

3D ARCHITECTURAL FORM STYLE TRANSFER THROUGH MACHINE LEARNING

HANG ZHANG¹ and EZIO BLASETTI²

^{1,2}*Stuart Weitzman School of Design, University of Pennsylvania*

^{1,2}{kv333q|ezioblasetti}@gmail.com

Abstract. In recent years, a tremendous amount of progress is being made in the field of machine learning, but it is still very hard to directly apply 3D Machine Learning on the architectural design due to the practical constraints on model resolution and training time. Based on the past several years' development of GAN (Generative Adversarial Network), also the method of spatial sequence rules, the authors mainly introduces 3D architectural form style transfer on 2 levels of scale (overall and detailed) through multiple methods of machine learning algorithms which are trained with 2 types of 2D training data set (serial stack and multi-view) at a relatively decent resolution. By exploring how styles interact and influence the original content in neural networks on the 2D level, it is possible for designers to manually control the expected output of 2D images, result in creating the new style 3D architectural model with a clear designing approach.

Keywords. 3D; Form Finding; Style Transfer; Machine Learning; Architectural Design.

1. Introduction

With regard to the form-finding of architectural design, It would be quite enlightening to study the simulation or integration of different architectural styles through machine learning, which could produces the form of architecture from the “mind” (or the algorithm) of artificial intelligence that go beyond human thought patterns, and open up more possibilities for aesthetic exploration and innovative design. Several related studies have been made, such as Huang et al. (2018) applied pix2pixHD to recognize and generate architectural drawings, also mark rooms with different colors to identify and reorganize different functional divisions within the plan drawings. Campo et al. (2019) applied neural-style transfer to architectural drawings, exploring the style-transformation possibilities of artificial intelligence and automation with respect to architecture.

However, these studies are confined to the 2D level and do not involve the information of the 3D space, making it difficult to explore the spatial form-finding of style mixing. More recently, there have been several attempts to explore 3D architectural forms based on machine learning. Sousa et al. (2019) introduced a methodology for generation, manipulation, and form-finding of structural typologies using variational autoencoders. Zheng (2020) proposed an interesting

method regarded to 3DGS (3D Graphic Statics) that quantifying the design preference of forms using machine learning. Zhang (2019) applied StyleGAN to train 2D architectural plan or section drawings, exploring the intermediate state between different input styles then generating serialized transformation images accordingly to build a 3D model.

Nevertheless, it is still very hard to directly apply 3D Machine Learning to architectural design, as those previous approaches are more or less suffered from the extreme limitation of the overall resolution of generated results. Therefore, based on these previous study of recognition and classification of architectural styles and 3D architectural form-finding through machine learning, this article mainly introduces 3D architectural form style transfer on 2 levels of scale (overall and detailed) through multiple methods of machine learning algorithms which are trained with 2 types of 2D training data set (serial stack and multi-view) at a relatively decent resolution (in terms of the Lenth*Width*Height, ranges from 1024*1024*1024 pixels in overall scale and 2048*2048*1024 pixels in detailed scale).

2. Method

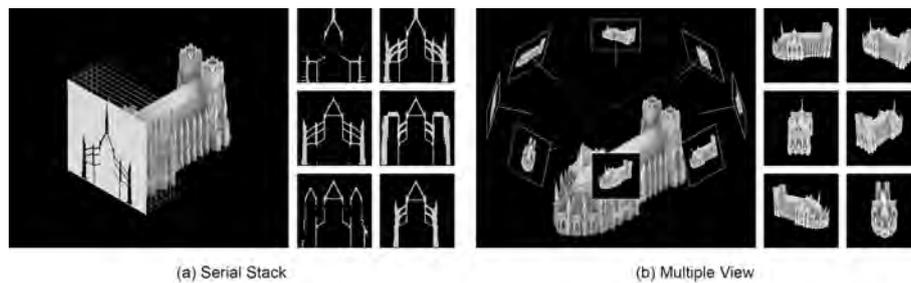


Figure 1. These are two different methods for deconstruction of the “Notre-Dame de Reims” 3D model : (a) From images stack to 3D. (b) From multi-view images to 3D. In each category, the left side is the schematic of the cutting principle, and the right side is the resulting sample.

In order to convert the 3D data of the architectural model into 2D pixel images, as is shown in Figure 1, two methods for deconstruction and reconstruction of the 3D model are given: serial stack and multi-view. From images stack to 3D form, style transfer basically processes on the z-axis (plan), x-axis and y-axis(section) through the decomposition of a linear way. The method of serial stack transfer the style conditions on each 2D layer of plan or section and integrates them to construct 3D. From. multi-view images to the 3D form, it is a method of capturing 2D pixel information of corresponding angles by surrounding the 3D model. Through the method of 3D reconstruction by giving a series of calibrated 2D images presented by Kolev et al. (2012), the authors simply extracted multi-view rendered images from the content 3D model, then reconstructed the new model from style-transferred results with the same density and position. This approach will avoid the direct calculation of massive complex 3D spatial data, and the

process of style transfer could stay at the 2D level, which significantly reduces the training time and equipment requirements.

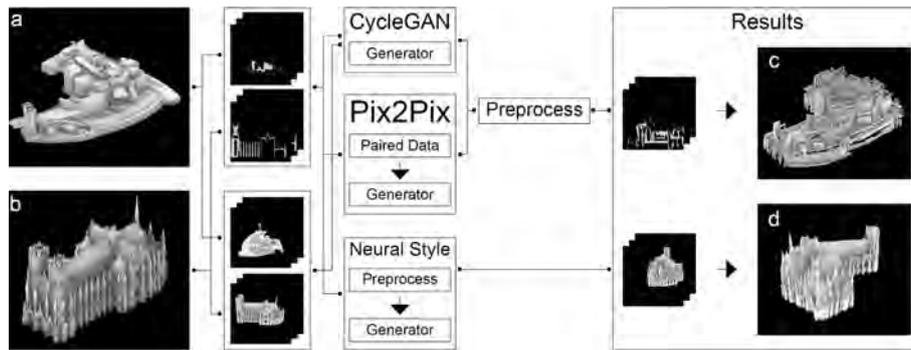


Figure 2. These are Preprocess-based workflow with different algorithms and data types : (a) Input Model A. (b) Input Model B. (c) Results from Serial Stacks. (d) Results from Multi-view.

Although methods have been given to decompose and reassemble the 3D model, the machine is hard to understand the spatial sequence rules in architectural images and the relationships within adjacent front and rear contained by spatial continuity in the original buildings. As shown in Figure 2, in order to record the rule of pixel transition between these 2D images then establishing the corresponding information correlation, during the style transfer process the authors preprocessed the data by calculating the transforming tendency of pixels before and after each of the single image based on the previous study from Ruder et al. (2016) which proposed the video style transfer, as a result of maintaining the consistency between each layer (in videos this refers to the frames). If the timeline before and after each frame in the video is the third axis that crosses the 2D screen, then the progressive sequence of these sliced stack or multi-view images is a similar third axis for a 3D building that is decomposed into serial 2D sequence images. The continuity of the third axis of 3D architectural space somehow bears a fascinating resemblance to the continuity of the video timeline.

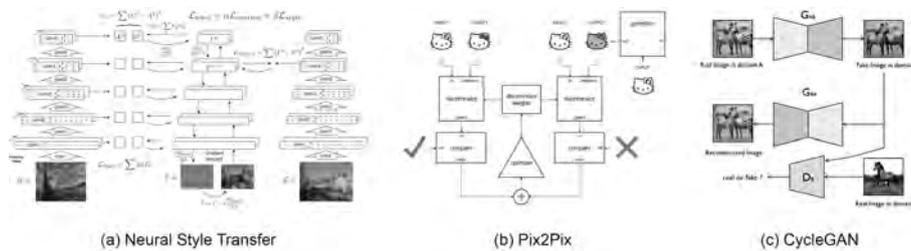


Figure 3. These are the main process of the machine learning algorithm this article will use : (a) Neural Style Transfer. (b) Pix2Pix. (c) CycleGAN.

Figure 3 shows the main structure of algorithms this paper will use. When the demand of design is to make the style transfer on the detailed level, the generated results need to retain the original structural shapes and spatial characteristics but integrate or synthesize the features from a new style into its local construction and organization. Gatys et al. (2016) presented a 2D style transfer of artistic paintings through neural networks named “Neural Style Transfer”. This algorithm will remain the original style’s general spatial organizational framework while its composition is imbued with a new style’s architectural language. As for style transfer on the overall level, which means the original style of the original model will no longer be recognized but transformed into a new or hardly identified style, this goal requires the training of our own model from a specific design approach. For now, there are two state-of-the-art style transfer GANs on 2D level: pix2pix and CycleGAN. the authors utilize both pix2pix and CycleGAN to train the decomposed 2D architectural images with paired or not paired target-style data, linking the corresponding connection between the content model (original style) and style model (target style), as a result of generating a brand new 3D form from these previous two styles.

3. Results from Serial Stacks

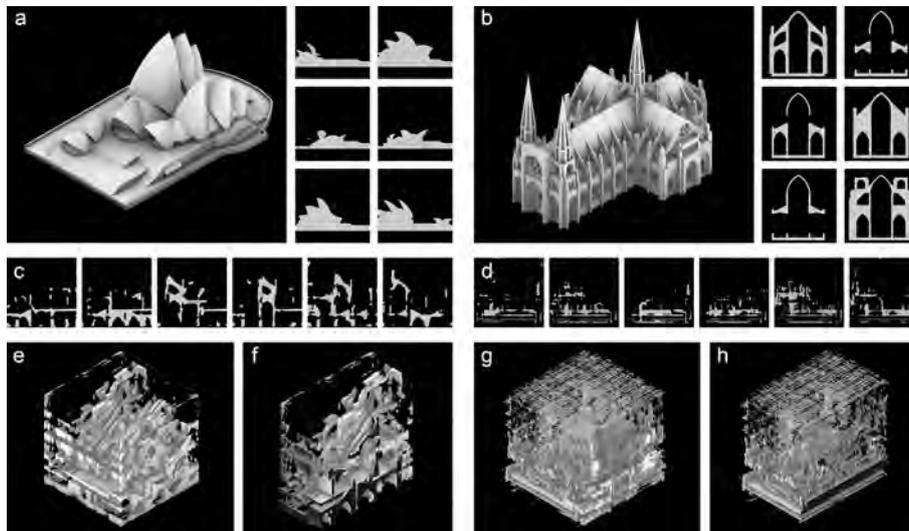


Figure 4. These are style transfer results through CycleGAN with and without continuity PRE (preprocess) : (a) Original input Model. (b) Target input Model. (c) 2D Results with PRE. (d) 2D Results no PRE. (e) 3D Result with PRE. (f) Section Model with PRE. (g) 3D Result no PRE. (h) Section Model no PRE.

3.1. OVERALL LEVEL

The transfer of styles at the overall level aims to explore the potential possibilities of innovation for architectural spatial design under the collision of various forms of

styles, especially how to break through the inherent spatial structure characteristics of various styles that go beyond human thought patterns. According to the modification of input data, algorithm logic, and preprocessing methods, the oriented style transfer can be realized with specific design ideas and the generated results can be controlled within a certain expectation range.

3.1.1. Continuity Preprocess

As mentioned in the method section, the preprocess of pixel continuity on the training data will help ensure the smooth combination of spatial sequences in the progress of 3D model decomposition and composition, otherwise, the generated results will have a large number of fragments, which are chaotic and disordered. Figure 4 shows the style transfer results through CycleGAN with and without continuity preprocess. While in the absence of this preprocessing method, the generator of the whole algorithm produces highly random results because there is no correlation information between the data, resulting in the chaos of the final 3d model.

3.1.2. Paired and Unpaired

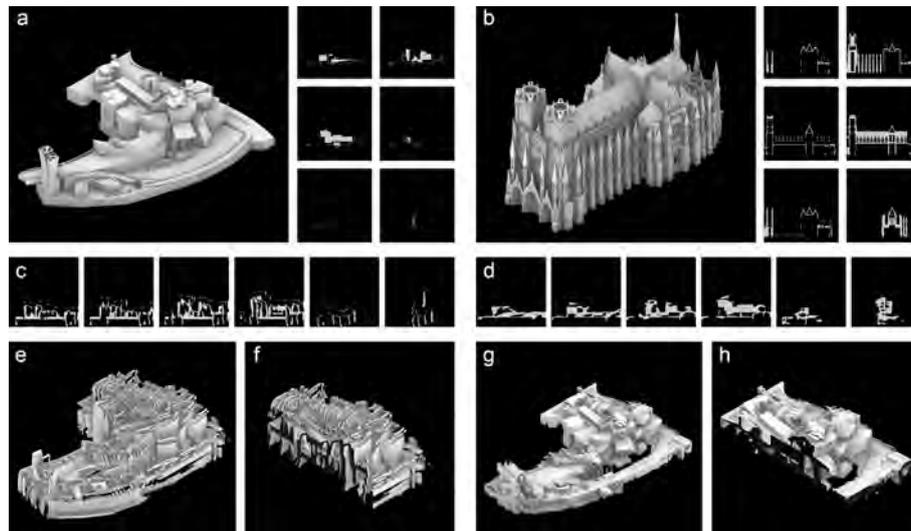


Figure 5. These are style transfer results from CycleGAN (unpaired) and Pix2Pix (paired) : (a) Original input Model. (b) Target input Model. (c) Image Results from CycleGAN. (d) Image Results from Pix2Pix. (e) 3D Result from CycleGAN. (f) Section Model from CycleGAN. (g) 3D Result from Pix2Pix. (h) Section Model from Pix2Pix.

While converting the input 3D model data into a 2D image sequence, the way of organizing the acquisition of data by style transfer will affect both the preparation of the data set and the specific GAN algorithm the authors use. In terms of the overall scale, this organizational structure is mainly reflected in the paired and

unpaired, which correspondingly refer to the Pix2Pix and CycleGAN. Figure 5 shows the results from CycleGAN and Pix2Pix which have exactly the same input 3D model. The difference between them in the data set is that Pix2Pix matches the cutting images of the two groups of models at the same position one by one, while CycleGAN has no such correspondence. And it is CycleGAN's relatively free and flexible data structure that makes its style transfer results richer and more diverse than Pix2Pix.

At the same time, it is obvious to notice that the result of this round is significantly different from that of Figure 4 in scale. Depending on the different scale of the input style data, the resulting refactoring will be able to produce disruptive changes based on the corresponding scale relationship between the original style and target style. In the whole training process, the network absorbs the spatial organization rules in the target style's own scale, meaning that the generated result will erase the original style scaling information and reorganizes completely according to the spatial scale of the target style.

3.1.3. Multiple Axis

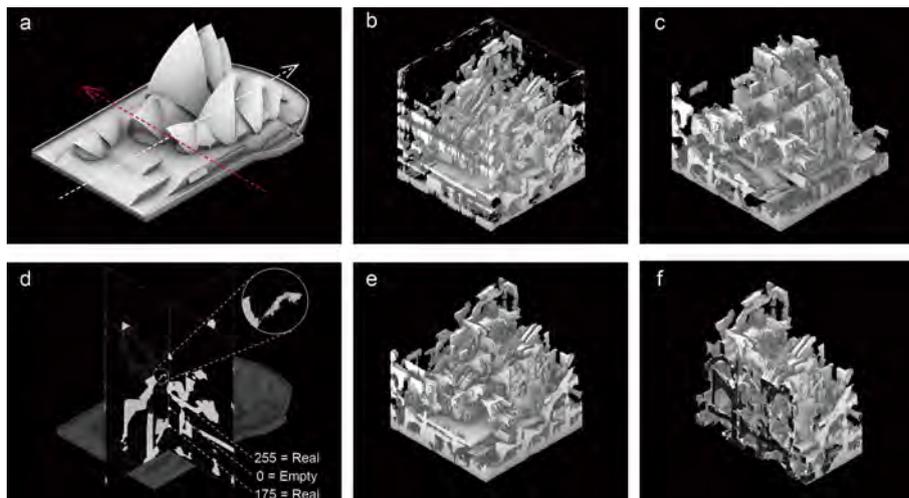


Figure 6. These are the method and results of multiple axis through CycleGAN : (a) Two style transfer Direction: Red – First, White – Second. (b) The First Direction Result. (c) The Second Direction Result. (d) Method of Value Calculation. (e) Result of Multiple Axis. (f) Section Model of Multiple Axis.

It is not feasible to simply piece together the generated results of three axes, which will produce a bloated and jumbled geometry. Benefited from the previous work that each image on linear sequence has their own serialization tag which indicating their respective positions in space, The pixel information owned by each 3d coordinate point is sampled and averaged from the matrix data at its specific position in the corresponding ordinal image. As a result, these 2D serial image

sequences will be compiled into the RGB data value of each coordinate point in the spatial point cloud. Through Boolean filtering of RGB data, that is, if the value is greater than the set condition (in this article it is 128), it is real otherwise is empty, so as to determine whether each point in 3d space is the spatial entity of the final model.

Figure 6 shows the method and results of multiple axis through CycleGAN. The results in the first direction are from Figure 4. The authors tried to train an additional 3D model from the other side and obtained the combined model through the method of pixel value calculation mentioned above.

3.2. DETAILED LEVEL

Compared with the overall level, the transfer of style in a detailed level focuses more on the organizational relationship and language of local architectural components, therefore its effect on spatial morphology was more conservative. The target style will attach its pattern, organizational logic and geometric characteristics to the skeleton of content style and then spread its growth, which meanwhile will not destroy the original architectural spatial structure.

3.2.1. Continuity Preprocess

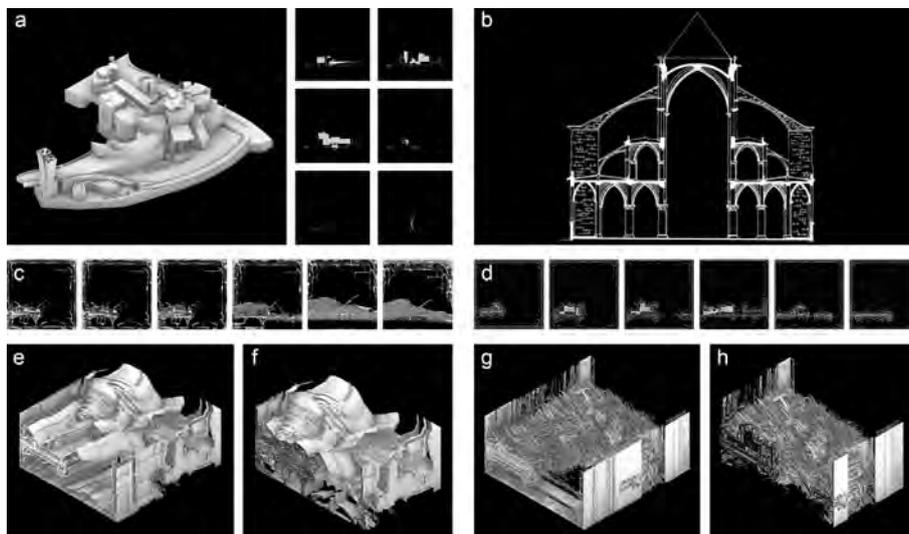


Figure 7. These are style transfer results (cropped) with and without continuity preprocess through Neural Style Transfer : (a) Original input Model. (b) Target input Model. (c) Image Results with Preprocess. (d) Image Results without Preprocess. (e) 3D Result with Preprocess. (f) Section Model with Preprocess. (g) 3D Result without Preprocess. (h) Section Model without Preprocess.

It will be much harder for “neural style transfer” (the algorithm used for detailed level) to achieve the convergence of 3D form because the randomness of this

method is so high that the results will be completely different if there is a slight change in input data, which will aggravate the difficulty of spatial continuity maintenance. As a result, different from the overall level, the preprocess of continuity will be embedded in the training process of the neural network, rather than given to the final generator of the algorithm after the training.

Figure 7 shows the results with and without continuity preprocess (The noise-filled edges of the model have been trimmed because they obscure the full view of the model). Without preprocess, gothic drawing transfers its style to make the generated model look like random weed. Given the contextual information, the result contains the explicit gothic element, such as an arched corridor.

4. Results form Multiple Views

The data set of multiple views directly maps the shape characteristics of the model in various parts of the space, so the results obtained from it will gain more transferability in the overall appearance although the possibility of exerting influence on the internal space is sacrificed.

4.1. PAIRED AND UNPAIRED

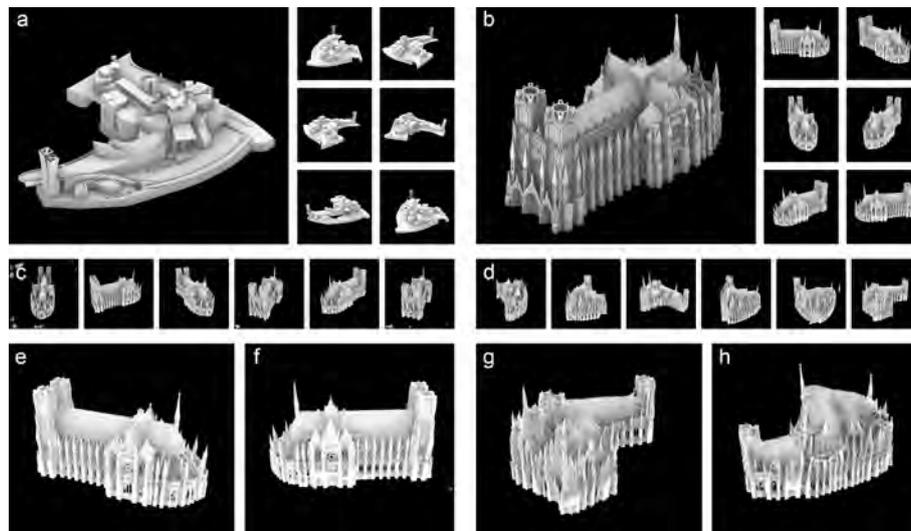


Figure 8. These are style transfer results from CycleGAN (unpaired) and Pix2Pix (paired) : (a) Original input Model. (b) Target input Model. (c) Image Results from CycleGAN. (d) Image Results from Pix2Pix. (e) 3D Result from CycleGAN. (f) 3D Result from CycleGAN. (g) 3D Result from Pix2Pix. (h) 3D Result from Pix2Pix.

As is been shown in Figure 8, The application of Pix2Pix to the data set of multiple views seems to indirectly prove the strict correspondence of this algorithm, and the generated results are almost the same as the target model, losing the characteristic of style transfer. CycleGAN, by contrast, does a good job of transferring the

morphological qualities of gothic, such as its columns, spires, and roofs, to the overall contours of “Guggenheim Museum Bilbao”.

4.2. COLOR TAG

Although the black and white sequence images obtained through the cutting model can be used for the preparation of training data simply and efficiently, the three-channel image learning ability possessed by the GAN system does not fully play its role. we can generate corresponding results in different colors for the components of different architectural elements in the 3D architectural model. This method of color tag will give the GAN system the ability to identify different architectural elements, so as to realize the oriented style transfer of specific parts or partitions.

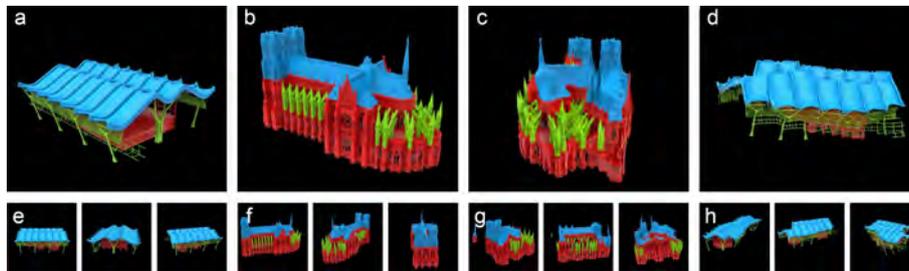


Figure 9. These are results of style transfer through CycleGAN with color tagged data set : (a) Style A Model. (b) Style B Model. (c) 3D Result from Style A to B. (d) 3D Result from Style B to A. (e) Data Set Sample of Style A. (f) Data Set Sample of Style B. (g) Image Results from Style A to B. (h) Image Results from Style B to A.

Figure 9 shows the results through CycleGAN with color tagged and double direction of style transfer. CycleGAN has the ability of bidirectional style transfer, so in this round of training, architectural elements of the corresponding colors of Style A and B have realized directional Style conversion to each other, such as the transformation of roof Style (Blue), and the mutual influence between flying buttress and steel frame structure (Green).

5. Conclusion

In terms of the basic logic, the style transfer on a detailed level is completely different from the overall. Compared to these GAN system we used for overall level style transfer, Its training time is almost negligible due to the avoidance of training through complex adversarial networks. Whereas, relatively, since there are no purpose-specific original style and target style data sets designed in advance, its training results are more difficult to predict. But this is also an interesting part of machine learning. It can always break out of the limitation of human thinking through unexpected results, while these accidents are controlled within a reasonable range so that we can explore more possibilities of architectural forms and structures with the help of it. Other types of similar generative algorithms, such as shape grammar, are a little bit monotonous because their carefully designed

and scrutiny of logic will always accurate to realize what we expect, in the form of tend to be predictable. The most interesting part of the style transfer is that, although the generated result has been permeated and dominated by target style, more or less it still retains the outline of the original style, which is more like the fight and game between original style and target style, to find their respective balance points in chaos and order, leading to the final synthesis and coexistence. In the collision and fusion of different styles, these aesthetics created by machine, instead of artificial, will break the traditional image inherent in architectural style and elements, expanding the broader boundary of design thinking.

Furthermore, for the time being, style transfer at the 3D level still remains a complex and lengthy process. Starting from 2D to build 3D is always a compromise because almost all the results have obvious deficiencies, for example, from multi-view, the result almost lost all interior details, and from the serial stack, the generated results are still more or less with traces of uniaxial slicing. Accordingly, the future work can focus on how to further optimize the structure of the network and workflow while maintaining the advantage of 2D machine learning network HD resolution, so as to obtain a more accurate and detailed 3D spatial information for innovative design. Or perhaps in the future, with the development of 3D GAN and computer technology, such as quantum computing, we could have the opportunity to directly achieve fast and efficient style transformation on the 3D level with a promising value of details.

References

- Del, D.C.M., Manninger, M.S., Sanche, S.M. and Wang, W.L.: 2019, The Church of AI—An examination of architecture in a posthuman design ecology., *Proceedings of CAADRIA 2019*.
- Gatys, G.L.A., Ecker, E.A.S. and Bethge, B.M.: 2016, Image style transfer using convolutional neural networks., *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- Goodfellow, G.I., Pouget-Abadie, P.A.J. and Mirza, M.M.: 2014, Generative adversarial nets., *Advances in neural information processing systems*, **2014**, 2672-2680.
- Huang, H.W.X. and Zheng, Z.H.: 2018, Architectural drawings recognition and generation through machine learning., *Proceedings of ACADIA 2018*.
- Isola, I.P., Zhu, Z.J., Zhou, Z.T. and Efros, E.A.A.: 2017, image-to-image translation with conditional adversarial networks., *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Ruder, R.M., Dosovitskiy, D.A. and Brox, B.T.: 2016, Artistic style transfer for videos., *German Conference on Pattern Recognition 2016*.
- Sousa, S.J.P., Xavier, X.J.P. and Castro, C.H.: 2019, Deep Form Finding - Using Variational Autoencoders for deep form finding of structural typologies., *Proceedings of eCAADe 2019*.
- Zhang, Z.H.: 2019, 3D Model Generation on Architectural Plan and Section Training through Machine Learning., *MDPI Technologies*, **7(4)**, 82.
- Zheng, Z.H.: 2019, Form Finding and Evaluating Through Machine Learning: The Prediction of Personal Design Preference in Polyhedral Structures., *Proceedings of the 2019 Digital FUTURES*.
- Zhu, Z.J.Y., Park, P.T., Isola, I.P. and Efros, E.A.A.: 2017, Unpaired image-to-image translation using cycle-consistent adversarial networks., *Proceedings of the IEEE international conference on computer vision*.