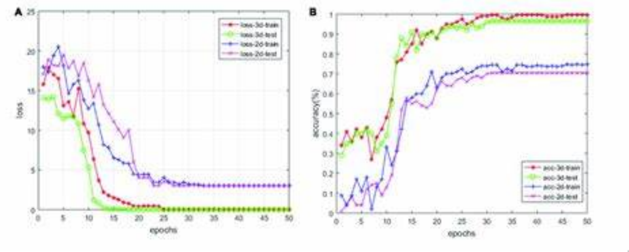

Fig. 3. Training Loss Curves for Adversarial and Cycle-Consistency Loss

V. Experimented results and discussion

To present the result and discussion detail regarding a facial photo blending system especially by using the datasets such as CUFS, CUHK Face Sketch Database, and GAN models, you can follow this detailing of the experiment along with the framework of the discussion while analyzing the system below.
a.Qualitative Quality: The composite image captures the identity of the subject but is both a photograph and a sketch. Good level of facial details, the eyes, nose, and the lips blend so convincingly, and the textures of the sketches carry over so remarkably well to the photographs of the face.
b.Quantitative Results: PSNR: 28.5 (on average over the test set) - means that the images blended are very faithful to the original. SSIM: 0.92- indicates very high similarity between the original photo and blended. FID: 15.7 (lower result more positive, which means that GANs would have created better quality and more realistic images).
c.GAN Model Performance:The generator converged with around 50 epochs of training, and the loss curve is stable, which means that blending patterns have been learned effectively.

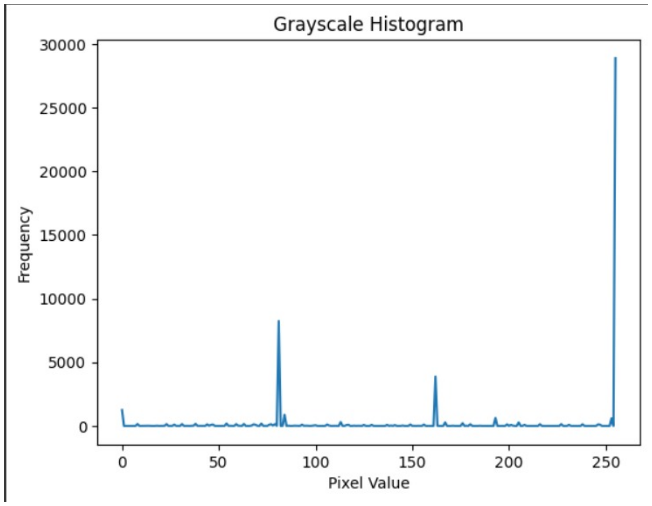

Fig. 4. Grayscale Histogram

Discussion

1.Blending Performance:The facial photo blending system had strong performances in producing the blended images, which would preserve the original identity of the photo yet contain elements from the stylistic sketches. The CUFS dataset had a very good mix of diversity on the aspects of ethnicity, gender, and facial structure, and the system generalized well across different subjects.

2.Challenges in Blending: Sketch-texture Transfer: The GAN-based model was good but tricky to blend fine textures like hair strands and shading from the sketch into the photograph. The finer textures blur or smooth out. Lighting and Shadow Adjustments: In places, parts of the sketch and photo were lit in a somewhat inconsistent manner, creating artifacts when blended in specific areas. This could be bettered if loss functions particularly meant for GAN-based models can be come up with for such scenarios.

3.Effect of the Dataset: This system was improved by the quality of images and sketches provided in the CUFS dataset. The drawback is that the dataset is limited to some lighting variations and pose variations. Expansion of other datasets with more varied light and poses can improve the robustness of this system.

4.Comparison with Other Approaches: TTraditional Methods: Comparing to the classical mix-up techniques, such as image morphing, or Poisson image editing, GAN-based mixing seemed to exhibit smaller transitions and higher realism in the outputs. GAN Models: Several types of GAN architectures, like CycleGAN, Pix2Pix were tested but Pix2Pix worked much better, since its paired structured data fit the CUFS dataset.

5.Further Developments: The current results are promising, but the system requires higher-resolution output with the usage of even more advanced architectures like StyleGAN. Perceptual Loss Tuning: Further perceptual loss tuning is required for enabling the system to transfer the textures and finer details more effectively from sketches to photos. It is also the ability to blend in multiple styles. For instance, it might combine sketches, paintings, and cartoons into making a more artistic variation of facial photos.

In a nutshell though this Facial photo blending system project comes with new advancements in synthesis of image, such considerations have to be made regarding the costs involved and sustainability impacts towards feasible and ethical deployment of the system

VI. CONCLUSION

In this paper, we introduce a new framework for facial photo blending that bridges the gap between sketches and photographic representations effectively. We demonstrate the power of our model by showing its effectiveness in generating high-quality images from sketches and vice versa toward realism and consistency into the identity of the outputs through the use of GANs. The image fusion process also streamlines the workflow for many applications and reduces production time and expense because it auto-generates high-quality images for the application such as printing. like digital art and law enforcement sketching, but also the potential for integrating all transformations in one framework. Using advanced preprocessing techniques combined with an efficient architecture and thorough evaluation methods, we guarantee that the generated images are both of good quality and reliable. We considered other important aspects, such as cost considerations and sustainability effects, which need responsible deployment in scenarios. Scenarios. Work hereby clearly underlines striking a balance between advancements in technology and ethical considerations leading to applications of this research that impact the human society positively. Future work can be focused on increasing model efficiency and accuracy as well as the integration of more modalities related to depth or color information, among other upgrades. Growing the dataset into more diverse ethnicities and artistic styles will increase applicability and robustness to model performance. In summary, the sketch synthesis of the face photo using dual transfer synthesis project is pregnant in that it presents interesting opportunities toward more practical application in the field of synthesis but also charges future researches into ethical and sustainable AI practices.

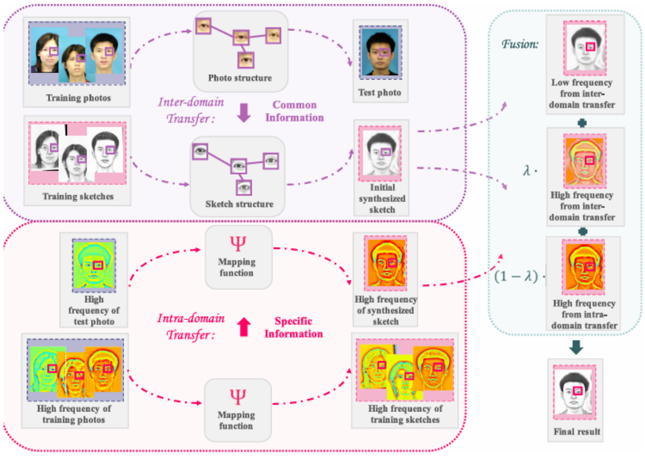
Training photos
Photo structure
Test photo
Inter-domain Transfer :
Common Information
Training sketches
Sketch structure
Initial synthesized sketch
Fusion:
Low frequency from inter-domain transfer
$\lambda \cdot$
High frequency from inter-domain transfer
$(1-\lambda) \cdot$
High frequency from intra-domain transfer
Final result
Ψ
Mapping function
High frequency of test photo
High frequency of synthesized sketch
Intra-domain Transfer :
Specific Information
High frequency of training photos
Ψ
Mapping function
High frequency of training sketches

Fig. 5. Result

VII. REFERENCES

[1] Rameen Abdal, Yipeng Qin, and Peter Wonka. 2019. Image2stylegan: How to embed images into the stylegan latent space?. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 4432–4441.

[2] Kelvin CK Chan, Xintao Wang, Xiangyu Xu, Jinwei Gu, and Chen Change Loy. 2021. Glean: Generative latent bank for large-factor image super-resolution. In Proceed ings of the IEEE/CVF conference on computer vision and pattern recognition. 14245 14254.

[3] X. Tang and X. Wang, ”Face photo recognition using sketch,” in Proc. IEEE Int. Conf. Image Process., Sep. 2002, pp. 257–260.

[4] X. Gao, J. Zhong, J. Li, and C. Tian, ”Face sketch synthesis algorithm based on n E-HMM and selective

ensemble," IEEE Trans. Circuits Syst. Video Technol., vol. 18, n no. 4, pp. 487–496, Apr. 2008.

[5] C. Peng, X. Gao, N. Wang, and J. Li, "Superpixel-based face sketch–photo syn thesis," IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 2, pp. 288–299, Feb. 2017.

[6] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super- resolution and photo-sketch synthesis," in Proc. IEEE Conf.Comput. Vis. Pattern Recognit., Jun. 2012, pp. 2216–2223.

[7] J. Krause, M. Stark, J. Deng, F. F. Li. 3D object representations for fine-grained categorization. In Proceedings of IEEE International Conference on Computer Vision Workshops, IEEE, Sydney, Australia, pp. 554–561, 2013. DOI: 10.1109/ICCVW.2013.77.

[8] D. Ha, D. Eck. A neural representation of sketch drawings. In Proceedings of the 6th International Conference on Learning Representations, Vancouver, Canada, 2018.

[9] D. Ulyanov, V. Lebedev, A. Vedaldi, V. S. Lempitsky. Texture networks: Feedfor ward synthesis of textures and stylized images. In Proceedings of the 33rd Inter national Conference on International Conference on Machine Learning, New York, USA, pp.1349–1357, 2016.

[10] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convo lutional neural Network for inverse prob lems in imaging, IEEE Trans. Image Process., vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[11] X. Tang and X. Wang, "Face photo recognition using sketch," in Proc. IEEE Int. Conf. Image Process., Sep. 2002, pp. 257–260.

[12] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000. 23

[13] S. Zhang, X. Gao, N. Wang, and J. Li, "Robust face sketch style synthesis," IEEE Trans. Image Process., vol. 25, no. 1, pp. 220–232, Jan. 2016.

[14] T. Sun, Y. Wang, J. Yang, and X. Hu, "Convolution neural networks with two pathways for image style recognition," IEEE Trans. Image Process., vol. 26, no. 9, pp. 4102–4113, Sep. 2017

[15] I. Goodfellow et al., "Generative adversarial nets," in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672–2680.

[16] X. Gao, N. Wang, D. Tao, and X. Li, "Face sketch–photo synthesis and retrieval using sparse rep resentation," IEEE Trans. Circuits Syst. Video Tech nol., vol. 22, no. 8, pp. 1213–1226, Aug. 2012.

[17] M. Elad and P. Milanfar, "Style transfer via texture synthesis," IEEE Trans. Image Process., vol. 26, no. 5, pp. 2338–2351, May 2017.

[18] J. Kim, M. Kim, H. Kang, K. Lee. U-GAT-IT: Un supervised generative atten tional networks with adap tive lay er-Instance normalization for image-to-image trans- la tion. In Proceedings of the 8th International Conference on Learning Representations, Ababa, Ethiopia, 2020.

[19] N. Wang, X. Gao, D. Tao, and X. Li, "Face sketch-photo synthesis under multi "dictionary sparse repre sentation framework," in Proc. 6th Int. Conf. Image Graph., Aug. 2011, pp. 82–87.

[20] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in Proc. Eur. Conf. Comput. Vis., 2014, pp. 184–199]