Hindawi
Scientific Programming
Volume 2023, Article ID 9443704, 12 pages
https://doi.org/10.1155/2023/9443704

Research Article

Generative Deep Learning for Visual Animation in Landscapes Design

Peter Ardhianto,[1] **Yonathan Purbo Santosa**,[2] **Christian Moniaga**,[3] **Maya Putri Utami**,[1] **Christine Dewi**,[4] **Henoch Juli Christanto**,[5] **and Abbott Po Shun Chen**[6]

[1]*Department of Visual Communication Design, Soegijapranata Catholic University, Semarang 50234, Indonesia*
[2]*Department of Informatics Engineering, Soegijapranata Catholic University, Semarang 50234, Indonesia*
[3]*Department of Architecture, Soegijapranata Catholic University, Semarang 50234, Indonesia*
[4]*Department of Information Technology, Satya Wacana Christian University, Salatiga 50711, Indonesia*
[5]*Department of Information System, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia*
[6]*Department of Marketing and Logistics Management, Chaoyang University of Technology, 168 Jifeng E Road, Taichung 413310, Taiwan*

Correspondence should be addressed to Abbott Po Shun Chen; chprosen@gm.cyut.edu.tw

Received 29 January 2023; Revised 16 June 2023; Accepted 15 July 2023; Published 8 August 2023

Academic Editor: Jianping Gou

The biggest challenge for architecture designers is the time required for the design process. Especially landscape architects who have different work limits from architects in general. In contrast to architects in general, who are assisted in producing design plans by building standards, building requirements, and space programs that adapt to the type of project being undertaken. At the same time, some design jobs demand high-productivity landscape animation presentation in a short time. The long process involved in designing animation often makes it difficult for designers to produce optimal work. This study proposes generative zooming animation with artificial intelligence support to shorten the designer's work process and energy optimization. Deep learning with Vector Quantized Generative Adversarial Network and Contrastive Language-Image Pre-Training was used to generate alternative landscape designs from text prompt-based and compile them in animation. Our experiment shows that one frame can be generated roughly in 3.636 ± 0.089 s, which is significantly faster than the conventional method to create animation. Moreover, our method is able to achieve a good-quality image, which scored 3.2904 using inception score evaluation. The effectiveness of deep learning in visual landscape and animation creation can help designers speed up the design process. Furthermore, working time efficiency without compromising design quality will increase designer productivity and economic growth.

1. Introduction

Architectural design works in a structured and relatively long work pattern [1]. Starting from discussing the basic concept behind the design request, the visualizing design, and the presentation of the work to revision is a circle of interrelated processes. This linkage often also needs to be more systematic due to the different perceptions between the designer and the employer. The other side of the length of a designer's work pattern is energy consumption which increases significantly. Digital design is expected to have the lowest greenhouse gas emissions and the smallest environmental impact [2]. This increase needs to be addressed with changes in work methods and the use of technology for a lower energy future. In architectural design, landscape design needs more exploration. The arrangement and modification of features in a landscape, urban area, or garden are referred to as landscape design. It entails the creation of urban and rural landscapes through the planning, designing, and managing of open spaces [3]. Landscape design connects the worlds of landscape architecture and garden design. It will help to improve and save the environment in the long future, before the development era of artificial intelligence (AI). The software that is often used is SketchUp and enhanced with the Vray or Enscape rendering engines [4]. The combination of this software can help transform design ideas into more realistic landscape planning drawings [5]. The development of drawing software makes the resulting images more realistic, resulting in the emergence of ideas to make design revisions even higher. Moreover, the revision process is an obstacle for architects when implementing conventional design processes. Thus, the amount of human resources, energy, as well as detailed revision processes carried out on the drawing software makes work efficiency increasingly lag [6].

According to Reijers et al. [7], the work pattern, or workflow management, is the denser (efficient), the better. The central premise of implementing workflow management is that easier work coordination will provide a higher quality of service delivered. Moreover, Reijers said that if workflow management is appropriately done, it will provide process efficiency and become more flexible. A designer's workflow for producing a visual environment is lengthy in concept design strategy [8]. By using AI, the exploration of concept architecture landscape design has a paperless solution and needs less energy. Along with the development of AI in deep learning methods, the work process of an architectural designer can also be made more accessible. One of the derivatives of AI technology that can be utilized in the efficiency of a designer's workflow management is the creation of AI-based generative deep learning [9]. Generative deep learning may influence designers to put forward initial concept ideas until the design is finally finished.

Generative deep learning has many advantages in reducing the workflow in the digital design field [10]. The capability to train algorithms to generate additional data based on existing sample data distinguishes generative adversarial networks (GANs) [11]. The GAN examines and comprehends data distribution and filters or adjusts multiple inputs by identifying and preserving the sample data in the training sample [12]. Many artists and designers will benefit from this neural network model's expanded creative area and alternatives [13]. Because of their remarkable mechanism connecting visual objects to abstract texts supported by computing power and the big-data era, Vector Quantized Generative Adversarial Network (VQGAN) and Contrastive Language-Image Pre-Training (CLIP) have shown significant potential in conceptual architectural design where their esthetic ingenuity and spatial perception can meet that of professional human architects [14].

VQGAN is able to generate high-resolution and detailed images because it learns to generate an image by distinguishing visual parts (context of the image) from the training sample. Given an architectural image sample, VQGAN finds distinguishing features such as walls, windows, grass, and other features, which are then encoded in its latent space to preserve the details of the generated images. The generated images will be evaluated by a discriminator, which determines if the generated image is a real image from the training sample or a fake image from the generator [15]. By doing so, the generator is expected to learn to deceive the discriminator resulting in a believable image. On the other hand, CLIP learns to pair images with a snippet of a text caption. Given several images and several text captions, CLIP trains an image encoder and a text encoder. Each encoder produces a vector and combines it with the dot product to calculate the likelihood of image and caption pairs. Given both models, we can use CLIP to generate a text-encoded vector from a given text prompt and VQGAN generated image to calculate the likelihood of the pair [10]. By training both models together, CLIP will act as the discriminator and evaluate whether VQGAN is generating images with high likelihood given a text prompt. Thus, the combination of both models is called VQGAN + CLIP. In addition, by exploring novel applications of the VQGAN + CLIP framework that extends beyond existing literature. The model may demonstrate the effectiveness in generating interactive animations, enabling users to influence the generated content through user-controlled attributes dynamically. This functionality introduces new avenues for creative expression and interactive media generation. Furthermore, this study includes a comprehensive analysis of the strengths and limitations of the VQGAN + CLIP framework in the context of animation creation. Through rigorous experimentation and comparative analyses of inception score (IS), this study examines the effects of different parameters of each ID prompt that can offer prompt insights into the underlying mechanisms of the framework and serves as a guiding resource for future research in digital architecture.

By using generative deep learning to create multiple visual landscapes and animation, a designer's energy and time will be significantly reduced and will further support energy efficiency. Furthermore, the use of deep learning-based animation, which can reduce design work time, can have implications for the productivity of designers in the architecture industries and boost economic growth. This research is an initial study in generative images and animation by utilizing the speed and performance of generative deep learning.

1.1. Related Work. AI has helped human life a lot [16, 17]. AI plays a role in almost all design studies, such as architecture and visual communication design [18]. Steenson, 2022 in his book "Architectural Intelligence: How Designers and Architects Created the Digital Landscape," discusses how designers and architects create digital landscapes [19]. This book explores the work of four architects in the 1960s and 1970s who incorporated interactive elements into their work. The four architects studied in this book, Christopher Alexander, Richard Saul Wurman, Cedric Price, Nicholas Negroponte, and the Massachusetts Institute of Technology, incorporated technology, including computational and AI, into their work and influenced digital design practice from the late 1980s to until now. Alexander, long before his famous 1977 book A Pattern Language, used computation and structure to visualize design problems; Wurman proposed the notion of "information architecture." Price designed some of the first intelligent buildings, and Negroponte is experimenting with how people experience AI, even at an architectural scale. Steenson explores how these architects pushed the boundaries of architecture and how their technological experiments pushed the boundaries of technology. What do computational, cybernetics, and AI researchers have to gain by engaging with architects and architectural issues? And what are the new spaces emerging in this collaboration? At times, writes Steenson, the architects in this book characterize themselves as anti-architect and their work as antiarchitect. The projects examined in this book mainly do not result in buildings being built but in design processes and tools, computer programs, interfaces, and digital environments. Alexander, Wurman, Price, and Negroponte laid the foundations for our contemporary interactive practices, from information architecture to interaction design, from machine learning to AI-assisted smart cities.

It is necessary to have an optimal visual concept from the designer. AI with the deep learning method is one of the main components in generating visual ideas and concepts. Wang [20] uses deep learning in computational art design that can recognize images in digital media environments to be used to recreate classic animated images. This study analyzes animated images and components needed to form classified characters. This study uses a modification method in the form of the non-uniform rational basis spline curve that is used to increase the diversity of the generated population to build rich and diverse components to produce various animation image styles [20].

Although image generation makes it easier for designers to process their ideas, the need for 3D visualization also plays a role in design projects. Sinenko et al. [21] carried out an automation process with image generation and then added organizational management to the program. This research studies the 3D software used in drawing the master plan and its elements. This study emphasizes increasing the level and quality of design work and broadening the user community. Furthermore, online design ideas emerge from analyzing the use of virtual technology to visualize design solutions in organizational and technological construction operations. In addition, this study describes the approach used to solve the visualization problem of corporate and technical solutions comparatively. The conclusion obtained from this study is that using 3D software makes the design more expressive and can increase its competitiveness in the construction field. 3D software also shows a photorealistic view of future projects, thereby increasing investor and consumer interest in project realization [21]. 3D-based image generation will also be developed in this study so that the animation visualization series can be clearer in design projects.

Wang and Ma [22] observe that AI is not just technical support for works of art but can become a new art form. This study found that the increasing number of AI artworks appearing on the art market makes AI artworks categorized as a new artistic form integrating contemporary art and technology. The focus of this research is to examine the connotations and developmental status of AI artwork, seeking the core values of AI artwork based on comparing the procedures and methods of art creation. This study analyzes AI artwork from production, circulation, consumption, and feedback [22]. This research helps reduce text prompts to generate AI-based images and animations, so they can find and recommend the proper steps for designers to use deep learning in architectural animation. The literature review found that designing AI-based animations needs to consider visual concepts, image generation, 3D models, and framing processes.

AI-based animations also use generative art as their process, where some researchers study about generative art and how it would disrupt the process of visual design. Singh et al. [23] explored the use of generated decorative art as a source of inspiration for design tasks. They found that the abstract images used by generative deep learning can provide different visual stimuli to inspire a design task and guide the direction for a given graphic design project [23]. GAN [24, 25] has two parts that are simultaneously trained, namely generative (G) and discriminative (D). The discriminatory model is intended to determine whether a sample is based on real data or false data. The generative model captures certain target information distribution to puzzle the discriminative model [26]. The D model is a binary classifier that classifies the G model's data as either (realistic) or not (unrealistic) in the training system. G minimizes its loss function by supplying data that D classifies as real, as modeled by Equation (1) [27].

$$\begin{aligned} \text{minmax}\,(D, G) = {} & E_{x-pdata(x)}[logD(x)] \\ & + E_{x-p(x)}[\log(1 - D(G(z)))]. \end{aligned} \tag{1}$$

Nassery et al. [28] use generative art in the application of architectural works. The result shows that the process allows for the creation of almost infinite numbers of variants for design and more artistic work. Artut et al. [29] state that generative art can be combined with futurism art, and this could result in exploring more esthetic constituents and procedural methodologies for generative art. Kim et al. [30] discuss the use of generative art in evaluating fashion designs. The result shows that generative art can be used to produce creative ideas in modern fashion based on shape analysis of previous designs. Furthermore, the concept of design art generation will be used in this study with the popular model approach, namely VQGAN and CLIP [15, 31].

This study improves previous work's method by eliminating manual steps and improves interactivity between users and technology so that it would be more effective and efficient. Previous work was divided into two sections, which are manual and software use. First, in the manual section, design requirements are used as the base for searching for ideas, which after that uses the idea of creating concept and sketches it on paper. Second, from paper sketches of idea concept, designer will make a concept design for the work and make another paper sketches of design. From the design made on paper sketches, then designer will use a computer application like sketch up to make visualization and software development like auto cad, BIM, etc., to create a better presentation of the idea. In this section, software development is looping steps that can be repeated several times in order to create the best visualization. In the previous work, a computer is used to help designer make better visual for presentation [32]. In the proposed method, the use of a computer shortens the steps where from design requirements, designer will think about the prompt to visualize

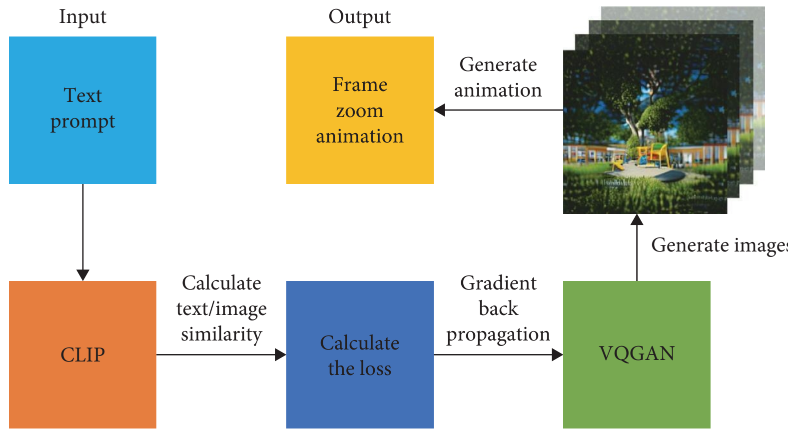


Figure 1: Research steps using Vector Quantized Generative Adversarial Network (VQGAN) and Contrastive Language-Image Pre-Training (CLIP) for visual and animation landscape design.

the requirements and then use AI to generate it. After that, design ideation and visualization as looping steps in order to create the best visualization require shorter steps which can increase the efficiency of time. Therefore, the proposed method eliminates manual steps and improves interactivity between users and technology, while at the same time, it creates high-quality solutions [33].

2. Materials and Methods

In this study, a generative deep learning algorithm applied the VQGAN and CLIP in producing visuals through text (Figure 1). VQGAN is a generative neural network that produces images based on the underlying distribution of images (but not from a prompt), primarily consisting of two deep learning models: the generator and the discriminator. The generator is trained to generate fake data points from a latent variable generated by an encoder that translates image input into the underlying distribution of vectors, and the discriminator identifies the data points as either real or fake [10]. Once the generator can deceive the discriminator indefinitely, the model is ready to be used, and the image encoder is replaced with a random number generator given some distribution parameters. CLIP is another neural network that can determine how well the caption (or prompt) matches the image [15]. The zooming effect is given for early studies in creating design animations that combine frame by frame from the visual generation produced by VQGAN and CLIP [34].

In this study, we use codebook size 16384 from ImageNet as pretrained data in generating images. The size of the VQGAN codebook means that a larger codebook results in a much better representation [35]. For the prompt text analysis process, we used CLIP which is capable of producing images of high visual quality from text prompts of significant semantic complexity. CLIP is a model trained to assess the fit of a caption, compared to other captions in a set, to an image. CLIP is capable of zero-shot learning, enabling it to perform well even on unknown data. When applied in VQGAN-CLIP, CLIP is able to assess the quality of generated images compared to a user-inputted caption, and the outputted scores can be used as weights to "guide" the learning of the VQGAN to more accurately match the subject matter through recursive iteration [15].

The research steps to be carried out in this study are as follows: first, we retrain the VQGAN continuing the checkpoint of ImageNet with codebook size 16,384 using architectural datasets obtained through Kaggle, which contains images of various architectural styles, both classic and modern, consisting of 5,000 images of different picture of building with 25 different styles [36]. In the experiment, the error comes to equilibrium at around three epochs, so we decided to stop early at Epoch 6, achieving validation reconstruction loss at 0.45059. We then used the trained model to generate animation videos in the later step. Second, the preparation of the algorithm and the installation of the required packages through the Google Colaboratory; third, prompt text analysis for the visual concept creation process; fourth, making a design prototype using VQGAN + CLIP; fifth, applying the zoom effect as an animation generator tool on the VQGAN + CLIP architecture. In training VQGAN + CLIP model, we train on top of the available pretrained model-based notebook; this was based on Katherine Crowson [15]. The model was trained using Google's Colaboratory on December 2022 with Python version 3.8.16 with the help of PyTorch version 1.13.0 [37] and Numpy version 1.21.6 [38]. We use random generator seed 123 and optimizer Adam with a learning rate of 0.1, and the remaining parameters are default to the PyTorch library.

2.1. Network Architecture. VQGAN learns the quantization of the vector in latent space by using a transformer. In Figure 2, $\tilde{z}$ is the distribution parameters of image patches. By using these distribution parameters, the transformer calculates the probability of the next image patch given the previous image patch, hence the name quantization. The result z_q is the latent variable used by the decoder to recreate the fake image. The discriminator will detect if the fake image indicates that the image is unnatural.

CLIP task is to calculate image similarity to a given text prompt. To achieve that, CLIP was trained by using a text prompt and image as the input, encoding both inputs into a matrix representation and classifying the results if both

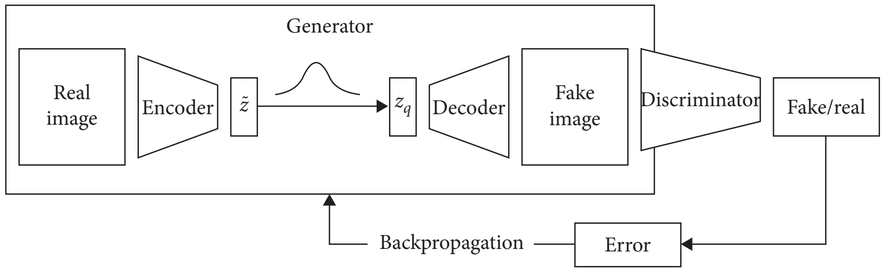


FIGURE 2: The network architecture of Vector Quantized Generative Adversarial Network (VQGAN).

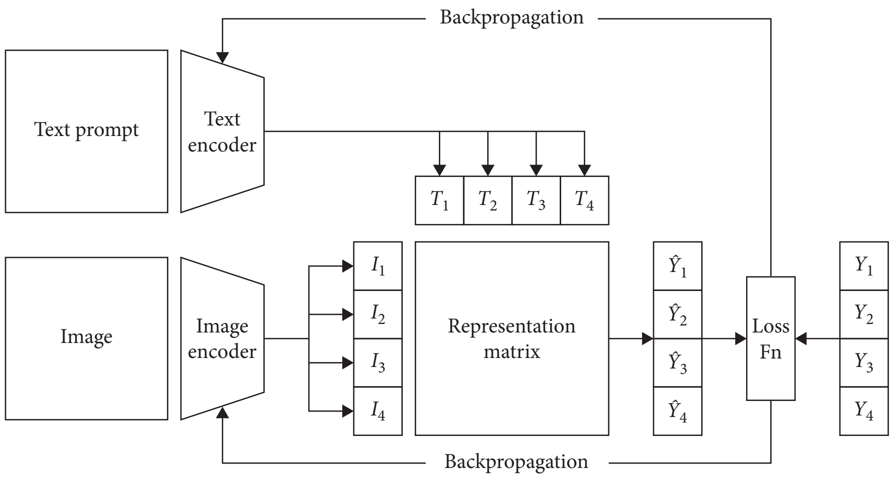


FIGURE 3: The network architecture of Contrastive Language-Image Pre-Training (CLIP).

contain a particular object (Figure 3). The matrix representation is a dot product of image embedding and text embedding, which contains the representation of both inputs. By classifying each row of the representation matrix, CLIP produces a vector of probabilities that can be compared with the ground truth label. After the training, the representation matrix will be able to identify if a given image and a given text prompt are similar or the opposite.

Combining both VQGAN and CLIP, we can use VQGAN to generate an image and evaluate the similarity of the generated image with a text prompt using CLIP. In VQGAN + CLIP, VQGAN acts as the generator, while CLIP acts as the discriminator. As a result, the generator should learn how to generate an image that satisfies the text prompt with high similarity.

Furthermore, the prompt used to generate landscape architectural images was *"modern kindergarten school area landscape architecture*, *kids' playground*, *seating areas*, *perspective view*, *sunny weather*, *cinematic photo*, *highly detailed*, *cinematic lighting*, *ultra-detailed*, *ultrarealistic*, *photorealism*, *octane render."* Image size 512×512 pixels, 12 frames per second. While the parameters in the animation use 0.95 zooming modes, shifting the y-axis to 5 pixels. The generative deep learning with VQGAN and CLIP was performed in the cloud using Google Colaboratory. The Google Colaboratory, often known as Colab, is a cloud-based service. Colab is a machine learning and deep learning platform based on the Jupyter Notebook [39]. Google Colab allows free GPU NVIDIA-SMI 460.32.03 access during runtime.

3. Results

3.1. Render Time. In order to measure the time needed for generating an animation video, we conduct experiments on several text prompts. We use eight different text prompts to measure the generation time, as presented in Table 1. We use ID for easier reading of results since some text prompts are quite long.

When generating video using VQGAN + CLIP, we train the model 10 iterations for producing one frame. We measure starting from the training process until one frame is generated. For each text prompt, we generate 72 frames; thus, there will be 72 measurements each. The results are presented in Figure 4. Overall, the average time needed to generate one frame is 3.636 ± 0.089 s meaning for 72 frames it requires roughly 261.792 s which is significantly faster than generating animations in conventional ways (from sketching, animating, until rendering).

3.2. Image Generation Quality. In order to measure the image quality generated by VQGAN + CLIP, we use IS, which compares the entropy between the generated images and the pretraining dataset and measures the ability to generate varying images by assessing the distribution of existing objects in the generated image according to InceptionV3 model [40]. IS is preferred to have a larger positive number. To evaluate our approach, we use the same text prompt, as described in Table 1 and added several combinations of different cases

Table 1: Text prompts and its ID.

ID	Text prompt
A1	Modern kindergarten school area landscape architecture, kids' playground, seating areas, perspective view, sunny weather, cinematic photo, highly detailed, cinematic lighting, ultra-detailed, ultrarealistic, photorealism, octane render
A2	Modern kindergarten school area landscape architecture
A3	Kids' playground
A4	Sitting areas kindergarten playground
A5	Perspective view kindergarten playground
A6	Lighting kindergarten playground
A7	Sunny weather kindergarten playground
A8	Cinematic photo, kindergarten playground

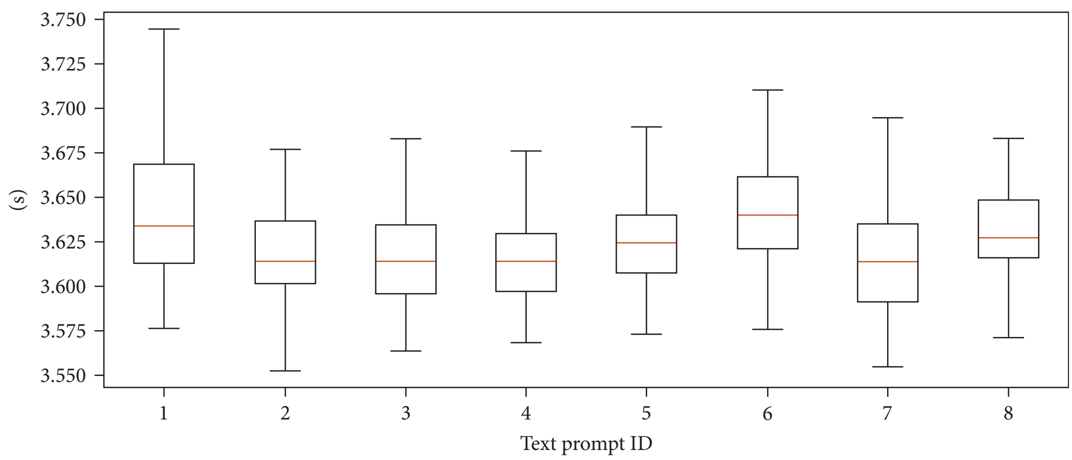


Figure 4: Measurements of time required to generate 72 frames.

based on the architecture style dataset classes were added. The added prompts are described in Table 2.

The dataset includes 72 frames generated for each text prompt, which are subsequently assessed using the IS. The obtained results are presented in Table 3, facilitating further analysis. It is important to note that the text prompts listed in Table 1 are not present within the architecture style dataset, while those featured in Table 2 can be identified by their corresponding class names in the dataset. Upon comparing the outcomes depicted in Table 3, it becomes evident that images imbued with architectural styles tend to exhibit lower IS when juxtaposed with text prompts lacking style information.

For qualitative evaluation, we investigate only for case A. The results of the generative images in this study were used at 300 iterations, producing 300 images. The generated image results have similar visuals but are different in each frame. Figure 5 describes the process from the initial to the final iteration. Furthermore, samples of generative images using the animation effect zoom 0.95 and shifting the y-axis 5 pixels make the resulting image have a zoom-out impact (Figure 5(a)–5(f)).

The results also show a high probability of human perception of the mentioned condition given the text prompt. To evaluate thoroughly, the text prompt will be divided into several parts. Each part that is easy to observe will be evaluated on three generated images of the VQGAN and CLIP model compared with the top four Google image search results. The Google image was searched on December 14, 2022, from Semarang City, Indonesia. Image search results might differ depending on the location it is searched for and different Google search history.

The first condition is "*modern kindergarten school area landscape architecture.*" These conditions were used as a Google search query, and as a result, all four images contain some form of the grass field, brick-colored walls, and some structure for toddlers to use. Although the underlying architecture and landscape appear different, it is easy to see that the generated image still resembles a kindergarten school area with its building color and grass field (Figure 6).

The second condition was "*kids' playground.*" The top four Google image search contains a kid's playground set. In the generated image, there is also some form of playground set that resembles a castle and a slide (Figure 7).

The third condition was "*seating area.*" The term seating area is far too general, so the keyword "kindergarten playground" was added to find the top four Google image searches containing "seating area kindergarten playground." In the generated image, it is hard to see whether the seating area exists or not (Figure 8).

The fourth condition was "*perspective view.*" The top four Google image search contains "perspective view." The

Table 2: Added text prompts and its ID.

ID	Text prompt
B1	Romanesque church area landscape architecture, city center, perspective view, cloudy weather, cinematic photo, highly detailed, cinematic lighting, ultra-detailed, ultrarealistic, photorealism, octane render
B2	Romanesque church area landscape architecture, city center
B3	Romanesque church area landscape architecture, perspective view
B4	Romanesque church area landscape architecture, cloudy weather
B5	Romanesque church area landscape architecture, cinematic photo
B6	Romanesque church area landscape architecture, highly detailed
B7	Romanesque church area landscape architecture, cinematic lighting
B8	Romanesque church area landscape architecture, ultra-detailed
B9	Romanesque church area landscape architecture, ultrarealistic
B10	Romanesque church area landscape architecture, photorealism, octane render
C1	Postmodern, lake side, perspective view, overcast, cinematic photo, highly detailed, cinematic lighting, ultra-detailed, ultrarealistic, photorealism, octane render
C2	Postmodern, lake side
C3	Postmodern, perspective view
C4	Postmodern, overcast
C5	Postmodern, cinematic photo
C6	Postmodern, highly detailed
C7	Postmodern, cinematic lighting
C8	Postmodern, ultra-detailed
C9	Postmodern, ultrarealistic
C10	Postmodern, photorealism

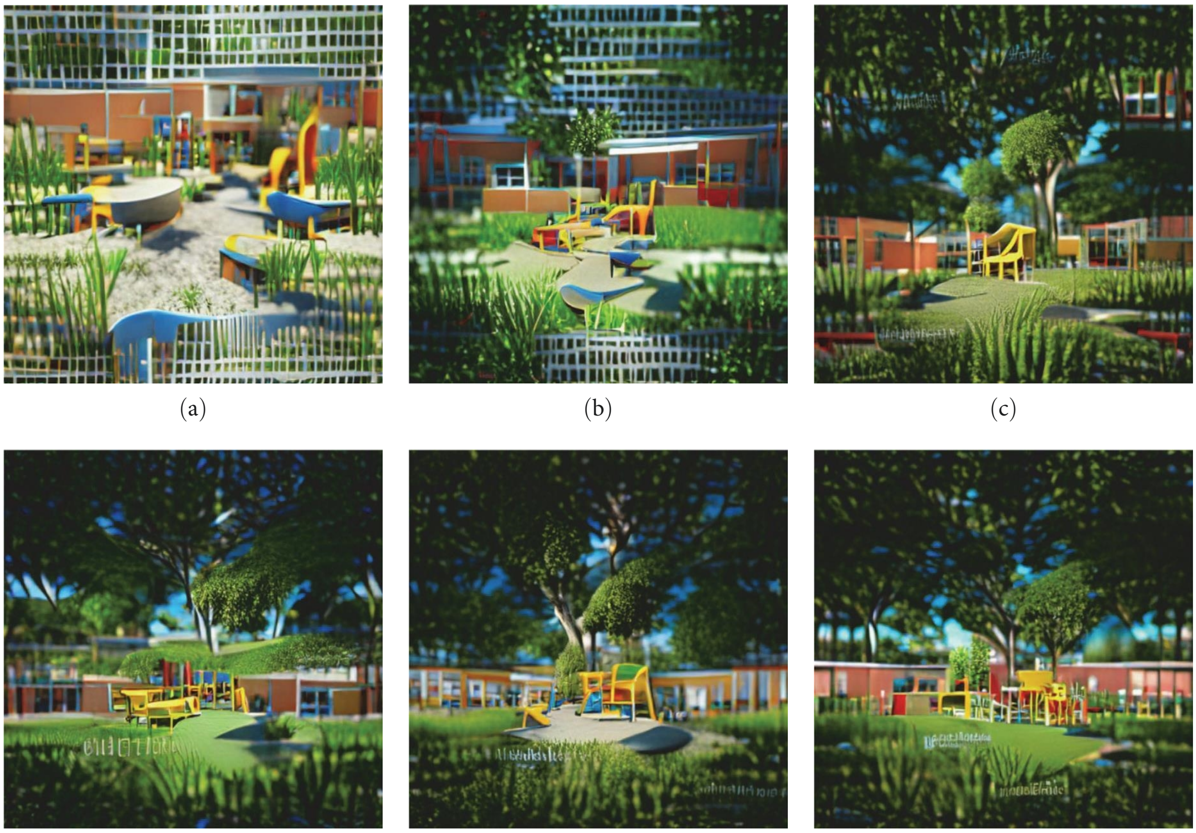
Table 3: Inception score for each text prompt.

Text prompt ID	Inception score A	Inception score B	Inception score C
1	3.4815	2.8816	2.1954
2	3.4415	2.5006	2.5119
3	3.5597	2.9206	3.1883
4	3.0212	2.3372	2.6195
5	2.9978	3.4886	2.9805
6	3.1705	2.3612	2.3383
7	2.9779	2.5733	2.3712
8	3.6729	2.8551	3.0752
9	–	2.5022	2.3094
10	–	3.4664	3.1766
Average	3.2904	2.7887	2.6766

generated image is clearly shown with multiple objects appearing larger or smaller depending on the distance to the point of view, as described with the Google image search results (Figure 9).

The fifth condition was "*sunny weather.*" The top four Google image search containing "sunny weather" was too general, so the keyword "*kindergarten playground*" was added to find the specific results. In the generated image, a similar weather setting was also observed with sunlight and harsh shadows (Figure 10). The rest of the conditions, "*cinematic photo*, *highly detailed*, *cinematic lighting*, *ultra-detailed*, *ultra-realistic*, *photorealism*, *octane render*," was an artistic description of architectural lighting effect for visual rendering technique.

3.3. Generative Animation. The results of an entire frame that has been generated are then converted to mp4 video format with a duration of 25 s. Generative deep learning automatically sequences from the first frame to the last frame in video formats without compressing the resolution. The video makes a zoom animating effect with various alternative landscape visual designs (Figure 10). According to Figure 11(a), in the initial frame, the generative animation shows a close-up zoom, starting from the middle of the landscape design. According to Figure 11(b), in the middle frame, the generative animation depicts the middle perspective with the surrounding environment according to the prompt text (Figures 6–8). Furthermore, Figure 11(c) shows the final frame, which shows the farthest distance from the perspective landscape design



(a) (b) (c) (d) (e) (f)

Figure 5: The samples of the visual landscape of 300 iterations are based on deep generative learning using the VQGAN and CLIP models. (a) Iteration: 20, loss: 0.646378 (1.10it/s), (b) iteration: 38, loss: 0.659106 (1.09it/s), (c) iteration: 140, loss: 0.669689 (1.09it/s), (d) iteration: 200, loss: 0.656828 (1.09it/s), (e) iteration: 260, loss: 0. 674269 (1.07it/s), and (f) iteration: 300, loss: 0. 0.669074 (1.09it/s).

(a) (b)

Figure 6: The first condition of generated image compared with Google image search "*modern kindergarten school area landscape architecture.*"

with lighting conditions and perspective according to the text prompt (Figures 9–10).

4. Discussion

The deep learning performance in generating visuals and animation showed excellent results. VQGAN and CLIP performed well in various generated images with IS score of around 3.2904 s (Table 2). Furthermore, the animation compilation showed a complete frame containing 300 images from each frame with a zooming feature in a 25 s-video format. The generative zooming animation may help the designer ideate, open the alternative objects and layout, determine the tone and manner of visual design, and open the digital design possibility in trial and error using the text prompt.

The generated 72 frames for each text prompt are then measured with IS, and the results are presented in Table 3 for

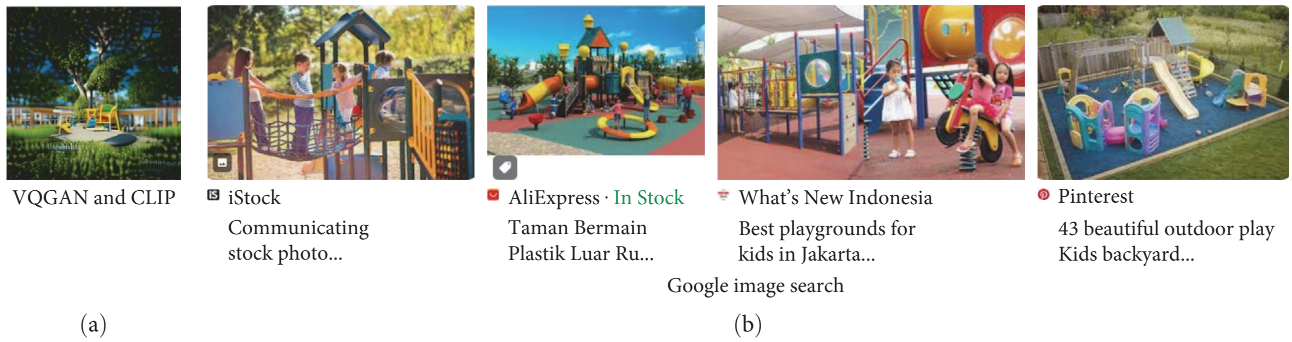


Figure 7: The second condition of generated image compared with Google image search "*kids' playground.*"

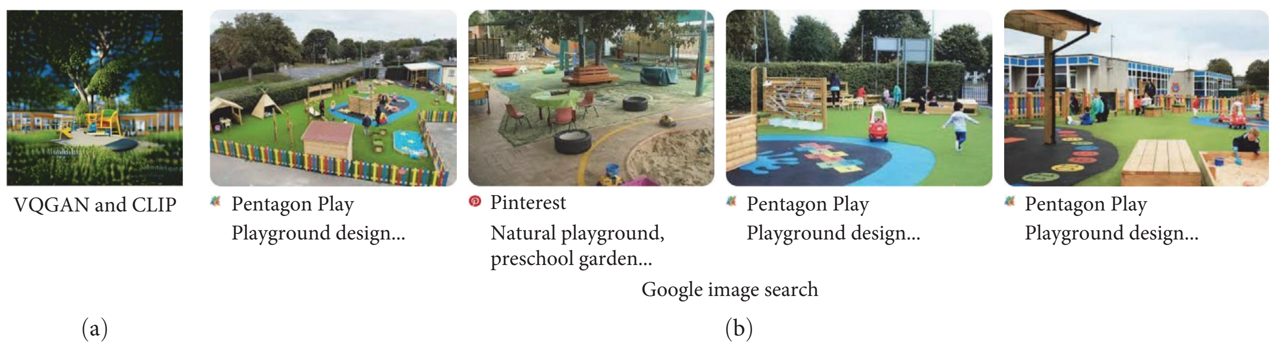


Figure 8: The third condition of generated image compared with Google image search "*sitting area kindergarten playground.*"

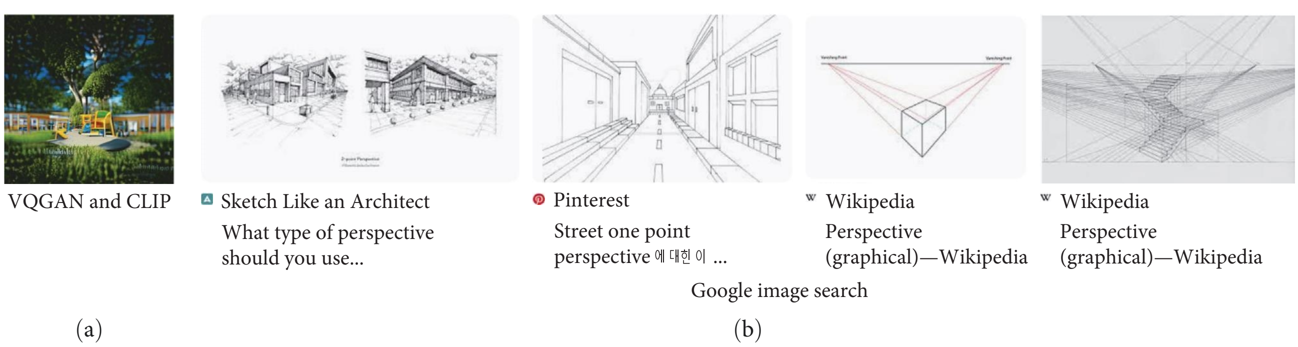
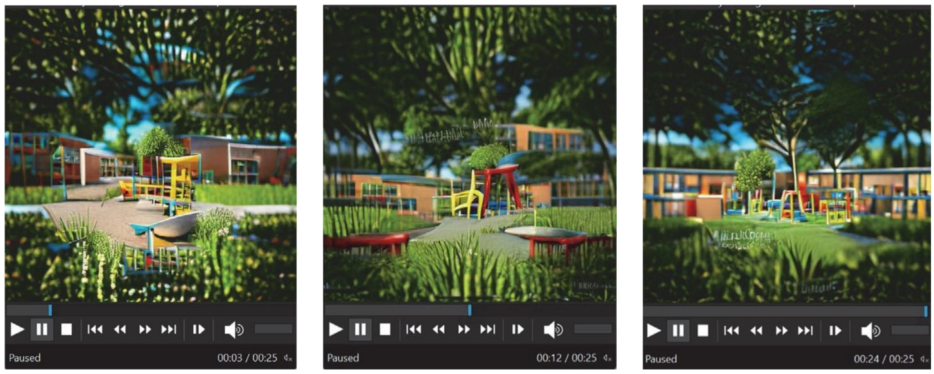


Figure 9: The fourth condition of generated image compared with Google image search "*perspective view.*"

Figure 10: The fifth condition of generated image compared with the Google image search "*sunny weather.*"



Early (a) Middle (b) End (c)

Figure 11: The samples preview animation videos from VQGAN and CLIP text-to-images with zooming animation from the early, middle, and ending frames.

further analysis. Text prompts in Table 1 do not exist in the architecture style dataset, while the text prompts in Table 2 exist in the architecture style dataset noted by the class names in the said dataset. Comparing the results in Table 3, the IS for pictures with architectural styles has lower values compared to the text prompt with no style information. We suspect that it might be due to the reconstruction loss that we achieve during the training of VQGAN using the architectural style dataset. Because of that, we recommend the upcoming research to study the effects of lower reconstruction loss of VQGAN with the IS of the generated results. Thus, now we are left at evaluating the results using qualitative means.

Furthermore, the designer's work system is currently often used with the help of mobile edge computing (MEC). One example of software used by landscape architects in creating design is Sketch Up. The current development of Sketch Up has adopted the MEC and IoT systems where all design element assets are represented by simple vectors whose original models are stored in the cloud during the design process. The original assets that represent the original model, such as trees, people, vehicles, and so on, will be seen in detail when the rendering process is complete. This process has the intention that computer performance is not too heavy because the assets used during the design process have a very low level of detail [41]. Nevertheless, energy consumption remains a clear issue to be overcome on mobile device networks, such as MEC environments [42]. Furthermore, IoT devices need to handle lots of data, which is also energy-consuming. Thus, reducing energy consumption in networks with IoT devices is a goal worth exploring.

Energy consumption caused by computer performance in interpreting designs will increase significantly when product designs are revised. Energy consumption can increase several times according to the number of design revisions or changes. AI can help optimize energy. Generative design processes in AI can help leapfrog conventional designer workflows. Optimization and energy efficiency will increase when designers only need to determine keywords (prompts) that match the expected design formulation, and AI will help generate the intended image interpretation.

According to Table 1, the study demonstrates how this collaboration between AI and designers can continue indefinitely to meet design goals through the prompt ID. The designer oversees deep learning throughout the entire design process, including ideation, even if deep learning can execute design production tasks. [43]. According to Zeng et al. [44]. Designer professionals remain responsible for addressing queries, creating directions, and establishing the framework. Deep learning can therefore assume some of the burdens of producing a wider variety of design forms than could reasonably be accomplished by a design team or even generate the visual layout [44]. The time, resources, and electricity needed by a designer to produce animations will be greatly decreased by using deep learning, thus reducing environmental impact. Additionally, deep learning-based animation, which can speed up design work, may impact designers' productivity in the creative industries and economic growth.

The development of text-to-images-based deep learning will undoubtedly speed up the design process (Figure 4 and Table 2). This study demonstrated the positive effects of deep learning using VQGAN and CLIP in generating multiple visual alternatives but maintaining the style and tone in each frame generated by deep learning (Figure 5). These advantages will benefit the designer landscape to select and compare the visuals that meet the design goal from generated images (Figures 6–10). With VQGAN and CLIP models, designers will spend less time on the most time-consuming design tasks [45]. Since design requires knowledge, esthetic and artistic considerations are given. Based on his expertise, experience, and added creative awareness, the designer crafts his creations [46].

Based on Figure 11, the alternatives of 300 frames have been compiled in mp4 video format. The zooming animation can help the designer understand the sequence landscape view more clearly and find the suitable perspective view in landscape design [47]. A suitable perspective view is advantageous to understand the application of perspective architecture, environment, and lighting. Furthermore, the zooming animation may help present the landscape architectural design beside the landscape mockup test [48].

However, the current work has several limitations that could be used to develop future improvement directions. First, the animation features employed in landscape design are currently confined to zooming and y-axis movement, limiting the diversity and dynamism of the visualizations. To expand the range of animation capabilities, investigating scalable vector graphics and nonautoregressive object prediction techniques, as proposed in reference [49], holds promise. Second, a thorough examination of the style and tone manifested by the visuals generated using VQGAN + CLIP especially when adding training data is imperative. By delving deeper into this aspect, we aim to gain a comprehensive understanding of their characteristics and explore techniques such as stable diffusion and text-to-image networks, as mentioned in reference, to gain finer control over the style and visual tone of the generated designs. Tackling these limitations will propel the advancement of our approach and contribute significantly to the landscape design visualization domain. [50].

5. Conclusions

This study aimed to apply a generative deep learning network to visualize and animate the landscape design. According to our results, VQGAN and CLIP are suitable to assist landscape designers in generating images based on a prompt text by employing 0.95 zooms and lifting the y-axis by 5 pixels. Our experiment shows that one frame can be generated roughly in 3.636 ± 0.089 s which is significantly faster than the conventional method to create animation. Moreover, our method is able to achieve a good quality image which scored 3.2904 using IS evaluation. As a result, the efficiency of generative deep learning in the construction of visual landscapes can assist designers in accelerating the design process, which may also boost designer productivity due to multiple alternatives design in the same style and tone. The designer can use generative zooming animation to develop ideas, explore different materials and layouts, choose the design esthetic, and explore the potential of digital design through trial and error.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors wish to express gratitude to the Head of the Human, Environment, and Technology Research Center of Soegijapranata Catholic University, Satya Wacana Christian University, Chaoyang University of Technology Taiwan, and Atma Jaya Catholic University of Indonesia for the support of this study. This research was funded by The Office of Research and Community Services Soegijapranata Catholic University Indonesia (00745/B.7.4/LPPM/12/2022), and this research is partially supported by the National Science and Technology Council, Taiwan (Grant number: MOST-111-2637-H-324-001-). The funding agency was not involved in data collection, analysis, or interpretation.

References

[1] M. W. Tracey and A. Hutchinson, "Developing designer identity through reflection," *Educational Technology*, vol. 53, no. 3, pp. 28–32, 2013.

[2] E. Benson and Y. Perullo, *Design to Renourish: Sustainable Graphic Design in Practice*, CRC Press, 2017.

[3] K. Irvine, H. H. Loc, C. Sovann et al., "Bridging the form and function gap in urban green space design through environmental systems modeling," *Journal of Water Management Modeling*, vol. 29, Article ID C476, 2021.

[4] F. Setyaningfebry, S. Felasari, and B. Michelle, "Developing a website for rendering based on renderers' experience," *Grafica*, vol. 11, no. 21, pp. 23–36, 2023.

[5] C. A. Paulson, *A study of the adaptation of parametric computer design among landscape architecture professionals in Texas*, 2017, University of Texas Arlington, http://hdl.handle.net/10106/26801.

[6] W.-L. Wu, Y. Hsu, Q.-F. Yang, and J.-J. Chen, "A spherical video-based immersive virtual reality learning system to support landscape architecture students' learning performance during the COVID-19 era," *Land*, vol. 10, no. 6, Article ID 561, 2021.

[7] H. A. Reijers, J. H. M. Rigter, and W. M. P. van der Aalst, "The case handling case," *International Journal of Cooperative Information Systems*, vol. 12, no. 3, pp. 365–391, 2003.

[8] P. Jones and M. Morrison, "Mind-mapping in design culture: a tool for ideation in graphic design education," 2021.

[9] P. Khatri, "The future of automatically generated animation with AI," in *Deep Learning in Gaming and Animations*, pp. 19–36, CRC Press, 2021.

[10] S. Shahriar, "Gan computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network," 2021.

[11] C. Dewi, R.-C. Chen, Y.-T. Liu, and H. Yu, "Various generative adversarial networks model for synthetic prohibitory sign image generation," *Applied Sciences*, vol. 11, no. 7, Article ID 2913, 2021.

[12] S.-W. Park, J.-S. Ko, J.-H. Huh, and J.-C. Kim, "Review on generative adversarial networks: focusing on computer vision and its applications," *Electronics*, vol. 10, no. 10, Article ID 1216, 2021.

[13] N. Meng, J. Yang, and H. Wang, "Icon art design with generative adversarial network under deep learning," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 3499570, 9 pages, 2022.

[14] R. Wang, "Collaborated machine learning-based design solution to conceptual design of architecture," 2022.

[15] K. Crowson, S. Biderman, D. Kornis et al., "Vqgan-clip: open domain image generation and editing with natural language guidance," 2022.

[16] P. Ardhianto, J.-Y. Tsai, C.-Y. Lin et al., "A review of the challenges in deep learning for skeletal and smooth muscle ultrasound images," *Applied Sciences*, vol. 11, no. 9, Article ID 4021, 2021.

[17] P. Ardhianto, B.-Y. Liau, Y.-K. Jan et al., "Deep learning in left and right footprint image detection based on plantar

pressure," *Applied Sciences*, vol. 12, no. 17, Article ID 8885, 2022.
[18] A. Raina, C. McComb, and J. Cagan, "Learning to design from humans: imitating human designers through deep learning," *Journal of Mechanical Design*, vol. 141, no. 11, Article ID 111102, 2019.
[19] M. W. Steenson, *Architectural Intelligence: How Designers and Architects Created the Digital Landscape*, The MIT Press, 2022.
[20] F. Wang, "Computer art design based on artificial intelligence," *Cluster Computing*, vol. 22, pp. 13881–13887, 2019.
[21] S. Sinenko, T. Poznakhirko, and V. Obodnikov, "Automation of visualization process for organizational and technological design solutions," *MATEC Web of Conferences*, vol. 270, Article ID 05008, 2019.
[22] Y. Wang and H. Ma, "The value evaluation of artificial intelligence works of art," in *2019 International Joint Conference on Information, Media and Engineering (IJCIME)*, pp. 445–449, IEEE, Osaka, Japan, 2019.
[23] D. Singh, N. Rajcic, S. Colton, and J. McCormack, "Camera obscurer: generative art for design inspiration," 2019.
[24] I. Goodfellow, "Nips 2016 tutorial: generative adversarial networks," 2016.
[25] C. Dewi, R.-C. Chen, Y.-T. Liu, X. Jiang, and K. D. Hartomo, "Yolo V4 for advanced traffic sign recognition with synthetic training data generated by various GAN," *IEEE Access*, vol. 9, pp. 97228–97242, 2021.
[26] G. Wei, M. Luo, H. Liu, D. Zhang, and Q. Zheng, "Progressive generative adversarial networks with reliable sample identification," *Pattern Recognition Letters*, vol. 130, pp. 91–98, 2020.
[27] C. Dewi, R.-C. Chen, Y.-T. Liu, and S.-K. Tai, "Synthetic data generation using DCGAN for improved traffic sign recognition," *Neural Computing and Applications*, vol. 34, pp. 21465–21480, 2022.
[28] F. Nassery and P. Sikorski, "New posibilities of using processing and modern methods of the "generative art" graphics in architecture," *Technical Transactions/Czasopismo Techniczne*, no. 4, pp. 25–33, 2015.
[29] S. H. Artut, "Futurism art and its significance to computational generative art," 2018.
[30] D. O. Kim and J. H. Choi, "Analysis of fashion design reflected visual properties of the generative art," *Journal of the Korean Society of Clothing and Textiles*, vol. 41, no. 5, pp. 825–839, 2017.
[31] M. J. Buehler, "Deepflames: neural network-driven self-assembly of flame particles into hierarchical structures," *MRS Communications*, vol. 12, pp. 257–265, 2022.
[32] K. Gallagher, A. Hatch, and M. Munro, "Software architecture visualization: an evaluation framework and its application," *IEEE Transactions on Software Engineering*, vol. 34, no. 2, pp. 260–270, 2008.
[33] W. Liu and F. Tao, "Art definition and accelerated experience: temporal dimension of AI artworks," *Philosophies*, vol. 7, no. 6, Article ID 127, 2022.
[34] J. Oppenlaender, "Prompt engineering for text-based generative art," 2022.
[35] P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12868–12878, IEEE, Nashville, TN, USA, 2021.
[36] Z. Xu, D. Tao, Y. Zhang, J. Wu, and A. C. Tsoi, "Architectural style classification using multinomial latent logistic regression," in *Computer Vision-ECCV 2014: 13th European Conference Part*, vol. I 13, pp. 600–615, Springer, Zurich, Switzerland, 2014.
[37] A. Paszke, S. Gross, F. Massa et al., "PyTorch: an imperative style, high-performance deep learning library," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pp. 8026–8037, The ACM Digital Library, 2019.
[38] C. R. Harris, K. J. Millman, S. J. van der Walt et al., "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, 2020.
[39] J. P. Gujjar, H. R. P. Kumar, and N. N. Chiplunkar, "Image classification and prediction using transfer learning in colab notebook," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 382–385, 2021.
[40] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," *Advances in Neural Information Processing Systems*, 2016.
[41] S. R. Sarangi, S. Goel, and B. Singh, "Energy efficient scheduling in IoT networks," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pp. 733–740, The ACM Digital Library, 2018.
[42] J. C. S. dos Anjos, J. L. G. Gross, K. J. Matteussi, G. V. González, V. R. Q. Leithardt, and C. F. R. Geyer, "An algorithm to minimize energy consumption and elapsed time for iot workloads in a hybrid architecture," *Sensors*, vol. 21, no. 9, Article ID 2914, 2021.
[43] Y. Lee, "Design participation tactics: the challenges and new roles for designers in the co-design process," *CoDesign*, vol. 4, no. 1, pp. 31–50, 2008.
[44] Z. Zeng, X. Sun, and X. Liao, "Artificial intelligence augments design creativity: a typeface family design experiment," in *Design, User Experience, and Usability. User Experience in Advanced Technological Environments. HCII 2019*, vol. 11584 of *Lecture Notes in Computer Science*, pp. 400–411, Springer, Cham, 2019.
[45] V. Liu, J. Vermeulen, G. Fitzmaurice, and J. Matejka, "3DALL-E: integrating text-to-image AI in 3D design workflows," 2022.
[46] Y. He, B. Camburn, J. Luo, M. C. Yang, and K. L. Wood, "Visual sensemaking of massive crowdsourced data for design ideation," in *Proceedings of the 22nd International Conference on Engineering Design (ICED19)*, pp. 409–418, Cambridge University Press, Delft, The Netherlands, 2019.
[47] Y. Chen, "Creation techniques of perspective effect drawing of garden landscape based on CAD technology," *Computer-Aided Design & Applications*, vol. 17, no. S2, pp. 158–168, 2020.
[48] J. A. Rahman, N. K. Amer, N. Atef, and M. A. Halim, "Stingless bee nest for housing area in Malaysia," *Design Ideals Journal*, vol. 1, no. 2, pp. 14–19, 2019.
[49] A. Carlier, M. Danelljan, A. Alahi, and R. Timofte, "Deepsvg: a hierarchical generative network for vector graphics animation," 2020.
[50] Z. J. Wang, E. Montoya, D. Munechika, H. Yang, B. Hoover, and D. H. Chau, "DiffusionDB: a large-scale prompt gallery dataset for text-to-image generative models," 2022.