Application of 3D Image Generation Algorithm for Museum Space Design

Lu Wang[1] and Jian Wang[2,*]

[1]School of Design and Art, Jingdezhen Ceramic University, Jingdezhen, Jiangxi 333403, China, aolifu80@163.com
[2]School of Design and Art, Jingdezhen Ceramic University, Jingdezhen, Jiangxi 333403, China, 003733@jdu.edu.cn

Corresponding author: Jian Wang, 003733@jdu.edu.cn

Abstract. Museum cohesion of human history and civilization is the crystallization of people's wisdom. It has an essential significance for people to trace history, understand history and culture, and enrich spiritual and cultural life. Museum space design should integrate information into the space and complete the story. It is necessary to coordinate the relationship between virtual images and spatial language, complete mutual interaction and practice, and comprehensively explain the content provided, which places higher demands on museum designers. This paper introduces 3D image generation technology into museum space design and proposes a two-stage image generation method based on a generative adversarial network (GAN) to enhance the image generation capability of a GAN. The experimentFal results show that the proposed algorithm can improve the number of generative patterns of the model, generate high-quality samples, and provide technical support for museum space design.

Keywords: Computer-aided technology; image generation; museum space design; Generative adversarial network.
DOI: https://doi.org/10.14733/cadaps.2023.S8.24-32

1 INTRODUCTION

Museums are places where objects representing natural and human cultural heritage are collected, stored, exhibited, and researched, positioning the marks of human cultural development and bearing the burden of popularizing traditional culture and development history. In designing the display space in museums, the characteristics of the objects in the space should be fully revealed and echoed so that people can find the features of the objects when they are viewing them. Each museum has many display spaces, and each space has a specific theme and cultural relics; in the process of design for each display space, take into account the museum's design work and theme, the location of the details of the design, so that it echoes the overall, through the space to highlight the value and characteristics of the objects, so that the effect is very efficient to improve.

The design method of a spatial narrative is introduced into the display space design. As Kim et al. [1] stated, a specific narrative meaning is conveyed by relying on exhibits by setting up time, space, audience, scenes, and the interrelationship between them. A two-way communication network between curators and audiences is woven.

Museums, in general, are mainly designed in space plus designers for display techniques and display methods to highlight the characteristics of their exhibits and the narrative of space, and this way of expression and presentation is very logical, relevant, and analytical, thus making the spatial design of museums a more intense experience and feeling, As Baradaran Rahimi et al. [2] proposed. Spatial form design can sometimes replace the role of text. Suppose it can achieve the role of text transmission through intuitive feelings. In that case, the spatial design of museums will be relatively easy, so the design of display space must first structure the space, which also puts forward higher requirements for designers. The emergence of computer vision technology provides a new technological path for museum spatial design, which can reduce manual dependence, improve spatial design accuracy, and take spatial narrative into account.

Therefore, this paper attempts to introduce 3D image generation technology into museum spatial design and proposes a two-stage image generation method based on a generative adversarial network (GAN) to improve the image generation capability of a GAN in response to the problems of training instability and pattern collapse of image generation technology.

2 RELATED WORK

2.1 Image Generation

In the field of computer vision, image generation is a hot research topic. Image generation techniques can solve the problem of collecting and processing the available image datasets and reduce the cost of acquiring them. Besides, image generation techniques have many application scenarios, such as converting text into a specific image; also converting one style of image into another style; also restoring images with missing parts of the content intact, and also converting poor quality low-resolution images into more detailed high-resolution images [3].

Image generation algorithms have long been developed and widely used; for example, image processing software can generate the desired images, but this method requires manual processing, is time-consuming, and cannot generate images in large batches. In deep learning, image generation algorithms are more intelligent and can handle large-scale tasks. Jing et al. [4] argued that the standard image generation algorithms in the field of deep learning are Pixel CNN, Variational Auto-Encoder (VAE), and GAN. These three methods have their advantages and disadvantages. Still, GAN has a better performance, which can fit random noise to the corresponding image, generate the related content according to the semantics, and change the style of the image, converting the low resolution to the corresponding high-resolution image. The other two methods have some disadvantages, for example, Pixel CNN cannot operate in parallel, and its computational cost is high when used to generate large graphics; VAE learns the model for one problem only, and sometimes the sampling samples are blurred in the image training. GAN, on the other hand, is able to learn sample features more effectively, automatically generates clear samples, and also uses parallel computation for a shorter training time.

2.2 Generative Adversarial Networks

The idea of GAN comes from zero-sum games. The network's performance is improved by continuously playing the game, consisting of a generator and a discriminator. The generator learns the true distribution and generates the accurate distribution from the prior distribution. The discriminator acts as a "money checker" to identify the actual and generated samples. After the judgment, the result is communicated to the generator to define its direction of improvement better and create better images. The two can continuously train against each other, with the discriminator enhancing its recognition ability and the generator generating higher quality samples.

Eventually, the generated samples are true enough that the discriminator cannot distinguish between the generated samples, and the training reaches a Nash equilibrium state. GAN have been used in many fields, such as style conversion, image complementation [5], image hyper-segmentation [6], and other imaging fields; as well as natural language processing [7], and GAN has excelled in all areas with their outstanding capabilities.

Although GAN performs well, they have drawbacks and are difficult to train.GANs are designed as a model in which the generators and discriminators are continuously optimized to a Nash equilibrium state. Yamada et al. [8] emphasized that although GAN has theoretically optimal solutions, in practice, they suffer from many problems, such as training instability, gradient disappearance, and pattern collapse. Further, Wośko et al. [9] proposed that the optimization goal of the original GAN is the cross-entropy between the accurate data and the generated data, and by continuously reducing the cross-entropy, the gap between the generated and real data distribution becomes smaller, and the generated data becomes more like the actual data. If the loss of minimizing the generative model becomes a fixed constant when the discriminator is optimal, this leads to the disappearance of the gradient of the generative model in training, and the whole network cannot be further updated and optimized; if the discriminator is poorly trained, and the discriminator gives "encouragement" to the poorly generated samples, then the generator will continue to If the discriminator is poorly trained. The discriminator gives "encouragement" to the bad generated samples, then the generator will continue to generate similar bad samples, resulting in poor sample generation without the discriminator and generator knowing each other; only when the discriminator is trained just right, and its discriminatory ability is just good, then the whole network can keep optimizing in the right direction and generate samples close to the real ones, as Fand et al. [10] said. Therefore, one of the reasons why GANs are challenging to train is because the actual training is difficult to grasp in terms of proportionality.

3 METHODOLOGY

GAN has an excellent performance in image generation. Still, it suffers from training instability, and sometimes the quality of the generated images varies and lacks diversity, so it needs to be optimized. Currently, there are two main approaches to optimize GAN, one is to improve and refine the network structure, and the other is to optimize the loss function. However, there are a few approaches to optimize GANs from the training process. Therefore, the two-stage model in this paper first learns the features of an image by a feature generation network and then generates the image by another GAN, expecting thus to improve the generative power of the model.

3.1 Feature Generation Model (Stage 1)

In the first stage of the model, a feature generation model needs to be trained, which is also a generative adversarial model in nature, and the learning objectives of the generator change from images to features. As shown in Figure 1, the image is transmitted to the feature capture network. Then the features are output, the generator learns the features, and the discriminator discriminates between the generated features and the real features of the generator. The two are continuously played and evolved.

This paper uses a ResNet-based classification network as a feature capture network. Theoretically, the more layers of the neural network, the better the performance of the network and the more features are extracted, but in the experiments, it is found that the accuracy of the model decreases as the depth of the neural network increases. ResNet modifies the network structure by using a new cross-layer connection, which can use a deeper network structure and improve the gradient disappearance problem. The core component of ResNet is the residual module, the structure of which is shown in Figure 2, and the network is transmitted in the direction of an increasing number of layers and depth from top to bottom. The residual block adds an extra line to the upper network to connect to the lower network so that the information from the upper layer can be transmitted directly to the lower layer. This ensures that the information passed by the increased number of layers is valid.

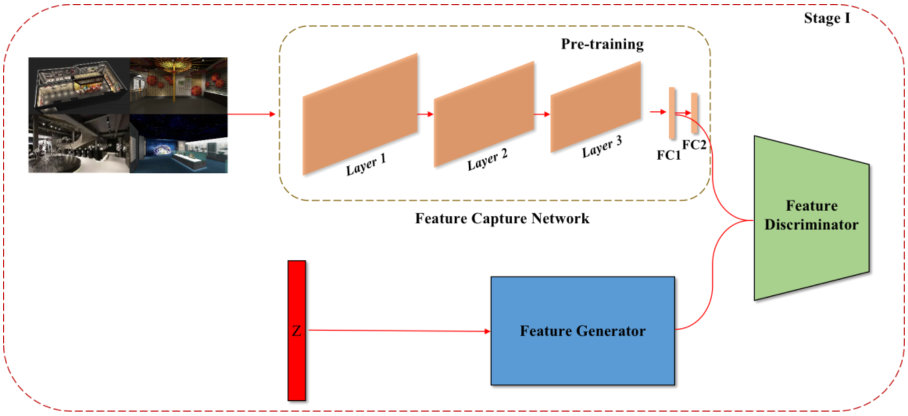

Figure 1: Feature generation model.

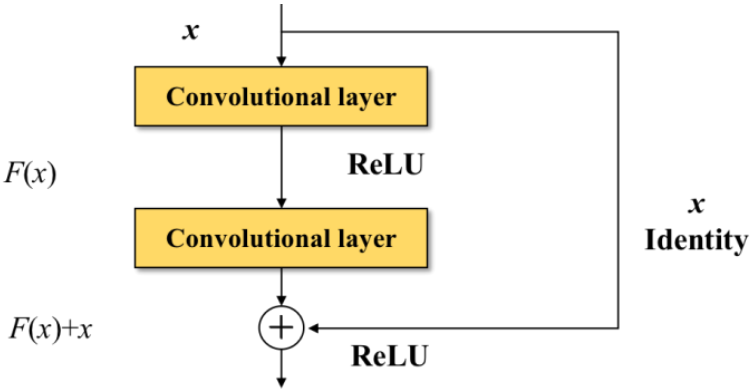

Figure 2: Residual block structure.

In addition to the above residual module, there is a bottleneck layer to design a deeper network. As shown in Figure 3, the basic residual module block is on the left and bottleneck layer bottleneck is on the right, which has one more convolutional layer and uses 1 × 1 convolutional layer. In the bottleneck, the 64-dimensional input is convolved with two 3 × 3 convolutions and then output; in the bottleneck, the 256-dimensional input is convolved with a 1 × 1 convolution, 3 × 3 convolutions, and 1 × 1 convolution and then output. bottleneck design can reduce the parameters by half compared to block, so bottleneck is more suitable for designing deeper networks. The bottleneck layer is used in deep networks such as ResNet-50/101/152. ResNet with bottleneck can design profound networks, and the increase in the number of parameters is not significant in practical training. ResNet, with a deeper network structure, has the same number of parameters as VGG, and ResNet has become a standard base network model for deep learning because of its deep network depth, a small number of parameters, and stable training.

The following experiments use the feature generation network mentioned above. We also use an Adam optimizer with a learning rate of 2×10^{-4}, parameters $\beta_1 = 0.0$, $\beta_2 = 0.9$, and 100,000 iteration rounds.

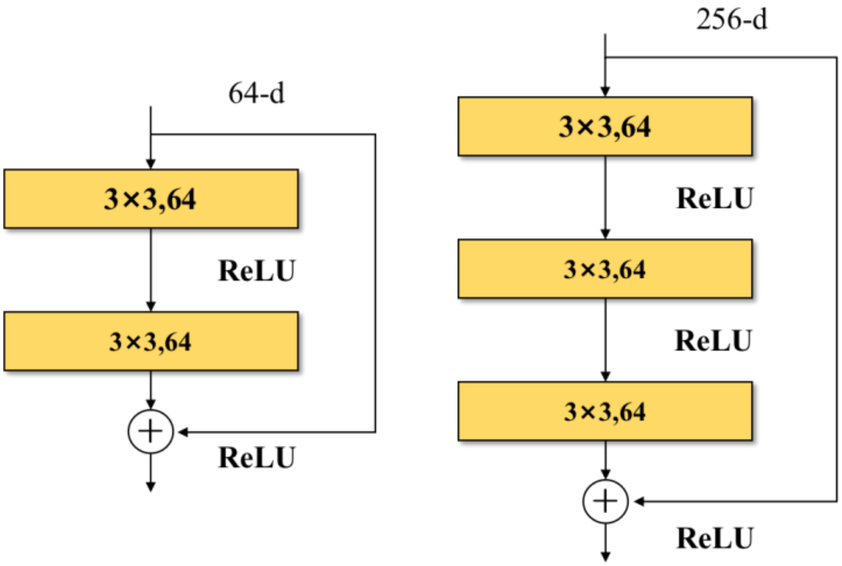

Figure 3: Block module and bottleneck module.

To ensure the experimental effect of the feature generation phase, the loss variation of the generator and the KL dispersion variation of the real and generated features are recorded in the experiment. The details of both are shown in Figure 4, and the model's training converges and is stable.

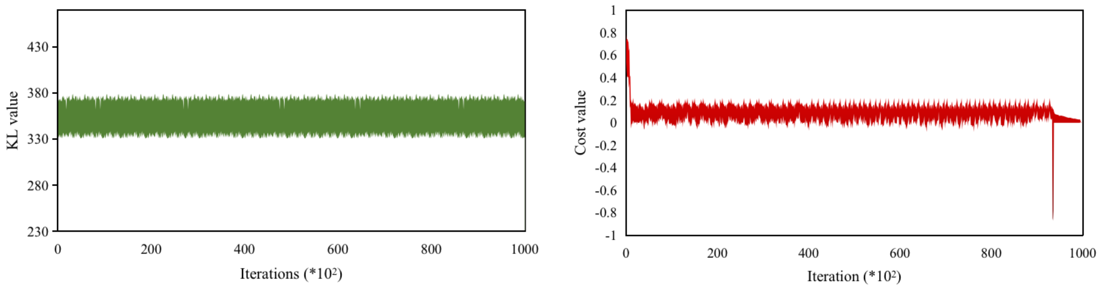

Figure 4: The loss of the generator in the feature generation phase and the variation of KL dispersion between the real features and the generated features.

3.2 Feature Generation Model (Stage 2)

After training the generator G_f that can generate features stably in the first stage, the second step, i.e., the image generation stage, can be performed. As shown in Figure 5, the generator and discriminator of the CT-GAN structure are used in this phase. Unlike the original CT-GAN, the feature generator G_f is added before the generator. Random noise z. After the feature generator G_f, the features are input to the image generator, and the discriminator discriminates between the real image and the generated image. The discriminator and generator structures in the image generation stage are consistent with the CT-GAN structure, using residual blocks, down-sampling, and Dropout in the discriminator of the model, and residual blocks, up-sampling, and other structures and methods in the generator.

The Adam optimizer is used in the training with a learning rate of 2×10^{-4}, parameters $\beta_1 = 0.0$, $\beta_2 = 0.9$, and 100,000 iterations. After two stages of training, the first and second stage models are combined to obtain the final model.

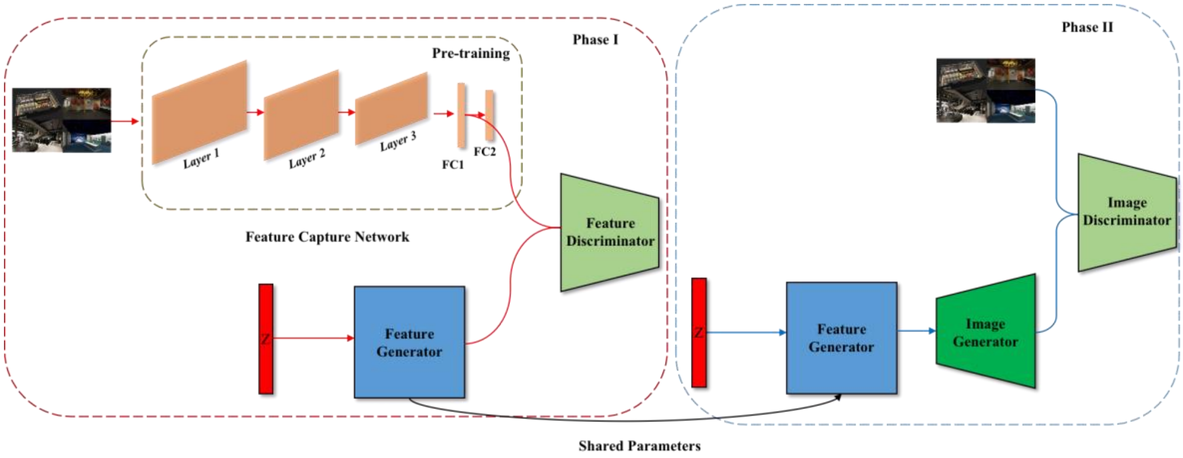

Figure 5: Two-stage generation model.

4 RESULT ANALYSIS

4.1 Loss Function and Training Method

The feature generation model in the first stage and the image generation model in the second stage are both GAN models, so the network structure and training methods are similar in this paper. Both networks are based on CT-GAN and adopt the same loss function as CT-GAN.

In the second stage of image generation, for example, when training, the network first trains the generator and inputs the random noise z to the feature generator G_f to obtain the feature $G_f(z)$, and then generates the image. The discriminator then discriminates the generated image from the real image as in the classical GAN network.

4.2 Feature Dimension Optimization

The CIFAR-10 classification network used in this paper is a ResNet network modified for the CIFAR-10 dataset. The dimension of the feature map before the fully connected layer becomes 256 dimensions after the pooling layer so that this 256-dimensional feature can be used as the learning objective of the first stage of the feature generation network. On the other hand, it is also possible to pool the feature map into 128-dimensional features or reduce the dimensionality to 128-dimensions by methods such as fully connected layers. Since features of different dimensions contain different amounts of information and have an impact on the structure of the generative network, the following section explores the impact of 256- and 128-dimensional features on the final image generation.

The same loss function and training method as WGAN-GP are used, and the Adam optimizer with parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$, batch size of 256, the learning rate of 2×10^{-4}, and the number of iterations of 20,000 is adopted. As shown in Figure 6, we can see that the impact of feature dimension on IS is minimal, and the model in this paper can be slightly improved on the base network. Considering the generality and the practice adopted in most studies, as well as the fact that the input of the image generator used in the CT-GAN paper is 128-dimensional, the dimensionality of the features is chosen to be 128-dimensional in the subsequent study.

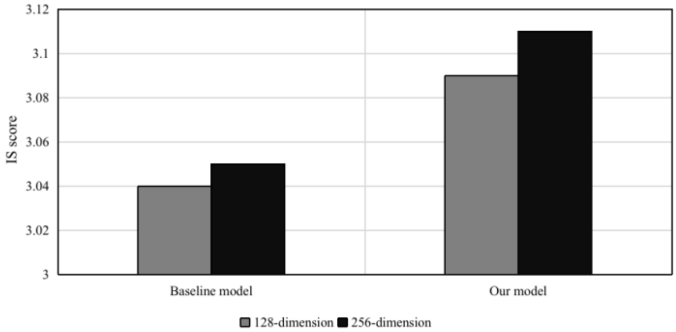

Figure 6: Effect of different intermediate feature dimensions on IS scores on small-scale WGAN-GP networks.

4.3 Feature Capture Network Optimization Analysis

In a classification network, the higher the classification accuracy of the network, i.e., the better the ability to recognize the images in the dataset; it can be assumed that the network with higher classification accuracy contains information in the feature map that is more effective for classification. Then, in the feature generation stage, the network with higher classification accuracy and the feature generator learns its features to achieve better results in the image generation stage.

The training was performed on the CIFAR-10 dataset with the following settings: SGD optimizer, the initial learning rate of 0.1, weight decay of 1×10^{-4}, the momentum of 0.9, training batch size of 128, test batch size of 100, and training period of 300. The final classification accuracies of different ResNet networks are shown in Figure 7. The Adam optimizer was used in training with a learning rate of 2 × 10-4 and parameters of $\beta_1 = 0.0$ and $\beta_2 = 0.9$ with 100,000 iterations. Figure 7 shows that the classification accuracy is higher when using ResNet-56 as the feature capture network in the feature generation phase compared with the ResNet-20 network. The IS score of the image generation in the second phase is higher, which indicates that the higher accuracy of the classification network in the feature generation phase is more beneficial and can improve the final image generation capability of the two-stage image generation model, so in the next step, we will try to using ResNet-110 network and compare it with ResNet-56 network.

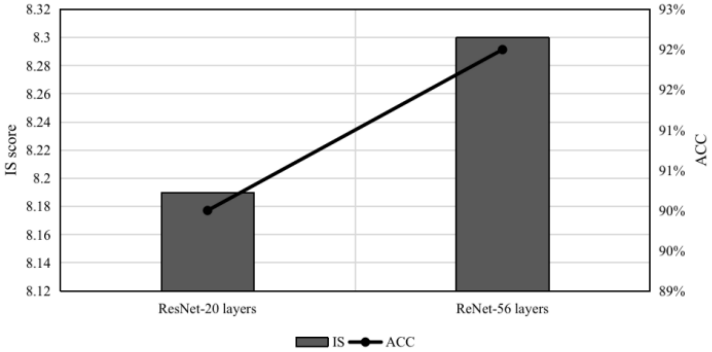

Figure 7: Comparison of classification accuracy and IS scores of different ResNet.

4.4 Base Model Selection Optimization Analysis

WGAN and WGAN-GP are two commonly used models in GAN. WGAN-GP solves the gradient disappearance problem of WGAN by gradient penalty and improves the model's ability to generate samples. CTGAN is an upgraded version of WGAN-GP with an additional CT term to enhance the Lipschitz continuity. In this subsection, we explore the performance of this model on WGAN-GP and CT-GAN and use it to select the appropriate model as the final model structure for this paper. The 128-dimensional features are used in the feature generation phase, and the feature capture network is based on ResNet-56. The loss function is consistent with CT-GAN, and the Adam optimizer is used in training with a learning rate of 2×10^{-4} and parameters $\beta_1 = 0.0$, $\beta_2 = 0.9$, and the number of iterations is 100,000.

The performance of this model based on CT-GAN and WGAN-GP on the CIFAR-10 dataset is obtained, as shown in Figure 8, where the IS scores of this method are improved on both different models, and the effect of using CT-GAN is better than that of WGAN-GP, which verifies the effectiveness and robustness of the method.

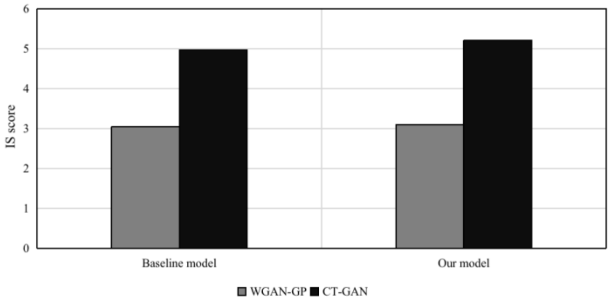

Figure 8: IS scores of WGAN-GP and CT-GAN on the CIFAR-10 dataset.

5 CONCLUSION

Museums are not only the symbols and signs of urban culture but also the base for people to understand and recover the history and play a vital role in meeting the cultural needs of contemporary people. The emergence of computer-aided technology, especially computer vision technology, has provided a new technical path for museum space design. However, little research has been done to discuss how to integrate computer vision technology with museum spatial design. Research in computer vision requires large-scale image datasets, but the acquisition of labeled images is difficult. Generating the images needed for research through technology is a convenient way to reduce labor costs and alleviate the shortage of high-quality image datasets. Among the many generation techniques, GAN has the advantages of high learning ability and good generation quality, which are very suitable for image generation work. However, there are many problems in GAN, such as an insufficient variety of generated samples, poor quality of generated samples, and unstable training. Therefore, this paper conducts a detailed study on the image generation problem of generative networks, propose a two-stage generation method, and demonstrate the method's effectiveness with experiments. Specifically, this paper presents a two-stage image generation method based on GAN. In the first stage, a classification network is trained to extract the features of the input image, then the parameters of the trained classification network are frozen, and the GAN is used to fit the features of the image after the classification network; in the second stage, the parameters of the feature generator in the first stage are frozen and added to a new GAN, so that the random input noise is firstly converted into an image by the feature generator and then by the image generator. The image generator is converted into an image, and then the discriminator distinguishes between generated and real images, thus aiding museum space design. The experimental results on the CIFAR-10 dataset show that the model using this paper has better image generation capability and higher image quality than the base model.

Lu Wang, https://orcid.org/0000-0002-2992-0877
Jian Wang, https://orcid.org/0000-0002-9674-8229

REFERENCES

[1] Kim, H.; Yeo, Y.: A study on public space design strategies in Japanese National Museums-focusing on The National Museum of Western Art and The National Art Center in Tokyo, Journal of Asian Architecture and Building Engineering, 18(2), 2019, 121-127. http://doi.org/10.1080/13467581.2019.1601567

[2] Baradaran Rahimi, F.; Levy, R.-M.; Boyd, J.-E.: Hybrid space: An emerging opportunity that alternative reality technologies offer to the museums, Space and Culture, 24(1), 2021, 83-96. http://doi.org/10.1177/1206331218793065

[3] Zhao, B.; Yin, W.; Meng, L.; Sigal, L.: Layout2image: Image generation from layout, International Journal of Computer Vision, 128(10), 2020, 2418-2435. http://doi.org/10.1007/s11263-020-01300-7

[4] Jing, B.; Ding, H.; Yang, Z.; Li, B.; Liu, Q.: Image generation step by step: Animation generation-image translation, Applied Intelligence, 52(7), 2022, 8087-8100. http://doi.org/10.1007/s10489-021-02835-z

[5] Hashimoto, Y.; Yamane, K.; Okada, N.; Tadatomo, K.: Growth of semipolar {20–21} GaN and {20–2–1} GaN for GaN substrate, Physica Status Solidi(b), 253(1), 2016, 36-45. http://doi.org/10.1002/pssb.201552271

[6] Asri, R.-I.-M.; Hamzah, N.-A.; Ahmad, M.-A.; Alias, E.-A.; Sahar, M.-A.-A.-Z.-M.; Abdullah, M.: Influence of growth temperature of p-GaN layer on the characteristics of InGaN/GaN blue light emitting diodes, International Journal of Nanotechnology, 19(2-5), 2022, 344-355. http://doi.org/10.1504/IJNT.2022.124514

[7] Al Taradeh, N.; Frayssinet, E.; Rodriguez, C.; Morancho, F.; Sonneville, C.; Phung, L.-V.; Maher, H.: Characterization of m-GAN and a-GAN crystallographic planes after being chemically etched in TMAH solution, Energies, 14(14), 2021, 4241. http://doi.org/10.3390/en14144241

[8] Yamada, H.; Chonan, H.; Takahashi, T.; Shimizu, M.: Comparison of Electrical Properties of Ni/n-GaN Schottky Diodes on c-Plane and m-Plane GaN Substrates, Physica Status Solidi (a), 215(8), 2018, 1700362. http://doi.org/10.1002/pssa.201700362

[9] Wośko, M.; Paszkiewicz, B.; Paszkiewicz, R.: AlGaN/GaN heterostructures electrical performance by altering GaN/sapphire buffers hrowth pressure and low-temperature GAN interlayers application, Crystal Research and Technology, 56(12), 2021, 2100090. http://doi.org/10.1002/crat.202100090

[10] Fang, W.; Gu, E.; Yi, W.; Wang, W.; Sheng, V.-S.: A new method of image restoration technology based on WGAN, Computer Systems Science and Engineering, 41(2), 2022, 689-698. http://doi.org/10.32604/csse.2022.020176