Naturality: A Natural Reflection of Chinese Calligraphy

Bingyuan Wang
Computational Media and Arts, Hong Kong University of Science and Technology (Guangzhou)
bwang667@connect.hkust-gz.edu.cn

Kang Zhang
Computational Media and Arts, Hong Kong University of Science and Technology (Guangzhou)
kzhangcma@ust.hk

Zeyu Wang*
Computational Media and Arts, Hong Kong University of Science and Technology (Guangzhou)
zeyuwang@ust.hk

Figure 1: An example of our artwork. Specifically, we input only one natural image (top left) and utilize the higher knowledge of CLIP model and generative process of diffusion model, to derive a set of images with natural transition. The final result is an AI artwork (bottom right) that is associated from the original scenery, which mimics the traditional creative process of Chinese calligraphy.

ABSTRACT

We present a machine learning-based interactive video installation powered by CLIP and diffusion models and inspired by the concept of naturality in traditional Chinese calligraphy. The artwork explores contemporary interpretations of this traditional concept through practical methods in Artificial Intelligence Generated Content (AIGC). Technically, the algorithms are based on state-of-the-art perceptual and generative models, incorporating multi-dimensional controls over text-to-image and image-to-image translation; conceptually, this real-time art installation extends the discussion brought by Xu Bing's pieces *Book from the Sky* and *Square Word Calligraphy*. The project explores the possibility of AIGC in bridging human creativity and natural randomness, as well as a shifting creative paradigm enhanced by AI knowledge, perception, and association.

*Corresponding author

VINCI 2023, September 22–24, 2023, Guangzhou, China

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

CCS CONCEPTS

• **Applied computing**; • **Arts and humanities**; • **Fine arts**;

KEYWORDS

Chinese calligraphy, naturality, AI art.

ACM Reference Format:
Bingyuan Wang, Kang Zhang, and Zeyu Wang. 2023. Naturality: A Natural Reflection of Chinese Calligraphy. In *Proceedings of The 16th International Symposium on Visual Information Communication and Interaction (VINCI 2023).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

This experimental art piece includes three sub-pieces with different but echoing settings. By manipulating a diffusion model and converting between different modalities, AI calligraphies are created from human pieces or natural sceneries, and AI paintings are generated from human calligraphy. The creative process is animated through frame interpolation and added with AI-generated audio content.

In addition to exploring the latent interpolation between AI calligraphy, human pieces, and natural sceneries, we suggest that this work is also connected with the creative and conceptual characteristics of calligraphy in traditional Chinese contexts. Specifically, the emphasis on naturality lay in the creative mind state, artistic presentation, and spiritual pursuit of traditional Chinese calligraphers and critics.

The setup of the installation and form of the results is a homage to Xu Bing's series of pieces, including *Book from the Sky* [1], *Square Word Calligraphy* [2], and the recent *Living Word* [3]. Xu Bing infuses his work with insights by differentiating between the form (features of shape and basic units) and the meaning of Chinese characters. We find that perspective still compelling in 2023, yet involved with the impact from digital tools.

By leveraging AI art and AI originality in the creative process, we want to manifest the concept of naturality in traditional Chinese calligraphy, visualize its involvement in the generation pipeline, and explore the possibility of AI in bridging natural imagery and human artwork. From our perspective, such artistic implementation not only furthers the exploration of human-AI collaboration in art, but also provides an approach to reflect on Chinese calligraphy in the digital age.

2 BACKGROUND AND MOTIVATION

2.1 Concept of Naturality in Traditional Chinese Culture

Naturality is a metaphysical concept deeply embedded in traditional Chinese culture and widely applied in different fields of art. From as early as Zhou Dynasty (1046-256 B.C.), people began to evaluate artistic pieces with nature-concerned descriptions as one of the highest standards. For example, the most melodious and emotional music is described as high mountains and flowing water (高山流水) or the sound of nature (天籁之音); the great buildings and gardens are embellished as walking firebirds and flying phoenixes (走鸾飞凤) or ghost's axe and a god's workmanship (鬼斧神工) for their ingenuity and spirituality; the body shape of renowned beauties that are full of expressiveness and tension are complemented as elegant as a startled swan (翩若惊鸿) and graceful as a wandering dragon (婉若游龙); even the delicious food is referred to as delicacies of the sea and the mountains (山珍海味).

Similarly, Chinese calligraphy has been embedded in such cultural environments since an early stage. As the ancients gradually became self-conscious about artistic aesthetics from around the Wei dynasty (A.D. 220), concepts related to naturality became even more popular for artists to pursue. On the other hand, the esteem for naturality also constitutes part of a wholistic value system typical in ancient China, where the Harmony between Man and Nature is a vital part. This explains the tendency for different art forms to be closely linked and interpret each other, while nature is often the essence, philosophical basis, and important reference behind them.

2.2 Concept of Naturality in Chinese Calligraphy

In traditional Chinese calligraphy, the concept of naturality is also deeply embedded in almost all the creation and evaluation processes. From historical recordings by distinguished calligraphers and influential critics, naturality is mainly manifested in the creative mind state of the artist, the target content and object of the creation; as well as the artistic presentation of final results (e.g., stroke, shape, structure, lines). The first part is mainly manifested in that the artist does not deliberately pursue elaborateness (期于如此而能如此者，工也。不期如此而能如此者，天也。——傅山《字训》) or peculiarity (诗不求工字不奇，天真烂漫是吾师。——苏轼《黄州寒食诗帖》) because they view randomness and unpredictability as a vital part of calligraphy aesthetics (人人细问此中妙，怀素自言初不知。——戴叔伦《怀素上人草书歌》). Some calligraphers are known for drinking before creation (张旭三杯草圣传,脱帽露顶王公前,挥毫落纸如云烟。——杜甫《饮中八仙歌》), others prefer to get immersed in an (imagined) natural environment (每静室僧趺，忘怀万虑，与碧虚寥廓同其流。——米友仁《米元晖画题》). In either way, it helps the creative subject forget about conventions (欲书先散怀抱，任情恣性，然后书之。——蔡邕《笔论》), get rid of apprehension(以其意不在书，天机自动，往往多入神解。——王澍《论书剩语》), and seek inspiration from the original state (必使心忘于笔，手忘于书，心手达情，书不忘想，是谓求之不得，考之即彰。——王僧虔《笔意赞》).

On the other hand, the content and shape of calligraphy works are expected to be a mimicry of nature. For example, in traditional specifications and theoretical tutorials, single strokes were often compared to clouds, vines, or thunders ("横"如千里阵云，"竖"如万岁枯藤，"捺"如崩浪雷奔。——卫夫人《笔阵图》); some calligraphers described the running script as crawling beasts, fighting birds, and cunning rabbits (兽鸟，志在飞移；狡兔暴骇，将奔未驰。——崔瑗《草书势》); others commented on outstanding historical works with dew, stone, snake, and other natural imageries (观夫悬针垂露之异，奔雷坠石之奇，鸿飞兽骇之资，鸾舞蛇惊之态，绝岸颓峰之势，临危据槁之形。——孙过庭《书谱》). Meanwhile, when facing outstanding calligraphy works with unwavering reputations, the commenters typically not only evaluate from an ontological perspective but also focus on manifesting the artist's state of creation and the spiritual pursuit (粉壁长廊数十间，兴来小豁胸中气。忽然绝叫三五声，满壁纵横千万字。——窦冀《咏怀素草书》), which converges to the consistency between calligraphy, human, and nature. In this respect, there is a gradual shift in perception from human calligraphy originated from and guided by nature (夫书肇于自然。——蔡邕《九势》) to the re-creation of nature by human practice (蔡中郎但谓书肇于自然，此立天定人，尚未及乎由人复天也。——刘熙载《书概》), which also reflects the change of artistic concept due to the development of socio-economic productivity.

2.3 Motivation

Compared to other artistic forms, calligraphy is purer in its content and less affected by pioneering concepts or artistic movements. Calligraphers are thus comparatively conventional, with less influence from contemporary Western aesthetics or experimental forms. However, according to the discussion above, a large proportion of distinguished historical or contemporary artists are willing to push forward the field by introducing variations and breaking away from rigid stereotypes. On the other hand, the primary goal of computer engineers is to fit AI with domain knowledge to better perform traditional tasks. However, such a process is driven by futuristic visions and cutting-edge tools, and the creative power is both learned from human pieces and generated with random inputs. Such correspondence creates a conversational relationship and presents a gap that may be filled with state-of-the-art AI technologies. Meanwhile, nature is the relentless inspiration for naturality and randomness, and can possibly bridge these two concepts.

Moreover, naturality also indicates a procedure that fuels inspiration by introducing variations to the creative process. By referring to existing artistic practices such as fractal art, we may view algorithmic randomness and subjective unpredictability as respective conceptual counterparts in AI and art. From our perspective, natural imageries and sceneries can also serve as a bridge between the artistic perception of naturality and the computational concept of randomness.

Furthermore, as audiences and researchers, we are constantly eager to see from the eyes of artists. Also, based on the value system and evaluation criteria in historical times, we wonder if we can apply current AI technologies to modify ordinary works and facilitate approaching such aesthetic realm. Finally, we want to mimic the artist's understanding and interpretation of natural sceneries, to derive a calligraphy work from an input image.

3 RELATED WORK

In this part, we discuss current AIGC technologies and their performance in calligraphy-related tasks (e.g., calligraphy generation and recognition). By evaluating existing methods and giving a clear definition of our target tasks, we filter out the state-of-the-art models most relevant to our goal for further experiments and reference.

3.1 Generation Task

In this project, we want to generate images from both historical pieces and calligraphy-related text descriptions. Such procedures mainly concern two basic tasks in AIGC: text-to-image conversion and image-to-image translation. In computer science, the former indicates generating corresponding images with text as the only or basic input; and the latter is a term proposed by Isola et al. [20], describing the process of translating an image into different representations (such as depth, façade, and edge map). These two approaches also represent two primary types of AIGC: unimodal and multimodal (cross-modal) generation [11]. As described in [8], methods in this field commonly follow a unified paradigm that incorporates representation learning and regeneration learning [39], with encoder-tokens-decoder as the basic framework. In 2021, Radford et al. proposed Contrastive Language Image Pretraining (CLIP) [30], which jointly trains an image encoder and a text encoder to predict the correct pairings of training examples. Such approach successfully constructs an aligned latent space and greatly facilitates content generation across different modalities. Combined with other advanced tools (e.g., Diffusion models), users are able to achieve impressive generative results.

3.2 Methods and Tools

The methods and tools of AI-Generated Content (AIGC) in the field of artificial intelligence have seen significant advancements in recent years [11]. Based on previous task analysis, we summarize those in the field of image generation and categorize them by different technical approaches.

Generative Adversarial Networks (GANs) [16] have been widely used for image generation. A GAN consists of a generator and a discriminator. The generator transforms random noise into new samples, while the discriminator distinguishes between real and generated images. GANs have been applied in various domains, including image restoration, style transfer, super-resolution, and image synthesis [14]. However, early GAN models had limitations in user control and image quality. To address these issues, Conditional GANs (CGAN) [25] were introduced to incorporate attribute information into the generator and discriminator. Furthermore, improvements such as applying convolutional neural networks (CNN) to GANs [31] have enhanced the stability and quality of generated images.

Text-to-image conversion using GANs was first proposed in 2016 [35]. Subsequent innovations in this field focused on network structure, text encoding, and datasets. StackGAN [46] introduced a two-level GAN with attention mechanism for generating high-resolution images. AttnGAN [44] utilized attention modules to focus on relevant words in natural language descriptions. Other models borrowed ideas from network architectures such as cyclic architectures [29] and style-driven approaches [21].

Auto-encoder-based methods, such as Variational Autoencoders (VAEs) [22] and VQ-VAE [40], have also been employed for image generation. Auto-encoders encode input data into a lower-dimensional representation and decode it to generate similar data. VAEs consider mean and standard deviation, while VQ-VAE uses discrete hidden variables. These methods have shown progress in image generation, sketch generation, and 3D point cloud reconstruction. Other methods like Neural Representation [17] and Reinforcement Learning [15] have demonstrated the ability to generate sketches and collaborate with computer drawing programs. Additionally, the use of large-scale pre-trained models like Google's Vision Transformer (ViT) series [12] has been explored for image generation.

In recent years, diffusion model [18] based on Langevin dynamics has gained popularity in the field of image generation. This model has shown significant improvements in generating high-quality images, especially after incorporating conditions and other enhancements. Various diffusion model-based approaches have emerged, focusing on semantic segmentation [9], point cloud processing [50], and video generation [19], etc. The diffusion model has also been integrated with models like VAE, GANs, and CLIP to achieve better results in AI painting.

Beyond 2D content generations, AIGC has expanded to include 3D and motion generation as well as rendering and camera path planning [41]. Text-to-3D generation [28] and 2D-to-3D conversion [33] have been explored, leveraging larger models and computational power. Motion and animation generation methods include text-based [38] and single-image-based [42] approaches. Rendering tools are utilized for different applications [24], lighting conditions [23], and artistic styles [48]. These advancements have also enabled automatic correction of photographs, high-resolution image conversion, stylization, video animation, etc.

Overall, AIGC in art has benefited from GAN-based methods, auto-encoder-based methods, and the diffusion model. Nowadays, the field is matching towards Large Language Models (LLM) such as ChatGPT [10] and AutoGPT [5]. These approaches have contributed to the generation of high-quality images, sketches, 3D models, and animations, opening up new possibilities for artistic expression and creativity.

4 METHODOLOGY

4.1 Design Choices and Decision

Based on the previous analysis, we view CLIP and diffusion model as the most appropriate method for image generation. For one thing, the CLIP model is pre-trained on 400 million (image, text) pairs including calligraphy as one of the queries [30] and can be fine-tuned to domain tasks. For another, within the generation procedure, output randomness can be achieved by controlling the random seed and noising-denoising steps, which caters to our expectations for AI creativity and originality.

We further investigate current diffusion models with state-of-the-art performance in different contexts. The first tier of diffusion-based commercial products includes DALL-E-2 [32] by OpenAI, Dreamstudio [36] by Stability.ai, and Imagen [37] by Google. In the field of art, Midjourney [4] achieved impressive results and significantly influenced the art field. There are also some models specifically adapted for the Chinese community, such as ERNIE-ViLG [13] and Taiyi-stable-diffusion [47]. However, our experiments show that they performed comparably to stable diffusion (v1.4) for Chinese calligraphy generation tasks.

On the other hand, as the noising and denoising results by step are still overly discrete, we want to add transitions and animations to better demonstrate AI perception. Besides this, we also want to combine background music to create a more immersive experience. Thus, we add a frame interpolation module and an image-to-music module to our final pipeline. To the authors' knowledge, there are many notable works in Frame Interpolation including SoftSplat [26], ABME [27], and FILM [7], and the latter achieves the most impressive results. Also, image-to-music is currently based on and related to image-to-text, text-to-audio, and other generative tasks. To combine the advantages of different models, we break it down into a two-step approach. We first conduct image-to-text conversion to leverage the powerful knowledge of CLIP. The text embeddings are then fed into a diffusion model for text-to-audio generation. In this way, we derive audio scripts from an input image with controllable length and style.

4.2 Approach Outline

Figure 2 illustrates the principle of diffusion model. By initializing from a random seed, the system go through an iterative noising and denoising process to gradually approach the generation target. As the input image and text prompt are fixed, the output is mainly influenced by the choice of different random seeds and the number of iteration steps. Accordingly, we leverage *random seed* as the source of randomness and richness of results, and *iteration step* to unveil the process of AI perception.

On the other hand, we also tested and compared the performances of different diffusion models mentioned in subsection 4.1 (Figure 3). Since the results did not differ significantly, we finally chose the most widely used stable diffusion (v1.4) for further experiments. Finally, we supplement the system with other AIGC networks, including frame interpolation and image-to-audio conversion. The aggregated pipeline is displayed in Figure 4.

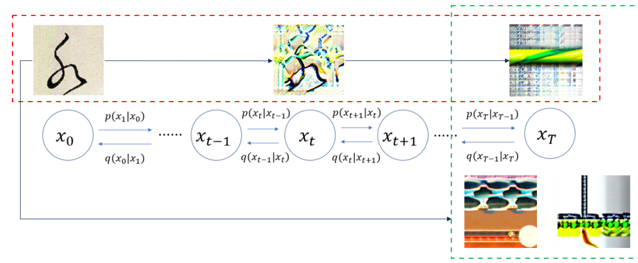


Figure 2: The principle of diffusion model and an example. We leverage the two parameters: *random seed* (green box) and *iteration step* (red box) to achieve different results.

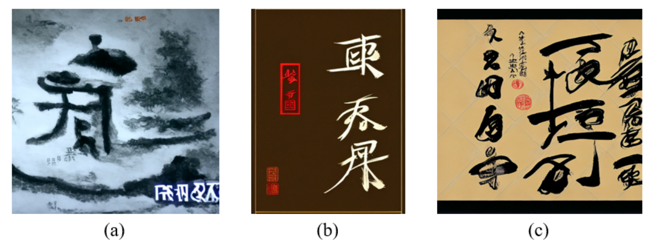


Figure 3: Comparison of Chinese calligraphy generated by different diffusion models: (a) Disco Diffusion; (b) Stable Diffusion; (c) Taiyi-stable-diffusion. We find the latter two have comparative performance.

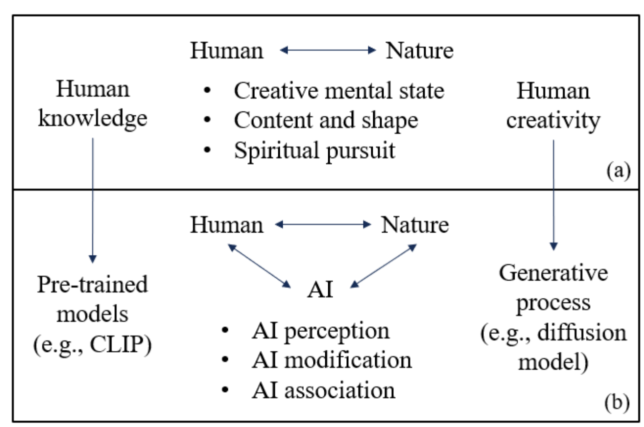
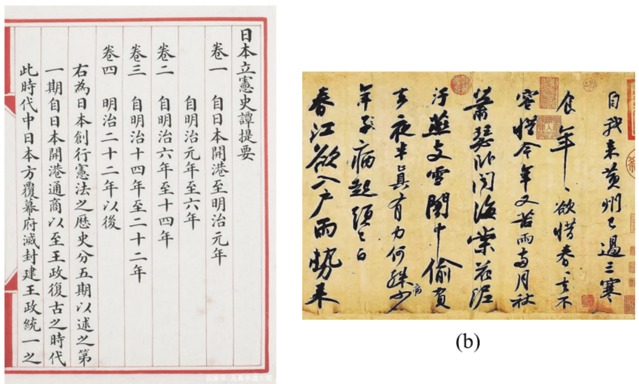


Figure 4: Traditional and contemporary conceptual paradigms of Chinese calligraphy creation. (a) Traditionally, the two major subjects are human and nature; (b) a possible approach with AI as a third component in the creative process. One-way arrows represent evolution and two-way arrows represent collaboration.



(a) (b)

Figure 5: Calligraphy as daily writing technique and an art form. (a) *Summary of Political Series*, an ordinary copybook by literati; (b) *Cold Food Observance* by Su Shi, known as the Third Greatest Lines of Writing in the World.

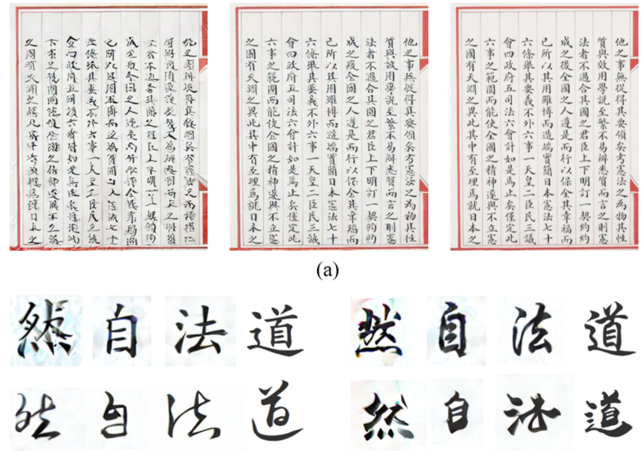
5 EXPERIMENTS AND RESULTS

Our experiments are majorly comprised of three different settings. In the first part, we use AI models to introduce deformation into human calligraphy pieces, to enhance variations within fixed shapes and rigid artworks. In the second part, we apply different levels of modification to single characters, to mimic natural concepts either in features or forms. Finally, we intend to realize the conversion of artistic expression from a natural image to calligraphy work, to recreate a creative process. In each experiment, the text prompt guiding the diffusion processes is "Chinese calligraphy." We also attempt to gradually alter the content and interpolate between steps. From our perspective, it is a way of showing expression consistency and utilizing different levels of AI knowledge.

5.1 Human and AI: Variation from Pseudo Randomness

The first set of experiments is inspired by the evaluation standard of traditional calligraphy, in the conversational context between ordinary copybooks and masterpieces. Before the invention of movable type, the main way to disseminate textbooks was by handwriting. As a result, there were many positions responsible for copying books at the court and in society before the Song Dynasty (around A.D. 1050). In consideration of speed and readability, they usually used lowercase letters, with little variations and innovations in the process. Despite this, calligraphy also existed as an art form, and the works of famous calligraphers were widely spread, studied, and sought after (Figure 5).

It is interesting to consider the factors that distinguish these two writing styles, which also concern the essence of artistic creations. Based on popular discourses of that time (subsection 2.2), we view that the manifestation of naturality has a surpassing contribution to the achievement of calligraphy works, beyond basic technical proficiencies. It thus becomes our goal to introduce random variations into ordinary human pieces.



(a)

(b)

Figure 6: Experiment 1. (a) By controlling the iteration step (from right to left: 0, 5, 10), we introduce variations to rigid copybooks; (b) we select some fixed fonts and use similar methods to modify them (each series from right to left: 0, 4, 7, 10). It is a process of interpolating between human calligraphy and AI calligraphy.

To achieve this, we use the original image of calligraphy work and calligraphy-related prompts as inputs, control the number of noising-denoising steps, and record the outputs from the diffusion model. An example result is shown in Figure 6-(a). Beyond historical pieces, we also want to apply similar methods to introduce variations to standard contemporary fonts (Figure 6-(b)). We view that the sequence from right to left not only represents the writing order of Chinese calligraphy but also symbolizes the evolution of human civilization and the gradual replacement of humans by AI. On the other hand, it indicates that human and AI are not separate, they may understand and complement each other. Also, there may be ways to combine human conventions and AI innovation, to help contemporary calligraphers push through the border.

5.2 Human and Nature: Figuration through Natural Concepts

The second set of experiments focuses on connecting the two subjects of human and nature, with the assistance of AI. Our primary goal is to utilize the higher knowledge of CLIP models, to generate images with consistent meaning from specific characters. However, our experiments show that such method can also be applied to modify the shape features of words when implemented with fewer steps. Thus, we conclude the two functions of association and modification into figuration, a process of deriving figurative images from abstract characters.

The experimental setting is similar to subsection 5.1, despite the text prompt is related to the actual meaning of input characters. Meanwhile, we majorly control the iteration steps in low-level modification and random seed in high-level association, to guarantee coherent transitions and diverse outcomes. Figure Figure 7

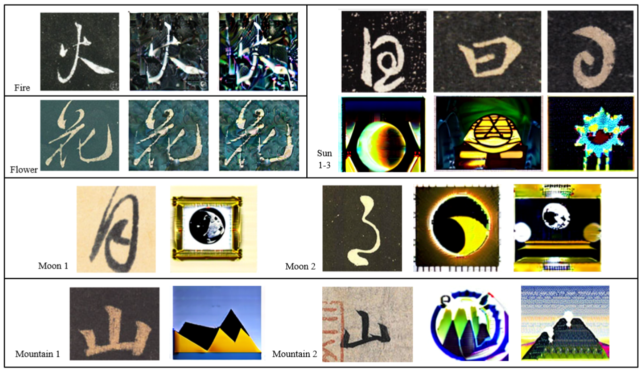


Figure 7: Experiment 2. From single characters, we conduct low-level modification by controlling iteration steps (top left, from left to right: 0, 5, 10) and high-level association by setting different random seeds (others). In each series, the word on the left indicates the character's meaning and is used as a text prompt. Different numbers indicate different writing styles of the same character.

displays the results of respective experiments, which grants different weights in interpolating between the shape and meaning of characters. Such attempts treat AI as an imaginative tool and source of inspiration. With its help, we may be able to recreate certain imageries or wholistic ideascapes from calligraphy to see from the eyes of calligraphers.

5.3 Nature and AI: Originality without Human Masterpiece

Based on the previous experiments, our final goal is to generate a calligraphy artwork from an image, to mimic the creative process of humans. In doing this, we select images of natural sceneries and feed them directly into the diffusion model. We then guide the CLIP model by entering calligraphy-related prompts. Finally, we set different iteration steps and record the process of generation.

Figure 8 demonstrates both a successful and a failed result. The major difference lies in the style and genre of the input image. We conclude that the CLIP model follows a similar perceptual pattern to humans. The latent space of the model acquires the semantic correlations between calligraphy and ideas such as idyllic landscapes and traditional Eastern paintings, which demonstrate our successful collaboration with AI models.

Finally, we would like to incorporate animation and interaction into the model. The former refers to interpolating between different steps with AI model, and the latter indicates creating an immersive experience for human savoring. Specifically, we want the scene to change with human involvement, to convey more philosophical meaning. By implementing an object detection module, the scenery will turn into calligraphy when someone walks past. Figure 1 displays an example of the transition process and final outcome. Artistically speaking, the sequence from left to right aligns with both the order in which the work is generated, the order in which people walk by, and the order in which the scenery is gradually transformed into calligraphy.

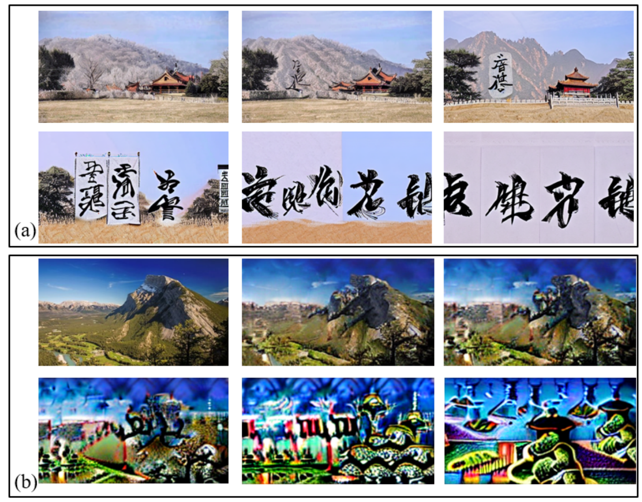


Figure 8: Experiment 3. A successful (a) and a failed (b) example of generating Chinese calligraphy. Although we input the same text prompt, corresponding outcome can only be achieved under certain images. This indicates possible alignment of human understanding and AI perception.

To summarize, we leverage the three components of human, nature, and AI in Figure 4. Each experiment focuses on bridging two of these three subjects, with the intervention of the third one. By doing this, we want to demonstrate the indivisibility of these realms, both technically and conceptually.

6 INSTALLATION

Based on the above image-based, text-guided generation results, we first conduct frame interpolation to derive a set of videos with smooth transitions. We follow the setting of FILM which processes each adjacent pair of images separately and connects them into a coherent video clip. For the interactive installation, we researched relevant human detection methods including face or motion features, body appearance, and deep learning-based modules [6], and decided on using You Only Look Once (YOLOv8) [34], a lightweight, fast model for target recognition. Specifically, we incorporate the module into all three settings in section 5. The image on the interactive interface stays in its *nature* state by default, and once people approach it transforms into the corresponding *human* state, the procedure of which is generated and interpolated by *AI*. In this way, we manage to place these three concepts and artistic realms into a conversational environment, to bridge either two of them with inspiration from the third one. Based on audience feedback, we selected some impressive visual effects and display them in Figure 1, Figure 6, Figure 7, and Figure 8.

7 DISCUSSION

7.1 Contributions

In this artwork, we attempt to approach spiritual naturality with natural inspirations, AI understandings, and computational randomness, in the specific field of Chinese calligraphy. To realize this, we connect state-of-the-art technologies with traditional understandings and design three interrelated experiments to externalize the process. Our results show that the knowledge of the CLIP model is to a large extent consistent with that of human, and diffusion model is able to introduce variation, figuration, and artistic creation into human and natural pieces. The generative process also indicates potential revolutions in the calligraphy creation paradigm and brings about new possibilities to use AI in learning and mimicking the naturality that calligraphers pursue.

On the other hand, unlike some previous works that introduce only a single dimension (such as emotion [45]) or several dimensions (such as style and structure [43]) into the calligraphy generation control, we are able to think about naturality in a more ontological way, to perceive variation as an *object*. Although the diffusion results lack clear guidance or precise control, they open up more space in directly processing the shape of characters, the features on paper, and the form of expression. By interacting with the installation, audiences are also involved in the discussion of content creation being revolutionized by advanced technologies and engaged in the reflection of calligraphy as a creative art form in the digital age.

7.2 Implication

Beyond the collaboration of human, nature, and AI, our installation and results have led to more discussions in the field of aesthetic appreciation and philosophical principles.

Just as the motivation proposed in subsection 5.1, naturality concerns the standards for beauty, creativity, and artistic value. In the contemporary context, these concepts are always in conversation with philosophical realms such as truth and goodness. From a deconstructionist perspective, the mainstream professional standards are also constantly challenged in their inconsistency with public intuition.

We conducted a set of comparative studies with our installation by controlling the variable of professional training. Specifically, we recorded different people's backgrounds in art, especially calligraphy, and asked whether they prefer the ordinary copybooks or masterpieces shown in Figure 5. The results show that people not professionally trained in calligraphy tend to prefer the copybook which they describe as "neat" and "graceful". However, those familiar with Chinese calligraphy almost unanimously choose the latter: *the Third Greatest Lines of Writing in the World*. Their comments vary from "spirituality" and "unrepeatability", to "untouchedness" and "true self".

We believe that the results of this experiment reflect the aesthetic enhancement that artistic training brings, which makes people more inclined to appreciate a piece of artwork from the spiritual perspective and more willing to empathize with the creator. On the other hand, just as one participant mentioned, naturality itself somewhat constitutes a paradox, "in the pursuit of a realm that does not deliberately pursue anything." Perhaps this is also what renders the concept with prolonged and relentless fascination.

7.3 Future Work

As we mentioned in the previous sections, the installation opens up some possibilities and spaces for further discussion. In this part, we summarize some potential approaches for technical improvements and ideas for artistic extensions.

From a technical standpoint, there are some advanced theories and tools that may be applied to our idea. For example, the original work of image-to-image translation [20] inspires us to incorporate different representations (e.g., black and white, contours, or skeletons) to the generation process; the recently proposed ControlNet [49] enables us to experiment on an all-in-one platform; we may also leverage relevant sub-tasks of image processing, such as inpainting and outpainting, to derive more diverse outcomes. Despite this, there are still limitations in existing tools to generate high-quality, meaningful Chinese calligraphy compared with human works.

Also, as our system incorporates generating images from written characters, we may consider different forms of input. Chinese calligraphy has undergone a development process in which the degree of hieroglyphics gradually diminished. Some earlier scripts such as Oracle and bell-cauldron inscription retain more practical features of the referred object. By converting them into corresponding images, we may retrieve the intended shape and forms, which is a process of "reversed" textual evolution. (Similar to the *Living Word* by Xu Bing [3]) It may also benefit some domain tasks such as inferring the meaning of ancient scripts.

On the other hand, our generative results include both individual imageries and holistic sceneries, with a lack of focus on the compositional process. It is worth thinking about how the concatenation and repetition of different characters form an integrated calligraphy work, and recreate the process artistically. We may borrow the concepts of imagery and ideascape from traditional Chinese paintings to facilitate such exploration.

From an artistic perspective, the conversation between effective meaning and aesthetic naturality is part of the dialectic between conformity and variation in art. Just as we mentioned in subsection 2.3, Chinese calligraphy is a comparatively conservative art field, where people typically place heritage and legacy over innovation. Accordingly, is there a uniform paradigm that may be calculated or visualized, where we can base the targeted randomness and unconformity? Can we solve the problem holistically or start from the basic parts such as single strokes or certain characters?

Finally, since Chinese calligraphy highlights the balance of inheritance and innovation, smoothness and change, it's important that we think about such thresholds and boundaries in the field. Specifically, we may consider the standards between artistic innovations and meaningless trials, which leads to another set of questions: How to judge rationale and constructive innovation? What is our standard? How to choose an appropriate standard? It's worth mentioning that such standards might also be discussed in different contexts, such as public or professional standards, content or aesthetic standards, human or AI standards, etc. From our perspective, it might be meaningful to compare and combine computational metrics and results from user studies, to facilitate the effective synergy between human, nature, and AI.

8 CONCLUSION

The aim of *Naturality* is to bridge human and nature, the two major components in the calligraphy creation process, with the help of state-of-the-art AI technologies. In traditional Chinese culture, naturality is an important artistic concept and spiritual realm. In calligraphy, it also symbolizes the distinction between daily writing techniques and a form of art. Through three sets of interrelated experiments and an interactive installation, we introduce AI as a third subject in contemporary calligraphy creation, demonstrate its capability in human-like perception, modification, and association, and discuss its possibility to facilitate achieving the pursuits of ancient calligraphers.

As a traditional and conservative art form, Chinese calligraphy is also facing the influence of perceptual and generative AI models in the digital era. Some contemporary artworks have explored the separation between form and meaning or dissolved the boundary between character and image, which manifests an ongoing paradigm shift. By experimentally engaging AI models in the whole creative process, we look forward to introducing more possibilities and discussions to AI-generated Chinese calligraphy, to push forward the aesthetic boundaries and edges of this field.

REFERENCES

[1] 1988. Book from the Sky. https://www.xubing.com/en/work/details/206?type=project#206

[2] 1994. Square Word Calligraphy. https://www.xubing.com/en/work/details/206?type=project#206

[3] 2022. Living Word. https://www.xubing.com/en/work/details/695?year=2022&type=year#695

[4] 2022. Midjourney. https://www.midjourney.com/

[5] 2023. Auto-GPT. https://github.com/Significant-Gravitas/Auto-GPT

[6] Mohd A. Ansari and Dushyant K. Singh. 2021. Human detection techniques for real time surveillance: a comprehensive survey. *Multimedia tools and applications* 80, 6 (2021), 8759–8808.

[7] Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni M. Farinella, and Tal Hassner. 2022. *FILM: Frame Interpolation for Large Motion*. Computer Vision - ECCV 2022, Vol. 13667. Springer, Switzerland, 250–266.

[8] Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni M. Farinella, and Tal Hassner. 2022. *MUGEN: A Playground for Video-Audio-Text Multimodal Understanding and GENeration*. Computer Vision - ECCV 2022, Vol. 13668. Springer, Switzerland, 431–449.

[9] Dmitry Baranchuk, Ivan Rubachev, Andrey Voynov, Valentin Khrulkov, and Artem Babenko. 2022. Label-Efficient Semantic Segmentation with Diffusion Models. arXiv:2112.03126 [cs.CV]

[10] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[11] Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun. 2023. A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. (2023).

[12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929 [cs.CV]

[13] Zhida Feng, Zhenyu Zhang, Xintong Yu, Yewei Fang, Lanxin Li, Xuyi Chen, Yuxiang Lu, Jiaxiang Liu, Weichong Yin, Shikun Feng, Yu Sun, Li Chen, Hao Tian, Hua Wu, and Haifeng Wang. 2023. ERNIE-ViLG 2.0: Improving Text-to-Image Diffusion Model with Knowledge-Enhanced Mixture-of-Denoising-Experts. arXiv:2210.15257 [cs.CV]

[14] Stanislav Frolov, Tobias Hinz, Federico Raue, Jörn Hees, and Andreas Dengel. 2021. Adversarial text-to-image synthesis: A review. *Neural Networks* 144 (2021), 187–209. https://doi.org/10.1016/j.neunet.2021.07.019

[15] Yaroslav Ganin, Tejas Kulkarni, Igor Babuschkin, S. M. Ali Eslami, and Oriol Vinyals. 2018. Synthesizing Programs for Images using Reinforced Adversarial Learning. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, 1666–1675. https://proceedings.mlr.press/v80/ganin18a.html

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative Adversarial Networks. *Commun. ACM* 63, 11 (oct 2020), 139–144. https://doi.org/10.1145/3422622

[17] David Ha and Douglas Eck. 2017. A Neural Representation of Sketch Drawings. arXiv:1704.03477 [cs.NE]

[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 6840–6851. https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf

[19] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J. Fleet. 2022. Video Diffusion Models. arXiv:2204.03458 [cs.CV]

[20] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. (2016).

[21] Tero Karras, Samuli Laine, and Timo Aila. 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).*

[22] Diederik P Kingma and Max Welling. 2022. Auto-Encoding Variational Bayes. arXiv:1312.6114 [stat.ML]

[23] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P. Srinivasan, and Jonathan T. Barron. 2022. NeRF in the Dark: High Dynamic Range View Synthesis From Noisy Raw Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 16190–16199.

[24] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2021. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *Commun. ACM* 65, 1 (dec 2021), 99–106. https://doi.org/10.1145/3503250

[25] Mehdi Mirza and Simon Osindero. 2014. Conditional Generative Adversarial Nets. arXiv:1411.1784 [cs.LG]

[26] Simon Niklaus and Feng Liu. 2020. Softmax Splatting for Video Frame Interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[27] Junheum Park, Chul Lee, and Chang-Su Kim. 2021. Asymmetric Bilateral Motion Estimation for Video Frame Interpolation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 14539–14548.

[28] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. 2022. DreamFusion: Text-to-3D using 2D Diffusion. arXiv:2209.14988 [cs.CV]

[29] Tingting Qiao, Jing Zhang, Duanqing Xu, and Dacheng Tao. 2019. MirrorGAN: Learning Text-To-Image Generation by Redescription. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).*

[30] Alec Radford, Jong W. Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. (2021).

[31] Alec Radford, Luke Metz, and Soumith Chintala. 2016. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv:1511.06434 [cs.LG]

[32] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv:2204.06125 [cs.CV]

[33] Daniel Rebain, Mark Matthews, Kwang Moo Yi, Dmitry Lagun, and Andrea Tagliasacchi. 2022. LOLNerf: Learn From One Look. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1558–1567.

[34] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

[35] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. 2016. Generative Adversarial Text to Image Synthesis. In *Proceedings of The 33rd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 48)*, Maria Florina Balcan and Kilian Q. Weinberger (Eds.). PMLR, New York, New York, USA, 1060–1069. https://proceedings.mlr.press/v48/reed16.html

[36] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695.

[37] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates,

Inc., 36479–36494. https://proceedings.neurips.cc/paper_files/paper/2022/file/ec795aeadae0b7d230fa35cbaf04c041-Paper-Conference.pdf

[38] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, Devi Parikh, Sonal Gupta, and Yaniv Taigman. 2022. Make-A-Video: Text-to-Video Generation without Text-Video Data. arXiv:2209.14792 [cs.CV]

[39] Xu Tan, Tao Qin, Jiang Bian, Tie-Yan Liu, and Yoshua Bengio. 2023. Regeneration Learning: A Learning Paradigm for Data Generation. (2023).

[40] Aaron van den Oord, Oriol Vinyals, and koray kavukcuoglu. 2017. Neural Discrete Representation Learning. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/7a98af17e63a0ac09ce2e96d03992fbc-Paper.pdf

[41] Zeyu Wang, Weiqi Shi, Kiraz Akoglu, Eleni Kotoula, Ying Yang, and Holly Rushmeier. 2018. CHER-Ob: A Tool for Shared Analysis and Video Dissemination. *J. Comput. Cult. Herit.* 11, 4, Article 18 (nov 2018), 22 pages. https://doi.org/10.1145/3230673

[42] Chung-Yi Weng, Brian Curless, and Ira Kemelmacher-Shlizerman. 2019. Photo Wake-Up: 3D Character Animation From a Single Photo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[43] Shan-Jean Wu, Chih-Yuan Yang, and Jane Yung jen Hsu. 2020. CalliGAN: Style and Structure-aware Chinese Calligraphy Character Generator. arXiv:2005.12500 [cs.CV]

[44] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. 2018. AttnGAN: Fine-Grained Text to Image Generation With Attentional Generative Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[45] Shaozu Yuan, Ruixue Liu, Meng Chen, Baoyang Chen, Zhijie Qiu, and Xiaodong He. 2021. Learning to Compose Stylistic Calligraphy Artwork with Emotions. In *Proceedings of the 29th ACM International Conference on Multimedia* (Virtual Event, China) *(MM '21)*. Association for Computing Machinery, New York, NY, USA, 3701–3709. https://doi.org/10.1145/3474085.3475711

[46] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N. Metaxas. 2017. StackGAN: Text to Photo-Realistic Image Synthesis With Stacked Generative Adversarial Networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

[47] Jiaxing Zhang, Ruyi Gan, Junjie Wang, Yuxiang Zhang, Lin Zhang, Ping Yang, Xinyu Gao, Ziwei Wu, Xiaoqun Dong, Junqing He, Jianheng Zhuo, Qi Yang, Yongfeng Huang, Xiayu Li, Yanghan Wu, Junyu Lu, Xinyu Zhu, Weifeng Chen, Ting Han, Kunhao Pan, Rui Wang, Hao Wang, Xiaojun Wu, Zhongshen Zeng, and Chongpei Chen. 2022. Fengshenbang 1.0: Being the Foundation of Chinese Cognitive Intelligence. *CoRR* abs/2209.02970 (2022).

[48] Kai Zhang, Nick Kolkin, Sai Bi, Fujun Luan, Zexiang Xu, Eli Shechtman, and Noah Snavely. 2022. ARF: Artistic Radiance Fields. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer Nature Switzerland, Cham, 717–733.

[49] Lvmin Zhang and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. arXiv:2302.05543 [cs.CV]

[50] Linqi Zhou, Yilun Du, and Jiajun Wu. 2021. 3D Shape Generation and Completion Through Point-Voxel Diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 5826–5835.

Received May 2023; accepted July 2023