Pix2Pix++: An Enhanced GANs based Model for Portrait to Pencil Sketch Translation

Humza Fazal Abbasi
Department of Computer Science
Bahria University
Islamabad, Pakistan
0009-0004-1881-0391

Merium Fazal Abbasi
Department of Electronics Engineering
CESAT
Islamabad, Pakistan
abbasimerium@gmail.com

Faizan Hamayat
Electrical Engineering Department
University of Engineering and Technology Taxila
Taxila, Pakistan
0000-0002-4417-9511

***Abstract*—The generation of sketches from portraits has been an active area of research for decades due to its applications in various sectors such as entertainment and education. However, due to numerous qualitative factors involved in sketching, programmatically generating sketches from images has been a challenging task. Recent advances in generative artificial intelligence (GAI) seem very promising in solving these challenges. In specific, generative adversarial networks (GANs) like conventional Pix2Pix, were specifically designed for accurate and efficient image-to-image translation tasks. In this research, inspired by GAI and original Pix2Pix, we proposed a GANs-based Pix2Pix++ technique for translating portrait images into sketches. We replaced U-Net with the U-Net++ for generator in Pix2Pix++ GAN and trained the proposed model on a self-collected dataset for sketch generation. Furthermore, we evaluated the proposed model's performance both qualitatively and quantitively and performed a comparative analysis with existing techniques. Proposed Pix2Pix++ GAN got MS-SSIM median scores of 84%, PSNR of 21.17, and MSE median score of 495.55. Experimental results showed our proposed technique outperforms the existing techniques in terms of capturing minute details, tonal variations, realistic appearance, and generalization.**

Keywords*—*GANs, Pix2Pix++, U-Net++, PatchGAN, portrait, sketch, image-to-image translation

I. Introduction

The advancements in computational technology and the emergence of Artificial Intelligence (AI) have transformed the way we create and interact with visual content. AI-based techniques are gaining momentum across all domains of visual content creation. This recent development has not only democratized art but it has also opened new avenues in the artistic realm [1]. Pencil sketching has always been considered an esoteric art. The object has to sit for hours to get their sketch drawn. The person drawing the sketch should either be a gifted person or should have gone through some rigorous training and can be expensive to hire. Therefore, not many people get their sketches drawn through traditional means despite their desire. Along with individual needs, there are some industries as well which rely on sketches such as, visual fiction, which originated in Japan and is now popular worldwide [2].

Efforts have been made for decades to draw artistic sketches using computer programs. Capturing all the minute details of a pencil sketch through a procedural program has been a challenging task. However, with the advent of AI, it has become possible to generate very realistic sketches with a fraction of the effort compared to procedural programs or traditional means [3]. The AI-generated sketches not only have artistic contributions but practical applications as well. They are particularly important for the digital content creation industry which deals with the production of games, animations and comics as this technology can assist artists to quickly transform the images into sketches. In education and research context, the easily available sketches, focusing only on the key areas of images, can highly facilitate the communication of complex data [4].

This research is focused on converting portraits into pencil sketches using an enhanced Pix2Pix GAN, namely Pix2Pix++, making novel contributions in three domains. First is the creation of dataset. A comprehensive and diverse paired dataset of sketches and images was not previously available. Therefore, we created our own paired data of images and pencil sketches. Our dataset is high quality and diverse. The constituent images have diversity in terms of gender, race and ethnicity. The second contribution of this research is the Pix2Pix++ model. All existing studies on generation of sketches from images using Pix2Pix, rely on vanilla U-Net generator. However, we upgraded our generator to U-Net++, which performs better in capturing the fine details [5]. The third contribution of this research is the quantitative and qualitative evaluation of proposed technique. To evaluate our generated sketches, we compared them with the ground-truth sketches using Multi-Scale Structural Similarity Index Measure (MS-SSIM). This paper is divided into six sections. After the brief introduction of the research area, the review of literature is presented. The third section is methodology which elucidates the employed Pix2Pix++ model and implementation strategy. The experiments and results section elaborates on the findings and performance evaluation of proposed technique. Fifth section is the discussion on results. Finally in the last section, it is concluded that our proposed technique outperforms the existing technique found in literature.

II. Literature Review

Before the popularity of AI, image-to-sketch conversion was primarily done through procedural methods that lie within the framework of Non-Photorealistic Rendering (NPR). The algorithms were procedurally designed to simulate various artistic styles, such as pencil or pen-and-ink drawings. Early works focused on generating line drawings and hatchings from 2D inputs and 3D models [6] [7] [8]. However, these procedural approaches often struggled with accurately capturing textures and tonal variations in a way that appeared natural and coherent [9]. With the advent of neural networks, specifically convolutional neural networks (CNNs), the landscape of image stylization was transformed. [10] proposed neural style transfer, enabling the transfer of deep texture statistics from a style image to a content image. This laid the groundwork for many subsequent developments.

However, neural style transfer also faced limitations in sketch generation. It often produced results that lacked the distinct hatching or outline characteristics of sketches [11]. To address these limitations, Generative Adversarial Network (GAN) based methods, including Pix2Pix were introduced. By using a generator and discriminator framework, the quality of generated sketches was iteratively improved. Pix2Pix is found to work well in this domain. The use of paired datasets not only enables precise learning of the mapping between images and sketches, but it also results in outputs that maintain the structural integrity of the original image while effectively emulating the desired sketch style [12]. With the incorporation of techniques like cycle-consistency loss and style control, efforts have been made to address the challenges related to unpaired data and fine-grained stylistic adjustments. Methods like CycleGAN have been explored for sketch generation in scenarios where paired data is not available. However, these approaches struggle to capture complex textural elements inherent in sketches [13]. Apart from that, efforts have also been made to enhance the controllability of the generated sketches. This has led to the development of models that allow for more nuanced style adjustments, enabling users to fine-tune the sketching process to achieve specific artistic effects [14]. Fig. 1 shows the summary of the techniques employed for image to sketch generation, mentioned in literature.

This research work demonstrates the generation of high-quality pencil sketches from portraits using Pix2Pix++. Generally, the availability of paired training data is the greatest obstacle in application of Pix2Pix GANs. However, we have generated the paired training data using a mobile application, which would convert images to sketches. Afterward, we manually adjusted the output of mobile application to get more realistic and high-quality training data. Another significant contribution of this study is the employment of U-Net++ generator, which outperforms the classical vanilla generator in capturing the minute details. Furthermore, the U-Net generator tends to introduce artifacts in the sketches, which is not the case with U-Net++. The comparison of our output sketches with the equivalent sketches generated from other approaches, highlighted the superior performance of our proposed Pix2Pix++ technique.

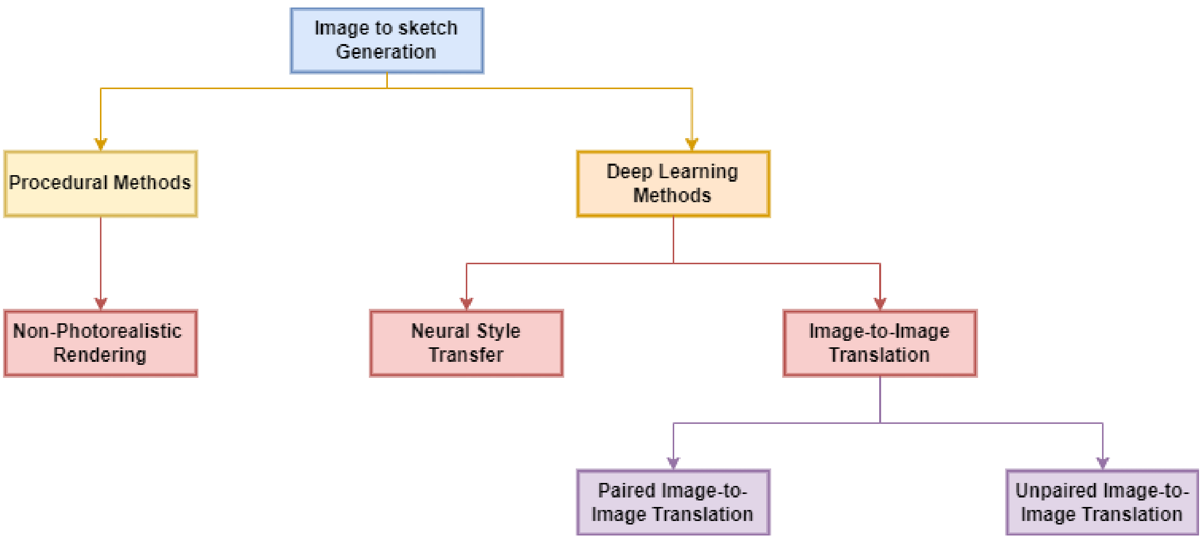

Fig. 1. Techniques for image to sketch generation mentioned in literature.

III. METHODOLOGY

Herein we proposed a GAN based model, called Pix2Pix++, for translating the portraits into pencil sketches. Similar to conventional GANs, the proposed model is comprised of a generator network and a discriminator network. Both of the networks are trained simultaneously in a competitive setting. The generator produces noisy data similar to the input in a particular distribution. The discriminator's job is to differentiate between real data from the actual distribution and the generated data produced by the generator. During training, the generator continues learning by generating more and more realistic data to deceive the discriminator. On the other hand, the discriminator keeps on getting better at identifying generated data from real data. The training of both networks continues until the generator becomes sufficiently capable of fooling the discriminator. The discriminator is unable to distinguish between the predicted data and the actual data. Thereby the generator achieves the desired outcome. [15]. The detailed descriptions of various components of Pix2Pix++ are as described in this section.

A. Generator

The generator takes an input image in one domain and transforms it into a corresponding output image in another domain. In the vanilla Pix2Pix GAN, it is Based on U-Net, a type of convolutional neural network, which features an encoder-decoder structure with skip connections. This allows the model to retain high-resolution details from the input image, thus maintaining the structural integrity of the original image [16]. However, we replaced this U-Net with U-Net++ which enhances the generator's ability to capture finer details. In U-Net, encoder and decoder layers are directly connected by skip connections. On the other hand, U-Net++ uses dense skip connections between immediate layers of encoder and decoder. This allows the network to transfer fine grained information between immediate layers more precisely. These dense connections will allow the generator to produce high resolution sketches compared to the vanilla U-Net. Another key difference between U-Net and U-Net++ is deep supervision. U-Net++ provides deeper supervision training as it can include auxiliary inputs at intermediate stages, leading to faster and more stable convergence [17]. However, in our case, the results produced in the absence of deep supervision were of higher quality. This could be credited to the fact that deep supervision introduces additional complexity to the loss function that may be unnecessary. Therefore, we omitted deep supervision in our final model.

B. Discriminator

The discriminator distinguishes generated images from real sketches. As in the vanilla Pix2Pix, we used PatchGAN as the discriminator. PatchGAN, unlike the conventional discriminators that evaluate the entire image as a whole, operates on a patch level. It assesses the realism of small patches (e.g., 70x70 pixels) within the image instead of the entire image. This approach allows the discriminator to focus on local structures. This is particularly beneficial for tasks requiring fine-grained detail, such as image translation and enhancement. By classifying each patch as real or fake, the PatchGAN discriminator enforces the generator to also consider patch level details. Along with capturing fine details, the PatchGAN architecture also maintains computational efficiency. These factors make PatchGAN a powerful tool for image-to-image translation tasks, where local realism is crucial [18].

C. Loss Function

In Pix2Pix++, we retained the original Pix2Pix loss function, utilizing a composite loss that combines adversarial loss with L1 loss to balance realism and fidelity to the input image. The adversarial loss is derived from the discriminator's feedback. It encourages the generator to produce outputs that are indistinguishable from real images, thus maintaining the overall realism of the generated images. However, adversarial loss alone tends to produce blurry results, particularly in image-to-image translation tasks where pixel-level accuracy is crucial. To address this challenge, Pix2Pix incorporates an additional L1 loss. The L1 loss penalizes the absolute differences between the generated image and the ground truth image. Therefore, it compels the generator to produce outputs that are not only realistic but also a close match to the input image in terms of structure and details. The mathematical expression of loss function is expressed in (1).

$$L_{pix2pix++} = L_{GAN} + \lambda L_{L1} \tag{1}$$

The adversarial loss, $L_{GAN}(G, D)$, is the standard loss used in GANs, where the generator, G(x), tries to generate images from the input images that the discriminator cannot distinguish from real images. $L_{GAN}(G, D)$ is mathematically defined in (2).

$$L_{GAN}(G, D) = E_{(x,y)}[log\, D\,(x, y)] + E_x[log(\,1 - D(x, G(x)))] \tag{2}$$

The generator G(x) tries to minimize this loss, while the discriminator D maximizes it. Along with L_{GAN}, there is L1 loss as well. The L1 loss ensures that the generated image G(x) is close to the ground-truth image, y, in a pixel-wise sense. $L_{L1}(G)$ is defined in (3).

$$L_{L1}(G) = E_{(x,y)}[\|y - G(x)\|_1] \tag{3}$$

L_{L1} penalizes the absolute differences between the generated image and the real image. It encourages the generator to produce outputs that are structurally similar to the ground truth. In the combined Pix2Pix loss function, the two types of loss are combined. λ in (1) is a weight that controls the relative importance of the L1 and L_{GAN}. The combined loss function $L_{pix2pix}(G, D)$ is expressed in (4).

$$L_{pix2pix}(G, D) = L_{GAN}(G, D) + \lambda L_{L1}(G) \tag{4}$$

This combined loss function is designed to make sure that the generator not only fools the discriminator by generating realistic images (adversarial loss) but also maintains fidelity to the input image by minimizing the difference between the generated and real images (L1 loss). In practice, λ is often set to a value like 100 to balance the two losses effectively. Putting it all together, Pix2Pix++ is a specialized architecture of Pix2Pix GAN, designed for image-to-image translation. Its generator is based on U-Net++, and its discriminator is PatchGAN. The loss function has two components i.e. adversarial loss and L1 loss. Fig. 2. illustrates the Pix2Pix++ architecture. In our case, the paired dataset was manually created. After normalization, the Pix2Pix++ architecture was implemented on the dataset. The results were then compared with the output of similar related works.

IV. Experiment and Results

A. Implementation Details

The model was trained on an NVIDIA Tesla P100 GPU for 150 epochs, using the Adam optimizer with a learning rate of $2x10^{-4}$. The L1 loss weight (`L1_lambda`) was set to 100, and a batch size of 16 was used for training. For the U-Net++ generator, we used VGG16 as an encoder with 5 VGG16 blocks. The decoder also consisted of corresponding 5 VGG16 blocks. Both the input and output images were resized to 256×256 pixels, with their pixel values normalized to the range of 0 to 1, instead of the -1 to 1 range used in the

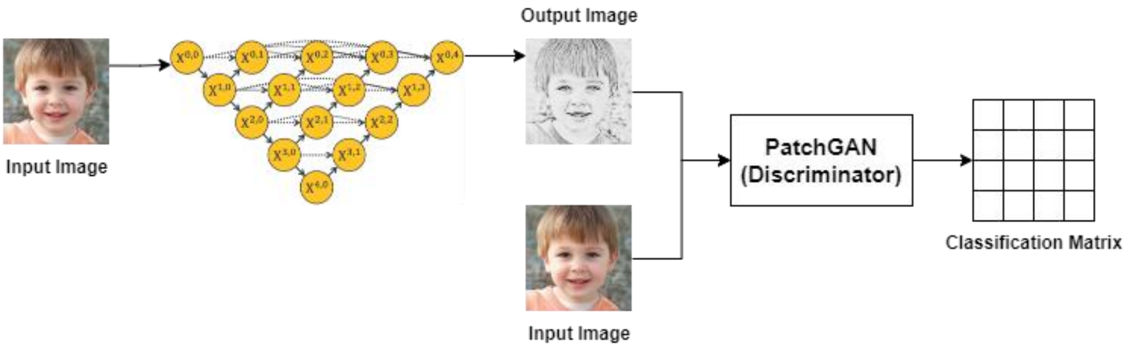

Fig. 2. Graphical presentation of architecture of proposed GAN-based Pix2Pix++ model.

original Pix2Pix paper. This adjustment was made to prevent the Generator from producing small negative values. Although setting the activation function of the final layer to sigmoid could have achieved a similar effect. However, we chose not to limit the Generator, allowing it to explore some negative values when necessary.

B. Dataset

The dataset [19] used in this study consists of 802 original portraits and their corresponding sketches. These portraits were sourced from websites such as Pexels.com, ThisPersonDoesNotExist.org, and Freepik.com. The sketches were generated using the mobile application "Photo Sketch Maker". However, the app's initial outputs were often suboptimal, particularly under varied lighting conditions, which led to inconsistent sketch quality. To address these issues, we manually adjusted the brightness and contrast of the sketches to enhance their realism and ensure consistency across the dataset. Through data augmentation techniques such as rotation and flipping, we expanded our dataset to 5,058 portrait-sketch pairs. The dataset was subsequently divided into train and test sets, comprising 4,299 and 759 pairs, respectively. Diversity was ensured in the dataset by considering portraits from different races, genders and age groups.

C. Evaluation Metrics

Quantitative evaluation of GANs like Pix2Pix presents unique challenges as these models generate outputs that cannot be easily measured by traditional metrics. In unsupervised GANs, the Fréchet Inception Distance (FID) is often employed to measure the fidelity and diversity of generated images [20]. However, in our study, the primary concern is the degree of similarity between the generated sketches and their corresponding original sketches. While human evaluation is a valuable approach for assessing this similarity, we sought a more quantifiable method. Therefore, we employed the Multi-Scale Structural Similarity Index Measure (MS-SSIM) to evaluate the generated sketches [21]. The MS-SSIM is mathematical expressed in (5).

$$MS-SSIM(x,y) = \prod_{j=1}^{M}[l_M(x,y)]^{\alpha_M}.\left[\prod_{j=1}^{M}[c_j(x,y)]^{\beta_j}.[s_j(x,y)]^{\gamma_j}\right] \quad (5)$$

MS-SSIM improves upon the traditional Structural Similarity Index Measure (SSIM) by evaluating image similarity across multiple resolutions. Furthermore, it is more robust and aligns with human visual experience providing an accurate assessment of visual similarity between two images [22]. Additionally, we also utilized Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE) to measure the pixel level similarity between the generated images and their corresponding original portraits. The MSE and PSNR are mathematically expressed in (6) and (7) respectively.

$$MSE_{(x,y)} = \frac{1}{m \times n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(x_{(i,j)} - y_{(i,j)})^2 \quad (6)$$

$$PSNR_{(x,y)} = 20.\log_{10}\left(\frac{(L-1)^2}{MSE_{(x,y)}}\right) \quad (7)$$

D. Performance Evaluation of Proposed Technique

We computed the MS-SSIM scores between real sketches in the test set and their corresponding sketches generated by vanilla Pix2Pix and Pix2Pix++. We then calculated the median of the MS-SSIM scores. Similarly, the median PSNR and MSE values were calculated. Table I. demonstrates that Pix2Pix++ improves over the original Pix2Pix model. Pix2Pix++ achieves superior results across all evaluation metrics.

TABLE I. QUANTITATIVE COMPARATIVE ANALYSIS OF PIX2PIX WITH PROPOSED PIX2PIX++ MODEL

Techniques	Performance Evaluation Metrics		
	MS-SSIM (median)	*MSE (median)*	*PSNR (median)*
Pix2Pix	84.4 %	608.73	20.28
Proposed Pix2Pix++	**88.4 %**	**495.55**	**21.27**

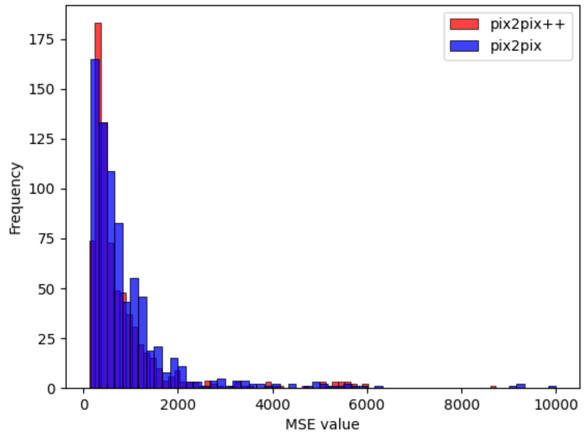

Fig. 3. MSE values of Pix2Pix and Pix2Pix++.

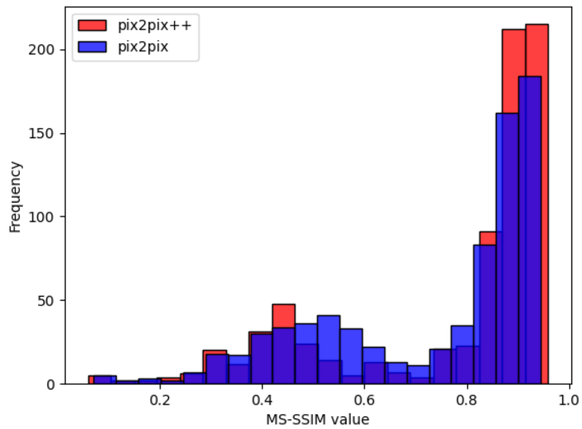

Fig. 4. PSNR values of Pix2Pix and Pix2Pix++.

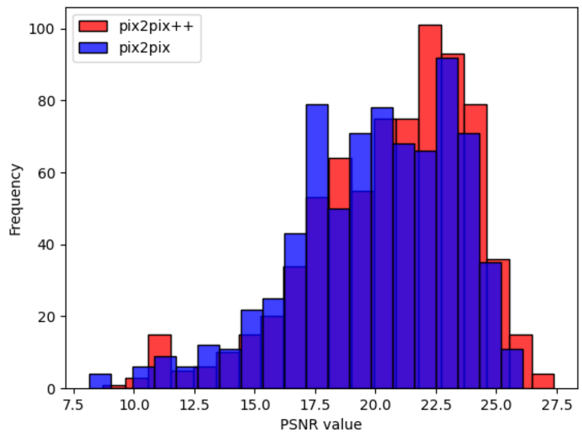

Fig. 5. MS-SSIM values of Pix2Pix and Pix2Pix++.

Fig. 3., Fig. 4. and Fig. 5. illustrate the frequency of MSE, PSNR and MS-SSIM values for both Pix2Pix and Pix2Pix++. It can be observed that Pix2Pix++ generates high quality sketches more frequently compared to Pix2Pix.

E. Performance Comparision with Existing Techniques

In order to compare our technique with the existing techniques we could not use quantitative measures as quantitative evaluation requires ground-truth sketches for reference. We applied existing techniques to four diverse portraits and compared their results with those of our model.

Portrait 1: Face image with less and clear details, good brightness/contrast and uniform light.

Portrait 2: Face image with a lot of distinguishable details (beard, turban, wrinkles) and proper lightening and contrast.

Portrait 3: Face image with low brightness/contrast value and unclear details.

Portrait 4: Face image with varying lightning/tones.

This was done to ensure that the methods could be compared in a wide range of circumstances to best evaluate its effectiveness. Fig. 6. demonstrates the performance of different techniques on the given portraits. The method proposed by Gupta et al [7], in its original form, struggles to capture even the basic facial details. With manual brightness, contrast and highlight correction; the sketches can be improved as a lot of details can be recovered. However, even after the manual adjustments/editing, the sketches still lack the shading which is a distinctive feature of pencil sketching.

The method introduced by [6] generates decent sketches for Portrait 1. For more complex portraits, not only artifacts are being introduced in the sketch, but a lot of facial details are lost as well. This is specifically the case for Portrait 2 and 3, where facial details are unclear due to low contrast/brightness. On the contrary, the shading quality in produced sketches is somewhat decent. The sketches produced with Neural Style Transfer (NST) [10] have a comparatively realistic appearance. The facial details are preserved very well even in difficult circumstances (Portrait 2,3,4). The shading quality is also quite appropriate. However, on closer inspection, one can observe that a lot of artifacts have been introduced especially in Portraits 2 & 3. We have also observed that finding the optimal style photo for each input portrait is a challenging task requiring much trial and error, as different style photos produce different sketches.

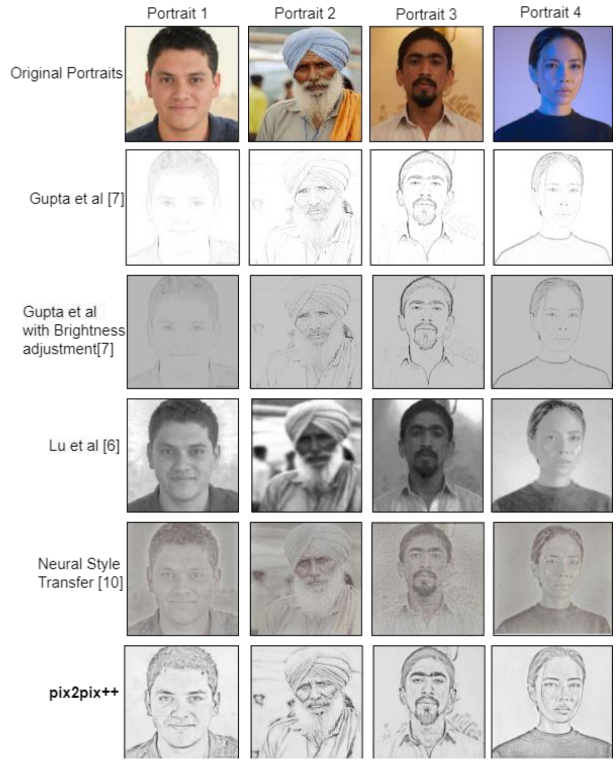

Fig. 6. Comparison of results from different methods.

The sketches produced by our model compare favorably to others. The details are preserved remarkably well, even in challenging scenarios. The shading closely resembles that of real pencil drawings, enhancing realism. Moreover, the sketches exhibit minimal to no artifacts, contributing to the overall quality and authenticity of the results. For reproducibility of results, the link for dataset and code are provided in the [19], and [23] respectively.

V. Discussions

The proposed model demonstrated impressive generalization capabilities. It was observed that the model can produce high-quality sketches even for images that deviated from the training data's distribution. The training data's distribution consisted of face portraits only. Despite this limitation, the model was able to generate visually appealing sketches for images outside the original dataset's scope. Some of the generated outputs on input images outside the training set distribution are given in Fig. 7. This illustrates the Model's robustness and adaptability. It further demonstrates that with appropriate fine-tuning on a specific dataset tailored for a particular task, the model can effectively learn the intricate features and patterns required for that task. It was observed that the vanilla Pix2Pix GAN required 500 epochs to generate artifact-free, high-fidelity outputs. In contrast, Pix2Pix++ achieved even better-quality results in just 100 epochs. This highlights the efficiency and superior performance of Pix2Pix++ compared to the vanilla Pix2Pix GAN.

Fig. 7. Generated sketches with distribution dissimilar from the training data.

979-8-3315-0869-2/24/$31.00 ©2024 IEEE

VI. CONCLUSION

In this research work, we proposed a portrait to sketch conversion method which is based on Pix2Pix++ GAN. The sketches generated by our proposed method are superior in comparison to the existing techniques available in literature. The output pencil sketches are very realistic as our model neither adds extra artifacts nor loses details. Our shading also appears authentic. Furthermore, the model is highly resistant to lightening and contrast variation. It not only performs well with different styles of input images but also gives good results when the input portrait belongs to a distribution on which the model was not trained, thus demonstrating the model's generalization capability. Furthermore, the proposed model generalizes significantly quickly compared to the vanilla Pix2Pix.

REFERENCES

[1] S. Park, "The work of art in the age of generative AI: aura, liberation, and democratization," *AI & society*, May 2024, doi: https://doi.org/10.1007/s00146-024-01948-6

[2] Zhu Mingrui, C. Liang, N. Wang, X. Wang, Z. Li, and X. Gao, "A Sketch-Transformer Network for Face Photo-Sketch Synthesis," Aug. 2021, doi: https://doi.org/10.24963/ijcai.2021/187

[3] X. Cai and B. Song, "Image-based pencil drawing synthesized using convolutional neural network feature maps," *Machine Vision and Applications*, vol. 29, no. 3, pp. 503–512, Jan. 2018, doi: https://doi.org/10.1007/s00138-018-0906-2

[4] P. Xu, T. M. Hospedales, Q. Yin, Y.-Z. Song, T. Xiang, and L. Wang, "Deep Learning for Free-Hand Sketch: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 285–312, Jan. 2023, doi: https://doi.org/10.1109/tpami.2022.3148853

[5] Works CitedBousias Alexakis, E., and C. Armenakis. "EVALUATION of U-Net and U-Net++ ARCHITECTURES in HIGH RESOLUTION IMAGE CHANGE DETECTION APPLICATIONS." The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLIII-B3-2020, 22 Aug. 2020, pp. 1507–1514, https://doi.org/10.5194/isprs-archives-xliii-b3-2020-1507-2020. Accessed 3 May 2022.

[6] C. Lu, X. Liu, and J. Jia, "Combining sketch and tone for pencil drawing production," *Non-Photorealistic Animation and Rendering*, pp. 65–73, Jun. 2012, doi: https://doi.org/10.5555/2330147.2330161

[7] R. Gupta and S. Subedi, "Image to Pencil Sketch Converter using Python," 10th International Conference on Computing for Sustainable Global Development (INDIACom),New Delhi, India, pp. 149–152, 2023.

[8] J. Wang, H. Bao, W. Zhou, Q. Peng, and Xu Yingqing, "Automatic image-based pencil sketch rendering," *Journal of Computer Science and Technology*, vol. 17, no. 3, pp. 347–355, May 2002, doi: https://doi.org/10.1007/bf02947313

[9] D. DeCarlo and A. Santella, "Stylization and abstraction of photographs," *ACM Transactions on Graphics*, Jul. 2002, doi: https://doi.org/10.1145/566570.566650

[10] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," *IEEE Xplore*, Jun. 01, 2016. doi: https://doi.org/10.1109/CVPR.2016.265. Available: https://ieeexplore.ieee.org/document/7780634

[11] Y. Li, C. Fang, A. Hertzmann, E. Shechtman, and M.-H. Yang, "Im2Pencil: Controllable Pencil Illustration From Photographs," *arXiv (Cornell University)*, Jun. 2019, doi: https://doi.org/10.1109/cvpr.2019.00162

[12] J. Henry, T. Natalie, and D. Madsen, "Pix2Pix GAN for Image-to-Image Translation," Aug. 2021, doi: https://doi.org/10.13140/RG.2.2.32286.66887

[13] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," Mar. 2017, doi: https://doi.org/10.48550/arxiv.1703.10593

[14] A. Bansal, S. Ma, D. Ramanan, and Y. Sheikh, "Recycle-GAN: Unsupervised Video Retargeting," *Springer eBooks*, pp. 122–138, Sep. 2018, doi: https://doi.org/10.1007/978-3-030-01228-1_8

[15] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: https://doi.org/10.1145/3422622

[16] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: a Review of Theory and Applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: https://doi.org/10.1109/access.2021.3086020. Available: https://arxiv.org/pdf/2011.01118.pdf

[17] Z. Zhou, M. Rahman, N. Tajbakhsh, and J. Liang, "U-Net++: A Nested U-Net Architecture for Medical Image Segmentation," *arXiv.org*, 2018. https://arxiv.org/abs/1807.10165

[18] C. Li and M. Wand, "Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks," *Computer Vision – ECCV 2016*, pp. 702–716, 2016, doi: https://doi.org/10.1007/978-3-319-46487-9_43

[19] M. Usman, H. Raees and H. F. Abbasi, "Portrait-Sketches-Paired-Dataset". Zenodo, Dec. 11, 2024. doi: 10.5281/zenodo.14391402.

[20] A. Borji, "Pros and Cons of GAN Evaluation measures: New Developments," *Computer Vision and Image Understanding*, vol. 215, p. 103329, Jan. 2022, doi: https://doi.org/10.1016/j.cviu.2021.103329

[21] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale Structural Similarity for Image Quality Assessment," *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, 2019, doi: https://doi.org/10.1109/acssc.2003.1292216. Available: https://ieeexplore.ieee.org/abstract/document/1292216/

[22] U. Sara, M. Akter, and M. S. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study," *Journal of Computer and Communications*, vol. 07, no. 03, pp. 8–18, 2019, doi: https://doi.org/10.4236/jcc.2019.73002

[23] H. F. Abbasi, "Pix2Pixpp Portrait to Pencil Sketch Conversion," Github, 2024, URL: https://github.com/Humzafazal72/pix2pixpp-portrait-to-pencil-sketch-conversion

979-8-3315-0869-2/24/$31.00 ©2024 IEEE