



Exploring in the Latent Space of Design: A Method of Plausible Building Facades Images Generation, Properties Control and Model Explanation Base on StyleGAN2

Shengyu Meng^{1,2}(✉)

¹ University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria
Shengyu.Meng@student.uibk.ac.at

² Guangxi Polytechnic of Construction, Nanning, Guangxi, China

Abstract. GAN has been widely applied in the research of architectural image generation. However, the quality and controllability of generated images, and the interpretability of model are still potential to be improved. In this paper, by implementing StyleGAN2 model, plausible building façade images could be generated without conditional input. In addition, by applying GANSpace to analysis the latent space, high-level properties could be controlled for both generated images and novel images outside of training set. At last, the generating and controlling process could be visualized with image embedding and PCA projection method, which could achieve unsupervised classification of generated images, and help to understand the correlation between the images and their latent vectors.

Keywords: Building façade generation · Architectural generative design · Image high-level property control · GAN · GAN explanation · StyleGAN · Latent space

1 Introduction

With the emerging of Generative Adversarial Network (GAN) based image generation method in recent years, many attempts have been made to apply GAN into architectural images and drawing generation research [1]. However, for the realistic building façade images generation task, most attempts faced different challenges, such as quality and controllability of generated image, and interpretability of model.

These challenges were due to various limitations, such as performance of the selected GAN model, the size of training dataset, the understanding of latent space, etc. In this paper, by training the state-of-the-art GAN based image generation model, StyleGAN2 [2], with high-resolution building façade image dataset, and exploring its latent space by applying PCA and GANSpace analysis, we could overcome above challenges in different extend [3].

In summary, the main functions and contributions of this paper are:

1. A StyleGAN2 model instance which could generate plausible building façade images without conditional input.
2. Introduce GANSpace and image embedding method to visualize the correlation between the generated building façade images and their corresponding latent vectors, which achieved unsupervised classification and high-level properties control of both generated and novel images.

2 Related Work

2.1 Image Generation Research via GAN in Computer Science

Generative Adversarial Networks (GAN) are a neural network architecture consist of generators and discriminators, and have shown the potential to generate novel image instances from the learned distribution of training set [1]. Recently, the derived research of GAN become the focus of image generation task in computer vision scope. Above research could be generally classified as supervised and unsupervised learning structure. The supervised GAN require conditional input in both training and inference process, such as Pix2Pix, Pix2Pix HD, GauGAN (require paired training sets), and cycleGAN (require unpair training sets but each set must be similar content) [4–7]. Because the supervised GAN require relatively small training set and less training resource, and achieve high quality output when inputs were appropriated, most architectural image generation research based on that. However, their performance and application were limited in the practical workflow because conditional inputs were required. In the other aspect, the unsupervised GAN models, such as DCGAN, BigGAN, StyleGAN require much bigger training sets (normally in millions) and higher training resource, which have been used in less research [2, 8–11]. However, because the unsupervised GAN model could generate diverse outputs without conditional input, it is more potential to apply into real task. In addition, the features of latent space they have brings the possibility to achieve further model explanation and semantic edition of generated images [3, 11].

2.2 Plan Drawing Generation Research

Most research of generating architectural plan drawing based on the supervised GAN model. Hao Zheng is one of the early researchers in this scope. In 2018, he pplied a conditional GAN, Pix2Pix to prove building plan, urban plan, and satellite images of city could be generated by given conditional input, such as footprints or color pattern images [12]. In the following research, he successfully generated plausible apartment plans, and explained the working principles [13–15]. In 2019, Stanislas Chaillou developed ArchiGAN also based on Pix2Pix, which could generate whole department building plan from building footprints [16]. In 2021, above research has been expanded to large scale plan drawing. Liu et al. applied Pix2Pix to generate campus layout by given campus boundary and surrounding roads [17]. Pan et al. implied GauGAN to generate plan of community by similar conditional input [18]. The outcomes images of above research

were normally significant when appropriated conditional input was satisfied. However, plan drawing are relatively simple to generate comparing with complex building façade images.

2.3 Building Facades and Other Perspective Architectural Images Generation Research

Similar with plan drawing generation, most previous research of building facades and other architectural perspective image generation required conditional inputs. In 2017, in the original paper of Pix2Pix, Isola et al. generated novel street scene and building façade images, but required street view and refine colour label as paired image inputs [4]. In 2019, Kyle Steinfeld developed GAN Loci upon both Pix2Pix and StyleGAN, when the Pix2Pix version required depth map as conditional input, and the StyleGAN version has only trained as a 512 pixel square unrefined instance, due to the limitation of computing resource and dataset [19]. Kelly et al. proposed the FrankenGAN which could generate 3D building model with detailed façade textures, but require mass 3D model as input [20]. Mohammad et al. attempted to generate novel building elevation design from AI generated datasets, but they only got low-resolution grayscale images [21]. In 2020, Chan et al. attempted to generate building facade images from hand sketch, but only got low-resolution output due to small dataset and small GAN architecture [22].

Different with previous research, In 2019, Bachl et al. developed City-GAN to synthesis novel city images from random input by learning from large street view dataset. City-GAN was developed upon the unsupervised DCGAN model with feeding additional label information to control the style of the generated city images, and allow simple interpolation between different styles. Nevertheless, the generated images were still with limited quality and resolution [23]. Chen et al. proposed another unsupervised model, embedGAN, which attempted to explore the property of latent space [24]. They embedded an interior image into the latent space as a starting point, and then guided that latent walk purposefully by a pretrained classification network, to regenerate the image into different decoration material and style. However, only trained image could be applied, and the image quality was not good enough.

3 Methodology

In this paper, the state-of-the-art GAN based image generation model, StyleGAN2 has been applied on the experiment [2]. In addition, a training set with 9772 building façade images of 1024×1024 resolution have been implemented. Because StyleGAN2 model generates images from random sampling vectors in the high-dimensional latent space, to explore and visualize the relations between the generated building façade images and corresponding latent vectors, the methods of dimensionality reduction, clustering and image embedding have been applied. Specially, by utilizing the principal component analysis (PCA) on the intermediate latent space W of StyleGAN2 model [3], this paper achieved high-level properties control of generated building facade images. In addition, even though StyleGAN2 does not have encoder network, projecting novel building facade images (outside from training set) into existing latent space has been achieved through

applying VGG16 pre-trained perceptual network. This method could locate the latent vector which could generate the most similar image with the target image [25]. Once the projection completed, we could control novel image as same as the generated one.

3.1 Training Building Facades Generation Model by StyleGAN2

3.1.1 Introduction of StyleGAN2

StyleGAN2 is the SOTA GAN based image generation model upgraded from StyleGAN, which was proposed by Nvidia company in 2020 [2, 11]. It has unique generator structure different from most GAN models, which provides better model performance and interpretability. Most GAN models, such as Pix2Pix and CycleGAN, have an image encoder to encode the input image as a latent vector, which used as the direct input of image synthesis network (decoder) [4, 7]. This structure requires images as input to generate others, and potentially limits the model performance, because the distribution of input images may not fit to the output images [11].

The style-based generator of StyleGAN2 could avoid using image as input. The synthesis network g of style-based generator begins with a learned constant number, and go through 18 layers to output as 1024 pixel square image. In each layer, a noise and a latent vector w will be inputted to adjust the style and content of the generated image. The latent vector w is an intermediate output in a 512-dimension latent space W , which was converted from a vector z in another randomly sampling 512-dimension space Z , by a 8-layer trainable fully connected network [11].

The improvement of Style-based generator in StyleGAN2 brings bellowing new features, which are the foundation of the further research in this paper [11].

1. State-of-the-art image generation performance.
2. Generate image from random vector without conditional input.
3. Latent space W is free from restriction and potential to be disentangle.
4. Unsupervised separation of high-level attributes of generated images.
5. Adjust the style and content of generated images by manipulating latent vectors.

3.1.2 Training Process

In this paper, an open source architectural style image dataset was firstly integrated with other 6000 building facades photos and renderings downloaded from internet [26]. Secondly, repeated, non-architectural and low-resolution images were removed. Finally, all the images will be converted to jpg format with RGB channel at 1024×1024 pixels. The final training set included 9772 architectural façade images from various styles. The training was proceeding in config-f (1024×1024 resolution) and with auto mirror augment function. The training has been running on a single NVIDIA Tesla V100 with 16G RAM, and continued about 816 h until 12240K images.

3.1.3 Generation Examples

After the training completed, this StyleGAN2 model instance could generate plausible building facades images with 1024×1024 resolution from random seeds (Fig. 1). The

generated images were similar with the training set images, but showing the mixing features from different examples rather than simply repeat (Fig. 2).

However, for some details in the generated image were still blurry, mismatched or missing. These may because of the relatedly small dataset, not enough original resolution of some training images (were enlarged) and not enough training time.



Fig. 1. Example of generated building facades images of the experiment in this paper (curated)

3.2 Exploration and Explanation of Latent Space

3.2.1 Visualizing High-Dimension Latent Space by PCA

The style-based generator will require an intermediate latent vector was input in the intermediate latent space W . The W was remapped from a 512-dimension randomly sampled latent space Z by 8 layers of trainable fully connected networks. To visualize the distribution of W and Z , the Principal Component Analysis (PCA) was introduced to reduce the dimensionality and projected the vectors in both spaces to 2D figure. The PCA method will firstly analysis the distribution of high-dimension latent space, then projects the vectors orthogonally with the principal axis into the low-dimension space, to present the main features in high-dimension space [27].

In this paper, for exploring both latent spaces, 2000 vectors have been randomly sampled in Z and remapped into W , and finally the generated building façade images. The vectors distributions in W and Z have been projected by PCA into 2D figure (Fig. 3). Because the vectors in Z were randomly sampled, its distribution was almost sphere. Relatively, the distribution in W has remarkable shape, which may reflect the features distribution of image contents.

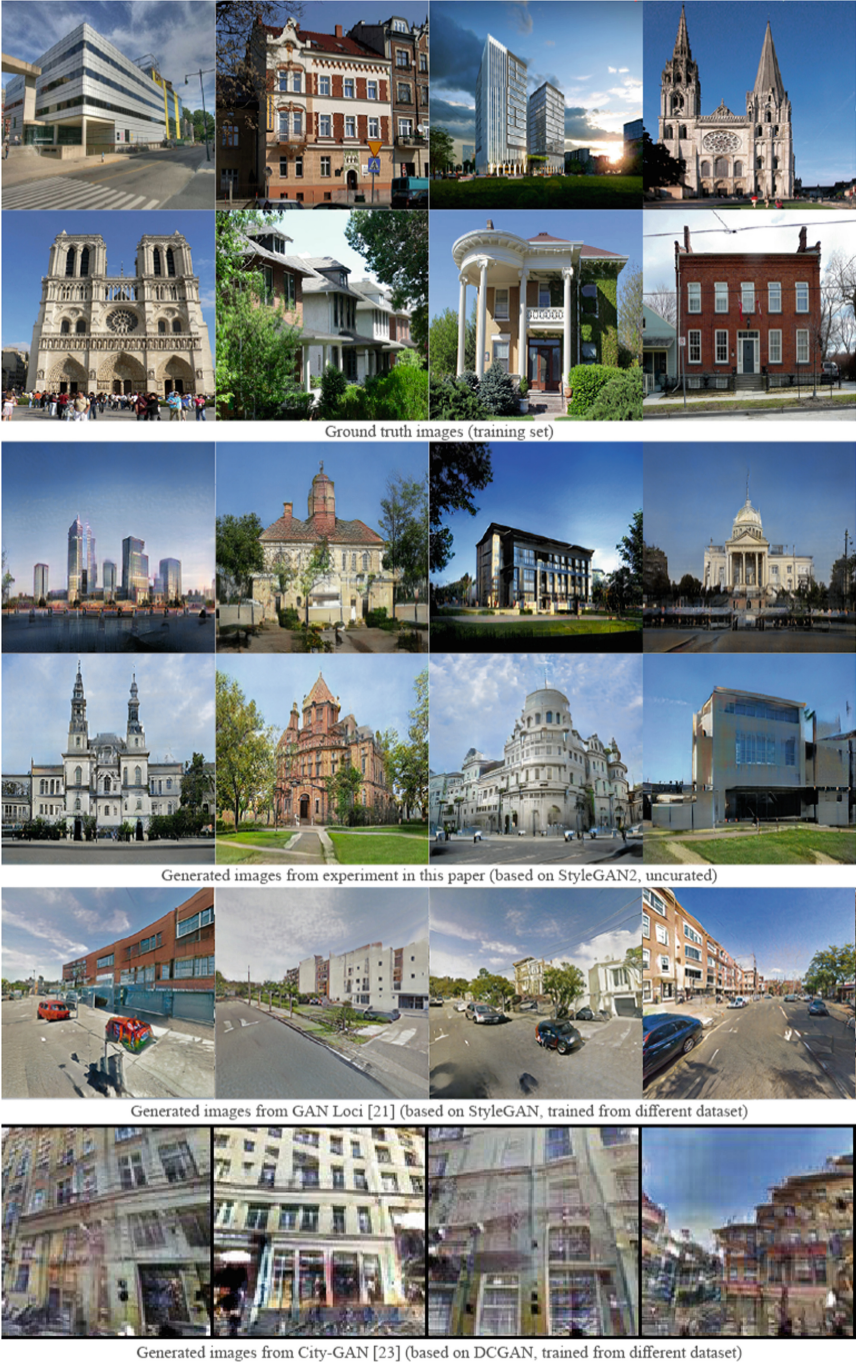


Fig. 2. Generated building facades images examples comparing with dataset and other models.

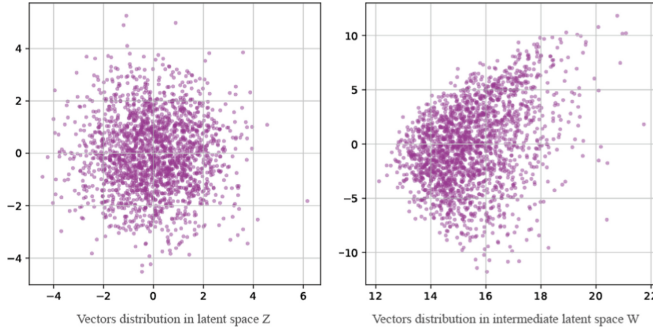


Fig. 3. Vectors distributions in latent space Z and W

3.2.2 Explanation of StyleGAN2 Model: Images Embedding and Clustering in Latent Space

To prove previous hypothesis and visualize the correlation between images and their latent vectors w in space W, 2000 generated images examples have been embedded at the projected locations of corresponding w as dots (Fig. 4). To avoid too much overlap, only about 10% images thumbnails have been shown. In addition, the unsupervised K-means clustering algorithm has been applied to cluster the w vectors into 4 types, which were marked as colours of the dots and frames of thumbnails.

It could be observed that the images in same clusters shared similar features. Moreover, some features will show linear change alongside certain direction. For example, the height of buildings was descending from the top to the bottom in the Fig. 4. These proposed the hypothesis the generated image's high-level property could be controlled by moving its vector w along certain principal axis.

3.3 High-Level Prosperity Control

3.3.1 GANSpace Method

Above hypothesis has been proved by the research GANSpace [3]. In that research, the principal axes V_n will be firstly computed by analyzing the Latent space W via PCA (the max amount of V_n is equal to the dimension of latent space, 512), then the modified vector w' can be computed from original vector w by the Eq. 1 below, where the x is the scale parameter customized by user [3]:

$$w' = w + V_n * x \quad (1)$$

3.3.2 High-Level Property Control

In this paper, the control process of principal axes 0 was visualized by setting series of equidifferent parameters (Fig. 5). Because axes 0 was assumed as the foremost axis of latent space W, it should present the most significant diversity of the training sets, which was from high-rise modern building to low-rise traditional residential house. The

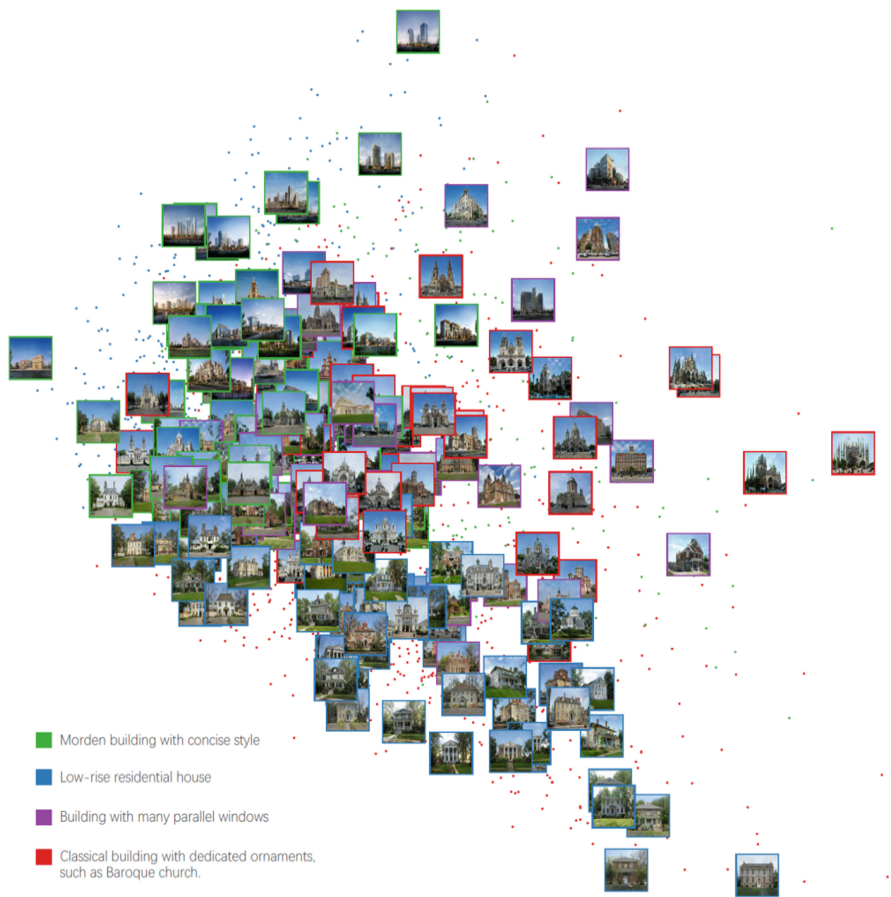


Fig. 4. Image embedding with corresponding vectors in latent space W

modified images showed the gradual change when it moved along axis 0, and represented similar features when passing nearby images. In addition, when the modified w' was approaching the border of W , the generated images become implausible, because the StyleGAN2 model have not enough training in these areas.



Fig. 5. Visualizing high-property control by principal axes 0 in scale parameter range $[-25, 25]$.

More examples controlled by different main axis could be seen on Fig. 6. Ideally, each axis would control one significant feature. However, these features were still partly disentanglement in this experiment. That may because of the insufficient quantity and diversity of the images in training set.



High-level properties control by main axis 2: from residential house to church



High-level properties control by main axis 5: control the amount of slope roofs on the building facade

Fig. 6. Examples of the high-property control by main principal axes 2 and 5.

3.4 Project Novel Image into Existing Model Instance

3.4.1 The Projection Method

The image's latent vector w in StyleGAN2 model is necessary when ccontrol its high-level property. However, because StyleGAN2 does not have encoder network, the novel image (outside of training set) cannot be directly encoded as w . To solve this problem, Image2StyleGAN algorithm employed a pretrained VGG16 perceptual model. In details, a latent walk will firstly start from the average w , then the VGG16 model will be applied to compute the loss between the output image of present w and the target image. Finally, gradient descend could be used to guide the latent walk to get close to the w which could generate the most similar output image with the target novel image, in the latent space W of existing model instance [25].

3.4.2 Projection and Control of Novel Image

In this paper, a white vacation house outside of training set has been projected into the w space, and the whole process could be observed by PCA projection (Fig. 7). It could be found that the moving interval of projecting vector was keep decelerating, because the computed loss of gradient descending was reducing progressively. The final project result image was not totally as same as the input target image, but very close. After that,

the projected image could be controlled with high-level property by principal axis, just like the generated images (Fig. 8).

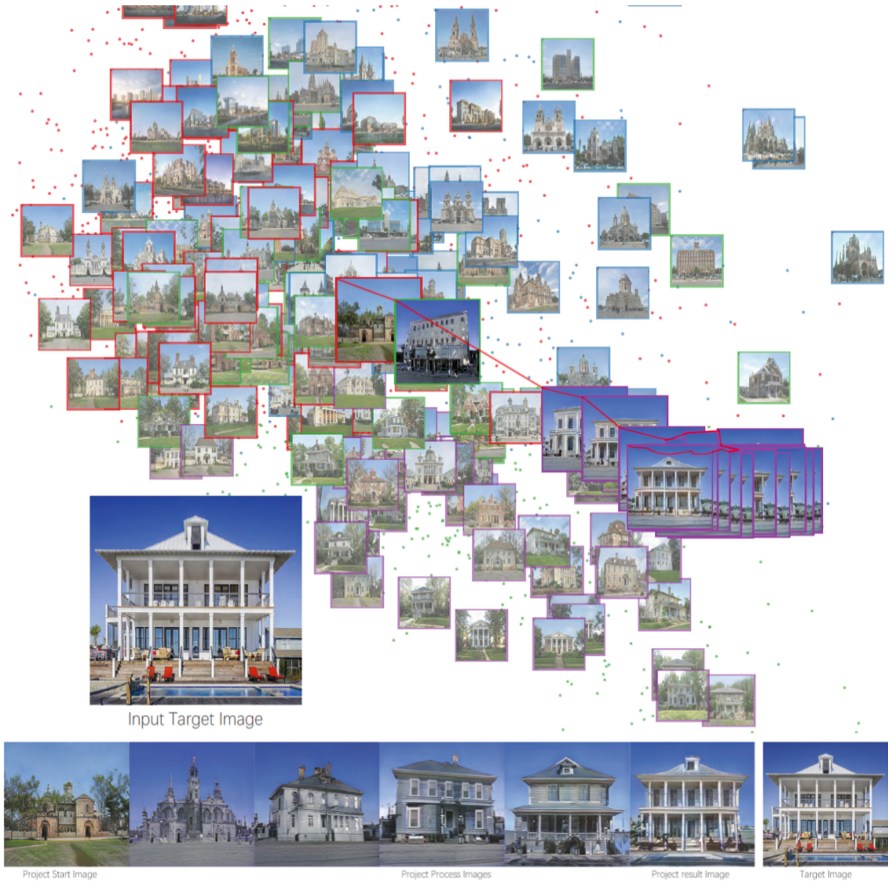


Fig. 7. Visualizing the novel image projection process in latent W space.



Fig. 8. High-level property control of projected image by principal axes 0: from high-rise modern building to low-rise residential house

4 Conclusion

This paper is aiming to remove the obstacle of applying GAN based image generation model into generative design workflow. By integrated series of SOTA method from computer vision scope, this research has improved the quality of generated building façade images, and visualized the correlation between generated images and their feature vectors in latent space. In addition, by analyzing and manipulating the latent space of the trained model, high-level property control has been achieved for the generated images and the novel images.

However, some details of generated image were still blurry or mismatched, and the property control has not achieved completely disentanglement. Both of that are possibly because of the insufficiency of the quantity, quality and diversity of training set images. The present training set is less than 10K images and part of them have been enlarged, when the original StyleGAN2 research applies 200K full high-resolution training set images. Better performance may realize with larger training set and longer training time.

Acknowledgement. The author would like to express the appreciation to Professor Claudia Pasquero for her teaching in patience, and Hao Zheng for his kindly help. This research is supported by The Project to Improve the Academic Ability of Junior Faculty in the Higher Education of Guangxi (Grant No. 2021KY1159), and the Innovation and Research Project of Professional Education in Guangxi (Grant No. GXGZJG2019A014).

References

1. Goodfellow, I.J., et al.: Generative adversarial networks. In: *Advances in Neural Information Processing Systems*, p. 27 (NIPS 2014) (2014)
2. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. In: *CVPR 2020*, pp. 8107–8116 (2020)
3. Härkönen, E., Hertzmann, A., Lehtinen, J., Paris, S.: GANSpace: discovering interpretable GAN controls. In: *34th Conference on Neural Information Processing Systems (NeurIPS 2020)* (2020)
4. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 5967–5976 January 2017
5. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional GANs. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 8798–8807 (2018)
6. Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y.: Semantic image synthesis with spatially-adaptive normalization. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2332–2341, June 2019
7. 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. 2242–2251, October 2017

8. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings, 1–16 (2016)
9. Donahue, J., Simonyan, K.: Large scale adversarial representation learning. arXiv 1–11 (2019)
10. Zhang, H.: 3D model generation on architectural plan and section training through machine learning (2019). <https://doi.org/10.3390/technologiesxx010005>
11. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 4396–4405, June 2019
12. Zheng, H.: Drawing with bots: human-computer collaborative drawing experiments. In: Learning, Prototyping and Adapting, Short Paper Proceedings of the 23rd International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA), pp. 127–132 (2018)
13. Huang, W., Zheng, H.: Architectural drawings recognition and generation through machine learning. In: Recalibration on Imprecision and Infidelity, Proceedings of the 38th Annual Conference of the Association for Computer Aided Design in Architecture, ACADIA 2018, pp. 156–165 (2018)
14. Zheng, H., Huang, W.: Understanding and visualizing generative adversarial networks in architectural drawings. In: 23rd International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA) (2018)
15. Zheng, H.: Apartment floor plans generation via generative adversarial networks. In: 25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), vol. 25, p. 10 (2020)
16. Chaillou, S.: ArchiGAN: a generative stack for apartment building design. In: NVIDIA Corporation (2019). <https://devblogs.nvidia.com/archigan-generative-stack-apartment-building-design/>
17. Liu, Y., Luo, Y., Deng, Q., Zhou, X.: Exploration of campus layout based on generative adversarial network. In: Yuan, P.F., Yao, J., Yan, C., Wang, X., Leach, N. (eds.) CDRF 2020, pp. 169–178. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4400-6_16
18. Pan, Y., Qian, J., Hu, Y.: A preliminary study on the formation of the general layouts on the northern neighborhood community based on GauGAN diversity output generator. In: Yuan, P.F., Yao, J., Yan, C., Wang, X., Leach, N. (eds.) CDRF 2020, pp. 179–188. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4400-6_17
19. Steinfeld, K.: GAN Loci imaging place using generative adversarial networks. In: ACADIA 2019, pp. 392–403 (2018)
20. Kelly, T., Guerrero, P., Steed, A., Wonka, P., Mitra, N.J.: Frankengan: guided detail synthesis for building mass models using style-synchronized GANs. In: SIGGRAPH Asia, 2018 Technical Papers SIGGRAPH Asia 2018 (2018). <https://doi.org/10.1145/3272127.3275065>
21. Mohammad, A., Beorkrem, C., Ellinger, J., Charlotte, U.N.C.: Hybrid elevations using GAN networks. In: ACADIA 2019, pp. 370–379 (2019)
22. Hin, Y., Chan, E., Spaeth, A.B.: Architectural visualisation with conditional Generative Adversarial Networks (cGAN). What machines read in architectural sketches. In: Proceedings of the 38th eCAADe Conference, vol. 2, pp. 16–18 (2020)
23. Bachl, M., Ferreira, D.C.: City-GAN: learning architectural styles using a custom conditional GAN architecture (2019). arXiv preprint <http://arxiv.org/abs/1907.05280>
24. Chen, Z., Huang, W., Luo, Z.: embedGAN: a method to embed images in GAN latent space. In: Yuan, P.F., Yao, J., Yan, C., Wang, X., Leach, N. (eds.) CDRF 2020, pp. 208–216. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4400-6_20
25. Abdal, R., Qin, Y., Wonka, P.: Image2StyleGAN: how to embed images into the StyleGAN latent space? In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4431–4440, October 2019

26. Xu, Z., Tao, D., Zhang, Y., Wu, J., Tsoi, A.C.: Architectural style classification using multinomial latent logistic regression. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 600–615. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_39
27. Shlens, J.: A tutorial on principal component analysis (2005). <http://arxiv.org/abs/1404.1100>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

