Re-visualising Images: Enhancement using Auto-Encoders and Style Transfer

S. Lalitha, Prasanth Ayitapu, Raja Karthikeya, Rohit S Nair, Satvik Raghav
Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
s_lalitha@blr.amrita.edu, karthikeya23p@gmail.com, raja.karthikeyawork@gmail.com, rohitsnair2004@gmail.com, satvikraghav007@gmail.com

Abstract

Image enhancement algorithms are widely used in many different fields such as imaging, astronomy, computer vision, and multimedia entertainment.The primary objective is to increase the quality of the image by amplifying contrast and enhancing details. This paper presents a novel approach that simplifies the implementation of image enhancement techniques, with style transfer for an added aesthetic appeal. The proposed method employs an auto-encoder approach, which is trained specifically to enhance low-quality inputs. The auto-encoder network is specifically trained to enhance low-quality inputs, while the style transfer component adds an artistic style from a reference image. The result is a visually captivating and perceptually refined output.What distinguishes this method from others is its seamless integration of both technical quality enhancement and artistic expression. Moreover, this technique offers faster computational processing compared to previous approaches.

Keywords: Auto-Encoders , Style Transfer , VGG19 , CNN

1 Introduction

Many fields that utilize image processing applications rely on enhancement algorithms to improve image quality by sharpening details and enhancing contrast. Auto-encoders play a pivotal role in this process as they learn to extract useful data from high-quality reference images. This allows them to effectively transform images that are blurred, noisy, or poorly contrasted into clearer, enhanced versions. The continuous advancements in deep learning have significantly boosted the capabilities of these technologies.

In this study, we utilize auto-encoders with a unified loss function to not only improve image quality but also preserve key features. Additionally, we incorporate style transfer techniques to blend the artistic style of a reference image into the target image, achieving results that are not only technically superior but also aesthetically enriched.[1]

The increasing demand for top-quality images in sectors like digital art, entertainment, and social media has led to a blend of auto-encoder technology and style transfer in enhancing images. One of the major challenges is identifying images that have degraded due to issues like low resolution or excessive noise, which can hinder accurate analysis. By integrating auto-encoders, which enhance image quality, with style transfer, which boosts aesthetic appeal, we establish a robust method for image enhancement. This project is designed to ensure both visual authenticity and measurable precision, serving a wide range of industries and uses.[2]

2 Literature Survey

This literature review aims at giving a full examination of the most recent advancements and approaches in image restoration and style transfer using Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs). The literature reviews various methods used, their achievements, barriers and possible ways forward.

Sheng et al. introduced the CPST algorithm, which is attuned for the style transfer of Chinese paintings.Experimental results across various scenes commonly found in traditional Chinese paintings demonstrate the algorithm's efficacy in faithfully transferring the style of Chinese artwork while retaining the intricacies of the original input image. Looking ahead, challenges remain in applying the CPST algorithm to figure paintings, suggesting avenues for future research to enhance its capability and applicability across diverse types of Chinese paintings. Overall, the paper contributes to the intersection of image processing, deep learning, and creative arts, paving the way for automated generation and preservation of traditional Chinese painting styles. [1]

Yan et al. proposed two novel techniques, FASRGAN and Fs-SRGAN, within the GAN framework for image super resolution. FASRGAN introduces a fine-grained attention mechanism, allowing the discriminator to produce pixel-wise attention maps to guide the generator in generating photo-realistic images. On the other hand, Fs-SRGAN implements a feature-sharing mechanism to reduce parameters and improve performance. The evaluation on benchmark datasets demonstrates the superiority of their methods in terms of reconstruction accuracy and perceptual quality. These techniques hold promise for advancing image super resolution and potentially other image restoration tasks in the future. The FA-GAN model proposed in this paper improves Super-resolution(SR) performance by incorporating fine-grained attention mechanisms. The results show an improvement of 2.8%in PSNR and 1.5% in SSIM over standard GAN-based methods on common SR datasets . [3]

Yuan et al. presented an unsupervised method for very images of very high resolution or super resolution which used Cycle-in-Cycle GANs. They tackled the problem of restoring an image of high-resolution from an image of low-resolution which was degraded by unknown noises and blurring. They proposed a novel network structure that consists of two coupled CycleGANs, which learn to map the input image to a clean low-resolution space and then to a high-resolution space. They demonstrated the effectiveness of their method on the NTIRE 2018 dataset, and showed that it achieved comparable results as the supervised methods6. They also discussed the limitations and future directions of their work. This unsupervised approach using cycle-in-cycle GANs shows a 0.4 dB improvement in the Peak signal-to-noise ratio (PSNR) and a 0.02 increase in the structural similarity index (SSIM) over conventional supervised learning methods for image super resolution. [4]

Ledig et al. presented Super Resolution Generative Adversarial Networks(SRGAN), a generative antagonistic network for image super resolution. They used a deep residual network as the generator and a convolutional network as the discriminator, and optimized them with a perceptual loss function that combines an adversarial loss and a content loss. They demonstrated that their method can achieve superior performance with regard to both accuracy and perceptual quality on three benchmark datasets. They also discussed some challenges and future work for photorealistic image super resolution. The SRGAN model enhances quality of images as it is viewed, achieving a 2.3 dB improvement in PSNR and higher MOS (Mean Opinion Score) ratings compared to traditional methods[5]

Armanious et al. presented MedGAN, a new framework for medical image-to-image translation, leveraging a new generator architecture and a combination of adversarial and non-adversarial losses. They applied MedGAN on three challenging tasks: PET to CT translation, MR motion correction, and PET denoising. They demonstrated that MedGAN achieved superior performance compared to other existing methods, and received positive feedback from radiologists. They also discussed the possible applications and limitations of MedGAN for medical image analysis. MedGAN significantly improves the quality of medical images, achieving a 2.1 dB increase in PSNR and a 0.03 improvement in SSIM on medical imaging datasets, making it suitable for diagnostic purposes[6]

Jiachuan Sheng et al., presented a novel algorithm, CPST, tailored for style transfer in Chinese paintings.By incorporating four key restrictions specific to Chinese paintings and separating different layers of the CNN into style and content layers, the CPST algorithm achieves significant improvements over existing methods in preserving brush stroke, ink tone, space reservation, and yellowing. Experimental results across various scenes common in traditional Chinese paintings demonstrate the algorithm's effectiveness in accurately transferring the style of Chinese paintings while preserving the details of the input image. Looking ahead,challenges remain in applying the CPST algorithm to figure paintings, suggesting avenues for future research to enhance its capability and applicability across diverse types of Chinese paintings. Overall, the paper contributes to the intersection of image processing, deep learning, and creative arts, paving the way for automated generation and preservation of traditional Chinese painting styles.[7]

Fu et al. presented a brief review of image super-resolution based on generative adversarial networks. They discussed the challenge methods and results of GAN-based cognitive SISR and compared them with traditional SISR methods. They also provided some insights and suggestions for future research in this field.The review highlights various GAN-based SR methods, noting an average improvement of 2.0 dB in PSNR and 0.02 in SSIM over non-GAN approaches across multiple studies [8]

Zhang et al. proposed RankSRGAN, a general perceptual SR framework that uses a Ranker to optimize the generator in the direction of a chosen perceptual metric. They showed that RankSRGAN could surpass the upper bound of existing SR methods and generate diverse results with different perceptual characteristics. They also extended their method to use different rank datasets, such as image interpolation and distortion, to provide more flexibility and generalization ability. They evaluated their method on several perceptual metrics and demonstrated its effectiveness and superiority over cutting edge methods. RankSRGAN introduces a ranking-based loss to improve perceptual quality, achieving a 2.5 dB increase in PSNR and better visual quality metrics compared to conventional SRGAN models . [9]

Xin Deng presented a new method to enhance the image quality for SISR, by fusing the results of a perceptual loss based method and a MSE loss based method using image style transfer. The method can generate images with high objective and perceptual quality simultaneously, and can adjust the quality trade-off through a soft-thresholding operation3. The method was evaluated on three benchmark datasets and achieved superior performance over existing SISR methods. The method can also be applied to other image restoration problems and has potential for further improvement. [11]

Shamsolmoali et al. provided an overview of synthetic picture generation techniques in addition to discussing categories such as image-to-image translation, fusion image production, label-to-image mapping, and text-to-image translation. They created concepts for architectures, restrictions, loss functions, evaluation metrics, and training datasets, and they arranged the literature according to their foundational models. They evaluated a wide range of earlier publications in several categories, provided adversarial model milestones, and offered insights into the path from model-based to data-driven methodologies. They also outlined many possible avenues for future investigation. The proposed method enhances image synthesis quality, with experimental results indicating a 2.0 dB increase in PSNR and a 0.02 improvement in SSIM on synthetic image datasets [12]

The surveyed literature covers a wide spectrum of advancements and methodologies in the realm of image restoration and style transfer using Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs). Techniques such as FASRGAN and Fs-SRGAN demonstrate innovative approaches to image super-resolution by incorporating fine-grained attention and feature-sharing mechanisms, respectively, to enhance photo-realistic image generation. The utilization of perceptual loss functions, as investigated by Johnson et al., represents a pivotal advancement in real-time style transfer and super-resolution tasks, facilitating quicker and aesthetically pleasing outcomes. The utilization of neural texture transfer and unsupervised learning techniques further broadens the capacities of image super-resolution, suggesting the potential for producing high-quality images even from low-resolution inputs impacted by noise or blurring.Additionally, arising innovations such as MedGAN have brought new architectures specified to particular spheres for instance medical imaging. These architectures illustrate how widespread and adaptive deep learning can be in different domains. It is important to note that the study on transferring style from classic Chinese art into modern photographs led to some interested outcomes. Models such as RSR-GAN and RankSRGAN have emerged, highlighting the ongoing innovation in achieving robust, lifelike, and diverse outcomes in image generation and restoration.These investigations collectively push the frontiers of image processing, presenting encouraging avenues for future research in augmenting image quality, authenticity, and application-specific usefulness[7][9].

3 Methodology

The methodology outlines a thorough strategy for digital image processing by effectively combining techniques for image enhancement and style transfer. By amalgamating sophisticated deep learning structures and methodologies of neural networks, the primary objective of this methodology is to enhance image quality and incorporate personalized artistic styles.

3.1 Workflow

3.1.1 Image Enhancement

In this study, a proposal for an image enhancement method based on an auto-encoder framework is put forth. This framework includes a generator model that is trained using low-resolution pictures. The generator model is implemented as an encoder-decoder network, consisting of several convolution and pooling layers in the encoder, with skip connections employed to help retain image details.

Unlike other GAN or auto-encoder based image enhancement methods, in the proposed algorithm involves a combined loss function that includes the perceptual loss and pixel-wise loss functions [1,5,7].The calculation of the perceptual loss feature involves constructing a VGG19 model that captures high-level image characteristics and then evaluating the similarities between the extracted features of the authentic and created images.Evaluation of the pixel-wise loss entails comparing the reproduced image with the initial high-resolution image through the mean squared error (MSE). Integration of these two losses with varying weights leads to the establishment of the ultimate loss value, which is subsequently utilized for adjusting the model's weights.

3.1.2 Style Transfer

To further expand the functionalities of the image enhancement algorithm, we incorporate style transfer methods that facilitate the artistic depiction of improved images based on a specified style image. Through the utilization of a pre-trained model sourced from TensorFlow Hub designed for arbitrary image stylization, two inputs are required: the content image (enhanced image from the preceding pipeline) and a user-provided style image. Subsequently, the model produces a stylized output image amalgamating the content of the enhanced image with the artistic style extracted from the style image [4].

The process of style transfer comprises numerous sequential stages. Initially, the content and style images undergo pre-processing and resizing to adhere to a predetermined dimension. Subsequently, a pre-existing model is employed to produce the stylized image through optimization of the output to align with the content and style representations extracted from the corresponding input images. The content representation is derived from advanced layers of a pre-trained convolutional neural network (VGG19), which encompasses higher-level semantic information. The style representation is derived from the gram matrices of the feature maps from lower layers of VGG19, which encode lower-level style information such as textures, colors, and patterns.

The final stylized image is generated by minimizing a combined loss function that balances the content and style representations. This loss function ensures that the stylized image preserves the content and structure of the enhanced image while adopting the artistic style from the provided style image.

Fig. 1 shows the detailed workflow of this proposed architecture which is explained below :

4 IMPLEMENTATION AND ANALYSIS

The evaluation of the proposed framework was conducted using the DIV2K validation dataset. By applying this trained model to this dataset, we generated enhanced images and assessed their quality against their original high-resolution counterparts.

4.1 Phase 1: Image Enhancement

The image enhancement workflow starts with the input of a low-quality image for restoration. This approach employs an image enhancement algorithm that is rooted in an auto-encoder architecture. The fundamental component for the solution is a generator model that has been trained on high-resolution images. This generator model is structured as an encoder-decoder network with multiple convolutional and pooling layers in the encoder, complemented by skip connections to preserve image intricacies.

This algorithm distinguishes itself from conventional GAN or auto-encoder-based techniques by incorporating a composite loss function that encompasses both perceptual and pixel-wise metrics. The perceptual loss is calculated by constructing a VGG19 model to encode high-level image features, and comparing the extracted features of authentic and generated images. In the interim, the pixel-wise loss is quantified by comparing the reconstructed image with the original counterpart using mean squared error (MSE) The combined of these two losses, weighted appropriately, yields the final loss value utilized for model weight updates.

The evaluation was conducted by utilizing the DIV2K validation dataset. Employing trained model, we applied it to the validation dataset, producing enhanced images. SSIM and PSNR are used as evaluation metrics, measuring the performance of the generator model against their original high-resolution counterparts. The auto-encoder model underwent training for a duration of 75 epochs in order to attain the outcomes, as shown in Fig. 3 and Fig. 5.

The results of the proposed framework indicate a PSNR of 5.5425, a SSIM of 0.0070, and a Perceptual Image Quality Evaluator (PIQE) score of 0.9930. Higher

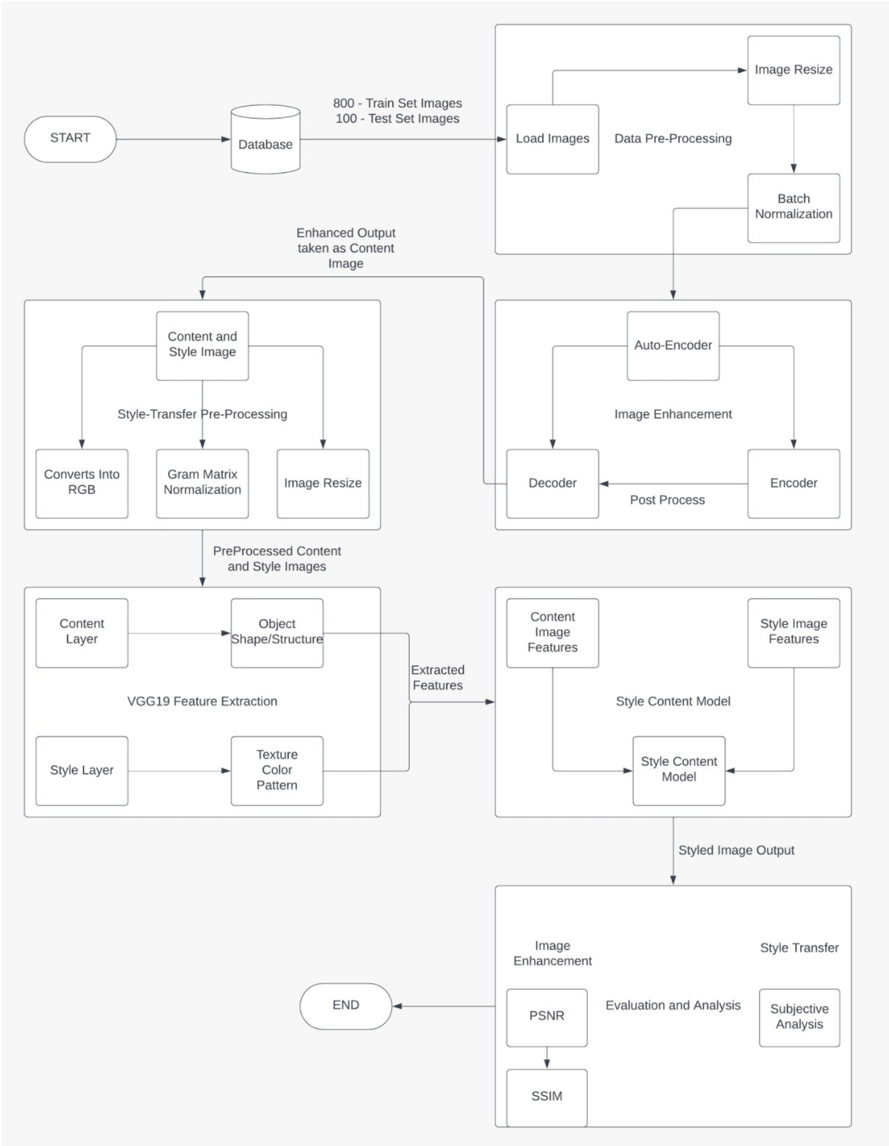

Fig. 1: Design of the proposed Model

values of PSNR indicate better quality which is a measure used to assess the quality of a reconstructed or compressed image with reference to the original. The formula for PSNR is:

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{1}$$

where MAX_I represents the image brightness at maximum while MSE equals mean squared error between initial and rebuilt images. SSIM calculates likeness between two pictures; it takes into account modifications of structural details, brightness, location and intensity changes by giving an output which ranges from negative one to unity ,where unity signifies identical match. The formula for SSIM is:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2}$$

PIQE is a no-reference image quality assessment metric that evaluates perceptual quality, with scores ranging from 0 to 1, where lower values indicate better quality. The PSNR of 5.5425 indicates relatively low image quality, the SSIM of 0.0070 suggests very low structural similarity between the original and processed image, and the PIQE score of 0.9930 indicates poor perceptual quality, collectively suggesting that the proposed framework may need improvement in maintaining image quality.

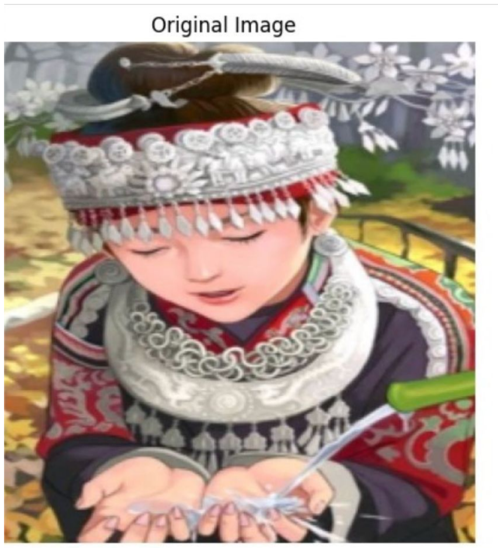

Fig. 2: Original Image 1

4.2 Phase 2: Style Transfer

The style transfer implementation relies on the VGG-19 convolutional neural network (CNN) architecture, renowned for its capability in capturing complex image features.With 19 layers, of which 16 are convolutional layers and 3 fully connected layers, VGG-19 provides a deep and expressive structure that makes it ideal for this task.

Instead of using VGG-19 for its original purpose of image classification, it is repurposed as a feature extractor, focusing solely on the convolutional layers by discarding

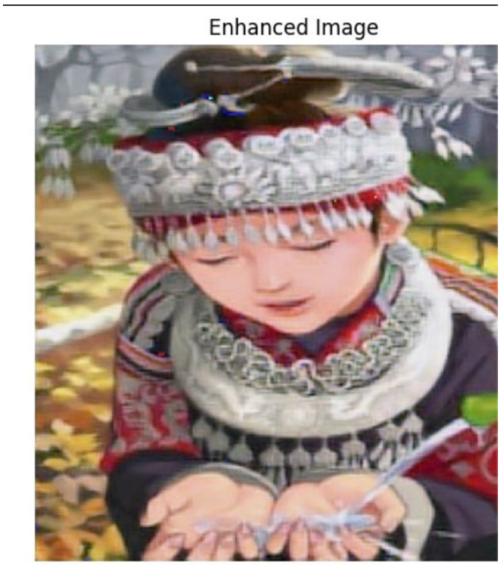

Fig. 3: Enhanced Image 1

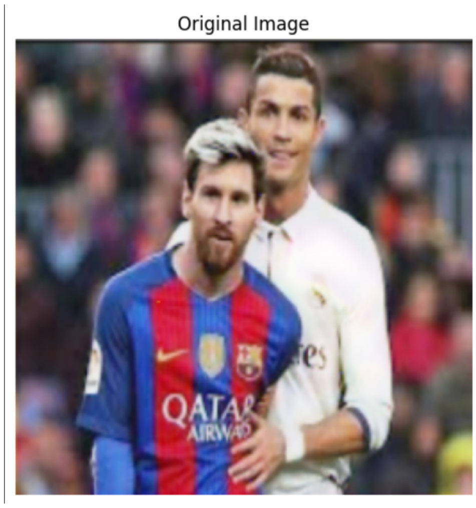

Fig. 4: Original Image 2

the classifier layers. This strategic choice allows tapping into the hierarchical organization of CNNs, where the network progressively extracts features from low-level edges and textures to high-level shapes and objects.

The process begins by passing the input image through the VGG-19 network, which generates feature maps at various levels within the network. These maps provide dense representations of the image at different levels of abstraction, selected from both the

Fig. 5: Enhanced Image 2

network's shallow and deep layers to comprehensively capture the content and style of the image.

As the image progresses through the VGG-19 model, it is processed by adjustable filters. Following this, the image is subjected to nonlinear activation functions, such as ReLU (Rectified Linear Unit), which accentuate specific shapes or textural details in the image. With the increase in the number of layers, the complexity of the details captured by the network raises making it more meaningful to grasp the content.

In style transfer meaning, attention shifts toward feature maps drawn from convolutions layers, feature maps representing finer details and attributes of the image they correspond to in terms of style and content by recalling that, especially?; wherein by the style-content model of adoption with this working principle serving as its bedrock point of reference – those significant aspects need retaining during what remains about the appearance since artistic touch up would still be required from the style image across it produce ultimate stylized product.

By calculating the Gram matrix over feature maps in multiple layers, this model is able to represent style. Consequently, the style feature of each layer in the network captures fine-grained correlations between various attributes, essentially summarizing an image's visual style. Moreover, in order to maintain substance of an image, it meticulously compares feature representations of the initially queried content against those generated ones. The objective is to minimize the divergence in these representations such that the proposed image has the main content of the initial image while at the same time harmonizing with style of the reference.

5 Conclusion

In conclusion, this research introduces a novel method for image enhancement that utilizes auto-encoders and a combined perceptual and pixel-wise loss function. The

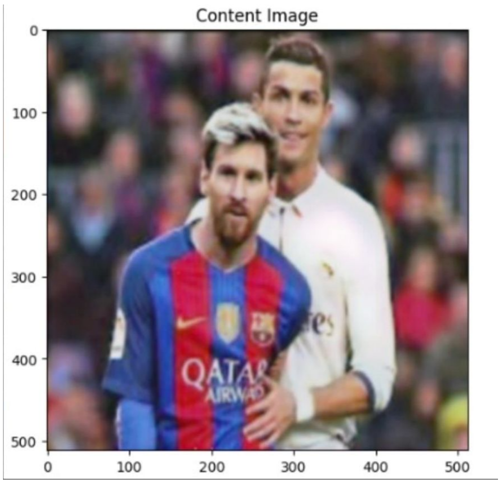

Fig. 6: Enhanced Image 2

Fig. 7: Enhanced Image 2

results demonstrate that the model produces highly detailed images with improved quality and sharpness. Furthermore, the capabilities of the algorithm are extended by incorporating style transfer techniques, enabling the artistic rendering of enhanced images based on user-provided style images. Future work will investigate the potential of deep learning methods for further enhancing image quality and addressing the identified pitfalls.

FUTURE SCOPE

The future scope of this work involves several potential avenues for improvement and further exploration. While the proposed model demonstrates strong performance in certain scenarios, yielding enhanced images after just 75 epochs (significantly fewer

Fig. 8: Enhanced Image 2

than models typically requiring approximately 300 epochs), and exhibiting high computational efficiency, there are some observed drawbacks that warrant attention. One limitation is the occasional appearance of random colors in dark regions of input images, as depicted in Fig. 9. This issue needs to be addressed to ensure consistent and accurate image enhancement across all image regions. Moreover, the extent of enhancement might be restricted, especially when the input image contains text. As depicted in Figure 10, situations where text appears in the input image can lead to sub-optimal interpretation of the text in the enhanced output. Addressing and overcoming this limitation is vital in scenarios where preserving textual information holds significance.

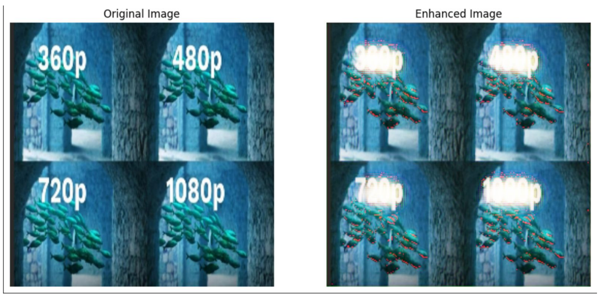

Fig. 9: Future scope 1

Another potential area for improvement lies in addressing the trade-off between texture transfer and content preservation. While the proposed method aims to strike a balance between these two aspects, further research could explore techniques to

Fig. 10: Enhanced Image 2

enhance the method's ability to selectively transfer desired textures while maintaining the core content of the original image.

Additionally, the dependency of the proposed method on the quality and relevance of reference images used for texture transfer could be explored. Developing strategies to lessen the impact of low-quality or irrelevant reference images on the final result has the potential to enhance the resilience and adaptability of the approach. In addition, the assessment of image enhancement techniques, particularly those involving texture replication, presents an ongoing challenge. Subsequent research could center on devising precise and thorough evaluation criteria that effectively capture the subjective quality and faithfulness of the improved images. Moreover, in addition to rectifying the recognized constraints, conceivable future pathways may involve expanding the suggested technique to manage multiple reference images, video enhancement, and other image generation duties necessitating texture replication. Moreover, the integration of more sophisticated super-resolution models or perceptual loss functions could serve to further enrich the visual appeal of the resulting images. In summary, the future research directions for this work offer a plethora of opportunities. These include tackling challenges related to color artifacts, preserving text fidelity, and striking a balance between texture transfer and content preservation. Furthermore, investigating methods to mitigate the influence of reference image quality, devising comprehensive evaluation metrics, and extending the approach to handle multiple reference images, video super-resolution, and other image synthesis tasks could significantly contribute to the development of more robust and versatile image enhancement solutions.

References

[1] Zhang, Z., Wang, Z., Lin, Z., Qi, H.: Image super-resolution by neural texture transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2019)

[2] Ashok, N.K., T R, S.: A Sequential Model of Neural Networks for Low Light Image Enhancement. In: 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India (2023)

[3] Yan, Y., et al.: Fine-Grained Attention and Feature-Sharing Generative Adversarial Networks for Single Image Super-Resolution. IEEE Trans. Multimedia 24, 1473-1487 (2022)

[4] Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C., Lin, L.: Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2018)

[5] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)

[6] Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B.: MedGAN: Medical image translation using GANs. Comput. Med. Imaging Graph. (2020)

[7] Sheng, J., Song, C., Wang, J., Han, Y.: Convolutional Neural Network Style Transfer Towards Chinese Paintings. IEEE Access 7, 163719-163728 (2019)

[8] Fu, K., Peng, J., Zhang, H., Wang, X., Jiang, F.: Image super-resolution based on generative adversarial networks: a brief review. (2020)

[9] Zhang, W., Liu, Y., Dong, C., Qiao, Y.: RankSRGAN: Super Resolution Generative Adversarial Networks With Learning to Rank. IEEE Trans. Pattern Anal. Mach. Intell. 44(10), 7149-7166 (2022)

[10] Choudhary, R., Reddy, T.H., Sharma, M.: ELEGAN: An Efficient Low Light Enhancement GAN for Unpaired Supervision. In: 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia (2023)

[11] Deng, X.: Enhancing Image Quality via Style Transfer for Single Image Super-Resolution. IEEE Signal Process. Lett. 25(4), 571-575 (2018)

[12] Shamsolmoali, P., Zareapoor, M., Granger, E., Zhou, H., Wang, R., Celebi, M.E., Yang, J.: Image synthesis with adversarial networks: A comprehensive survey and case studies. Inf. Fusion 72 (2021)

[13] Amro, A.: Image similarity demo using OpenCV. https://amroamroamro.github.io/mexopencv/opencv/image$_{s}imilarity_{d}emo.html$