
Generative Adversarial Network Algorithms in Art: Machine Vision in Generation of Collage Art

Man Lai-Man Tin

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

ABSTRACT

The paper proposes artistic and computational approaches to investigate the ability of machine learning to synthesize and manipulate the image dataset into artwork creation. By using the Generative Adversarial Network (GAN), it is observed how the machine algorithms are able to learn artistic styles and manipulate relevant pictures to generate digital artifacts, in particular, the images generated through latent space interpolation. Referring to the artwork of Pablo Picasso, the paper also aims at observing the image manipulation being generated by GAN in order to understand and compare machine vision with human vision through the means of collage and artwork creation. And finally, to explore the process of seeing based on the phenomenology of embodiment, trying to understand how the objects could be visible to us through the machine and artificial intelligence without being 'bodily involved in the world'.

Keywords: Artificial intelligence, Creative AI, Machine learning, Generative adversarial network, GAN, Image generation, Collage

INTRODUCTION

The use of AI and machine learning technology in creative tasks is one of the domains of art and technology. The application of AI as an art tool for artists to create artwork and explore creative practice leads to a new perspective on art and creativity - automation, augmentation, and machine-oriented creation. Given the nature of intelligence and learning ability of AI, the Creative AI is no longer the same as the conventional art tools, artists to a certain extent are working with data throughout the training and manipulation process of machine learning, and tend to aim at the artistic perspective in the selection of outcomes. Machine learning in image manipulation has been widely used and applied, the ability to further consolidate and assemble the images in composing new work is examined in this paper. Among the art creations, collage is an art form that requires the technique of reassembling pieces of elements to create a new whole, given the complication of interpreting and manipulating the various elements in collage creation, it is, therefore, a good attempt in testing the ability of machine in generating collage, trying to explore the possibility of performing more complicated creation in terms of visual and element manipulation, combination,

composition, etc. With reference to the 20th-century regular collage artwork, for example, the work created by Pablo Picasso, the new whole creation by machine learning is juxtaposed with the collage of Pablo Picasso, trying to further investigate the nature and mechanism of machine vision in comparison with human vision. The machine-generated collage is further examined based on the theory of phenomenology of perception in order to understand how things could be visible to us in relation to disembodiment under the age of machine intelligence.

DATA-DRIVEN UNCONDITIONAL GENERATIVE IMAGE MODELING

A machine learning framework Generative Adversarial Network (GAN) was used in this paper in order to perform the training between two neural networks - generative and discriminative networks. The GAN architecture allows the neural networks on one hand to constantly generate new data based on the training datasets, and on the other hand, to discriminate the new data from the original ones. This framework allows the machine to perform unsupervised learning and automatic generation of 'realistic' data.

Given that the ability of GAN to generate images that are authentic to human has been proven by various researches, for example, Karras Tero et al. in their research successfully introduced a new and high-quality architecture to generate human faces by using alternative generator architecture for generative adversarial networks (Karras et al., 2019), the next question that we must ask is about the application of machine learning and the image modeling framework to generate art. This paper is in contrast to many of the generation methods and applications that focus on learning a particular artistic style, for example, Leon A. Gatys et al. successfully implemented the idea and technology of merging an artistic style into an image in their research (Gatys et al., 2016). Arbitrary selection of various styles of work as initial datasets in order to allow artificial intelligence to perform unpredictable style creation. Being inspired by Pablo Picasso and his collage 'Bowl with Fruit, Violin, and Wineglass' (1913), the GAN is used to further investigate its ability to generate collages based on the elements related to the artwork title, trying to see whether AI and machine learning can generate similar collage work with the datasets.

In this paper, besides pictures of the artwork 'Bowl with fruit, violin and wineglass', four aspects of pictures, namely bowl of fruit, violin, wineglass and collage, were also used respectively as the datasets for GAN training in the machine learning tool RunwayML and its remote GPU. Pictures of disparate elements were collected from the internet through keyword search on Google, which aim at investigating the ability of GAN algorithms in processing various information, and how machine learning can synthesize and produce new combinations under data-driven unconditional generative image modeling conditions.

GENERATION OF IMAGES AND COLLAGES

'Bowl with fruit, violin and wineglass' was used as the keyword to collect relevant pictures from the internet (see Figure 1). These pictures served as



Figure 1: Some of the pictures related to ‘Bowl with fruit, violin and wineglass’ downloaded from the internet.



Figure 2: Images generated by GAN based on the training of the dataset in Figure 1.

the first stage of GAN training to synthesize the disparate visual information to generate new sets of pictures with relevant visual elements. It is observed that some of the generated images (see Figure 2) are of high imitation of the original training dataset in terms of the visual elements, presentation, composition, etc., in particular, those with obvious visual composition, form and color such as a bowl with fruit from the training dataset. While the violin is comparatively difficult to train and being generated in the outcomes at this stage, which shows the need for further training in order to obtain work with comprehensive elements.

To allow GAN to continue to perform data-driven unconditional training and generate new datasets through the learning process, ‘Bowl with Fruit’, ‘Violin’, and ‘Wineglass’ related pictures were collected respectively from the internet and served as the datasets for training GAN in the second stage. Respective elements can be found in the new images (see Figure 3), showing the learning progress of AI and its ability for image manipulation and combination.

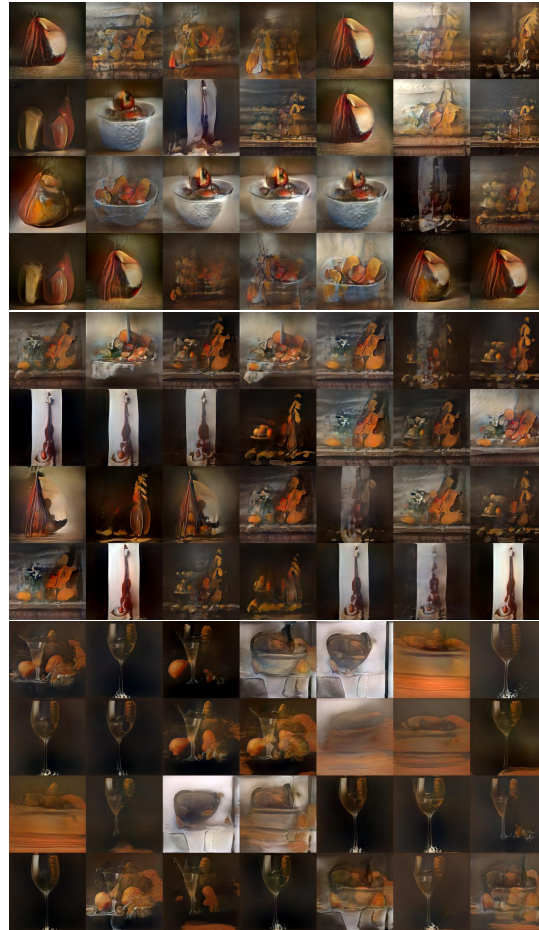


Figure 3: Continue data-driven unconditional training on 'Bowl with Fruit', 'Violin', and 'Wineglass' related pictures based on the first stage of training as presented in Figure 2.

The next attempt is to train the GAN about the style of collage in terms of how disparate and fragmented subject matters can be combined together in a particular form. Referring to Picasso's collage 'Bowl with Fruit, Violin, and Wineglass' (1913), the collage method encourages new images through fragmented cutting and collided combination. 'Collage by Picasso' related pictures were collected through the keyword search on the internet for GAN to continue the training. Results show that AI is able to learn and present its understanding of collage and visual combination (see Figure 4). However - most of the details of the subject matters previously trained are replaced by the new visual elements, the pictures look abstract to spectators, the anticipation of a similar visual presentation to Picasso's 'Bowl with Fruit, Violin, and Wineglass' cannot be achieved.

MACHINE VISION VS HUMAN VISION

Based on the results as shown in Figure 4, the GAN was instructed to perform the training again by applying the datasets generated in the second stage



Figure 4: Third training on ‘Collage by Picasso’ related pictures based on the second stage of training as presented in Figure 3.

of training to the third stage of training in order to explore the possibility of obtaining different collage results with the details of the subject matters remain, so as to make sense of the generated images to the artist (author) and spectators.

The latent space interpolation images generated throughout the training reveal the learning process and visualize how the machine can synthesize and manipulate data through automation. This makes AI an unconventional tool for art creation - on one hand minimizing the intervention of human in the image manipulation and combination, i.e. towards machine-oriented creation, on the other hand, augmenting and expanding artists’ perspective about art creation through datasets selection, training instruction, as well as outcomes selection.

It is observed that the violin elements can be presented visually in the fourth training, while the elements of the bowl with fruit and wineglass previously trained can generally be maintained. This result presents the possible direction of training that all the subject matters can be presented in the picture with AI’s new understanding and method to combine and present the visual elements. The latent space interpolation images present various results of generation throughout the training. To balance the presentation of elements of the bowl with fruit, violin and wineglass as a new whole, a careful selection of work among the generated pictures that can keep the originality without being too similar to the original datasets, and at the same time, to keep some details of subject matters is needed based on the artistic perspective of the artist (author). A picture was selected from the GAN training to represent the image manipulation generated by machine learning. By comparing it with Picasso’s collage ‘Bowl with Fruit, Violin, and Wineglass’ (see Figure 6), we can shed light on the difference between machine vision and human vision.

The visual presentation of the two collages is very different in terms of the use of visual elements, composition, way of combination, etc. It involves



Figure 5: Fourth training on 'Bowl with Fruit, Violin and Wineglass' related pictures based on the second stage of training as presented in Figure 3.

various factors such as the training dataset selection, number of training steps, training instruction, result selection, etc. which need further investigation in another research. The main attempt in this research, however, is to examine the ability of AI and machine learning in synthesizing and combining various information and subject matters to generate a new whole in unsupervised and automated conditions. The new image generated by GAN as presented in Figure 6 can be perceived as a kind of digital collage (and to a certain extent with the feature of photomanipulation) compare with the regular collage, it is somehow between traditional still life and unconventional visual expression which allows us to echo Lev Manovich's idea about '20th century visual art styles tolerate more randomness and less precision' when he investigated AI creations (Manovic, 2018), and reveals the characteristic and ability of machine vision of 'come up with something new, surprising and of value' (Sautoy, 2019).

PHENOMENOLOGY OF PERCEPTION IN AI

Phenomenology is widely used to interpret artificial intelligence and its possibility of development, the result in this research is interpreted in two aspects of contemporary philosophy of art, according to the theory of Maurice Merleau-Ponty's phenomenology of perception. One aspect is the concept of extension of human perception - Merleau-Ponty spells out that tools as extensions of body perception, the extension of bodily perception to perform data training and image manipulation allows machines and artificial intelligence to become part of the bodily and phenomenal field. Merleau-Ponty holds that perception opens a window to determinate things, perception in



Figure 6: Comparison of 'Bowl with Fruit, Violin and Wineglass' collage generated by GAN in 2022 (left) and work of the same title created by Pablo Picasso in 1913 (right, retrieved from Philadelphia Museum of Art).

this case also opens a window to Creative AI. The digital collage (with the feature of photomanipulation) generated by GAN (see Figure 6) presents the machine vision that can emancipate the limitation of the conventional perspective of image combination and composition made by human, suggesting a new perspective in shaping contemporary art.

On the other hand, the images generated by machine and artificial intelligence is performed by means of disembodied cognition - compared with human, the artificial intelligence is inherently limited in its ability to perceive the environment through the body - the 'bodily being in the world' (Merleau-Ponty, 2012), which is constantly generated in the dimensions of time and space as the 'ability to see is intertwined with my ability to move' (Quigley, 2013) cannot be performed in this case. This situation to a certain extent explains the differences in the visual presentation between the works made by Pablo Picasso and GAN of the same title.

CONCLUSION

Inspired by Pablo Picasso's artwork, this paper demonstrates the ability of GAN in synthesizing and manipulating data through automation to reassemble various elements and create a new whole of digital collage (with the feature of photomanipulation). By comparing the works made by human and GAN, it is observed that the results can be very different in which both human and machine vision were revealed in the works respectively. Based on Merleau-Ponty's Phenomenology of Perception, the extension of bodily perception and disembodied cognition situation of the machine help us shed light on the phenomenon of the differences between human and machine-made art.

REFERENCES

- Gatys, L., Ecker, A. and Bethge, M (2016). A Neural Algorithm of Artistic Style. *Journal of Vision*, 16(12), p. 326.
- Images of Bowl with fruit, violin and wineglass. (n.d.) Available at: [https://www.google.com/search?q=Bowl\protect\\$\relax+\\\$with\protect\\$\relax+\\\$fruit,\protect\\$\relax+\\\$violin\protect\\$\relax+\\\$and\protect\\$\relax+\\\$wineglass&source=lnms&tbm=isch&sa=X&ved=2ahUKEwiMqKeCp4v3AhVIQPUHHUISBUIQ_AUoAXoECAEQAw&biw=1792&bih=854&dpr=2](https://www.google.com/search?q=Bowl\protect$\relax+\$with\protect$\relax+\$fruit,\protect$\relax+\$violin\protect$\relax+\$and\protect$\relax+\$wineglass&source=lnms&tbm=isch&sa=X&ved=2ahUKEwiMqKeCp4v3AhVIQPUHHUISBUIQ_AUoAXoECAEQAw&biw=1792&bih=854&dpr=2) (Accessed 16 March 2022).
- Karras, T., Laine, S. and Aila, T (2019). “A Style-Based Generator Architecture for Generative Adversarial Networks”, in: *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 43, No. 12.
- Manovich, L (2018) *AI Aesthetics*. Strelka Press.
- Merleau-Ponty, M (2012). *Phenomenology of Perception*. London and New York: Routledge.
- Picasso, P. R. y (1913). *Bowl with Fruit, Violin and Wineglass*, Philadelphia Museum of Art. Available at: <https://philamuseum.org/collection/object/53855>. (Accessed March 16, 2022).
- Quigley, T (2013, May 5). Review of Summary: Maurice Merleau-Ponty, “Eye and Mind” (1964).
- Sautoy, D. M (2019). *The Creativity Code: How AI is Learning to Write, Paint and Think*. London: 4th Estate.