



A Survey on 3D Style Transfer

Qifeng Zhu^(✉), Min Sun, and Jiang Wang

Suzhou City University, Suzhou, China
qifeng_zhu@szcu.edu.cn

Abstract. Image style transfer is a popular and widely studied task in computer vision, and it aims to apply the style of the source image to the target while the target remains its original content. Style transfer is widely used in creating new images in 2D, but style transfer in 3D images still has many challenges. In this paper, we summarize the major existing methods of 3D style transfer, including traditional and neural network based approaches. Moreover, we discuss the application field and the future research direction in 3D style transfer.

Keywords: 3D Images · Style Transfer · Neural Networks

1 Introduction

Style transfer is widely researched by many researchers. The traditional way to make an image style transfer is by picking up the texture information of a sample, then applying the texture to the target image. In 2001, Efros et al. [1] created a style transfer image using a quilting and texture synthesis algorithm. Although it works well on texture transfer, it has a limitation in extracting the deep feature information and realizing the style transfer in a more complex situation. With the development of machine learning, Gatys et al. [2] proposed an original idea of style transfer based on Convolutional Neural Networks. Since then, picking up feature maps by using convolutional networks to realize the 2D image style transfer has been expanded upon. However, there are few studies on 3D style transformation because the 3D graphics structure is complex, it contains spatial information. Style transfer methods in the 2D field cannot be directly used for 3D style transfer, and further changes and optimization are needed. This paper introduces the 3D style transfer method in both traditional ways and based on machine learning. We discuss the advantages and limitations of each method and the applications in the 3D style transfer field.

2 3D Style Transfer

2.1 Non-neural Network 3D Style Transfer

Before Gatys proposed [2], the traditional way to make 3D style transfer was through non-neural networks. For instance, Zheng et al. [3] make style transfer by decomposing the 3D model into parts, then restructuring them with similar visual structures. In contrast,

Ribeiro et al. [4] resemble the individual parts based on style features. Ma et al. proposed an analogy-driven 3D style transfer [5] to apply the style feature of an example to the target in the 3D field. This method is inspired by cognitive science; it works well in creating 3D models for geometric shapes. The framework of the method is illustrated in Fig. 1.

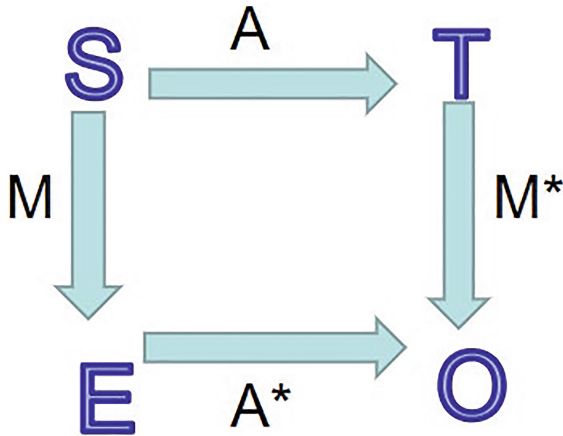


Fig. 1. Analogy relationships [5]

S, T, and E represent the source, target, and example input sets. O is the model of our goal. M is the result of computing the exemplar-source correspondence, and A is the result of computing the source-to-target transformations [5]. In the certain shape analogy quiz setup, M^* is similar to M; according to this approximate A^* by MAM^{-1} . This method can capture style features in fine-level, such as geometric textures or shapes, but it is limited to shapes with clearly separable [5].

2.2 3D Style Transfer Through Machine Learning

With the development of machine learning, researchers try to get a better style transfer result through neural networks. There are two different directions in neural network 3D style transfer. One is the “Decompose-assemble model” which has two steps: transform the 3D model into 2D images and then recreate the 3D model by processing style-transferred 2D images. The other is directly making a 3D style transfer through a 3D neural network without the deconstruction and reconstruction.

Decompose-assemble Model. Cutting the 3D model into plans or capturing the 2D pixel information of corresponding angles by surrounding the 3D model is an effective way of translating the 3D model to 2D images [6]. The style transfer process based on image iteration is proposed by Gatys, captures feature map information through VGG(Visual Geometry Group Network), creates Gram matrix as style feature from feature map of content and style image, computes the content loss and style loss, and optimizes the loss through gradient descent method. Formula (1) express the total loss

fraction in [2]; α and β are the weighting parameters for content and style reconstruction [2].

$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G) \quad (1)$$

Although the image iteration-based style transfer has an excellent visual effect, the optimization is through back propagation, which is slow and costly. Since the adversarial generative network is created by Goodfellow [7], it is also used in the style transfer field. Especially, Zhu et al. proposed an unsupervised adversarial generative network, CycleGAN, which has two pairs of generators and discriminators [8]. The CycleGAN realizes bidirectional domain conversion to transfer style between unpaired examples. The final step is the reconstruction 3D model from style-transferred 2D images. There are multiple ways to reassemble the 3D model, one of the ways Ren used in [9] is to get grayscale value from stylized images, then relocate the pixelized points to the original level in the spire volume according to the grayscale value, apply the smooth algorithm for final form modification [9].

Towards 3DNet. Different from 2D image data, which is usually represented as pixels, the representations of 3D images shape are point clouds, meshes, et; as a result, users, to make a style transfer through a 3D neural network directly, need to choose or create a 3DNet first. Timo [10] summarized common shape representations of 3D style transfer, which correspond to different 3DNet, including voxel, point cloud, surface mesh, and octree graph. Voxels are the straightforward extension of 2D pixels [10]; the detail information relates to the resolution size of voxels, which is less efficient than surface mesh and octree graph [10], while a point cloud is a simple list of 3D points in space. Jo [11] utilized PointNet [12] for realize 3D style transfer directly. However, different from the original style transfer system, he chooses the Hausdorff distance and the Chamfer distance as metrics to consider for loss functions, $d(x, y)$ which is the Euclidean distance [11].

$$d_H(X, Y) = \max \left\{ \max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y) \right\} \quad (2)$$

$$d_C(X, Y) = \sum_{x \in X} \min_{y \in Y} d(x, y) + \sum_{y \in Y} \min_{x \in X} d(x, y) \quad (3)$$

In addition, 3DNet style transfer based on Gram Matrix is also effective; however, the regular Gram Matrix in 3DNet style transfer has noisy results. To improve the result, Timo proposed standardization of Gram Matrix. According to the experiment, the style loss value is smaller for the small cube than for the sphere to avoid size changing by using the corresponding activation map [13].

$$G_{ij,s}^l = \frac{G_{ij}^l - \mu(G_{ij}^l)}{\sigma(G_{ij}^l)} \quad (4)$$

$$E_{l,s} = \sum_{(i,j)} \left(G_{ij,s}^l - A_{ij,s}^l \right)^2 \quad (5)$$

Here, μ and σ denote the mean and standard variation of a given tensor [14].

Either style transfer based on voxel-Gram Matrix or PointNet depends on generating labels to supervise, Mattia et al. proposed an unsupervised 3DNet style transfer model [15], which addresses by learning style-dependent generative models and adopt AtlasNet [16] and MeshFlow [17] decoders to generate style-dependent output distributions [15].

3 Application of 3D Style Transfer

3.1 Medical Data

For a long time, machine learning based on 2D images has been a powerful tool for diagnosing or evaluating diseases. Compared to the 2D model, 3D has a better visual effect to help doctors judge the lesion.

Therefore, a large 3D medical image data set is essential. Traditional 3D medical images can be obtained from thick 3D planes sampled from manually labeled anisotropic volumetric scans [18, 19]. However, we can now use 3D style transfer to obtain multi-modal 3D medical image datasets. For instance, Cesare Magnetti et al. slice each XCAT volume along the z-axis and retain style from a collection of 15 unique CT images per volume, and content from a window of ± 3 slices along the z-axis [20], to obtain abundant 4-Chamber 3D view datasets.

3.2 Virtual and Augmented Reality

With the development of augmented reality, 3D style conversion is also applied in augmented reality. Like the popular augmented reality game Pokémon GO, people can enrich more pets and increase play ability through 3D style transfer. Also, in virtual reality development, 3D scenes make people immersive, and furniture styles based on 3D style transfer are diversified, which is convenient for interior designers to create and design. Moreover, 3D style transfer could also be used to create human body shapes and help users customize the character images of the virtual world [21], thereby enhancing the social presence in virtual games and collaborative environments [15].

3.3 Animation Modeling

Animation modeling is perhaps the most prominent among the many application domains that 3D style transfer targets. With the rapid development of 3D animation and the game industry, there is a huge demand for 3D scene modeling. In particular, rich architectural scenes are usually large-scale and rich in details, and manual rendering modeling takes a long time. The 3D style conversion can change the shape of the building and change the texture details of the surface. Various combinations can produce multiple architectural styles, significantly reducing the workload of animation modeling. With the rise of the meta-universe, the establishment of a large number of character models can rely on 3D style conversion technology to achieve rapid and effective changes. For instance, Joao et al. [22] realized that 3D Human Shape Style Transfer includes skeletal motion and shape motion through the Generative Adversarial Network.

4 Future Research Direction of 3D Style Transfer

4.1 Algorithm and Parameter Optimization

In order to obtain stylized and ideal image results, it is necessary to manually adjust the parameters, especially based on the model optimization method. After each adjustment of the model parameters, the model needs to be retrained. Moreover, the model of a neural network can be optimized; Timo proposed normalization standardization to improve the result with less noise and optimized the activate function with $\text{elu}(x)$ to enhance the neural work quality. Therefore, algorithm and parameter optimization is of great significance to the further development of image 3D style transfer.

4.2 Model Optimization

In contrast to 2D images, where color and texture are the dominant artistic style features, the representations of 3D images are point clouds and meshes. Therefore, the directly 3d style transfer changes the shape of the original 3D model but lacks texture detail information. Currently, the 3D style transformation needs two networks to realize the transformation of texture and shape, respectively. The final results are superimposed together. Finding new 3D data patterns or optimizing existing models, creating new 3D neural networks, which can transform the structure and surface details of 3D models at the same time, is also a research direction for future exploration.

4.3 Algorithm Evaluation

As a derivative of artistic creation, 3D style transfer is usually difficult to have an objective evaluation standard. Some researchers have invited volunteers to score the images after the style change. This method has two disadvantages. One is that there are not enough volunteers and the sample size is not large enough to be representative. On the other hand, the results are affected by personal preferences and a lack of persuasiveness and objectivity. Mattia et al. proposed 3D-LPIPS and STS, two new metrics used to measure 3D perceptual similarity and style transfer effectiveness [15], which is instructive to the objective quantitative analysis of 3D style transformation. A standard evaluation method is helpful in understanding how to improve the existing style transfer algorithm in the 3D field.

5 Discussion

3D style transfer has been a new research field of image migration in recent years. Especially compared with traditional methods, 3D style transfer based on neural networks has more flexibility and style diversity. Among the above studies, we summarized the advantages and limitations of each method as well as the major loss functions used in 3D style transfer (Tables 1 and 2).

Table 1. The advantages and disadvantages of the method we summarized to use in the 3D style transfer field.

Method	Advantages	Disadvantages
Based on analogy-driven [5]	Capture fine-level style features.	Constrained by rigid similarity transformation [5].
Decompose-assemble model	Well done in the texture transfer of the 3D model.	Lack of structure transfer.
Supervised 3DNet style transfer model	Well done in the structure transfer of the 3D model.	A large number of Labels need to be generated.
Unsupervised 3DNet style transfer model	Based on classification, there is no need to generate a large number of labels.	Classification could be attribute to errors, thus affecting the final result.

Table 2. An overview of major loss functions used in the 3D style transfer field.

Paper	Loss	Description
Gatys et al. [2]	Gram Loss [2]	The style loss based on Gram-based style representations is also useful in the 3D field [23].
Jo et al.[11]	Hausdorff and Chamfer Loss [11]	Reflect the similarity of 3D shapes.
Li and Wand [24]	Adversarial Loss [24]	More efficient retention of coherent textures in complex images compared to Gram Loss [23].

6 Conclusions

The methods of 3D style transfer are comprehensively described, including traditional image segmentation, fusion, and convolutional neural networks. In this paper, the current application fields of 3D style transfer are analyzed and summarized, as well as possible research hotspots in the future. Image style transfer based on deep learning is a hot research field with rapid development. This paper classifies and describes the current research work according to the main principles of image style transfer, which can help beginners and researchers in this field grasp the current research direction and deepen their understanding of the research.

Acknowledgment. This work was supported partly by the 2021 National pre-research project of Suzhou City University (2021SGY010).

References

1. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 341–346. (2001)
2. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint [arXiv:1508.06576](https://arxiv.org/abs/1508.06576) (2015)
3. Zheng, Y., Cohen-Or, D., Mitra, N.J.: Smart variations: Functional substructures for part compatibility. In: Computer Graphics Forum, pp. 195–204. Wiley Online Library, (2013)
4. Ribeiro, P., Pereira, F.C., Marques, B.F., Leitão, B., Cardoso, A., Polo, I., de Marrocos, P.: A Model for Creativity in Creature Generation. In: GAME-ON, pp. 19–21. (2003)
5. Ma, C., Huang, H., Sheffer, A., Kalogerakis, E., Wang, R.: Analogy-driven 3D style transfer. In: Computer Graphics Forum, pp. 175–184. Wiley Online Library, (2014)
6. Zhang, H., Blasetti, E.: 3D architectural form style transfer through machine learning. In: Proceedings of the 25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), pp. 659–668. (2020)
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in neural information processing systems* 27, (2014)
8. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232. (2017)
9. Ren, Y., Zheng, H.: Voxel-based 3D Neural Style Transfer. In: Proceedings of the 25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), pp. 619–628. (2020)
10. Friedrich, T., Aulig, N., Menzel, S.: On the potential and challenges of neural style transfer for three-dimensional shape data. In: Rodrigues, H.C., et al. (eds.) *EngOpt 2018*, pp. 581–592. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-97773-7_52
11. Mazeika, J., Whitehead, J.: Towards 3D Neural Style Transfer. In: *AIIDE Workshops*. (2018)
12. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 652–660. (2017)
13. Friedrich, T., Menzel, S.: Standardization of gram matrix for improved 3D neural style transfer. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1375–1382. IEEE, (2019)
14. Friedrich, T., Wollstadt, P., Menzel, S.: The effects of non-linear operators in voxel-based deep neural networks for 3D style reconstruction. In: 2020 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1460–1468. IEEE, (2020)
15. Segu, M., Grinvald, M., Siegwart, R., Tombari, F.: 3dsnet: Unsupervised shape-to-shape 3d style transfer. arXiv preprint [arXiv:2011.13388](https://arxiv.org/abs/2011.13388) (2020)
16. Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: A papier-mâché approach to learning 3d surface generation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 216–224. (2018)
17. Gupta, K.: Neural mesh flow: 3d manifold mesh generation via diffeomorphic flows. University of California, San Diego (2020)
18. Alansary, A., et al.: Automatic view planning with multi-scale deep reinforcement learning agents. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) *MICCAI 2018*. LNCS, vol. 11070, pp. 277–285. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_32

19. Lu, X., et al.: Automatic view planning for cardiac MRI acquisition. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6893, pp. 479–486. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23626-6_59
20. Magnetti, C., Reynaud, H., Kainz, B.: Cross Modality 3D Navigation Using Reinforcement Learning and Neural Style Transfer. arXiv preprint [arXiv:2111.03485](https://arxiv.org/abs/2111.03485) (2021)
21. Yifan, W., Aigerman, N., Kim, V.G., Chaudhuri, S., Sorkine-Hornung, O.: Neural cages for detail-preserving 3d deformations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 75–83. (2020)
22. Regateiro, J., Boyer, E.: 3D Human Shape Style Transfer. arXiv preprint [arXiv:2109.01587](https://arxiv.org/abs/2109.01587) (2021)
23. Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M.: Neural style transfer: a review. IEEE Trans. Visual Comput. Graph. **26**, 3365–3385 (2019)
24. Li, C., Wand, M.: Precomputed real-time texture synthesis with markovian generative adversarial networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 702–716. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_43