Machine Learning Aided 2D-3D Architectural Form Finding at High Resolution

Hang Zhang(✉) and Ye Huang

University of Pennsylvania, Philadelphia, USA
kv333q@gmail.com, 392057135@qq.com

Abstract. In the past few years, more architects and engineers start thinking about the application of machine learning algorithms in the architectural design field such as building facades generation or floor plans generation, etc. However, due to the relatively slow development of 3D machine learning algorithms, 3D architecture form exploration through machine learning is still a difficult issue for architects. As a result, most of these applications are confined to the level of 2D. Based on the state-of-the-art 2D image generation algorithm, also the method of spatial sequence rules, this article proposes a brand-new strategy of encoding, decoding, and form generation between 2D drawings and 3D models, which we name 2D-3D Form Encoding WorkFlow. This method could provide some innovative design possibilities that generate the latent 3D forms between several different architectural styles. Benefited from the 2D network advantages and the image amplification network nested outside the benchmark network, we have significantly expanded the resolution of training results when compared with the existing form-finding algorithm and related achievements in recent years.

Keywords: Machine learning · Architectural design · Form finding · 3D · GAN

1 Introduction

The amazing development in machine learning neural network algorithms in recent years brings us brand-new tools in design with the help of high-performance graphic cards. Many design issues can be solved by those new machine learning algorithms. Some of them are working pretty well such as the Deep Learning and GAN system.

However, most of the relative works about machine learning applications in the architecture field are working in 2D. One possible reason for the lack of a 3D architecture machine learning algorithm is the comparatively lagging development of 3D machine learning algorithms. Compared with 2D images, the complexity of 3D form issues increases dramatically, not only because of the new z-dimension but also because of different methods of 3D form representation such as point-cloud, voxel, and mesh.

When it comes to the architecture field, architects always face the issue of 2D and 3D. Traditionally, architects have used standardized 2D architectural drawings, such as floor plans, elevations, and sections, to represent 3D architectural forms. These 2D drawings, however, are limited to describe general information of the 3D building. Therefore, it is

P. F. Yuan et al. (Eds.): CDRF 2020, *Proceedings of the 2020 DigitalFUTURES*, pp. 159–168, 2021.
https://doi.org/10.1007/978-981-33-4400-6_15

obvious that the simulation and reconstruction of the 3D architectural model require 3D spatial data as the basis. In terms of the 3D model's generation, instead of using emerging 3D machine learning algorithms which has poor performance at present, we are thinking of using the more sophisticated 2D algorithm which has astonishing performance in the 2D image generation and style transfer, to help us generate 2D architectural drawings and then, use them to construct corresponding 3D architectural models.

2 Relative Work

More recently, there have been several attempts to explore three dimensional architectural forms based on machine learning. Kyle et al. presented their results of generative architectural design by transforming the 3D model data into 2D multi-view data through the neural network which is trained to perform image classification [1]. Through the concept of a 3D-canvas with voxelized wireframes, Sousa et al. introduced a methodology for generation, manipulation and form finding of structural typologies using variational autoencoders, a machine learning model based on neural networks [2]. Zheng proposed an remarkable method regarded to 3D Graphic Statics (3DGS) that quantifing the design preference of forms using machine learning and finding the form with the highest score based on the result of the preference test from the architect [3]. Zhang 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 [4].

Nevertheless, it is still very hard to directly apply 3D Machine Learning on the architectural design, as most of those previous approaches are all suffered from the extreme limitation of the overall resolution of generated results.

3 Method

This article mainly uses StyleGAN as the base network of form-finding training [5]. Compared to the other GANs, StyleGAN typically talented in generating three categories of results: (a) similar fake images, (b) style-mixing images, (c) truncation trick images. The similar fake images share common features with these input original images but look like another brand new designs. The style-mixing images are the result of 2 similar fake images. One has its basic content information and the other one transfers its image style to the content one. The truncation trick images are a series of images continuing transforming style from one type to another one.

In order to translate 3D problems into 2D, as Fig. 1 shows, we start from sections of 3D models because sections contain very rich information about the target building not only its outline but also its interior space. Instead of simply slicing the model to get just a few key sections, we start with slicing the target models for 64 times to get 64 section pieces, leading to the snapshot of each piece in the resolution of 128 * 128. After getting 64 drawings, we array them into an 8*8 grid one by one and finally get a 1024 * 1024 image, which contains the former target 3D model information in the resolution 128 * 128 * 64.

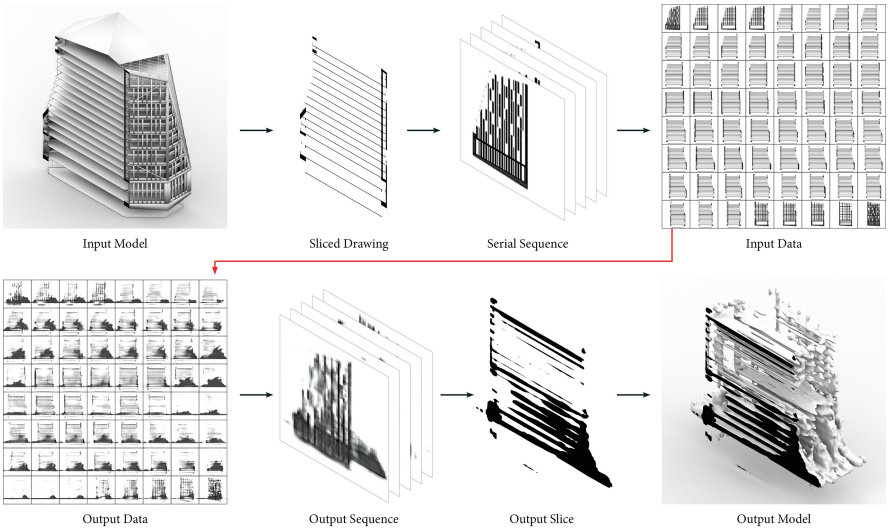

Fig. 1. These are 2D-3D form encoding and decoding work flow.

In addition, with the help of the Waifu2X for image super-resolution [6] and the Pix2Pix for image-to-image translation [7] as extra training networks, the pixel size of training model could be further expanded. Figure 2 shows the secondary network with these two layers of auxiliary model nested outside the primary network which will serve for the resolution magnification.

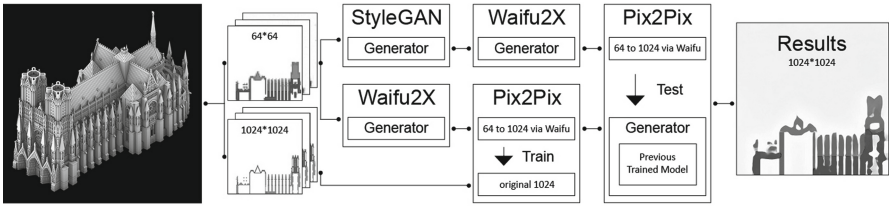

Fig. 2. These are whole process of multi-level training network.

The process on the top row is primarily the main network training via StyleGAN. Firstly, the 3D model was split into 256 slices of 2D bitmaps. While sections pieces are 4 times as large as before, the information of a single 2D layer was compressed to 64 * 64. When training finished, the pixel of generated results is processed by Waifu2X to enlarge by 16 times, which will lead to the 1024 * 1024 resolution. However, these results are distorted and blurred due to the lack of details. Here, we introduce the extra process on the bottom row. The original 3D model was sliced with up to 2000 layers in the 2D resolution of 64 * 64 and 1024 * 1024 respectively, then enlarging all 64 * 64 sequences to 1024 * 1024 via Waifu2X. The original information and enlarged information group are established on this basis, also be used as learning resources to provide image pairing

information for Pix2Pix. After obtaining pairing logic from the trained model, the pixel enlarged result under the main network process is fed back to the sub-network. At last, the final output result via translation of this paring trained rules will be generated.

4 Results

4.1 Training Data Preparation

In terms of train data, we pick several styles of 3D architectural models, all of which have either historical value or form value. We divide them into several groups in which has two counterpart models, to find out the possibilities of form synthesizing within these two buildings in the context of 2D-Image Encoding. The final paring is showed in Fig. 3, from A to F they are (A) Fallingwater. (B) Sydney Opera. (C) High Rise. (D) NYC Guggenheim. (E) LV Foundation. (F) Gothic Church. We tried the following combinations: (1) E and F. (2) B and D. (3) B and F. (4) A and F. (5) B and C.

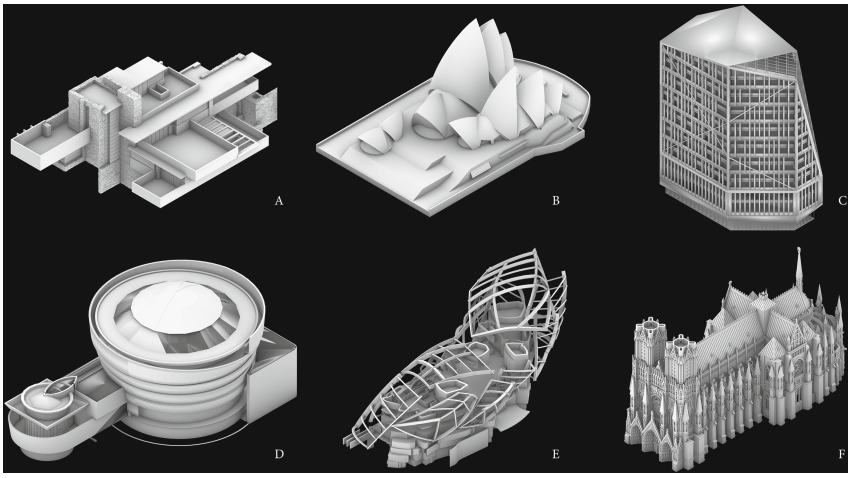

Fig. 3. These are resources of 3D models with different styles and forms.

Each 3D building model will be processed by one-way slicing, and the relative scale will be unified to fit the 2D neural network. In order to ensure that the model data mapped to 2D can be accurately decoded and restored to 3D, we fixed the position and order of each layer of 2D slices. Figure 4 shows the effect of reverse-compiling 2D mapping information into 3D.

4.2 Main Network Training

In the training of the core basic network, we used an 8 * 8 slice array, with a total of 64 layers of pixel data from the 3D model. Because the size of the whole 2D network is 1024 * 1024, every single image could have up to 128 * 128 pixels information. These

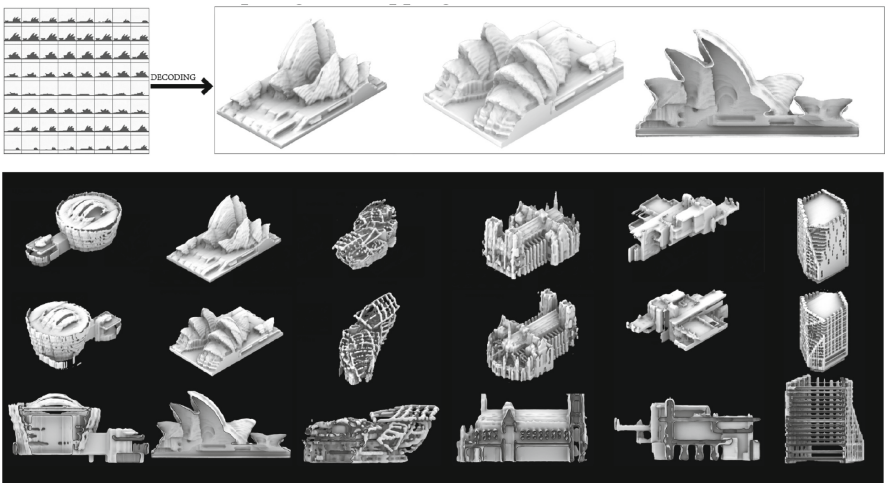

Fig. 4. These are the effect of decoding from 2D mapping information back to 3D model.

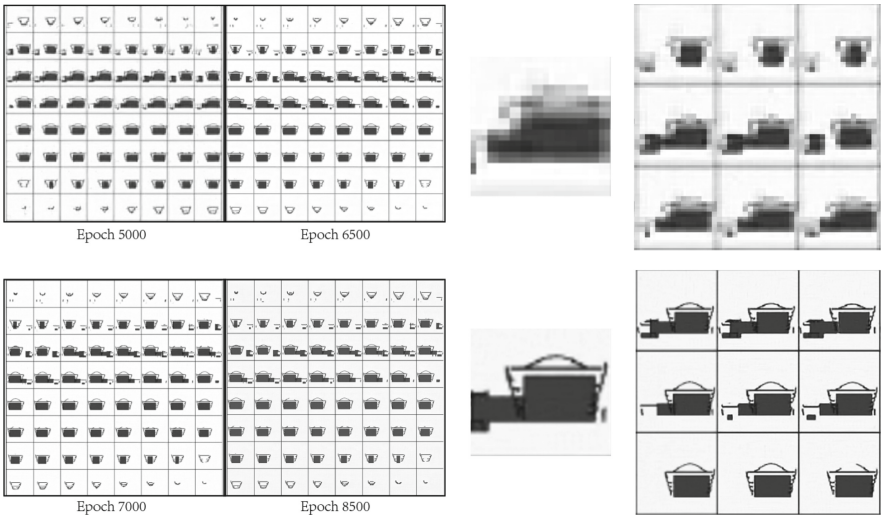

Fig. 5. These are training process snapshot of the main network.

images will function as the single-layer model information by using the pixels of the specific color gamut.

Figure 5 shows the progress snapshot during the training process. In the process of training, the whole network could keep the pixel contour information of each layer of pixel steadily, but it cannot converge to the pixel precision under the condition of insufficient training. Before 7000 kimgs, the reserved pixel area remained at least 4 * 4. As the compromise of training the 3D model through 2D is to increase the pixel limit and reduce the training time and hardware requirements, it is essential for StyleGAN

to be trained at least 8500 kimgs to achieve the pixel convergence of 1 * 1. When this condition is satisfied, the training results could be effectively reconstructed into a 3D model with an pixel accuracy of 128 * 128 * 64.

Figure 6 shows one sample of trained results via StyleGAN training on 2D networks, the styles are transferred from FallingWater to Gothic Church. The row from left to right illustrates the gradual integration of one architectural style into another. By observing the different generated data on the fixed slicing position, it can be found that the original architectural form will extend or shrink from the original pixel position of the original form when it is evolving. The intermediate state in the middle of this process is where a mixture of the two architectural styles happens.

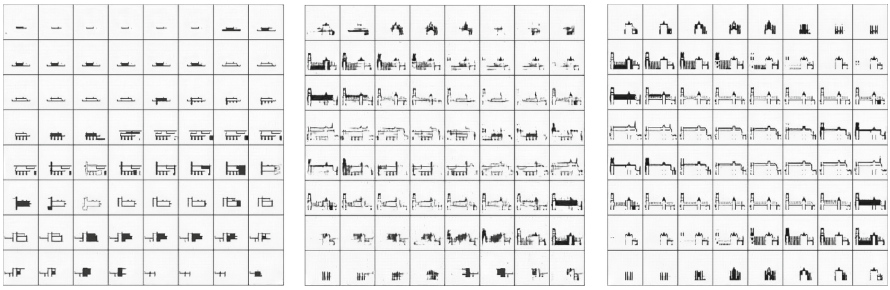

Fig. 6. These are 2D samples of trained results via StyleGAN training on 2D networks.

At this point, the training of the main network has been completed, but 2D slice images cannot directly explain the spatial sequence transformation under the influence of machine learning. Figure 7 shows the final 3D model based on the reverse compilation of the training results. From the top row to the bottom they are: NYC Guggenheim to Sydney Opera; LV Fondation to Gothic Church; Sydney Opera to Gothic Church; Fallingwater to Gothic Church; Sydney Opera to High Rise. On the whole, many pairs of architectural models of different styles shift to the other side of the style and synthesize the results of the remarkable intermediate state, and reach to the 1024 * 1024 * 64 pixels of resolution.

Compered with several previous similar projects [2, 8, 9], the generated results from our workflow are still competitive. Figure 8 shows the comparison of these 3 types of results. However, some of the results have serious noise and obvious lamination. This is because the limit of pixels makes the local details of the building lose information, and the insufficient number of vertically cut layers causes the information between layers to jump.

4.3 Multiple Network Training

As mentioned earlier, multiple nested networks produce two sets of parallel picture delivery workflows. Figure 9 presents the sample images during the whole training process: (a) 64 * 64 sliced layer from the original 3D model. (b) 1024 * 1024 sliced layer from the original 3D model. (c) 1024 * 1024 results from a via Waifu2X. (d) Final 1024 * 1024 results from c via Pix2Pix.

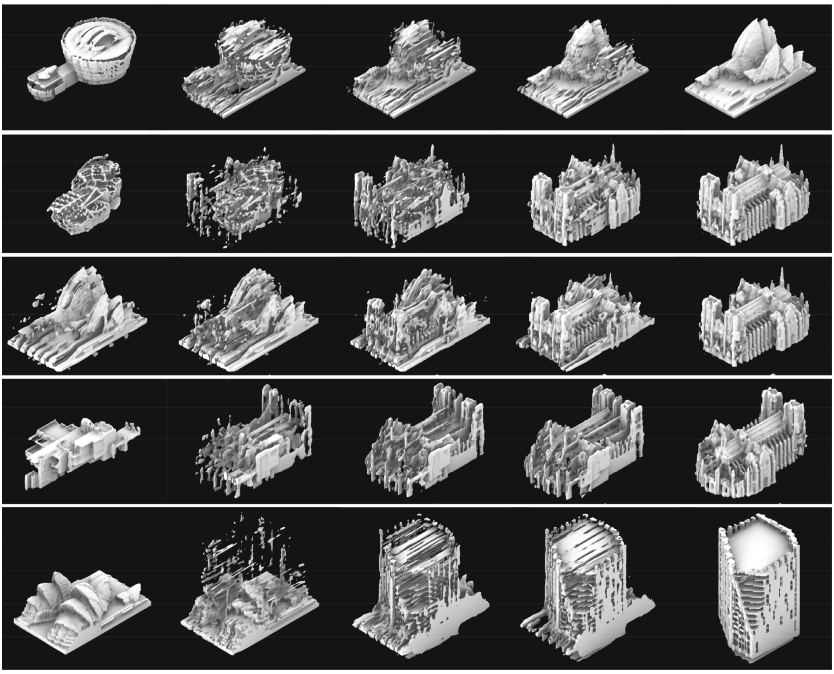

Fig. 7. These are the final 3D model based on the reverse compilation of the training results.

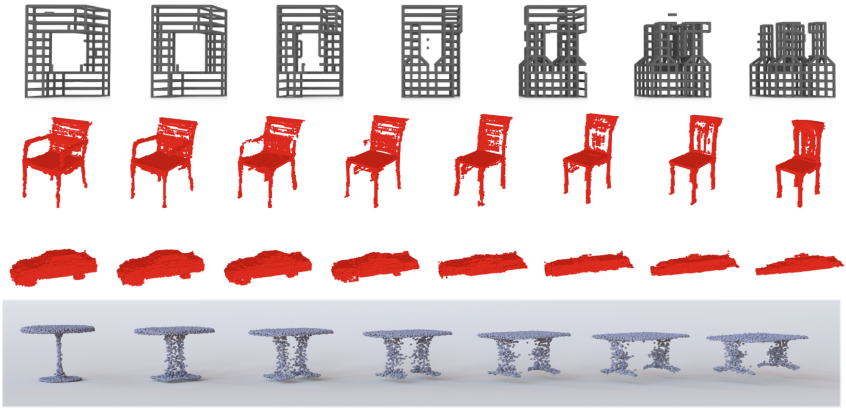

Fig. 8. These are comparison results of 3 other types of machine learning based projects. Top row: Sousa et al. with 15 *15 *12 resolution. Middle row: Jiajun et al. with 15 * 15 * 12 resolution. Bottom row: Wu et al. With total 2048 points.

Different from the previous 3d volume of 128 * 128 * 64, the result derived from multiple networks has a single-layer resolution of 1024 * 1024 and a total of 256 slices. Correspondingly, due to the disadvantages of the complex network, the generated results will produce more noise. Figure 10 shows the training results obtained based on multiple

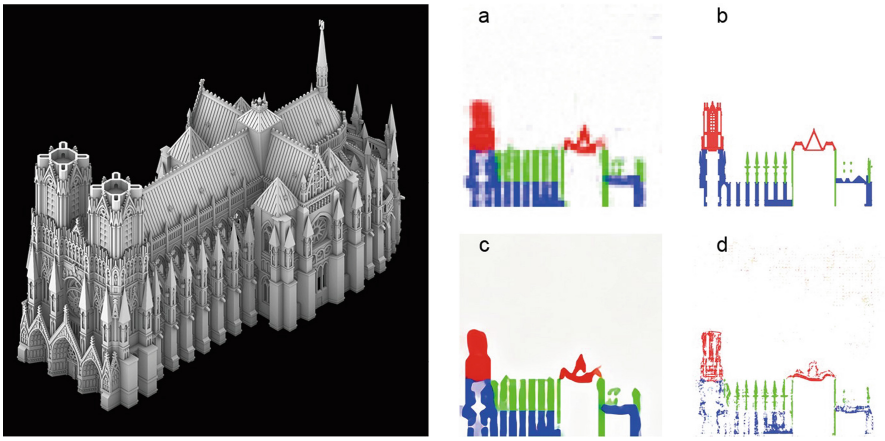

Fig. 9. These are 2D samples from Gothic Church model during the process of each networks.

networks. Due to the difficulty in controlling the noise, such nested rules are difficult to become an applicable exploration method. Of course, it is possible that training with the right parameters takes enough time to achieve more than expected results, but this goes against the principle that training time should not be too long.

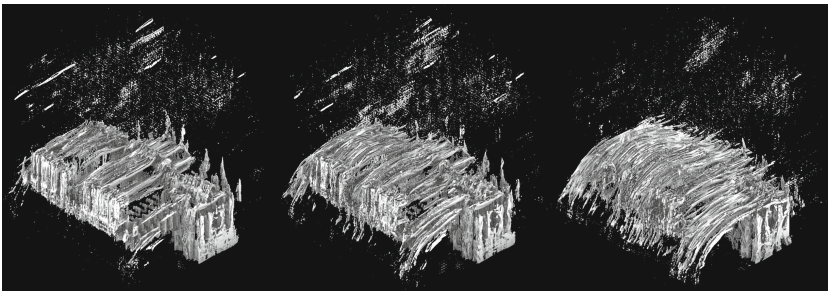

Fig. 10. These are the training results obtained based on multiple networks.

5 Conclusion

In this paper, by using StyleGAN, Waifu2X, and Pix2Pix, we successfully reconstruct existing building models through the 2D-3D en-coding strategy and realize style blending and generation of new architectural forms based on existing forms at the relatively high resolution. These combinations attempt to explore how to synthesize the morphological features among various styles and forms of several buildings with distinctive features, also leading to the innovative design through these unexpected forms the AI gives us. This style blending strategy provides totally an innovative method of architecture design which expands the boundary of form finding.

On the other hand, some predictable improvements are still needed for this encoding strategy. First, the running time issue is still a difficult problem we have to deal with. As mentioned before, the training process of those input lasts for 10 days in total with 1 single Nvidia RTX Titan. Even though the reconstruct result is satisfying enough for model decoding after 8000 kimgs (5 days), the running time is unacceptable for most designers, not to mention that nested networks are more time-consuming.

Besides, the resolution of this encoding is indeed way much higher than the existing 3D machine learning algorithm most of which are about 32 * 32 * 32 or 64 * 64 * 64. But our 2D-3D encoding strategy which has the resolution of 1024 * 1024 * 64 (or 1024 * 1024 * 256 on an unreliable state) still needs improvement to get a better representation of a building system.

Last, in most situation, 64 sections in one direction is a little bit redundant for a simple building. The traditional way of representing a simple building such as a three-layer house may only have 10–20 sections and many other plans or detail. Therefore, based on the conclusion above, our first next move about our strategy is to develop a better framework for building information representation. Based on the new framework, we can extract the crucial sections or plan information from input building models in a more efficient way, so as to reduce the number of sections. In this case, the running time will be the same but the building still conveyed clearly and the detail level of the final reconstruct model and style blending model will be increased a lot.

References

1. Kyle, S., Kat, P.: Adam M. fresh eyes a framework for the application of machine learning to generative architectural design, and a report of activities at smartgeometry 2018. In: Proceedings of the 18th International Conference, CAAD Futures 2019 (2019)
2. Sousa, J.P., Xavier, J.P., Castro Henriques, G.: Deep form finding - using variational autoencoders for deep form finding of structural typologies. In: Proceedings of the 37th eCAADe and 23rd SIGraDi Conference on Architecture in the Age of the 4th Industrial Revolution (2019)
3. Zheng, H.: Form finding and evaluating through machine learning: the prediction of personal design preference in polyhedral structures. In: Yuan, P., Xie, Y., Yao, J., Yan, C. (eds.) Proceedings of the 2019 DigitalFUTURES (2019)
4. Zhang, H.: 3D model generation on architectural plan and section training through machine learning. Technologies (2019)
5. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. CoRR, abs/1812.04948 (2018)
6. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2015)
7. Isola, P., Zhu, J.Y., Zhou. T., Efros, A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26, pp. 1125–1134 (2017)
8. Achlioptas, P., Diamanti, O., Mitliagkas, I., Guibas, L.: Learning representations and generative models for 3D point clouds. arXiv preprint arXiv:1707.02392 (2017)
9. Wu, J., Zhang, C., Xue, T., Freeman, B., Tenenbaum, J.: Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In: Advances in Neural Information Processing Systems, pp. 82–90 (2016)