Human Interaction and Emerging Technologies (IHIET 2023), Vol. 111, 2023, 602–608
https://doi.org/10.54941/ahfe1004067

Machine Vision, Prompts and Neural Network Structure in Art: Reverse Engineering in Image Generation

Man Lai-Man Tin

School of Arts and Social Sciences, Hong Kong Metropolitan University, Ho Man Tin, Kowloon, Hong Kong

ABSTRACT

The paper proposes the use of reverse engineering and artistic creation perspective to explore the convergence of machine vision and neural networks for generating images that closely resemble the original objects. With the rapid and significant development of machine learning and neural networks, these technologies have garnered significant attention in recent years due to their image generation and prompt text functions, which offer new possibilities. Rather than utilizing these technologies solely for creating new works, it is important to investigate their potential for generating images that resemble the original objects. By converging different technologies in the era of machine intelligence, it is possible to achieve greater flexibility and adaptability in image generation processes, leading to a wider range of potential outcomes. It is hoped that it can shed light on the possibility of generating images in particular using artistic approaches.

Keywords: Artificial intelligence, Creative AI, Image generation, Machine vision, Neural network, Reverse engineering

INTRODUCTION

Machine vision involves the integration of computer algorithms, machine hardware such as cameras and sensors, and sometimes artificial intelligence (AI) to perceive, inspect, analyse and identify the environment and objects, and perform tasks as per the setting and instruction, making this technology an important tool for machine vision orientated creative method. On the other hand, with the rapid development of artificial intelligence and deep learning technologies, the potential of using AI in image generation and artistic creation has drawn high attention lately. Among all, the use of prompts and neural networks in text-to-image, and image-to-image models aroused great responses. Everyone creates images and becomes 'creators'.

Given this situation and phenomenon, it is important to further study the possibility of using the technologies in image generation, in particular from the domain of art in order to generate a deeper and more meaningful discussion, as Nuria Rodríguez-Ortega has highlighted 'the analysis based on

the computational processing of images should be considered as a theoretical construction of what could, and perhaps should, emerge as a new image and visuality theory' (Kathryn Brown, 2020). Instead of using these technologies solely for creating new works, this paper aims to explore their potential and capability for generating images that can resemble the original objects through a reverse engineering approach. Some researchers have attempted to investigate reverse engineering in image generation by using techniques such as image-to-text and text-to-image to test the ability of artificial intelligence to generate images that are visually similar to the original ones. However, there has been limited research on applying reverse engineering to image creation and the comparison of images created by machine vision, making this area of study essential in this regard, in particular, to pave the way for future developments in how technologies can be used to restore original objects based on artistically manipulated images with a variety of visual presentations.

MACHINE VISION IN STILL LIFE DRAWING

Still life drawing is one of the traditional genres in art, which refers to the depiction of still objects such as fruits, flowers, and man-made things. The process requires not only rendering the objects through drawing, but also the artists' unique vision. It involves the observation and attention to details such as light and shadow, color, shape, perspective, shape and size, etc. and then making use of drawing techniques and expressive means to capture them on paper. As coined by Hui Yuk 'The understanding of technology is no longer a matter of a cultural critique of technology.' (Hui Yuk, 2016) This research attempts to include technology in art and vision and make use of machine vision technology and computer algorithms to perform still life drawing, which will be used for the next stage to test the prompt text and image generation functions.

In order to access the capability of the machine to do still life drawing, and increase the complexity and difficulty for prompt text and image generation to create new results, an artistic style was incorporated into the machine drawing in this paper. An open-source Max program 'Abstract image painter', a visual programming language commonly utilized in artistic creation and performance, was employed as the basis for the machine to do the drawing. The Max patch used was adjusted and modified as the drawing algorithm, upon visual presentation, it appears to resemble an impressionism expression visually.

Inspired by Andy Warhol, a prominent American visual artist and pop art master, the composition of still objects for this research was composed of a Campbell's soup can and three tomatoes arranged on a table covered with cloth (see Figure 1), a simple imitation of traditional still life drawing scene with the man-made object. The setting was composed of a computer connected to an external camera. The scene was observed through the camera and turned into the signal for the computer and algorithm to perform still life drawings (see Figure 2).

Figure 1: Composition of still objects for still life drawing through machine vision and algorithm.

Figure 2: Progress of still life drawings by machine vision and algorithm.

PROMPTS AND NEURAL NETWORK STRUCTURE

In recent years, with the advancements in machine learning technology, text-to-image and image-to-text models are increasingly utilized particularly in generative art. The proficiency and capability of AI-generated art with prompt engineering have been proven, and more studies on the capability of text-to-image and image-to-text tools were conducted, for example, Jonas Pppenlaender in his paper 'Prompt Engineering for Text-Based Generative Art' (2022) investigated the opportunities to apply text-based generative art in broader implications for creative work and interaction of humans with computers beyond the use case of text-based generative art, and the guidelines designed to assist better outcomes from text-to-image generative models 'Design Guidelines for Prompt Engineering Text-to-Image Generative Models' as per the research of Vivian Liu and Lydia B. Chilton (2021). These studies pave the way to exploring the possibilities for applying and improving computer analysis ability, prompt engineering and natural language processing (NLP) in text-based generative models. In this paper, after the formation of still life drawing by machine vision and algorithm, the utilisation of the neural network at this stage focuses on image-to-text prompt generation, in order to obtain text prompts for image generation in the next stage.

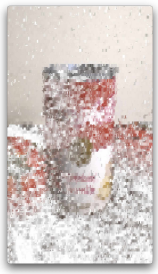
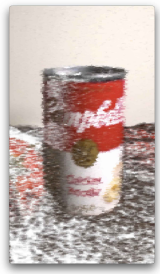
The text prompts were generated via Replicate, an open machine learning API, to match the given still life drawings. Three different stages of the still life drawings were selected throughout the drawing process to test the output of text prompts (see Table 1) for the next stage of image generation.

Table 1. Generation of text prompts.

Input - Still life drawings			
Output - **Text prompts generated**	a can of soda sitting on top of a table, a stock photo by Warhol, pixabay contest winner, net art, stockphoto, contest winner, 8k resolution	a can of soda sitting on top of a table, a raytraced image by Warhol, polycount, net art, anaglyph filter, ray tracing, 8k resolution	a can of soda sitting on top of a table, a raytraced image by Warhol, polycount, net art, anaglyph filter, physically based rendering, ray tracing

The three text prompts were analysed (see Table 2). It is observed that there are variations in text prompts based on different progress of still life drawing, which demonstrates the ability of the machine learning API to predict and generate texts based on the relevant contents of depiction.

Table 2. Matching of generation of text prompts to machine vision's still life drawings.

Still life drawings / Text prompts			
a can of soda sitting on top of a table	✓	✓	✓
a stock photo by Warhol	✓		
a raytraced image by Warhol		✓	✓
pixabay contest winner	✓		
polycount		✓	✓
net art	✓	✓	✓
anaglyph filter		✓	✓
stockphoto	✓		
contest winner	✓		
ray tracing		✓	✓
8k resolution	✓	✓	
physically based rendering			✓

The generated text prompts of the second and third images are similar, it is expected that the generated images in the next stage will be similar too, while the text prompts for the 1st image are, to a certain extent, quite different from the others.

REVERSE ENGINEERING IN IMAGE GENERATION

The text prompts obtained were used to generate images through a Latent Diffusion Model. It is proven the capability of enabling the production of high-resolution images from general conditional inputs, such as text, by converting diffusion models into generators, for example, the research about 'High-Resolution Image Synthesis with Latent Diffusion Models' by Robin Rombach et al. (2022) This paper aims to employ the text-to-image generation diffusion-based model to restore the original images of the still life drawing generated by machine vision based on the text prompts produced by the same still life drawing. The diffusion model used in this paper is Stable Diffusion v2-1 Model Card. The three text prompts were submitted to the stable diffusion model to perform latent diffusion in the latent space based on the encoder in pixel space, and tried to investigate the possibility of generating images that can resemble the original objects through a reverse engineering approach.

Three results were obtained in response to relevant text prompts generated based on the relevant still life drawings. A table comparing the drawings, text prompts, and images is presented below (see Table 3).

Table 3. Comparison of still life drawings, text prompts, and the results of the generated images.

Still life drawings			
Text prompts	a can of soda sitting on top of a table, a stock photo by Warhol, pixabay contest winner, net art, stockphoto, contest winner, 8k resolution	a can of soda sitting on top of a table, a raytraced image by Warhol, polycount, net art, anaglyph filter, ray tracing, 8k resolution	a can of soda sitting on top of a table, a raytraced image by Warhol, polycount, net art, anaglyph filter, physically based rendering, ray tracing
Generated images			

Although the primary object, a can of soda, is presented in all three images, it has been noted that the outcomes generated by the stable diffusion model are distinct from one another in terms of the scenery and details. The first generated image is characterized by an obvious setting, featuring a colorful background and a table or ground with wood texture. The second and third images generated look more similar, which echoes the analysis and prediction in the previous stage of this paper. The second image with a can in the middle of the picture bears a closer resemblance to Andy Warhol's Campbell's soup art compared with the first image. The style of the pull ring on the top of the can and the bright packaging color are reminiscent of soda drinks. The third image with a bigger can place in the center, displays the greatest similarity to Campbell's soup, such as the style of the pull ring and the can packaging, in contrast to the first and second images. The background of the second and third images are in grey color which perfectly matches the original setting of the still objects (see Fig. 1). Compared with the third image, the grey color in the second image appears duller. All images do not have elements and traces of the depiction of tomatoes.

Overall, the third image displays the greatest similarity to the still life drawing and closely resembles the original objects in contrast to the first and second images (see Figure 3).

Figure 3: Comparison of original still objects, still life drawing, and text-to-image generated image (left to right).

CONCLUSION

Through machine vision and algorithm, image-to-text and text-to-image machine learning technologies, the potential and capability for generating images that can resemble the original objects through a reverse engineering approach has been investigated. It has been noted that the outcome can generally resemble the primary object, Campbell's soup, which appears in different images. The level of image resemblance is satisfactory, and demonstrates the potential to use technologies to restore original objects with artistic manipulation or various visual presentation. Perhaps this result can also generate a deeper discussion about how 'unsupervised machine learning can expose limitations of such categories and suggest new ways of seeing culture' (Lev Manovich, 2019).

Despite the significant progress made in the paper, certain factors that influence the results still need to be investigated and addressed. These include the meticulousness of the still life drawing, the artistic expression used, the precision of text prompts generated, also the level of comprehension of the diffusion-based model.

REFERENCES

Brown, K. (2020). The routledge companion to digital humanities and art history. Routledge.

Hui, Y. (2016). On the existence of digital objects. University of Minnesota Press.

Liu, V., & Chilton, L. B. (2021). Design Guidelines for Prompt Engineering Text-to-Image Generative Models (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2109.06977

Manovich, L. (2019). AI aesthetics. Moscow: Strelka Press.

Oppenlaender, J. (2022). A taxonomy of prompt modifiers for text-to-image generation (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2204.13988

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-Resolution Image Synthesis with Latent Diffusion Models (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2112.10752

YouTube. (2020, April 14). Abstract image painter -- max MSP tutorial. YouTube. https://www.youtube.com/watch?v=BUhVNIOJTLU&t=14s