# A Sketch-Based Generation Model for Diverse Ceramic Tile Images Using Generative Adversarial Network

**Jianfeng Lu[1,*], Xinyi Liu[1], Mengtao Shi[1], Chen Cui[1,2] and Mahmoud Emam[1,3,*]**

[1]School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, 310018, China
[2]Key Laboratory of Public Security Information Application Based on Big-Data Architecture, Ministry of Public Security, Zhejiang Police College, Hangzhou, 310000, China
[3]Faculty of Artificial Intelligence, Menoufia University, Shebin El-Koom, 32511, Egypt
*Corresponding Authors: Jianfeng Lu. Email: jflu@hdu.edu.cn; Mahmoud Emam. Email: memam@ai.menofia.edu.eg

**Abstract:** Ceramic tiles are one of the most indispensable materials for interior decoration. The ceramic patterns can't match the design requirements in terms of diversity and interactivity due to their natural textures. In this paper, we propose a sketch-based generation method for generating diverse ceramic tile images based on a hand-drawn sketches using Generative Adversarial Network (GAN). The generated tile images can be tailored to meet the specific needs of the user for the tile textures. The proposed method consists of four steps. Firstly, a dataset of ceramic tile images with diverse distributions is created and then pre-trained based on GAN. Secondly, for each ceramic tile image in the dataset, the corresponding sketch image is generated and then the mapping relationship between the images is trained based on a sketch extraction network using ResNet Block and jump connection to improve the quality of the generated sketches. Thirdly, the sketch style is redefined according to the characteristics of the ceramic tile images and then double cross-domain adversarial loss functions are employed to guide the ceramic tile generation network for fitting in the direction of the sketch style and to improve the training speed. Finally, we apply hidden space perturbation and interpolation for further enriching the output textures style and satisfying the concept of "one style with multiple faces". We conduct the training process of the proposed generation network on 2583 ceramic tile images dataset. To measure the generative diversity and quality, we use Frechet Inception Distance (FID) and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) metrics. The experimental results prove that the proposed model greatly enhances the generation results of the ceramic tile images, with FID of 32.47 and BRISQUE of 28.44.

**Keywords:** Ceramic tile pattern design; cross-domain learning; deep learning; GAN; generative adversarial networks; ResNet Block

## 1 Introduction

Nowadays, people are seeking for computer-based home architectural interior finishes due to the rapid development of technology. Ceramic tiles are one of the most indispensable materials for interior decoration that are usually derived from natural textures which are difficult to make any changes on it. So, in industry, the product patterns can't meet the design requirements in terms of diversity and interactivity. With the development of modern computer graphics and photo realistic drawing tools, researchers have applied some methods based on texture generation for designing ceramic tile images to improve design efficiency and save more time [1]. However, these methods have some disadvantages such as: possible deformation of the generated textures, and existence of artifacts at the seams of the stitched textures. Some traditional methods have been developed to control the generated images using parametric textures, but they are limited to some categories as: fluid textures and marble textures. Recently, researchers are focus on how to find suitable features for image texture description and parameterized features to design ceramic tile images. Another aspect of the research is how to effectively represent the designer's ideas interactively. In recent years, the designers' hand-drawn sketches have been used as a representation of the design idea, and then the sketch features are employed to control the textures of the generated images. Whereas, retrieving the sketch features is still a key problem. To realize a mapping between sketch and actual image, Hu et al. [2] used the bag-of-words model to represent the sketch edge detection and developed the invariant features in the image domain and the sketch domain. Furthermore, these methods usually require elaborated feature representations and involve more complex post-processing steps such as: graph cutting synthesis and gradient domain hybrid, to ensure the authenticity of the composite image. Additionally, a large dataset is required, which increases the complexity. Moreover, the similarity between style and texture allows using style as a texture and complete the conversion from image to image by style transfer. Portilla et al. [3] used statistical information of local features to represent the texture of the original image, and then applied image resampling to obtain new texture to generate the required stylization image. However, when their method was applied to generate ceramic tile images, some important textures are identified as noise and cannot be mapped. Furthermore, a sketch ceramic tile texture contour is also difficult to fit with real ceramic tile image, so it is difficult to realize interactive design requirements.

With the widespread applications of deep learning in the field of image generation and classification, deep learning-based methods are received the researchers' attention. Generative Adversarial Networks (GAN) [4] is usually used to generate image contents. It can be used to generate highly realistic human faces or natural images. However, utilizing GAN in the field of ceramic tile images design is still a challenging task, and one of the main problems is that the images generated by GAN-based methods generally have high non-controllability and randomness. So, it is difficult to represent the designer's ideas to generate the images. Therefore, the problem of generating realistic texture images has a significant research value and a potential large-scale of applications. In this paper, we propose a generation method which takes the designer's hand-drawn sketches as an input for the latest GAN framework and redesigns the network loss function to enable it to automatically generate ceramic tile images that match the sketch texture morphology. Consequently, the proposed method can solve the problem of interactivity and diversity in ceramic tile image design. The main contributions of this paper can be summarized as follows:

- We propose a sketch-based generation model for generating diverse ceramic tile images based on sketch images using GAN, which can effectively express the design ideas over the interactivity of the final images. On this basis, the generation network can generate ceramic tile images with similar texture patterns to the hand-drawn sketches.

- We establish ceramic tile images with diversified distribution and the corresponding sketch images dataset. Furthermore, the generation network is pre-trained.
- We interpolate the latent space of the GAN network to obtain more diverse ceramic tile texture images.

This paper is organized as follows: related works are presented in Section 2. Section 3 describes the proposed sketch-based generation model for diverse ceramic tile images. Experimental results and analysis are then discussed in Section 4. Finally, Section 5 concludes the paper and presents the limitations of the proposed model in addition to the potential future directions for improvement.

## 2 Related Works

In this section, we introduce some sketch extraction and image content generation approaches. Sketch extraction methods are used to extract textures from the ceramic tile images, and image content generation is then applied to synthesize ceramic tile images that express the designer's ideas.

Generally, sketch textures can simplify and abstract the real images. Chen et al. [5] reduced the redundant edge information of the image by binarization and morphology to abstract the real image. Besides, feature representation is also an important area where sketch textures can be applied. Chen et al. [6] and Xu et al. [7] used hand-drawn sketches to represent texture features of human faces, cats, and clocks. For sketch texture extraction, traditional image classification and generation algorithms are employed by using first-order and second-order operators. Sobel et al. [8] detected the image edges by calculating the approximate gradient of the change in image brightness values. Their method was relatively small in computation and fast in recognition, but it is not good enough for image recognition tasks. Canny algorithm [9] was then integrated and optimized based on the predecessors. It has a better contour extraction effect, but the extraction parameters need to be optimized for different images. Due to the efficiency of deep learning approaches in complex parameter optimization, it can be applied in the field of end-to-end image generation [10–12]. Li et al. [13] proposed an automatic generation method for high-quality graffiti drawing based on deep learning. In recent years, many deep learning-based network structures as: Scribbler [14], UGATIT [15], pix2pix [16] and SketchyGAN [5] have been proposed successfully. Additionally, pix2pix network provides a straightforward method for converting from image to image, which has the advantage of being more versatile and a good performance in image conversion. The pix2pix network uses UNet [17] network as a generator, and PatchGAN as a discriminator. In traditional GAN, L1 or L2 loss functions are used, which can lead to blurring in the output images and loss in the high frequency information of the images. Therefore, to obtain the output image with better details, the input image can be segmented first into multiple patches by using PatchGAN. Furthermore, PatchGAN judges the truth of each patch separately and finally takes their average as an output of the discriminator. Pix2pix network provides a direct method to convert one image into another image. Therefore, it can be applied to the sketch texture extraction for ceramic tile images.

In image content generation, Tero et al. [18–20] have proposed ProGAN, StyleGAN and Style-GAN2 networks, respectively. The core idea of ProGAN [18] lies in progressive training, where the network starts with very low-resolution images (e.g., $4 \times 4$) and adds one higher resolution layer at a time, image texture details are added to it until the generator can generate a high-resolution target image. However, ProGAN has some limitations in controlling specific features of the generated images, and it is difficult to achieve fine-tuning of individual attributes. On this basis, StyleGAN [19] has been proposed. Generating intermediate latent vector $w$ by a series of affine transformations of random vector z, the space constituted by $w$ is called the latent space. It is used to control the style of generated

image, while the AdaIN (adaptive instance normalization) module is added to the generator structure to generate the image. By these measures, StyleGAN can modify the input of each level separately to control the visual features corresponding to that level without affecting the other levels. Since AdaIN destroys the inter-feature information and causes artifacts in the generated images, Tero et al. [20] adapted the AdaIN module and proposed StyleGAN2. This generation of StyleGAN has better stability during training and the number of optimizations for regular items has been reduced, which reduces the computational efforts to some extent.

In terms of controllability of the output image texture, Lu et al. [21] proposed multi-stage image generation network by adding multi-scale shortcut connections with attention weights to the styleGAN, to control image texture coarseness. Their model seems to be more effective, but it still needs further improvements in terms of controllability in texture and style of the generated image [21]. Furthermore, Wang et al. [22] proposed changing the weights of the original GAN model based on the hand-drawn sketches. The output of the model was utilized and matched with the hand-drawn sketches by cross-domain adversarial loss, to maintain the diversity of the original model and generated images quality. However, their model is only use thin-edge contours, which cannot successfully generate ceramic tile images as the original input sketch. Additionally, the overall image aesthetics and details were not well-handled. Therefore, this paper redesigns the ceramic tile texture extraction network based on Wang's et al. network, to achieve the separation of ceramic tile pattern texture regions, and more effectively guide the generated ceramic tile image texture form to approximate the input sketch and obtain the ceramic tile image that can reflect the designer's intention.

## 3 Sketch-based Generation Model for Diverse Ceramic Tile Images

### 3.1 Proposed Method

In this paper, a sketch-based method for diverse ceramic tile image generation is proposed, and the flow diagram of the proposed method is shown in Fig. 1, which is mainly divided into five parts:

- In part (I) Ceramic tile dataset generation: the ceramic tile dataset required for training is obtained by pre-processing the ceramic tile data provided by well-known ceramic tile manufacturers, such as Newpearl ceramics. The pre-processing operations are composed of some basic operations such as binarization and opening & closing morphological operations, to obtain the ceramic tile sketch dataset which will be used for the subsequent network training.
- In part (II) Sketch extraction network 'F': The obtained ceramic tile sketch dataset is trained through the network structure proposed in this paper to obtain the sketch generator F.
- In part (III) Ceramic tile image generation network 'S': The original ceramic tile dataset is trained through the StyleGAN2 network to generate the required ceramic tile image generator G.
- In part (IV) Sketch-based ceramic tile image generation network 'P': The generator F of sketch extraction network F and the generator G of ceramic tile training network S are entered to the sketch-based ceramic tile image generation network P. After iterative training with specified number of times, the generator G′ of the required specific sketch style is obtained.
- In part (V) Ceramic tile image diversity expansion: The completed generator G′ which trained in the sketch-based ceramic tile image generation network P is used to obtain more diverse ceramic tile texture images by using latent space perturbation and interpolation.
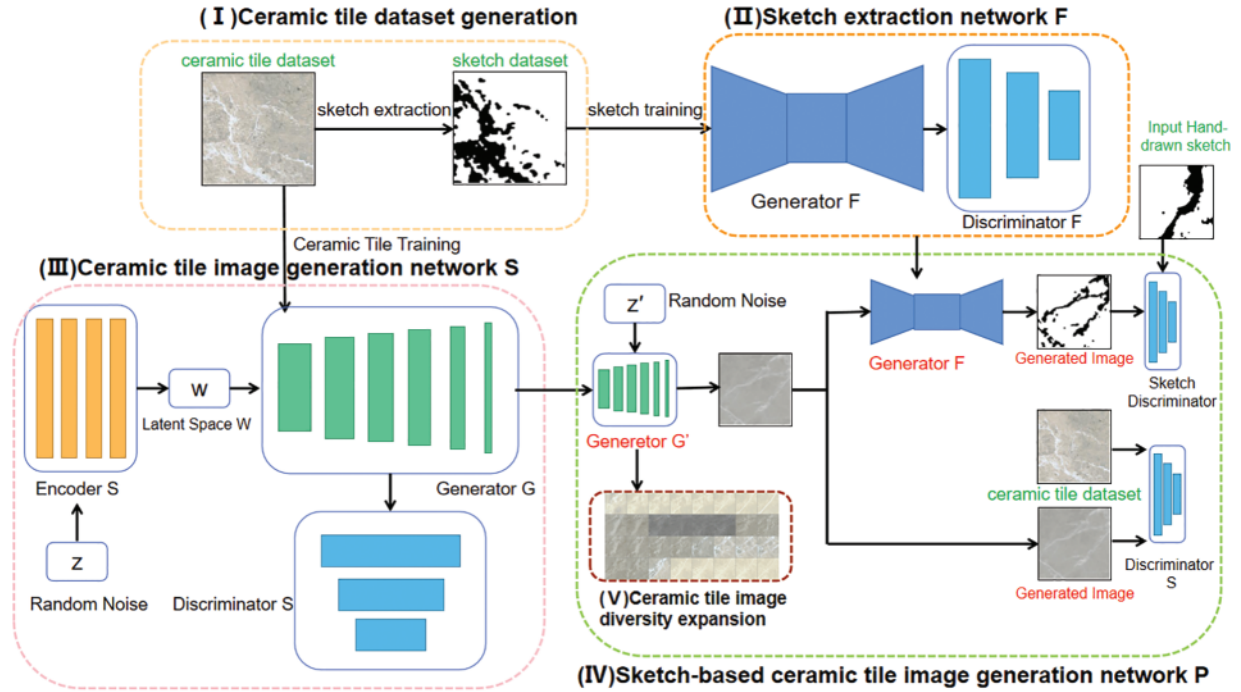
**Figure 1:** Flow diagram for the proposed sketch-based method for diverse ceramic tile image generation

### 3.2 Ceramic Tile Images Dataset

The dataset of the ceramic tile generation network is built independently by using ceramic tile data provided by ceramic tile manufacturers. Due to the very high resolution of the provided tile images and their irregular textures, we select only the tile images with more clear textures and diversity in styles. Then, we perform some basic pre-processing operations on the selected tile images as shown in Fig. 2.

From Fig. 2 part (I), we first reduce high resolution tile images isometrically by 2×, 4× and 8×, respectively. Then, the horizontal and vertical axes of the tile images are cropped to 256 × 256 pixels These ceramic tiles images are then rotated clockwise by 90°, 180° and 270°; respectively. Finally, the less-textured solid color blocks are eliminated, and the high-textured blocks are retained and saved in the dataset. Through the above-mentioned operations, we obtain 2583 ceramic tile images as the generation dataset. Fig. 3 shows some ceramic tile images from the generated dataset. This dataset will be used to conduct the training process of the ceramic tile generation network.

The above generated ceramic tile images dataset is processed twice to get the image pairs of the sketch and ceramic tile image, as shown in Fig. 2 part (II). Each image in the dataset is firstly converted to its grayscale version. Then, for each grayscale image we obtain contour and binary image. The contour image is obtained by extracting the texture contour using the canny operator, whereas the binary image is obtained by the binarization and inversion operations. The two obtained images (contour and binary) are subjected to the opening and closing morphological operations. The opening operation makes the contour of the image more smooth and eliminates the fine protrusions and it can be defined as in Eq. (1), where template B erodes image A, and then the result is expanded with template B.

$$A \circ B = (A \ominus B) \oplus B \tag{1}$$

whereas the closing operation can fill the tiny gaps (small holes) and it can be defined as in Eq. (2), where an expansion operation is performed on image A with template B, followed by an erosion operation on image B.

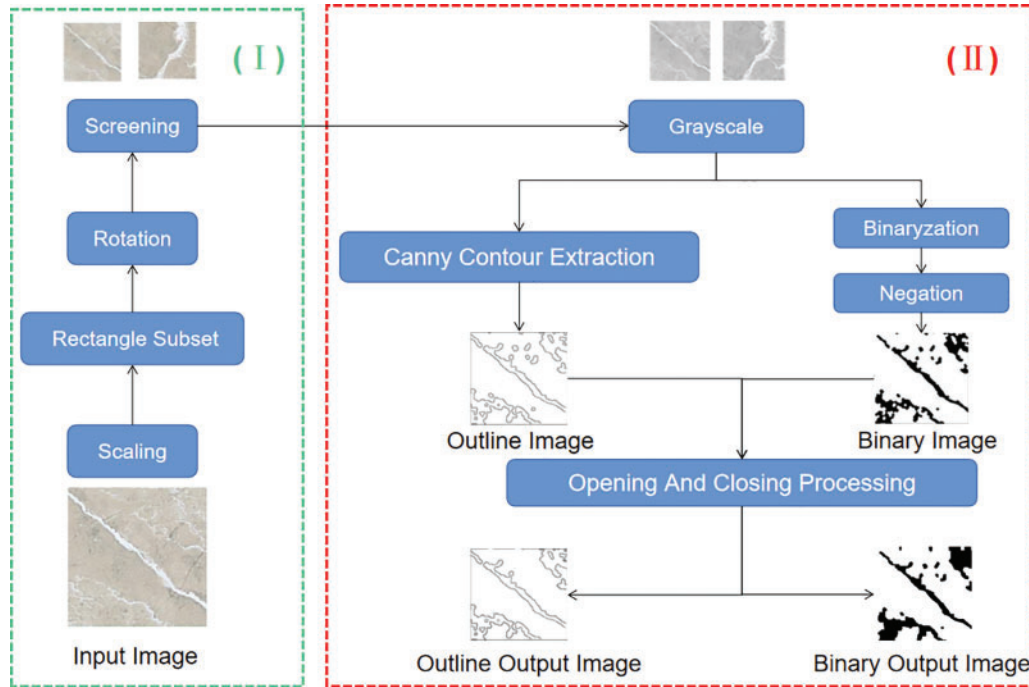$$A \cdot B = (A \oplus B) \ominus B \qquad (2)$$



**Figure 2:** Basic preprocessing operations for preparing the ceramic tile images dataset
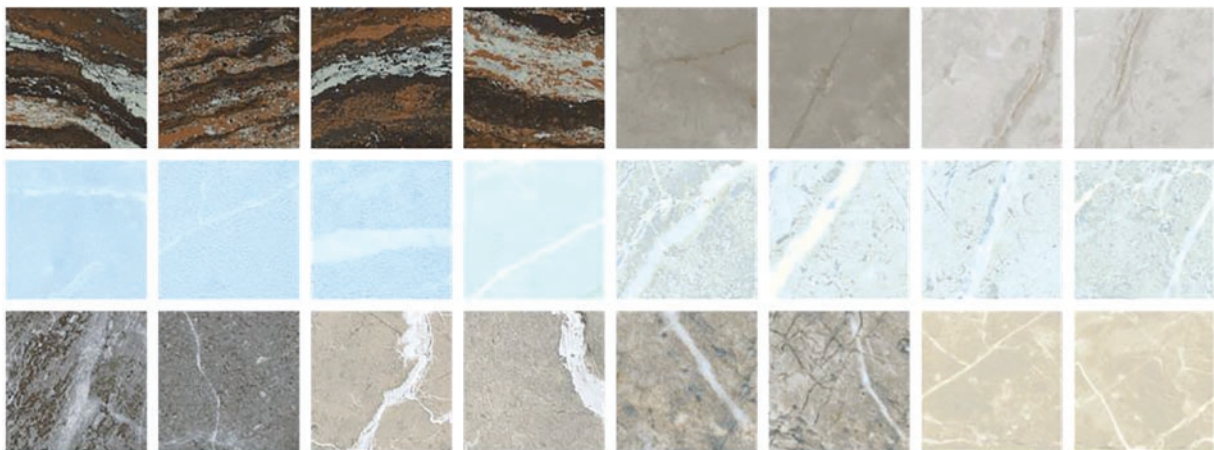


**Figure 3:** Some ceramic tile images from the dataset

After some preliminary experiments, we find that the undersized filters with size $3 * 3$ and $5 * 5$ are too small and produce more noise. Therefore, the proposed method uses a circular filter with a template size of $7 \times 7$. The binarized ceramic tile image is first subjected to an opening operation

to eliminate the excess noise and then a closing operation to fill the small cracks. The results of the opening and closing operations are shown in Fig. 4.
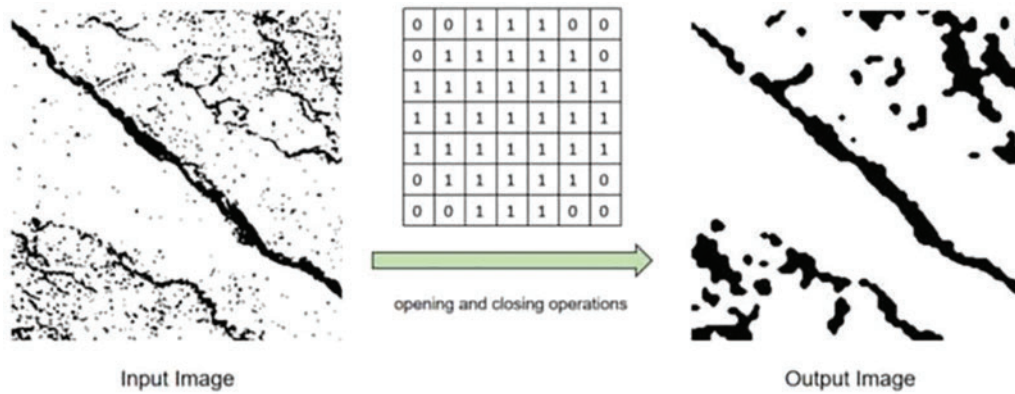


**Figure 4:** Schematic diagram of the opening and closing operations

Through the above steps, the obtained images will be used for the training of the sketch extraction network in the following Subsection 3.3. Some of the resulted images are shown in Fig. 5.
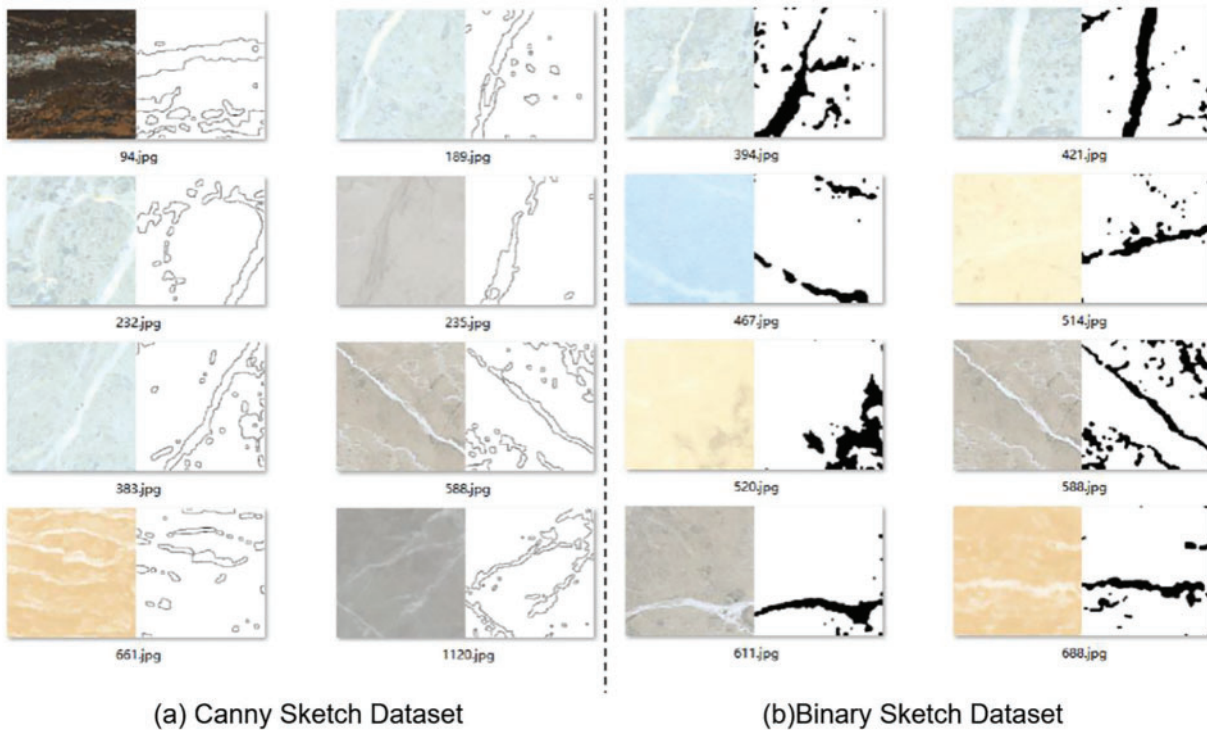


**Figure 5:** Graphical representations of some sketch images

### 3.3  Training of Sketch Extraction Network

Considering the versatility advantage of the pix2pix network and its excellent image generation results [10], the proposed sketch extraction network is based on the idea of pix2pix network. In pix2pix network, ResNet Block module is presented for the generator construction. Furthermore, the jump connection is used to make the information between residual blocks pass through the next layer without loss, which improves the flow of information and speeds up the convergence of the network. Fig. 6 shows the structure of the ResNet Block module. In this paper, the pix2pix network with the generator structure of UNet is called pix2pix-UNet, and with the improved ResNet Block module is called pix2pix-ResNet Block.
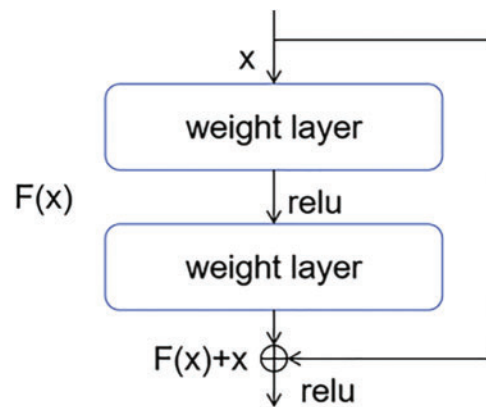


**Figure 6:** The structure diagram of the ResNet Block module

The generator model is shown in Fig. 7. Firstly, a $256 \times 256$ three-channel ceramic tile image is used as an input, and the number of channels is first increased to 64 layers by a convolutional layer with a convolutional kernel size of $7 \times 7$, and then followed by two convolutional downsampling layers with a convolutional kernel size of $3 \times 3$ and a step size of 2. Then, through 6 ResNet Block modules and join the skip connection. This stage does not change the size of the feature map or the number of channels. The image size is reduced to the original size of $256 \times 256$ by two transposed convolutional upsampling layers with a convolutional kernel size of $3 \times 3$ and a step size of 2. Finally, a convolutional layer with a convolutional kernel size of $7 \times 7$ and an activation function of *Tan* is used to output a single-channel ceramic tiled sketch image with a size of $256 \times 256$. Moreover, the discriminator is adopted from the original structure PatchGAN discriminator in pix2pix. The ceramic tile images and ceramic tile sketch pairs generated in Subsection 3.2 are then used as an input to the improved pix2pix network, and the trained generator is used for the next network training.

### 3.4  Construction of Generation Network

The proposed ceramic tile image generation network is inspired by Wang's et al. [22] network architecture, which modifies the network architecture of the sketch extraction module and proposes a new sketch form according to the characteristics of the ceramic tile images. Firstly, the ceramic tile designer inserts one or more sketches and then matches the sketches through a cross-domain adversarial loss, to modify the weight parameters of the ceramic tile generator network in reverse. The proposed generation network structure of the sketch-based ceramic tile image is shown in Fig. 8. Firstly, a noise Z is randomly generated, then a generated real ceramic tile image is obtained by pre-trained the ceramic tile image generator G of the StyleGAN2 network. The generated ceramic tile image is passed through the proposed generator F of the pre-trained sketch extraction network F, to get

the sketch corresponding to this image, and the distance between the ceramic tile image and the input sketch is calculated by sketch confrontation loss. At the same time, the distance between the generated ceramic tile image and the real image is calculated by image confrontation loss to reversely modify the parameters of the generator G. After certain iterative training, the ceramic tile image generator $G'$ based on the sketch matching is finally obtained, where X,Y are the real ceramic tile image domain and the generated ceramic tile sketch image domain, respectively; the ceramic tile image generator $G'$ represents the sketch-based ceramic tile image generator completed by training, while its initial weights are derived from the generator G of the StyleGAN2 network structure pre-trained from the real ceramic tile dataset before the network training. The generator F of the sketch extraction network F represents the proposed generator of the pre-trained sketch extraction network, but this generator is not involved in the training process, while the ceramic tile image generator $G'$ and two discriminators $D_X$ and $D_Y$ are involved in the training of the network weights.
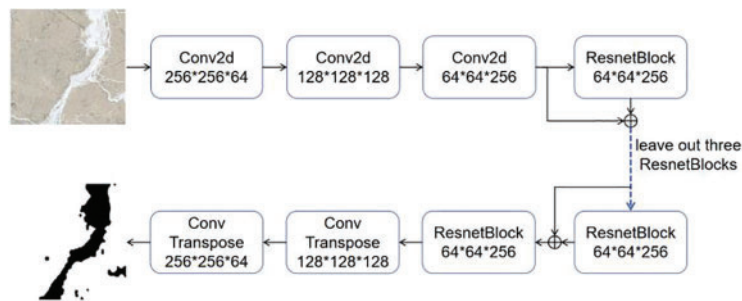


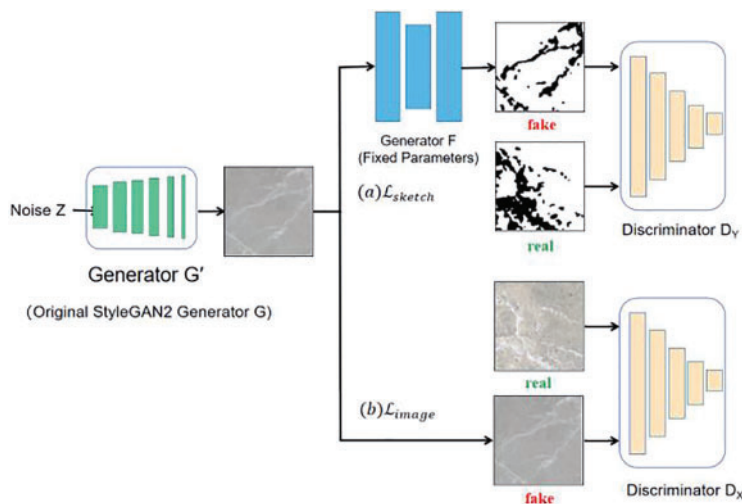**Figure 7:** The proposed network structure with ResNet Block generator



**Figure 8:** The structure of the proposed generation network for the sketch-based ceramic tile images

Since, the designers will not insert many ceramic tile images for training when designing ceramic tile images, they need only to assist the generation of ceramic tile images by inserting a small number of hand-drawn sketches. To solve this problem, the proposed method uses a pre-trained completed sketch extraction network to generate texture sketches in real-time. To bridge the gap between the sketch training data and the image generation model, the proposed method uses double cross-domain adversarial loss to encourage the generated ceramic tile image texture structure to match the texture

structure of the ceramic tile sketch, so the output image of the generator G will pass through the pre-trained ceramic tile sketch extraction network F before entering to the discriminator. The adversarial loss function for sketch matching is shown in Eq. (3), where z is the input noise vector, G is the ceramic tile image generator, F is the ceramic tile sketch generator, and $D_Y$ is the sketch discriminator.

$$\mathcal{L}_{\text{sketch}} = \mathbb{E}_{y \sim p_{\text{data}(y)}} \log\left(D_Y\left(y\right)\right) + \mathbb{E}_{z \sim p_{(z)}} \log\left(1 - D_Y\left(F\left(G\left(z\right)\right)\right)\right) \tag{3}$$

We observe that, using the loss function only on the sketch will lead to a sharp decrease in the quality and diversity of the generated ceramic tile images. To solve this problem, a second adversarial loss is added among the network as shown in Eq. (4), to compare the output results with the training set of the original model. A separate discriminator $D_X$ is used to match sketches while maintaining the image quality and diversity of the model output.

$$\mathcal{L}_{\text{image}} = \mathbb{E}_{x \sim p_{\text{data}(x)}} \log\left(D_X\left(x\right)\right) + \mathbb{E}_{z \sim p_{(z)}} \log\left(1 - D_X\left(G\left(z\right)\right)\right) \tag{4}$$

Finally, the loss function of this network is shown in Eq. (5), and $\lambda_{\text{image}}$ denotes the weight of the image against loss.

$$\mathcal{L} = \mathcal{L}_{\text{sketch}} + \lambda_{\text{image}}\mathcal{L}_{\text{image}} \tag{5}$$

After preliminary experiments, the best results are obtained at $\lambda_{\text{image}} = 0.7$. The network is trained to obtain an improved generator G′ with the minimum objective shown in Eq. (6).

$$\Theta' = \arg \min_{\Theta'} \max_{D_X, D_Y} \mathcal{L} \tag{6}$$

To prevent the proposed model from overfitting and speed up the fine-tuning, the generator parameters of the proposed method use a pre-trained ceramic tile image generator instead of randomly assigned initial weights. We experimentally observed that modifying the mapping network is sufficient to obtain the target distribution, which is a subset of the original distribution. Furthermore, to accelerate the training speed of the network and improve the reliability of the network, we use pre-trained ceramic tile image generator G and the generator of the sketch extraction network F. Additionally, the pre-trained ceramic tile image discriminator $D_X$ and $D_Y$ in StyleGAN2 are also used as initial weights, which will further improve the training efficiency of the network.

### 3.5 Expansion of Image Diversity Design

Usually, the design of ceramic tile images is not limited to the design of a specific texture, but also must generate more similar texture styles; such a feature is called "one ceramic tile styles with multiple faces". Inspired by the idea of Deep Generative Prior (DGP) to realize the design concept of multiple faces for one ceramic tile style, we generate natural and diverse ceramic tile images by perturbing the generated latent vectors and applying linear interpolation with different weights. This process can be summarized in two steps as follows:

- **Image W latent space perturbation method:** the random noise vector *z* is fed into the trained completed G′ model of generation network model to obtain the *w*-latent vector in the **W**-space domain. The random Gaussian noise is then added to the obtained *w*-latent vector, and the generated ceramic tile images are available through the generated image module, which usually close in style and have some reference significance.
- **Model interpolation method:** the two completed trained G′ models of Subsection 3.4 are linearly interpolated by different weights, where $\alpha \in [0,1]$ controls the interpolation, and the interpolation step is shown in Eq. (7), where $W_1$, $W_2$ are the two models to be involved in the

interpolation. A random noise vector is fed into the interpolated models with different weight parameters to obtain an image with a transition texture shape between the two models.

$$W = \alpha W_1 + (1 - \alpha)W_2 \tag{7}$$

Overall, the output generated images are structurally similar to the input images due to the existence of the proposed sketch loss function. Furthermore, the proposed sketch-based generation model for generating diverse ceramic tile images can effectively express the design ideas over the interactivity of the final images. Therefore, the generation network can generate ceramic tile images with similar texture patterns to the hand-drawn sketches.

## 4  Experimental Results and Analysis

### 4.1  Experiments Environment

The models for the experiments are built based on the Pytorch framework, and the server contains four 10-core CPUs, model Intel(R) Xeon(R) Silver 4210 CPU @ 2.20 GHz, 128 G of RAM, and GPU model GeForce RTX 2080Ti with 12 G of video memory. The dataset generation and pre-processing steps are done by using MATLAB 2016a.

### 4.2  Generation Network Performance

After selecting the data from the ceramic tile image dataset created in this paper and choosing the corresponding binarization threshold, the texture contour images by using Canny operator are extracted. We select a $7 \times 7$ circular template for the opening and closing operations to remove the redundant noise and smooth the contour image. Then by human screening, we finally obtained ceramic tile sketch pairs as a dataset for the comparison tests. The sketch dataset is randomly divided into 70% as a training set, 20% as a validation set and 10% as a testing set. To evaluate the effect of different generative adversarial models on the quality of the sketch generation images, we used CycleGAN and pix2pix models with UNet (pix2pix-UNet) as generators for the comparisons. The number of the training epochs are set to 200. The testing set is fed into the completed training network and some of the generated results are shown in Figs. 9 and 10. We find that the whole training time is 18 h.

By comparing the red dashed boxes in the Figs. 9 and 10, we can notice that the proposed method is more similar to the input labeled graph in terms of detail portrayal. Overall, the texture outlined by the proposed method is more detailed than others. Additionally, the structure of the proposed sketch extraction network outperforms the compared methods in terms of visual similarity regardless of using Canny or binarization sketches. Since CycleGAN does not fit the ceramic tile sketch image well, it can't be used as a generator for the sketch extraction network in the sketch-based ceramic tile image generation network.

To evaluate the effect of different types of sketches and sketch extraction network F on the sketch-based ceramic tile image generation network P, there are four sets of comparison experiments are performed (Canny Sketch with Pix2Pix-UNet, Canny Sketch with Pix2Pix-ResNetBlock, Binary Sketch with Pix2Pix-UNet and Binary Sketch with Pix2Pix-ResNetBlock). The two sketch datasets are first fed into the original pix2pix network and the generator-optimized pix2pix network. Then, the output of each sketch extraction network is fed into the sketch-based ceramic tile image generation network structure (as described in Subsection 3.4) for weight adjustment respectively.
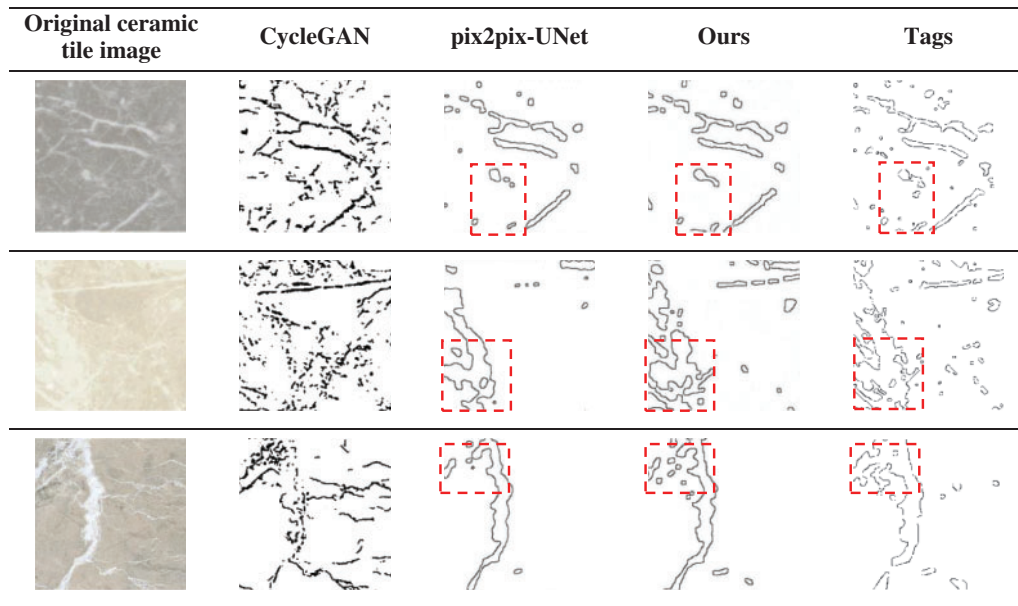
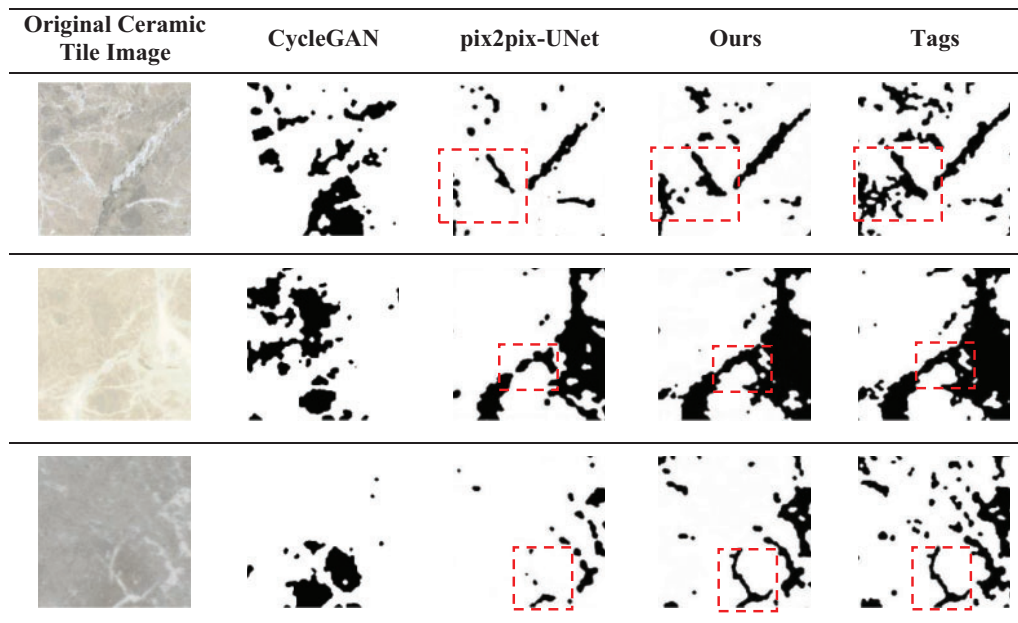**Figure 9:** Sketch results of the contour generated by different GANs



**Figure 10:** Sketch results of the binarization generated by different GANs

The experimental results of the input fine texture sketches are shown in Fig. 11. We can observe that the different styles of sketches (contour or binarized) have different guiding effects for the ceramic tile generator. Furthermore, the effect of binarization is significantly better than the contour-based sketch produced by Canny operator. Binarization-based sketch can better distinguish between major texture features and backgrounds. Moreover, the generation effect of the Canny operator sketches is

unstable. For example, the second ceramic tile image in the second row on the left, does not achieve the effect of sketch fitting. Obviously, the generator with ResNet Block leads to a slightly better effect than the generator with UNet structure. For example: comparing the first image in the third row on the left with the first image in the fourth row on the left, we can notice that the texture generated by the proposed method is closer to the input sketch texture. Additionally, the experimental results of the input coarse texture sketches are shown in Fig. 12. Obviously, the generator with ResNet Block has better detail portrayal ability compared with the original UNet structure. For example: most of the images in the fourth row are fitted to some extent with the detailed textures in the upper left and lower right corners of the sketch.
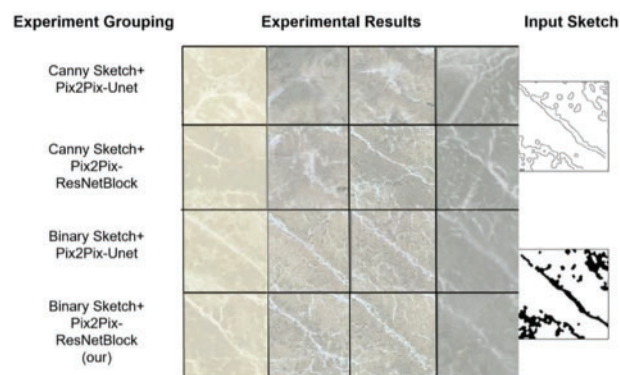


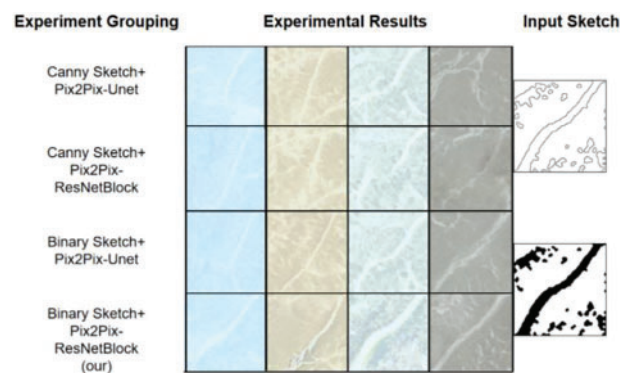**Figure 11:** Experimental results comparison for the input fine texture sketches



**Figure 12:** Experimental results comparison for the input coarse texture sketches

After an overall comparison, it is obvious that the optimized proposed method generates ceramic tile images with better quality results than those generated by the guidance before optimization. The main reason is that the ceramic tile data doesn't have corresponding semantic information and their contours may contain less information, which leak the sketch images. In order to solve these problems, this paper proposes to use a binarized ceramic tile sketch dataset, where the dark part of the sketch represents the texture and the white part represents the background. These binarized sketches can better guide the ceramic tile images to be generated in the desired direction of the designers.

To evaluate the quality and diversity of the proposed generative model, the distribution similarity of the two generated datasets of images is measured based on the Frechet Inception Distance (FID) metric between the generated images and the ceramic tile dataset. The FID is a commonly used metric

for measuring the generative diversity and quality of the GAN model. The smaller value of FID refers to the better performance. Additionally, we select Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) metric to evaluate the quality of the generated images, where smaller values of BRISQUE indicate better performances. To compare the quality of the generated models, four groups of experiments are performed as shown in Table 1. The ceramic tile image generation network generated by each group of experiments is randomly generated with 5000 images, the value of FID and BRISQUE are obtained and repeated 10 times and the best result among the 10 times is recorded as a quantitative indicator. Table 1 shows that the binarized sketch style has better guiding effects on the GAN generator compared with Lu et al.'s model [21] and other models. Overall, the binarization-based sketches are superior in terms of performance compared with the Canny sketches and Lu et al.'s model [21].

**Table 1:** Quantitative analysis results

|  | Sketch extraction network structure | FID | BRISQUE |
|---|---|---|---|
| Pre-trained models | / | 46.35 | 39.58 |
| Lu et al. [21] | StyleGAN-VGG19-UNet | 33.54 | 30.15 |
| Canny sketches | pix2pix-UNet | 47.28 | 40.41 |
|  | pix2pix-ResNetBlock | 43.52 | 38.12 |
| Binarization-based sketches (Ours) | pix2pix-UNet | 33.85 | 29.65 |
|  | pix2pix-ResNetBlock | **32.47** | **28.44** |

Furthermore, the effects of different adversarial loss functions on the generation model are presented as shown in Table 2. The quality and diversity of the generated images are reduced, if only the sketch adversarial loss function $\mathcal{L}_{\text{sketch}}$ is applied. Therefore, it can be concluded that the image adversarial loss function $\mathcal{L}_{\text{image}}$ is necessarily to be included in the network.

**Table 2:** Experimental results against loss of ablation

|  | Loss function | FID | BRISQUE |
|---|---|---|---|
| Pre-trained models | / | 46.35 | 39.58 |
| Lu et al. [21] | $\mathcal{L}_{GAN} + \mathcal{L}_{style}$ | 33.54 | 30.15 |
| Binarization-based sketches with pix2pix-UNet | $\mathcal{L}_{\text{sketch}}$ | 47.73 | 41.80 |
|  | $\mathcal{L}_{\text{sketch}} + \mathcal{L}_{\text{image}}$ | 33.85 | 29.65 |
| Binarization-based sketch with pix2pix-ResNetBlock (Ours) | $\mathcal{L}_{\text{sketch}}$ | 43.65 | 38.23 |
|  | $\mathcal{L}_{\text{sketch}} + \mathcal{L}_{\text{image}}$ | **32.47** | **28.44** |

To ensure subjectivity, a total of five sets of training results are collated into a questionnaire to evaluate the similarity of ceramic tile texture and sketch texture for the five sets of data. A total of 522 questionnaires were received. The evaluation indices are set to: very similar (5 points), similar (4 points), barely similar (3 points), not similar (2 points) and irrelevant (1 point). Finally, the average

value is obtained as a subjective score of similarity for the four comparison experiments. The results of the survey are shown in Table 3. It is observed that the ceramic tile images generated by the proposed method has the highest score compared with Lu et al. [21] and other models, which indicates that the proposed method is subjectively similar to the sketch styles in the generated ceramic tile images.

**Table 3:** Results of subjective evaluation indices for different sketch types and network structures

|  | Lu et al. [21] | Canny Sketch+ pix2pix-UNet | Canny Sketch+ pix2pix-ResNet Block | Binarization sketch+ pix2pix-Unet | Binarization sketch+ pix2pix-ResNet Block (Ours) |
|---|---|---|---|---|---|
| **Average score** | 3.60 | 1.75 | 2.87 | 3.45 | **3.97** |

### 4.3 Diversity Module Performance

In order to obtain natural ceramic tile images with similar and diverse textures, **W** hidden space perturbation is applied. Some of the experimental results are shown in Fig. 13. The first column shows the ceramic tile images generated by the random vector $z$. The columns from the second to fifth show all the results of adding a random Gaussian noise with the mean value of 0 and variance of 0.5 in the W space. We conclude that the perturbation in the W-space can moderately fine-tune the texture structure of the ceramic tiles.
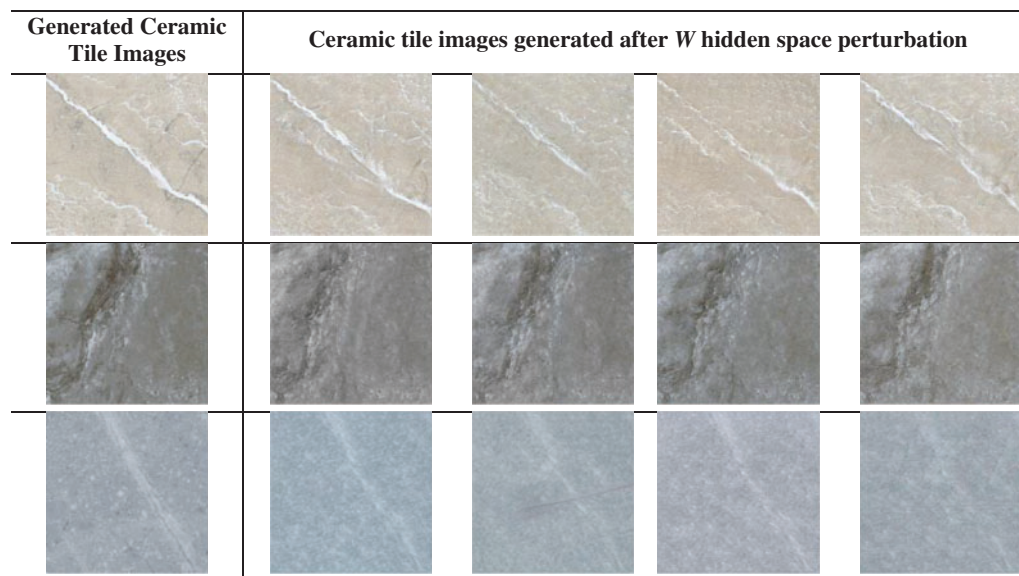


**Figure 13:** ceramic tile images generated after W hidden space perturbation

Due to the transitional nature of the linear interpolation methods for textures, we also perform model interpolation to obtain the transition images of ceramic tile textures by equating the weight parameter $\alpha$ into 8 parts and linearly interpolating Model I and Model II as shown in Fig. 14. Three sets of experiments are executed as follows: First, the transitions between sketches with similar directions but different thicknesses from top to bottom. Second, the transitions of textures with

different directions under fine textures. Finally, the transitions of textures with two different forms. Some of the experimental results are shown in Fig. 14. Each row is an input with the same noise to generate the transition effect of ceramic tiles. The first and last columns indicate the model guided by its sketch to generate the required interpolation. The second and ninth columns indicate the ceramic tile images generated by the same noise after applying model I and model II, respectively. The remaining columns in the middle indicate the transition images generated by noise $z$ under the interpolation generation model with different weighting parameters $\alpha$. We concluded that, the transition in the middle is more natural and smoother for sketch models with closer textures, and the more complex images tend to have larger transition magnitudes.
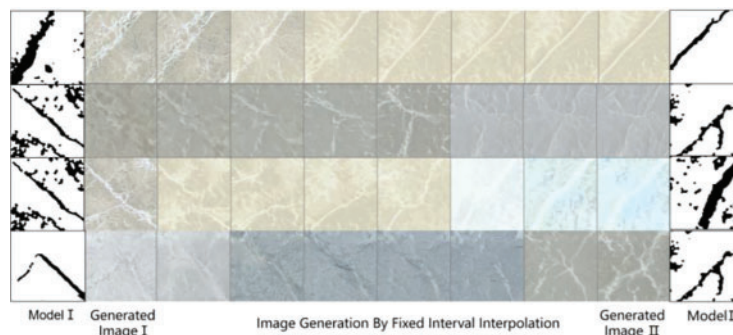


**Figure 14:** Interpolation results for the network model

## 5  Conclusion and Future Work

In this paper, we propose a sketch-based diverse generation method for ceramic tile images. The ceramic tile generation network and sketch extraction network are pre-trained first and then two adversarial loss functions have been added to adjust the weight parameters of the ceramic tile generation network. The designer obtains a series of ceramic tile design images by inserting a small number of ceramic tile design sketches into the training generative model, which often have the realistic marble pattern desired by the designer that realize the diversity of the design concepts. In the network design, the sketch extraction network is improved by introducing double cross-domain adversarial loss functions and proposing a new input sketch style based on the characteristics of the ceramic tiles. Furthermore, the proposed method can generate natural and diverse ceramic tile images by perturbing the generated latent vectors and applying linear interpolation with different weights. However, the proposed method for generating ceramic tile images still has some limitations such as the slow speed of the training network for ceramic tile texture diversity design requirements. The feature extraction network is added to improve the training network performance but, it slightly increases the complexity of the model. Therefore, the focus of our future research is on how to filter the redundant information and optimize the network structure to improve training speed of the network. Moreover, attention mechanisms [23] which are more instructive in practical applications, can be considered to improve network performance.

Security Information Application Based on Big-Data Architecture, Ministry of Public Security of Zhejiang Police College (Grant No. 2021DSJSYS002).

**Author Contributions:** study conception and design: J. Lu, X. Liu, M. Shi; data collection: X. Liu, C. Cui; analysis and interpretation of results: J. Lu, X. Liu, M. Shi, M. Emam; investigation: M. Shi, M. Emam; draft manuscript preparation: X. Liu. M. Shi, C. Cui, M. Emam. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The data were prepared and analyzed in this study and available upon a request to the corresponding author.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

[1] J. H. Chang, Z. H. Zhao, C. M. Jia, S. Q. Wang, L. B. Yang *et al.,* "Conceptual compression via deep structure and texture synthesis," *IEEE Transactions on Image Processing*, vol. 31, pp. 2809–2823, 2022.

[2] R. Hu, M. Barnard and J. Collomosse, "Gradient field descriptor for sketch based retrieval and localization," in *17th IEEE Int. Conf. on Image Processing (ICIP 2010)*, Hong Kong, China, pp. 1025–1028, 2010.

[3] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–70, 2000.

[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley *et al.,* "Generative adversarial nets," in *Advances in Neural Information Processing Systems (NIPS 2014)*, Montreal, Canada, pp. 2672–2680, 2014.

[5] W. Chen and J. Hays, "SketchyGAN: Towards diverse and realistic sketch to image synthesis," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2018)*, Salt Lake City, UT, USA, pp. 9416–9425, 2018.

[6] S. Y. Chen, W. C. Su, L. Gao, S. H. Xia and H. B. Fu, "DeepFaceDrawing: Deep generation of face images from sketches," in *ACM Transactions on Graphics*, vol. 39, no. 4, pp. 72:1–72:16, 2020.

[7] P. Xu, Z. Y. Song, Q. Y. Yin, Y. Z. Song and L. Wang, "Deep self-supervised representation learning for free-hand sketch," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, no. 4, pp. 1503–1513, 2021.

[8] I. Sobel and G. Feldman, "A $3 \times 3$ isotropic gradient operator for image processing," *A Talk at the Stanford Artificial Project*, vol. 1, pp. 271–272, 1968.

[9] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 679–698.

[10] H. Emami, M. M. Aliabadi, M. Dong and R. B. Chinnam, "SPA-GAN: Spatial attention gan for image-to-image translation," *IEEE Transactions on Multimedia*, vol. 23, pp. 391–401, 2021. https://doi.org/10.1109/TMM.2020.2975961

[11] J. Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. of the IEEE Conf. on Computer Vision (ICCV 2017)*, Venice, Italy, pp. 2242–2251, 2017.

[12] T. C. Wang, M. Y. Liu, J. Y. Zhu, A. Tao, J. Kautz *et al.,* "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2018)*, Salt Lake City, UT, USA, pp. 8798–8807, 2018.

[13] M. Li, Z. Lin, R. Mech, E. Yumer and D. Ramanan, "Photo-sketching: Inferring contour drawings from images," in *2019 IEEE Winter Conf. on Applications of Computer Vision (WACV 2019)*, Waikoloa, HL, USA, pp. 1403–1412, 2019.

[14] P. Sangkloy, J. W. Lu, C. Fang, F. Yu and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2017)*, Honolulu, HL, USA, pp. 6836–6845, 2017.

[15] J. Kim, M. Kim, H. Kang and K. Lee, "U-GAT-IT: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation," ArXiv Preprint ArXiv:1907.10830, 2019.

[16] P. Isola, J. Y. Zhu, T. Zhou and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2017)*, Honolulu, HL, USA, pp. 5967–5976, 2017.

[17] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*, vol. 9351, pp. 234–241, 2015.

[18] K. Tero, A. Timo, L. Samuli and L. Jaakko, "Progressive growing of GANs for improved quality, stability, and variation," ArXiv Preprint ArXiv: 1710.10196, 2017.

[19] K. Tero, L. Samuli and A. Timo, "A style-based generator architecture for generative adversarial networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2019)*, Long Beach, CA, USA, pp. 4396–4405, 2019.

[20] K. Tero, L. Samuli, A. Miika, H. Janne, L. Jaakko *et al.,* "Analyzing and improving the image quality of StyleGAN," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR 2020)*, Seattle, WA, USA, pp. 8107–8116, 2020.

[21] J. Lu, M. Shi, Y. Lu, C. C. Chang, L. Li *et al.,* "Multi-stage generation of tile images based on generative adversarial network," *IEEE Access*, vol. 10, pp. 127502–127513, 2022.

[22] S. Y. Wang, D. Bau and J. Y. Zhu, "Sketch your own GAN," in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision (ICCV 2021)*, Montreal, QC, Canada, pp. 14030–14040, 2021.

[23] J. Lu, H. Ren, M. Shi, C. Cui, S. Zhang *et al.,* "A novel hybridoma cell segmentation method based on multi-scale feature fusion and dual attention network," *Electronics*, vol. 12, no. 4, pp. 979, 2023.