A generative deep learning for exploring layout variation on visual poster design

Peter Ardhianto [a,1,*], Yonathan Purbo Santosa [b,2], Yori Pusparani [c,d,3]

[a] Department of Visual Communication Design, Soegijapranata Catholic University, Semarang 50234, Indonesia
[b] Department of Informatics Engineering, Soegijapranata Catholic University, Semarang 50234, Indonesia
[c] Department of Visual Communication Design, Budi Luhur University, Jakarta 12260, Indonesia
[d] Department of Digital Media Design, Asia University, Taichung 41354, Taiwan
[1] peter.ardhianto@unika.ac.id*; [2] yonathansantosa@unika.ac.id; [3] yori.pusparani@budiluhur.ac.id
* corresponding author

ARTICLE INFO

Article history
Received 2023-02-16
Revised 2023-04-18
Accepted 2023-05-20

Keywords
Visual Design;
VQGAN;
Generative Art;
Designer;
Dynamic Design

ABSTRACT

Layout variation is an essential concept in design and allows designers to create a sense of depth and complexity in their work. However, manually creating layout variations can be time-consuming and limit a designer's creativity. The use of generative art as a tool for creating visual poster designs that emphasize layout variety is explored in this study. Deep learning through generative art offers a solution by using an algorithm to generate layout variations automatically. This paper uses the VQGAN and CLIP approach to describe a generative art system, which renders images via a text prompt and produces a series of variations based on the zoom parameter 0.95 and shifts the y-axis 5 pixels. Our experiment shows that one frame can be generated roughly in 10.108±0.226 seconds, significantly faster than the conventional method for creating layouts on poster design. The model achieved a good quality image, scoring 4.248 using an inception score evaluation. The layout variations can be used as a basis for poster design visuals, allowing designers to explore different visual representations of layouts. This paper demonstrates the potential of generative art to explore layout variation in visual design, offering designers a new approach to creating dynamic and engaging visual designs.

1. Introduction

In visual communication design, producing posters often takes a lot of time and effort for designers when faced with many choices of design elements and layouts [1]. Experienced graphic designers spend a lot of time on repetitive manual layout designs. Existing templates are typically simplistic and unsuitable for most designs, lowering productivity and limiting inventiveness [2]. Poster layout design can be complicated since a poster needs to properly transmit information to its audience while being visually appealing and attention-grabbing [3]–[5]. Furthermore, a poster must convey a message or set of information while also being visually appealing. The difficulty is to strike a balance between the amount of information on the poster and the amount of space available for design components. Because the poster should be able to deliver the essential idea in a few seconds, it's critical to consider what information to add and what to leave out [6]. However, there are many different layout options when designing a poster and choosing the right one [7]. Moreover, poster design needs to be visually appealing, and finding the right balance of color, images, and other graphic elements can be difficult [8]. Deep learning with generative art is a form of digital art created using algorithms and code, allowing for the automatic generation of art without direct human intervention [9], [10]. It has grown in popularity recently, with artists and designers adopting generative techniques to create everything from animations to prints. Simultaneously, graphic design has always played a significant role in communication and marketing. Posters, for example, are prominent visual design media frequently used to promote events, products, or ideas. Layout variation, or displaying numerous layout variations, is an important part of visual design because it allows designers to create a sense of depth and complexity in their work.

Researchers have studied generative art and its potential impact on visual design. Singh *et al*. found that generative art's abstract images can provide different visual stimuli for design tasks and guide the direction of a given graphic design project [11]. Nassery and Sikorskii used generative art in architecture and discovered that it created almost infinite numbers of variants for design and more artistic work [12]. Artut believes that generative art can be combined with Futurism Art to explore more aesthetic constituents and procedural methodologies for generative art [13]. In addition, Kim and Choi discuss how generative art can evaluate fashion designs and produce creative ideas based on shape analysis of previous designs [14]. Otherwise, Mayahi and Vidrih examine the possible ramifications of generative AI on visual content marketing, with a focus on how it could revolutionize the industry [15]. AI-based generative art can be used to create personalized and engaging visual content at scale, as well as the potential ethical implications and limitations. Generative art through Vector Quantized Generative Adversarial Network (VQGAN) and Contrastive Language-Image Pre-Training (CLIP) models have become a popular tool for designers seeking to create alternative layout designs [16]. These models use deep learning methods to generate images based on textual prompts, allowing designers to explore new and unique visual possibilities. With VQGAN, the model is trained to generate images based on patterns and textures found in a dataset of images [17]. Conversely, CLIP uses natural language processing to understand the content of an image and generate similar images based on textual prompts [18]. Using these models, designers can experiment with new design concepts and generate novel layouts that would be difficult to achieve through traditional design methods. The possibilities for creative expression through these generative art techniques are seemingly endless and offer an exciting new avenue for designers to explore.

The combination of VQGAN and CLIP enables designers to create original graphics based on textual cues, giving them greater flexibility and creativity over the design process [19]. This allows designers to experiment with new design concepts and develop layouts that are thematically tied to the content of their poster, resulting in a more unified and visually appealing end product. Deep learning models such as VQGAN and CLIP enable designers to work more quickly and efficiently since the models can instantly provide a large range of visual possibilities based on textual suggestions [20]. This means that designers can experiment with numerous layout alternatives and make adjustments more quickly and easily than traditional design processes would allow. This study intends to investigate the usage of generative art as a technique for producing visual poster designs with layout diversity using VQGAN and CLIP. It intends to show the potential of generative art as a tool for experimenting with layout variation in visual design, providing designers with a new method to generate dynamic and appealing visual designs.

2. Method

2.1 Network Architecture

VQGAN learns vector quantization in latent space via a transformer. In Fig. 1, $\tilde{z}$ is the distribution parameters of image patches. The transformer estimates the probability of the next picture patch given the previous image patch using these distribution parameters, hence the name quantization. The result z_q is the latent variable used by the decoder to recreate the fake image. If the fake image suggests that the image is artificial, the discriminator will notice it.

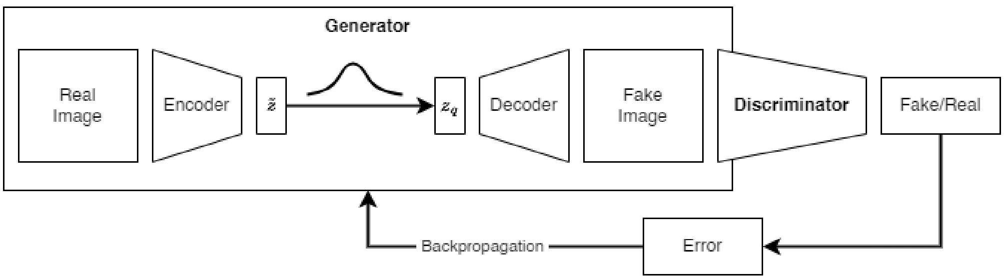


Fig. 1. The network architecture of Vector Quantized Generative Adversarial Network (VQGAN).

CLIP task is to calculate image similarity to a given text prompt. 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 contain a particular object (Fig. 2). 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.

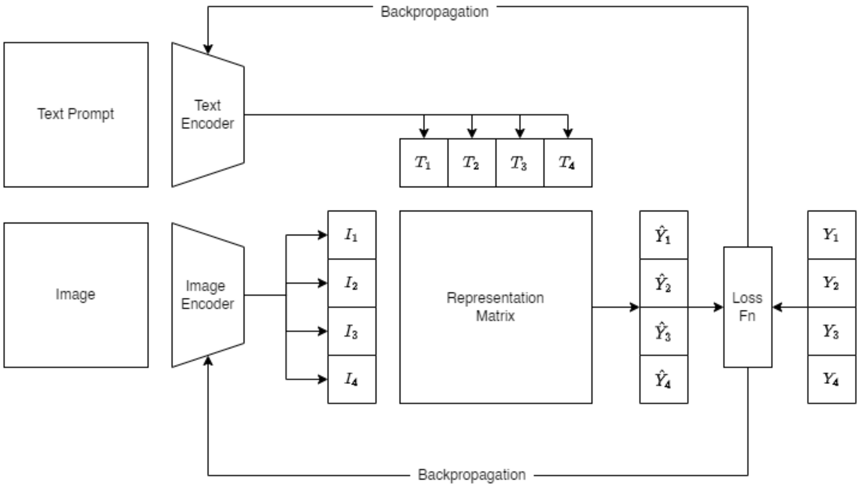


Fig. 2. The network architecture of Contrastive Language-Image Pre-Training (CLIP).

By combining VQGAN with CLIP, we may utilize VQGAN to generate an image and CLIP to assess the resemblance of the resulting image to a text prompt. VQGAN is the generator in VQGAN and CLIP, while CLIP is the discriminator. As a result, the generator should learn how to generate an image that is highly comparable to the text prompt.

2.2. Parameter Setting

Furthermore, the prompt used to generate landscape architectural images was *"[a poster design with scorpion fish made by origami, watercolor style, perspective view, cinematic photo, highly detailed, cinematic lighting, ultra-detailed, ultrarealistic, photorealism, octane render]'* with image size 512x512 pixels, ten frames per second. While the parameters in the generative images use 0.95 zooming modes, shifting the y-axis to 5 pixels. The generative deep learning was performed using Google Colaboratory. The Google Colaboratory is a machine learning and deep learning platform based on the Jupyter Notebook [21].

2.3. Evaluation

In this study, we used the inception score to evaluate generative deep learning with VQGAN and CLIP image quality. The inception score is a measure of the quality of generated images and is based on the idea that high-quality generated images should be both diverse and visually appealing and one common method to validate and evaluate generative deep learning [22]. In order to assess the quality of the model, a set of images should be generated using the VQGAN and CLIP models. These images can be generated based on a set of textual prompts or by randomly generating images using the model. Once the set of images has been developed, the inception score can be calculated. This involves using a pre-trained Inception model to classify the generated images and calculate a score based on the quality of the images. The inception score considers both the diversity and quality of the generated images, providing a comprehensive evaluation of the model's performance. The Inception score provides a numerical value that can be used to compare the performance of different VQGAN and CLIP models. A higher Inception score indicates that the generated images are of higher quality and more diverse, while a lower score suggests that the model may need improvement.

3. Results and Discussion

The generative art system allows for the automatic generation of layout variations, resulting in a wider range of design options and promoting creativity and diversity in the resulting designs. Using the generative art system can save time and effort for the designer by automating the generation of layout variations, allowing for faster design iteration and exploration (Fig.3). The study found that it took approximately 10.108±0.226 seconds to generate one frame using these models, which is significantly faster than the conventional method for creating layouts on poster designs (Table 1). This is a promising development because designers and artists can create high-quality layouts for their posters much more quickly than before [23]. These models can suggest design ideas that might not have occurred to the designer, leading to more innovative and original designs. However, these models rely on large amounts of data to generate their designs, so they may not be as effective for designing posters in niche or specialized fields. Using these models may require certain technical expertise that not all designers possess.

Table 1. Time to Generate 1 frame based on each text prompt

ID	Text Prompt	Average (second)	Standard Deviation
1	a poster design with scorpion fish made by origami	10.3430	0.4285
2	a poster design with scorpion fish made by origami, water color style	9.9915	0.1178
3	a poster design with scorpion fish made by origami, water color style, perspective view	10.0423	0.1606
4	a poster design with scorpion fish made by origami, water color style, perspective view, cinematic photo	10.1012	0.1143
5	a poster design with scorpion fish made by origami, water color style, perspective view, cinematic photo, highly detailed	10.0370	0.1502
6	a poster design with scorpion fish made by origami, water color style, perspective view, cinematic photo, highly detailed, cinematic lighting	10.2191	0.1249
7	a poster design with scorpion fish made by origami, water color style, perspective view, cinematic photo, highly detailed, cinematic lighting, ultra-detailed	10.0901	0.1614
8	a poster design with scorpion fish made by origami, water color style, perspective view, cinematic photo, highly detailed, cinematic lighting, ultra-detailed, ultrarealistic	10.1177	0.2827
9	a poster design with scorpion fish made by origami, water color style, perspective view, cinematic photo, highly detailed, cinematic lighting, ultra-detailed, ultrarealistic, photorealism	10.0760	0.1282
10	a poster design with scorpion fish made by origami, water color style, perspective view, cinematic photo, highly detailed, cinematic lighting, ultra-detailed, ultrarealistic, photorealism, octane render	10.0555	0.1348

The quality of the generated designs can vary depending on the quality and diversity of the initial dataset and the specific implementation of the VQGAN and CLIP models. However, in some cases, the system has been shown to generate high-quality designs that are similar in style to the original dataset. Overall, the generated images achieved a 4.247 inception score (Table 2). The inception score is a commonly used metric for evaluating the quality of generated images in deep-learning models, and a score of 4.247 is considered to be quite high [22]. This means that the images produced by the VQGAN and CLIP models are both diverse and realistic (Fig. 4).

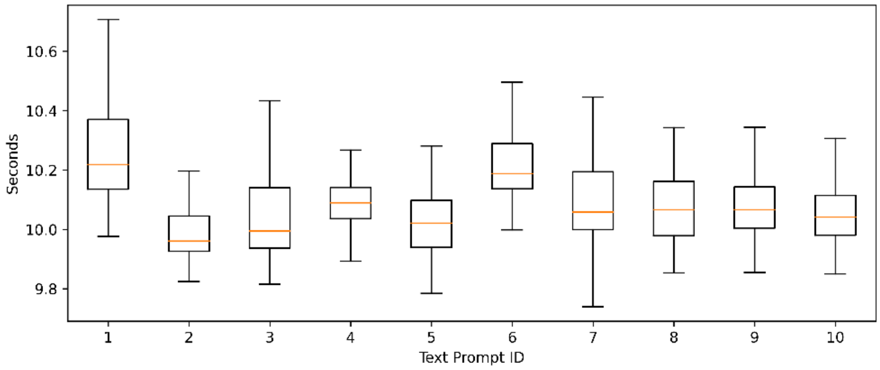


Fig. 3. Comparison of time to generate 1 frame

One of the reasons why VQGAN and CLIP models are so effective at generating high-quality images is that they can learn and understand the structure and patterns of the visual data they are working with [24]. Using a combination of unsupervised learning and supervised learning, the model can identify and extract relevant features from the image data and then use these features to generate new images that are both aesthetically pleasing and true to the underlying data distribution. Another advantage of VQGAN and CLIP models is that they are highly flexible and can be adapted to a wide range of use cases [25]. For example, these models have been used to generate artwork, produce realistic 3D models, and even create new musical compositions. While there are challenges to overcome when working with generative deep learning models like VQGAN and CLIP, the high inception score achieved by the study is a promising sign of what is possible.

Table 2. Time to Generate one frame based on each text prompt

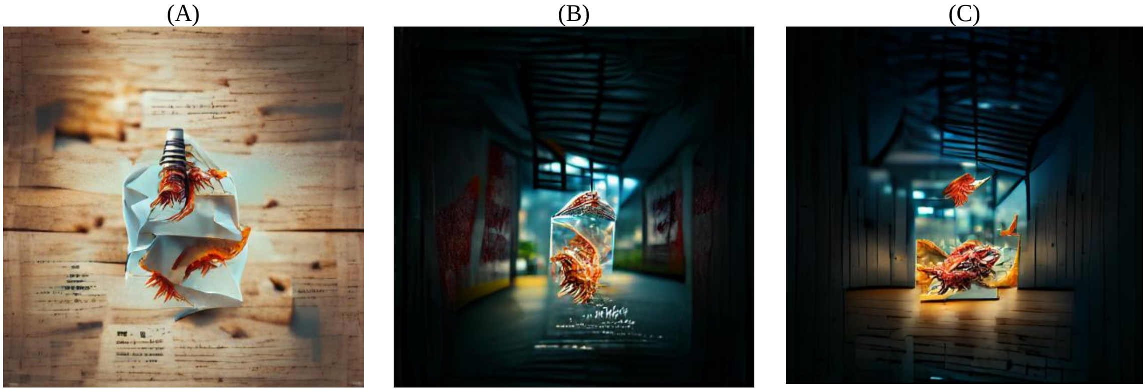
ID	Text Prompt	Inception score
1	a poster design with scorpion fish made by origami	2.3972
2	a poster design with scorpion fish made in origami, watercolor style	2.3085
3	a poster design with scorpion fish made by origami, watercolor style, perspective view	2.2444
4	a poster design with scorpion fish made by origami, watercolor style, perspective view, cinematic photo	2.3706
5	a poster design with scorpion fish made by origami, watercolor style, perspective view, cinematic photo, highly detailed	2.7885
6	a poster design with scorpion fish made by origami, watercolor style, perspective view, cinematic photo, highly detailed, cinematic lighting	2.9329
7	a poster design with scorpion fish made by origami, watercolor style, perspective view, cinematic photo, highly detailed, cinematic lighting, ultra-detailed	2.9091
8	a poster design with scorpion fish made by origami, watercolor style, perspective view, cinematic photo, highly detailed, cinematic lighting, ultra-detailed, ultrarealistic	2.9884
9	a poster design with scorpion fish made by origami, watercolor style, perspective view, cinematic photo, highly detailed, cinematic lighting, ultra-detailed, ultrarealistic, photorealism	2.9037
10	a poster design with scorpion fish made by origami, watercolor style, perspective view, cinematic photo, highly detailed, cinematic lighting, ultra-detailed, ultrarealistic, photorealism, octane render	3.9849

As depicted in Fig 2, the finding that the inception score tends to increase with more complex text prompts is a promising development for generative deep learning. It suggests that these models have the potential to become even more powerful tools for creating high-quality and diverse visual content in a wide range of applications. The more complex text prompts allow the models to capture more subtle relationships between different visual elements in the generated images [26]. This can lead to more nuanced and intricate patterns and designs, leading to higher inception scores.

Fig. 4. Comparison of Inception score for each text prompt

Increased unpredictability in the generated images may impact performance (Fig. 5). By giving the model more explicit and specific cues, it is forced to explore a broader range of visual possibilities, resulting in more diverse and engaging visuals. This, in turn, can result in higher inception scores since the model can capture a broader spectrum of visual elements. Yet, there are several drawbacks to this strategy. One restriction is that the generated images' quality strongly depends on the source dataset's quality and diversity. The generated variations may be limited or repetitive if the dataset is not diverse enough. Additionally, the VQGAN and CLIP model requires significant computing power and may be challenging to use for designers without a background in machine learning [27]. Traditional manual design processes are not replaced by generative deep learning. While technology can save time and encourage innovation, it cannot entirely replace the necessity for human involvement and decision-making in the design process. Also, the approach may not be appropriate for all forms of visual poster design, as some designs may necessitate greater control over specific design aspects. Using generative art in visual poster design utilizing the VQGAN and CLIP model offers various advantages and disadvantages that must be carefully examined. The generative deep learning on poster design is a promising method to encourage originality and diversity in visual design

(A) (B) (C)



iteration: 25, loss: 0.670601 iteration: 139, loss: 0.675493 iteration: 300, loss: 0.676877

Fig. 5. Three samples of generative deep learning using VQGAN and CLIP

4. Conclusion

The generative deep learning employing the VQGAN and CLIP model for exploring layout variation on visual poster design gives a unique and novel way to the design process. The technology enables the automatic development of layout alternatives, fostering creativity and diversity in the resulting designs. While this strategy has some advantages, such as saving time and encouraging creativity with 10.108±0.226 for one generative image, achieving a good quality image which scored 4.248 using an inception score evaluation. Generative deep learning may not be suited for all styles of visual poster design, and it does not eliminate the requirement for human input and decision-making in the design process. Exploring layout variation on visual poster design using generative deep learning using the VQGAN and CLIP model is a viable method to boost creativity and diversity in visual design. This strategy can revolutionize how we approach visual design in the future by offering designers a new approach to developing dynamic and involving visual design.

Acknowledgment

The authors wish to express gratitude to the Head of the Human, Environment, and Technology Research Center of Soegijapranata Catholic University for the support of this study.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.
Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] S. Guo *et al.*, “Vinci: An Intelligent Graphic Design System for Generating Advertising Posters,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–17, doi: 10.1145/3411764.3445117.

[2] M. Guo, D. Huang, and X. Xie, “The Layout Generation Algorithm of Graphic Design Based on Transformer-CVAE,” in *2021 International Conference on Signal Processing and Machine Learning (CONF-SPML)*, 2021, pp. 219–224, doi: 10.1109/CONF-SPML54095.2021.00049.

[3] M. Bakar Findikci, “Pierre Mendell’s Opera Poster Designs from Simile to Metaphor,” *J. Int. Soc. Res.*, vol. 14, no. 82, pp. 93–102, 2021.

[4] P. Ardhianto, “Tinjauan Desain pada Karya Poster; Studi Kasus Seri Poster ‘Marlyn Fish 2009,’” *Andharupa J. Desain Komun. Vis. Multimed.*, vol. 2, no. 01, pp. 15–22, Mar. 2018, doi: 10.33633/andharupa.v2i01.1040.

[5] D. Lopes, J. Correia, and P. Machado, “EvoDesigner: Towards Aiding Creativity in Graphic Design,” in *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)*, Springer, 2022, pp. 162–178. doi: 10.1007/978-3-031-03789-4_11

[6] C. F. Provvidenza, L. R. Hartman, J. Carmichael, and N. Reed, “Does a picture speak louder than words? The role of infographics as a concussion education strategy,” *J. Vis. Commun. Med.*, vol. 42, no. 3, pp. 102–113, Jul. 2019, doi: 10.1080/17453054.2019.1599683.

[7] G. Slingerland, M. Murray, S. Lukosch, J. McCarthy, and F. Brazier, “Participatory Design Going Digital: Challenges and Opportunities for Distributed Place-Making,” *Comput. Support. Coop. Work*, vol. 31, no. 4, pp. 669–700, Dec. 2022, doi: 10.1007/s10606-022-09438-3.

[8] C. Redies, M. Grebenkina, M. Mohseni, A. Kaduhm, and C. Dobel, “Global Image Properties Predict Ratings of Affective Pictures,” *Frontiers in Psychology*, vol. 11. pp. 1–16, 12-May-2020, doi: 10.3389/fpsyg.2020.00953.

[9] C. Dewi, R.-C. Chen, Y.-T. Liu, and H. Yu, “Various Generative Adversarial Networks Model for Synthetic Prohibitory Sign Image Generation,” *Appl. Sci.*, vol. 11, no. 7, pp. 1–15, Mar. 2021, doi: 10.3390/app11072913.

[10] P. Ardhianto *et al.*, “Deep Learning in Left and Right Footprint Image Detection Based on Plantar Pressure,” *Appl. Sci.*, vol. 12, no. 17, pp. 1–13, Sep. 2022, doi: 10.3390/app12178885.

[11] D. Singh, N. Rajcic, S. Colton, and J. McCormack, “Camera Obscurer: Generative Art for Design Inspiration,” in *8th International Conference, EvoMUSART 2019, Held as Part of EvoStar 2019, Leipzig, Germany*, 2019, pp. 51–68, doi: 10.1007/978-3-030-16667-0_4.

[12] F. Nassery and P. Sikorski, “New possibilities of using processing and modern methods of the ‘generative art’ graphics in architecture,” *Technical Transactions/Czasopismo Techniczne*, no. 4. pp. 25–33, 2015.

[13] S. H. Artut, “Futurism art and its significance to computational generative art,” in *XXI Generative Art Conference 2018*, 2018, pp. 338–345.

[14] D. O. Kim and J. H. Choi, “Analysis of Fashion Design Reflected Visual Properties of the Generative Art,” *J. Korean Soc. Cloth. Text.*, vol. 41, no. 05, pp. 825–839, Oct. 2017, doi: 10.5850/JKSCT.2017.41.5.825.

[15] S. Mayahi and M. Vidrih, “The impact of generative ai on the future of visual content marketing,” *arXiv preprint arXiv:2211.12660*. pp. 1–15, 2022.

[16] K. Crowson *et al.*, “VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance,” in *17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVII*, 2022, pp. 88–105, doi: 10.1007/978-3-031-19836-6_6.

[17] J. Yu *et al.*, “Vector-quantized image modeling with improved vqgan,” in *ICLR 2022*, 2021, pp. 1–17, doi: 10.48550/arXiv.2110.04627.

[18] P. Li, “Application of CLIP on Advanced GAN of Zero-Shot Learning,” in *2021 International Conference on Signal Processing and Machine Learning (CONF-SPML)*, 2021, pp. 234–238, doi: 10.1109/CONF-SPML54095.2021.00052.

[19] R. Wang, “Collaborated machine learning-based design solution to conceptual design of architecture,” Cornell Universit, 2022.

[20] V. Liu and L. B. Chilton, “Design Guidelines for Prompt Engineering Text-to-Image Generative Models,” in *CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–23, doi: 10.1145/3491102.3501825.

[21] J. Praveen Gujjar, H. R. Prasanna Kumar, and N. N. Chiplunkar, "Image classification and prediction using transfer learning in colab notebook," *Glob. Transitions Proc.*, vol. 2, no. 2, pp. 382–385, Nov. 2021, doi: 10.1016/j.gltp.2021.08.068.

[22] A. Obukhov and M. Krasnyanskiy, "Quality Assessment Method for GAN Based on Modified Metrics Inception Score and Fréchet Inception Distance," in *Proceedings of the Computational Methods in Systems and Software*, 2020, pp. 102–114, doi: 10.1007/978-3-030-63322-6_8.

[23] S. P. Porkodi, V. Sarada, V. Maik, and K. Gurushankar, "Generic image application using GANs (Generative Adversarial Networks): A Review," *Evolving Systems*. pp. 1–15, 30-Sep-2022, doi: 10.1007/s12530-022-09464-y.

[24] K. Yamamoto and K. Yanai, "Text-based Image Editing for Food Images with CLIP," in *Proceedings of the 7th International Workshop on Multimedia Assisted Dietary Management on Multimedia Assisted Dietary Management*, 2022, pp. 29–37, doi: 10.1145/3552484.3555751.

[25] R. Gal *et al.*, "An image is worth one word: Personalizing text-to-image generation using textual inversion," *arXiv preprint arXiv:2208.01618*. pp. 1–26, 2022.

[26] G. Couairon, J. Verbeek, H. Schwenk, and M. Cord, "Diffedit: Diffusion-based semantic image editing with mask guidance," *arXiv preprint arXiv:2210.11427*. pp. 1–21, 2022.

[27] K. Man and J. Chahl, "A Review of Synthetic Image Data and Its Use in Computer Vision," *J. Imaging*, vol. 8, no. 11, pp. 1–33, Nov. 2022, doi: 10.3390/jimaging8110310.