Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES 2023)
IEEE Xplore Part Number: CFP23AWO-ART; ISBN: 979-8-3503-9663-8

Efficient Artistic Image Style Transfer with Large Language Model (LLM): A New Perspective

Bo Pan*, YuKai Ke
Kyonggi University,Suwon,Gyeonggi-do,16227, Korea
panbo1693619720kr@hotmail.com

2023 8th International Conference on Communication and Electronics Systems (ICCES) | 979-8-3503-9663-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCES57224.2023.10192799

***Abstract*— With the development of intelligent information systems, image style transfer technology has been widely known. However, many image style transfer methods can only target one style, which is inefficient in application. In this study, the novel efficient artistic image style transfer with LLM is proposed. To be gin with the study, the related single style transfer models are reviewed to serve as the background. Then, the novel image demosaicing is designed to serve as the pre-processing for the complex images. As the core of the model, novel style transfer algorithm is proposed with the LLM. The novel neural network organization is designed and the core functions are optimized. Furthermore, to validate the performance, the visualzied style transfer test is conducted and the numerical simulation results on the efficiency is tested.**

Keywords*— *Large Language Model (LLM); Artistic Image; Style Transfer; Image Feature

I. INTRODUCTION

Style transfer [1, 2, 3, 4] is the integration of classical art form and artificial intelligence technology, which has great influence on both art and technology. In addition, the products with style transfer technology as the core have attracted a large number of the users in a short period of time, which also proves their broad application scenario. The traditional image style transfer method based on feature analysis only extracts the low-level image features of the image, rather than the high-level semantic information of the image. When stylising images with the complex colours and textures, the composite effect is very unsatisfactory. This makes it difficult to use in practical application scenarios. To deal with the challenge of the low efficiency, the state-of-the-art methods have been proposed, and the Figure 1 firstly show the pipeline.

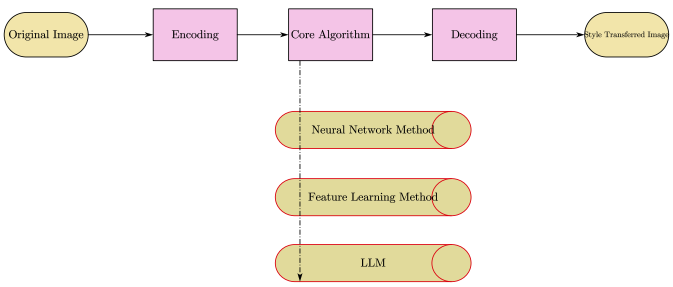

Fig. 1. The Common Image Style Transfer Pipeline

In [5], Zhan et al proposed the domain enhanced arbitrary image style transfer via the contrastive learning. The author obtains the stylized images offline by training a model that encodes style information or image reconstruction information. The commonly used method is single-network single-style fast style transfer and it is considered in the paper. The Contrastive Arbitrary Style Transfer (CAST) is proposed to effectively represent the information.

In [6], Gihyun et al proposed the Single Text Condition based image style transfer model. In the model, the users do not need to provide the reference image which will be suitable for the rea-time scenarios.

In [7], Tsu-Jui proposed the novel language-driven artistic style transfer scheme. The auhtors proposed the language based model to extract visual semantics from style instructions and provide the efficient results.

In [8], Shuai et al proposed the exemplar-based high-resolution portrait style transfer model. The authors proposed the novel DualStyleGA to generate the multiple styles. The color and also the complex structures are both considered to improve the transfer efficiency.

In [9], Hong et al proposed the deep attentive style transfer for images with wavelet decomposition. In the paper, the authors focused on the texture preservation in the style transfer tasks, the wavelet analysis model is used to further obtain the structural optimization.

In [10], Yingying et al proposed the image style transfer model with Transformers. The Authors pointed out that the traditional CNNs cannot obtain the feature effectively, two different transformer encoders are applied to the proposed model to construct the efficient algorithm.

Based on the mentioned innovative ideas, this paper then proposes the novel efficient artistic image style transfer with large language model with the new perspectives. In summary, our main contributions include but not limited to:

1) The novel image demosaicing is designed to serve as the pre-processing for the complex images.

2) The novel style transfer algorithm is proposed with the LLM.

3) The systematic experiment is conducted, the novel experiment data set is collected for the industry.

II. THE PROPOSED METHOD

A. The Image Demosaicing Pre-processing

The source of the image mosaic [11, 12, 13] can be considered as two types: one is the mosaic caused by the original image collection due to influence of the environment and the interference in the transmission process; the other is the mosaic of a single color value generated by the expansion of the image. In order to effectively improve image quality, it is necessary to distinguish between two types of phenomena and treat them separately.

In the designed algorithm, the object targeting is the first step, we describe the mosaic block area based on a simple feature: the area is rectangular, the edges form a rectangle, and the corner points of the edge are two pairs of antiradial points of a certain circle. Our method directly scans each pixel in the corner image, according to whether it is surrounded by the four white points that are symmetrical about its center and form the vertices of the rectangle determines whether it is the center point of the mosaic. Then, the removal this information is the key action with following steps:

1) The incomplete sampling values are divided into blocks, and these blocks will be demosaiced to obtain the blocks corresponding to the original image.

2) Using the optimised SL0-based demosaic optimisation reconstruction algorithm to solve the optimal value of the objective function, obtaining the sparse coefficients of the sparse coefficients in the three-dimensional space-spectral Fourier domain, and inversely transforming sparse coefficients to obtain the reconstruction of each three-dimensional image block image, combining the reconstructed images of each three-dimensional image block to obtain the reconstructed image of the original image.

The step 2 is the focused, the compressed sensing [14, 15] reconstruction algorithm for demosaics is considered. The smooth 0-norm (SL0) CS reconstruction algorithm transforms the problem of solving the minimum 0-norm into a smooth function optimization problem by introducing a sequence of the smooth functions. The SL0 is selected as the processing model and in the Figure 2, the complete steps of the original image demesaicing model is illustrated.

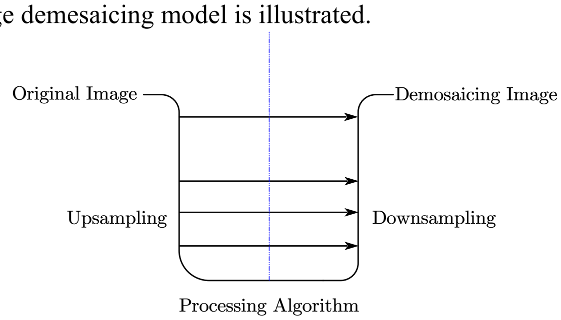

Fig. 2. The Process of the Image Demosaicing

The SL0 algorithm approximates the 0-norm through a 0-mean Gaussian function, that is defined as:

$$\|\theta_0\| = N - \sum_i^N \exp\left(\frac{\theta_i^2}{2\sigma^2}\right) \tag{1}$$

Where, the σ represents the similarity level, through the controlling of the parameter θ, the optimal value can be done. For this reason, on the basis of the original SL0 algorithm, the maximum number of iterations of the inner loop is limited to judge whether to further continue iterating, which has a fast convergence speed and high convergence accuracy. Therefore, consider defining a margin δ to judge whether it is necessary to continue to iterate to L times, that is:

$$\delta = y - A\theta \tag{2}$$

The initial of δ is set as 0, among them, while the margin of two adjacent internal iterations satisfies:

$$\|\delta - \delta_0\| < e \tag{3}$$

It indicates that there is no need to continue iterating under the current σ, then, after the repeat operation of the mentioned ideas, the optimal parameter can be found and the image demosaicing pre-processing can be done.

B. The Artistic Image Style Transfer with LLM

The network structure of this paper consists of three parts, one is a pre-trained encoder-decoder module which is used to extract image features and also reconstruct images, the style conversion module is used to learn linear transformation [16, 17, 18] between arbitrary styles, and the loss function network module is used to constrain rebuild the network to generate the high quality transferred images. Specifically, before the actual transfer algorithm is further implemented, forward-propagated neural network is trained on the style image, and each time the style transfer task needs to be completed, the target image is only needed to be forwarded on the trained neural network. The propagation calculation can achieve the desired effect and firstly, the style representtation should be modelled.

The desired style can be obtained from the style image using a specific texture information feature space. This feature space can be built on the output of the filter processing of any layer of the convolutional neural network. The expected value depends on the spatial extent of the feature map. For feature relationship, the lemma 4 defines the initial matrix:

$$G_i \in R^{N_t \times N_t} \tag{4}$$

There are several interrelationships between the different network layers. It can be used to get an accurate, multi-angle representation of the input image. Texture information can also be obtained, but the overall structure information of the image is not included. The feature is defined as:

$$G_{i,j}^t = \sum_k F_{i,k}^t F_{j,k}^t \tag{5}$$

Constructing a new image through the style representation of the selected image, and obtain information on the feature space of the style constructed between different layers of the convolution neural network to achieve the style reconstruction. In the Figure 3, the transfer neural network is presented.

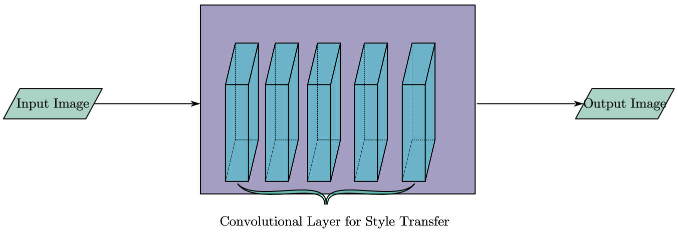

Fig. 3. The Image Style Transfer Neural Network

Let the $\vec{p},\vec{x}$ represent the original and generated images, respectively, the P^L, F^L represent the characteristic response. The error squared loss function between the two is as follows:

$$Loss_{content}\left(\vec{p},\vec{x},l\right)=\frac{\sum_{ij}\left(F_{ij}^{l}-P_{ij}^{l}\right)}{2} \tag{6}$$

The loss function is derived from the activation value of the l layer as:

$$\frac{\partial Loss_{content}}{\partial F_{ij}^{l}}=\begin{cases}0 & F_{ij}^{l}<0\\ \left(F^{l}-P^{l}\right) & others\end{cases} \tag{7}$$

The style transfer algorithm is to find an image with both content and style from white noise image, while the perceptual loss method regards generation problem as a transformation problem, and the generated image is obtained directly from the content image. The perceptual loss method does not need to retrain the convolutional neural network, and the output can be obtained by performing only one forward computation, and the time cost is reduced by two orders of magnitude, which can fully meet the requirements of real-time systems. Then, the algorithm is achieved and in the next section, the experiment will be conducted.

III. The Experiment

The operating system used in this work is Ubuntu 16.04, 32-bit, based on the PyTorch framework, version 1.2, and the programming language is Python. The data set we collected is divided into training set and test set, both of which include content pictures and style pictures. The style transfer task has no real labels and its performance depends on the user's satisfaction.

The style transfer result should have the perceptual semantics, spatial layout and artistic style of the style image.

However, style is a combination of several perceptual attributes, including global tone, local strokes, texture, line, shape, contrast between light and dark, positive and negative space. It is difficult to define and quantify intuitively, so qualitative assessment is widely used. Then, in the Figure 5, the fuzzy image generation result is presented. In the Figure 6 and 7, the style transfer result is presented.

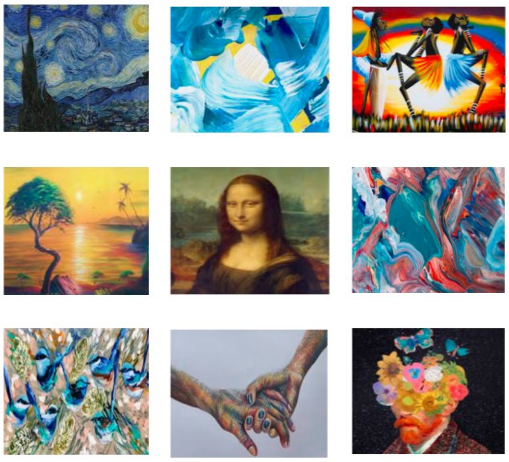

Fig. 4. The Collected Data Sets for the Artistic Images

Fig. 5. The Fuzzy Image Generation Result

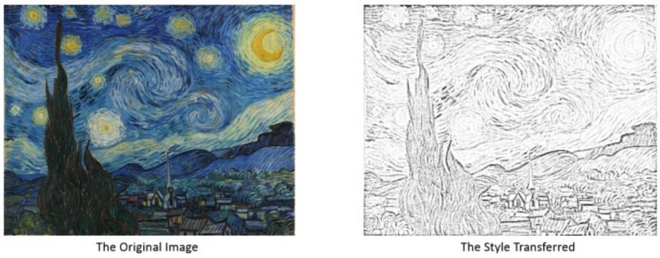

Fig. 6. The Style Transfer Result of Simple Scenarios

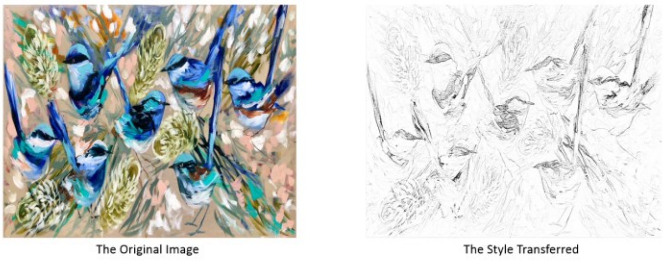

Fig. 7.The Style Transfer Result of Complex Scenarios

It can be seen from the results that the proposed model can transfer the styles effectively. For evaluating the efficiency, the style transfer time for the different images sizes are conducted and presented in the Table 1 and 2. Obviously, the time consuming isacceptable.

TABLE I. THE STYLE TRANSFER OF THE IMAGE SIZE 256*256 WITH 100 AND 500 ITERATIONS (S)

Test No.	*100 Iterations*	*500 Iteration*
1	3.15	14.26
2	3.22	15.20
3	3.15	14.99

Test No.	*100 Iterations*	*500 Iteration*
4	3.56	14.89
5	3.33	15.02
6	3.02	15.06
7	3.59	15012
8	3.55	15.23
9	3.62	15.11
10	3.71	15.21

TABLE II. THE STYLE TRANSFER OF THE IMAGE SIZE 512*512 WITH 100 AND 500 ITERATIONS (S)

Test No.	*100 Iterations*	*500 Iteration*
1	5.15	19.89
2	5.33	19.56
3	5.98	19.74
4	5.45	19.33
5	5.21	19.05
6	5.25	19.65
7	5.56	19.78
8	5.35	19.14
9	5.75	19.63
10	5.65	19.34

IV. CONCLUSION AND DISCUSSION

This paper introduces the application of the style transfer algorithm based on the LLM in scene processing, and further proposes a new idea for the image generation step. In the previous style transfer model, each model can only correspond to one style. When transferring images of other styles, the model needs to be retrained, which makes the application efficiency low. This study considers the demosaicing pre-processing to construct tyhe efficient system. In the future, the different applications will be tested and the comparison analysis wil be done.

REFERENCES

[1] An, Jie, Siyu Huang, Yibing Song, Dejing Dou, Wei Liu, and Jiebo Luo. "Artflow: Unbiased image style transfer via reversible neural flows." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 862-871. 2021.

[2] Chen, Haibo, Lei Zhao, Huiming Zhang, Zhizhong Wang, Zhiwen Zuo, Ailin Li, Wei Xing, and Dongming Lu. "Diverse image style transfer via invertible cross-space mapping." In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 14860-14869. IEEE Computer Society, 2021.

[3] Deng, Yingying, Fan Tang, Weiming Dong, Haibin Huang, Chongyang Ma, and Changsheng Xu. "Arbitrary video style transfer via multi-channel correlation." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 2, pp. 1210-1217. 2021.

[4] Kwon, Gihyun, and Jong Chul Ye. "Clipstyler: Image style transfer with a single text condition." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18062-18071. 2022.

[5] Zhang, Yuxin, Fan Tang, Weiming Dong, Haibin Huang, Chongyang Ma, Tong-Yee Lee, and Changsheng Xu. "Domain enhanced arbitrary image style transfer via contrastive learning." In ACM SIGGRAPH 2022 Conference Proceedings, pp. 1-8. 2022.

[6] Kwon, Gihyun, and Jong Chul Ye. "Clipstyler: Image style transfer with a single text condition." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18062-18071. 2022.

[7] Fu, Tsu-Jui, Xin Eric Wang, and William Yang Wang. "Language-driven artistic style transfer." In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVI, pp. 717-734. Cham: Springer Nature Switzerland, 2022.

[8] Yang, Shuai, Liming Jiang, Ziwei Liu, and Chen Change Loy. "Pastiche master: exemplar-based high-resolution portrait style transfer." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7693-7702. 2022.

[9] Ding, Hong, Gang Fu, Qinan Yan, Caoqing Jiang, Tuo Cao, Wenjie Li, Shenghong Hu, and Chunxia Xiao. "Deep attentive style transfer for images with wavelet decomposition." Information Sciences 587 (2022): 63-81.

[10] Deng, Yingying, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. "Stytr2: Image style transfer with transformers." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11326-11336. 2022.

[11] Chen, Zhuoqi, Zhaohui Chi, Karl B. Zinglersen, Ying Tian, Kaijia Wang, Fengming Hui, and Xiao Cheng. "A new image mosaic of Greenland using Landsat-8 OLI images." Sci. Bull 65, no. 7 (2020): 522-524.

[12] Lati, Abdelhai, Haitham Mekhermeche, Belkheir Benhellal, Mahmoud Belhocine, and Nouara Achour. "Novel LabVIEW implementation of FREAK based image mosaicing and object tracking algorithms." In 2020 International Conference on Electrical Engineering (ICEE), pp. 1-6. IEEE, 2020.

[13] Zhang, Ziyan, Lixiao Wang, Wenfeng Zheng, Lirong Yin, Rongrong Hu, and Bo Yang. "Endoscope image mosaic based on pyramid ORB." Biomedical Signal Processing and Control 71 (2022): 103261.

[14] Taha, Abdelrahman, Muhammad Alrabeiah, and Ahmed Alkhateeb. "Enabling large intelligent surfaces with compressive sensing and deep learning." IEEE access 9 (2021): 44304-44321.

[15] Wu, Yan, Mihaela Rosca, and Timothy Lillicrap. "Deep compressed sensing." In International Conference on Machine Learning, pp. 6850-6860. PMLR, 2019.

[16] Barone, Francesco, Marco Marrazzo, and Claudio J. Oton. "Camera calibration with weighted direct linear transformation and anisotropic uncertainties of image control points." Sensors 20, no. 4 (2020): 1175.

[17] Li, Xueting, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. "Learning linear transformations for fast image and video style transfer." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3809-3817. 2019.

[18] Lin, Rui, Jie Ran, King Hung Chiu, Graziano Chesi, and Ngai Wong. "Deformable butterfly: A highly structured and sparse linear transform." Advances in Neural Information Processing Systems 34 (2021): 16145-16157.