Neural Style Transfer: A Review and Analysis

Jiacheng Cao*
Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China
*Corresponding author: 1220022681@student.must.edu.mo

Abstract. Neural style transfer (NST) is a fascinating deep learning application that merges the style of one image with the content of another, resulting in visually compelling artistic images. This article reviews multiple research advances in the field of NST, including new style loss based on Sliced Wasserstein distance, VTNet model for video style conversion, Graph Neural Network based image style transfer method, multi style geometric deformation method, 3D neural style conversion, style transfer for 3D grids, traditional Chinese painting style transfer, and artistic conversion of QR codes. Through these studies, the author aims to improve the effectiveness and efficiency of image and video style conversion while maintaining the integrity of content information. The newly proposed style loss function utilizes SWD to improve computational efficiency and theoretically ensures the accuracy of style similarity. VTNet introduces prediction and generation branches to achieve high-quality, real-time video style conversion. In addition, the article also explores semi parametric methods based on GNN, multi style geometric transformations, and style transfer techniques for specific cultural artworks, demonstrating the potential application of neural style transfer in different fields and its contribution to artistic creation and design.

Keywords: neural style transfer; deep learning; Convolutional Neural Network.

1. Introduction

Neural style transfer (NST) is a class of software algorithms that allows us to transform scenes, change/edit the environment of a media with the help of a Neural Network. NST finds use in image and video editing software allowing image stylization based on a general model, unlike traditional methods. This made NST a trending topic in the entertainment industry as professional editors/media producers create media faster and offer the general public recreational use.

Li, Jie proposed a new style loss based on Sliced Wasserstein Distance for improving efficiency in calculating style similarity for the neural style transfer [1]. Traditionally, since WD has a very high computational complexity, most NST methods either simplify the distribution of style or use its lower bound for approximation, thus losing theoretical guarantees. This paper proposes a new style loss that is theoretically guaranteed by applying SWD and further introduces an adaptive sampling algorithm to optimize style transfer results [1]. The experiment shows that compared with existing methods, the proposed scheme can better maintain the similarity of style distribution and achieve significant improvement in visual effects. Besides, the article also discusses the comparison with other commonly used style loss functions, user research, and the effectiveness of training feed-forward style transfer networks, which confirms the effectiveness and superiority of the new method. At last, by analyzing quantitative evaluation indicators and time complexity, this paper proves that the proposed method can ensure high computation efficiency without sacrificing high-quality style conversion.

It is further proposed by Xu, Kai that VTNet, an unsupervised spatiotemporal convolutional neural network for high-quality, real-time, and highly coherent video style transfer, be introduced [2]. The framework consists of a prediction branch and a generation branch, whereby in this model, the pre-trained adversarial strategies leverage video content and dynamic information in the unlabeled video to guide the learning of the predictor [2]. Style conversion is conducted by the generation branch. The authors indicate a new coherence loss, which is incorporated into SCNet together with content loss and style loss to guide the whole network's learning process for making sure good coherence and consistent relationship between consecutive frames in the converted video [2]. Experimental results show that VTNet outperforms existing methods in removing flicker and maintaining content coherence in various styles.

Furthermore, Jing, Yongcheng proposes a Graph Neural Network (GNN)-based semi-parametric method for image style transfer with the aim of mitigating the problems faced by previous parametric and non-parametric style transfer methods, as these traditional methods lead either to local style pattern distortions due to global statistical alignment or unpleasant artifacts owing to inaccurate block matching [3]. Thanks to the employment of GNNs, precise content style correspondence at fine granularity levels in an approach proposed by the article produces more delicate, artifact-free image style conversion. Specifically, the author constructed a detailed GNN model, where local regions of content and style are treated as nodes in the graph, and then utilized attention mechanisms for heterogeneous message passing to achieve fine style content association from many to one in a self-learning manner. Besides, deformable graph convolution operations have been introduced to achieve cross-scale style content matching [3]. The experimental results verify that the proposed semi-parametric image styling method can handle complex style patterns well and maintain both global appearance consistency and exquisite details. Besides, by adjusting the number of edges in the inference stage, this method can also achieve diversified block-based stylization, and only one model can support the needs of different users. In summary, this work proposes a new arbitrary image stylization method, which, by combining the strengths of GNN, improves the quality of style conversion while enhancing flexibility and controllability.

This article reviews multiple research advances in the field of NST, including new style loss based on Sliced Wasserstein distance, VTNet model for video style conversion, GNN based image style transfer method, multi style geometric deformation method, 3D neural style conversion, style transfer for 3D grids, traditional Chinese painting style transfer, and artistic conversion of QR codes. In addition, it discuss the basic method,pros and cons and also some applications in our real life.

2. Overview of Related Research

NST is a technique that leverages deep learning, particularly convolutional neural networks (CNNs), to achieve image style transfer. It allows the artistic style of one image to be applied to another while preserving the latter's main content. The core idea behind NST is to train a neural network to separate and recombine the content and style of images.

The brief overview of how NST works are showed below:

1.Selecting a Network Model: Typically, a pre-trained convolutional neural network such as VGG16 or VGG19 is used. These networks have been trained on large-scale image datasets (e.g., ImageNet) and can effectively capture image features.

2.Defining Content Loss: This ensures that the generated image remains similar to the original content image in terms of high-level feature representation. Content loss is generally calculated using deeper layers of the network since these layers focus more on semantic information.

3.Defining Style Loss: This measures the difference in style between the generated image and the style reference image. Style loss is usually computed based on the Gram matrices derived from different layers' feature maps, which reflect the correlations between channels and thus capture texture, color, and other stylistic attributes of the image.

4.Optimization Process: A random noise image is initialized as the generated image. Then, through optimization algorithms like gradient descent, the pixel values of this image are iteratively updated to minimize both content and style losses. The result is a new image that combines the target content with the specified style.

5.Total Loss Function: In practice, content loss and style loss are combined, often with an additional total variation regularization term to ensure smoothness in the output image, forming a total loss function for optimization.

Not only dose NST find applications in the field of artistic creation, but also it inspires the development of various software tools and services that allow users to easily transform their photos into works featuring different artistic styles. Moreover, NST has spurred further research and technological advancements in image synthesis and editing.

2.1. Image Style Transfer via Multi-Style Geometry Warping

Image Style Transfer via Multi-Style Geometry Warping is a novel approach that intends to generate richer artistic effects with multi-style geometric warping. The method uses deep learning technology and automatically detects objects in the content image, further finding their corresponding art style images for processing geometric transformation [4]. The deformation process follows an improved neural network architecture from earlier work, which involves multi-step object detection, feature detection and matching, final synthesis, and style transfer of images with unique artistic styles while maintaining the original shape consistency of the content [4]. Experimental results illustrated that, compared with previous methods, the proposed can generate more artistic effects of diversity and detail, whereas there is also a limit in the impact of the alignment accuracy on the quality of the output. Hence, these methods can achieve some abstract image style transfer such as Van Gogh's Starry Night.

2.2. Voxel-Based Three-Dimensional Neural Style Transfer

Voxel-Based Three-Dimensional Neural Style Transfer represents a deep learning-based approach for three-dimensional neural style transfer, with the idea of transferring artistic styles from one three-dimensional object to another, thereby extending the application of 2D image neural style transfer to the field of 3D design [5]. Using a high-resolution voxel or polyhedral block representation, this can generate new shapes with highly detailed, visually appealing characteristics. What's more, 3D neural style transfer framework includes standardized gram matrix style loss, bipolar exponential activation function, and regularization terms specific to voxel data that improve the quality and aesthetic value of the generated 3D models [5]. The approach can be used for various content targets such as cars and rabbits. Not only can this method generate high-quality model, but also it can achieve various style transfer. However, there are some limitations. The computational complexity is high and requires a large amount of computing resources. And for some complex 3D models, there may be some inconsistencies or unnatural places. In the future, the method has great potential to enhance the automation design process in design and engineering.

2.3. Neural style transfer for 3D meshes

The NST for 3D meshes uses the features extracted from a pre-trained MeshNet model in a content and style way from the input mesh, used to guide the generated mesh for shape optimization [6]. Particularly, this approach first represents the given input mesh in the point-cloud form and then inputs that into the MeshNet for the extraction of features. The neural style transfer then iteratively refines the generated mesh M by minimizing the following loss function:

$$\mathrm{Loss}(\mathrm{C},\mathrm{S},\mathrm{M}) = \mathrm{L}_{c}(\mathrm{C},\mathrm{M}) + \lambda_r * L_s(S,M) + \lambda_r * \mathrm{L}_{c}(M) \tag{1}$$

The loss function includes two parts: one is content loss which is used to ensure that the generated grid is similar in content to the target grid:

$$\mathrm{L}_{\mathrm{C}}(\mathrm{C},\mathrm{M}) = \sum_{\mathrm{l}\in\{\mathrm{l_c}\}} ||\mathrm{F}^{\mathrm{l}}(\mathrm{C}) - \mathrm{F}^{\mathrm{l}}(\mathrm{M})||_2^2 \tag{2}$$

where $\{l_c\}$ is the set of layers in the feature extractor for content representation.

Another is style loss which is used to ensure that the generated mesh has the style features of the source mesh:

$$\mathrm{L}_{\mathrm{S}}(\mathrm{S},\mathrm{M}) = \sum_{\mathrm{l}\in\{\mathrm{l_s}\}} ||\mathrm{G}(\mathrm{F}^{\mathrm{l}}(\mathrm{S})) - \mathrm{G}(\mathrm{F}^{\mathrm{l}}(\mathrm{M}))||_2^2 \tag{3}$$

where $\{l_s\}$ represents the feature layers from MeshNet related to geometric style.

It uses gradient descent to minimize a loss function and output the final grid. Besides that, shape regularization in the proposed network enforces uniformity and quality of triangular faces in M during optimization. The smoothness of the predicted mesh M is further regularized by adding a shape regularization to the objective function:

$$L_r = \lambda_e * L_{edge} + \lambda_n L_{normal} + \lambda_l * L_{laplacian} \tag{4}$$

$$L_{edge} = \frac{1}{n_e} \sum_{e_i=(v,v')\in E} ||v - v'||^2 \tag{5}$$

$$L_{normal} = \frac{1}{n_e} \sum_{e_i\in E} 1 - \cos(\hat{n}, \tilde{n}) \tag{6}$$

This experiment shows that this approach has an effective ability to transfer several geometric styles from one style grid to other grids, and it can also preserve the content information of the original grid well. However, there are still some limitations of this method in learning and transferring texture details, which can be considered for further improvement of the model's detail capture capability. In the future, research on the ability of capturing and processing details and exploration to improve algorithm efficiency in how to reduce requirements of the calculation amount and storage space in application will be needed for actual production.

2.4. Chinese Character Style Transfer Model Based on Convolutional Neural Network

Chen, WR, Liu, CP and Ji, Y propose a Chinese character style transfer model based on Convolutional Neural Network (CNN), aiming to automatically generate new Chinese character fonts with specific styles to reduce the workload and improve efficiency of traditional manual font design. This model achieves style transition from one font to another by utilizing pre trained deep neural networks to extract and analyze features from two different font samples [7]. Specifically, a root mean square optimizer was used to automatically adjust the learning rate, and the average absolute error was used as the loss function to measure the difference between the generated font and the real font [7]. The experimental results show that the model can effectively convert Simkai font to Huangkaiti font, and the generated results are basically close to the manually designed font style, verifying the effectiveness and reliability of the method. In addition, the impact of different loss functions and total variation values on model performance, as well as the generalization ability of the model, were also explored, indicating that the model has good adaptability and is suitable for transfer tasks with multiple font styles [7]. The research in this article is of great significance for reducing font design workload and promoting the inheritance of Chinese character culture. However, the research also has some limitations. The size of dataset is limited and the dataset only contains a small number of Chinese character font samples, making it difficult to cover all the variations and characteristics of font styles. In addition, the model only considers pixel level features and ignores the importance of semantic and contextual information, which may result in a lack of coherence and readability in the generated results. In the future, there will be many research directions for this study, for example:

1.Explore how to apply the model in more application scenarios, such as in digital libraries, electronic publishing, and other fields;

2.Explore how to combine this model with other technologies such as natural language processing, image recognition, etc. to achieve more complex text processing and analysis tasks;

3.Explore how to apply this model to datasets in other fields, such as digital art, music, etc., to expand its application scope and influence.

2.5. Sound Transformation: Applying Image Neural Style Transfer Networks to Audio Spectograms

Liu, Xuehao proposes the method of using neural networks in the field of audio style transfer, whereby applying the techniques for image neural style transfer on audio spectrograms generates a new sound [8]. The authors start with an introduction on basic principles and related works of style transfer for images, which later became followed by a minute elaboration of how these have been applied in the audio field [8]. The two musical instruments, flute and keyboard, were selected as source signals in the experiment. Using their spectrograms, three different methods of neural style transfer processed them into a new type of sound effect that incorporates both keyboard style and flute characteristics [8]. Using these newly generated sounds for classification by applying the author's classification model, with t-SNE clustering analysis thereafter, revealed that those are not only clearly distinguished from the original instrument classes, but they also distinguish well among themselves, thereby attesting the effectiveness and creativity of the above method [8]. Moreover, there was a further verification, for difference both from the old and new, visual way. This article gives new ideas and possibilities on making unique sound effects using deep learning techniques. In the future, this could be expanded for a much broader dataset of, say, natural environmental sounds, and then apply it to more fields, such as speech synthesis and music creation.

2.6. Direction-aware neural style transfer with texture enhancement

Wu, Hao presents the neural style transfer method termed DaNST, which is performed by combining direction perception with texture enhancement [9]. Current neural style transfer methods suffer from NST due to the generated image having less detail and looking less natural. First of all, this work utilizes a direction-aware neural style transfer stage, leveraging two new loss networks for directional fields, respectively, to guide the generated synthesized images whose direction should keep consistent with that of the content image [9]. Subsequently, based on the synthesized image, a texture enhancement module was introduced to add more rich texture details in order to improve the quality and realism of the image further [9]. Besides, the paper proposes a simple interactive mechanism to allow users to intervene in the content image in view of their needs in order to control the direction and texture details of the final synthesized image for more flexible and personalized artistic creation [9]. The experimental results demonstrate that compared to the existing NST methods, the proposed DaNST can significantly improve the naturalness, detail richness, and directional consistency of images, especially when dealing with artworks characterized by complex texture and directional features. Therefore, this research has a great impact on art creation, video production, and image restoration.

2.7. Convolutional Neural Network Style Transfer Towards Chinese Paintings

Sheng, Jiachuan proposes an algorithm called CPST for the style transfer of traditional Chinese ink paintings, which uses machine learning technology to automatically create works with characteristics of Chinese painting [10]. Different from the traditional style transfer methods that only focus on photographs or Western paintings, CPST specially considers the characteristics of Chinese painting, including brushstrokes, ink diffusion, spatial preservation, and yellowing. Firstly, the algorithm classifies the reference ink painting as a fine or large brushwork and adopts different strategies to ensure the generated artwork not only displays the rich layers of color but the delicate and intricate strokes and composition technique in its creation [10]. Besides, four major constraints are introduced in the article, including stroke constraints, ink diffusion simulation, space reservation preservation, and reproduction of yellow tones, which can be of great help in enhancing the visual effect and artistic quality of the generated works [10]. The experimental results have verified that CPST is much better than other style transfer algorithms in reproducing traditional aesthetic features of Chinese ink painting and also has good generalization ability to input images with different themes and techniques. User research further confirms the advantages of CPST in maintaining traditional ink painting styles and demonstrates its potential application value in the field of artistic style transfer.

2.8. Style transfer for QR code

A novel two-dimensional code style conversion approach is put forward based on CNN to solve monotonous colors and poor visual effect of the traditional two-dimensional codes [11]. In this way, a system framework of the style transfer and its detailed network structure are brought up. It consists of a style transfer network and loss network. The adopted architecture by the style transfer network follows that of the deep residual network, thus includes six convolutional blocks along with five residual blocks, and the pre-trained VGG16 for the loss network [11]. In the process of training, the pre-trained structure of VGG16 is transformed to fit the QR code style conversion task, and standard positioning point correction is performed on the generated artistic style QR code to improve the recognition precision [11]. Experimental results demonstrate that the proposed style transfer method maintains the content of the original image, giving the QR code a better artistic aesthetic effect while keeping a good reading performance. Compared to the prior arts, the information fidelity and recognition rate are enhanced. In addition, the article also explores the importance of balancing content and style, and provides prospects for future research directions.

3. Conclusion

This article reviews and analyses multiple research advances in the field of NST and their basic methods, advantages and disadvantages and also potential application and value in the future. In conclusion, the field of neural style transfer has made significant progress in recent years, with researchers exploring novel techniques and applications for this technology. The use of deep learning-based approaches for three-dimensional neural style transfer has extended the application of 2D image neural style transfer to the field of 3D design, generating new shapes with highly detailed and visually appealing characteristics. Additionally, the use of standardized gram matrix style loss, bipolar exponential activation function, and regularization terms specific to voxel data have improved the quality and aesthetic value of the generated 3D models. These advancements in neural style transfer hold great potential for enhancing the automation design process in design and engineering, as well as improving the visual effects and cross-medium art conversion. However, challenges remain in terms of computational complexity and achieving consistency and naturalness in complex 3D models. As such, further research is needed to address these issues and continue to push the boundaries of what is possible with neural style transfer.

References

[1] Li J. Sliced Wasserstein Distance for Neural Style Transfer. Computers & graphics, 2022, 102 : 89–98.

[2] Xu K. Learning Self-Supervised Space-Time CNN for Fast Video Style Transfer. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 2021, 30: 2501-2512.

[3] Jing Y C. Learning Graph Neural Networks for Image Style Transfer. Computer Vision - ECCV 2022, PT VII, 2022, 13667: 111 - 128.

[4] Alexandru I. Image Style Transfer via Multi-Style Geometry Warping. Applied sciences, 2022, 12(12)

[5] Friedrich T. Voxel-Based Three-Dimensional Neural Style Transfer. Advances in Computational Intelligence: 16th International Work-Conference on Artificial Neural Networks, IWANN 2021, PT I, 2021, 12861: 334 - 346.

[6] Kang H Y. Neural Style Transfer for 3D Meshes. Graphical models. 2023, 129

[7] Chen W R. Chinese Character Style Transfer Model Based on Convolutional Neural Network. Artificial Neural Networks and Machine Learning -- ICANN 2022, PT IV, 2022, 13532: 558 - 569.

[8] Liu X H. Sound Transformation: Applying Image Neural Style Transfer Networks to Audio Spectograms. Computer Analysis of Images and Patterns: 18th International Conference, CAIP 2019, PT II, 2019, 11679: 330 - 341.

[9] Wu H. Direction-Aware Neural Style Transfer with Texture Enhancement. Neurocomputing, 2019, 370: 39–55.

[10] Sheng J C. Convolutional Neural Network Style Transfer Towards Chinese Paintings. IEEE access : practical innovations, open solutions, 2019, 7: 163719–163728.

[11] Li H S. Style Transfer for QR Code. Multimedia tools and applications, 2020, 79(45–46); 33839 - 33852.