Transformation of Landscape into Artistic and Cultural Video Using AI for Future Car

Mai Cong Hung [1], Mai Xuan Trang [2], Naoko Tosa [1], Ryohei Nakatsu [1]

[1] Kyoto University, Kyoto, 606-8501 Japan
[2] Faculty of Computer Science, Phenikaa University, Ha Noi 12116, Viet Nam
hungmcuet@gmail.com, trang.maixuan@phenikaa-uni.edu.vn,
tosa.naoko.5c@kyoto-u.ac.jp, ryohei.nakatsu@design.kyoto-u.ac.jp

Abstract. When autonomous driving comes to reality in near future, a space in a car would become a special space where people can enjoy various kinds of entertainment. This paper proposes one such entertainment in which the outside landscape is to be changed into moving art. Firstly, the outside landscape is captured by a camera set near to a driver's head. Then each frame of the captured landscape video is transformed into an artistic image by utilizing the image transformation capability of CycleGAN, one of the recent AI technologies. And the converted video is projected on the windshield. Landscape seen from side windows can be changed into artistic moving images in the same way. In this way, passengers, as well as a driver of a future car, can enjoy artistic moving images that are converted from outside landscape into art video with any art style.

Keywords: CycleGAN, Art style, Transformation of landscape, Autonomous driving.

1 Introduction

What would be an interior space of a future car that will adopt autonomous driving capability? If autonomous driving becomes possible, the interior space of a car will not be a simple space where passengers have to stay during their traveling time, but space where the time during driving can be spent for entertainment [1][2][3]. At that time, the windshield and side glass will be transformed into a screen for entertainment. What kind of content would be appropriate to be shown on that screen? Of course, showing a movie is an idea. Another idea is to use the scenery outside the car as a material for entertainment. At present, viewing the scenery outside the vehicle is what we experience while traveling by car. Driving while enjoying the beautiful scenery is a great pleasure when traveling by car. However, when the same landscape continues for a long time, or when driving in a trashed urban space, it is tiring to see the scenery outside the vehicle.

In such a case, is it possible to transform the scenery outside the car into something else? Art is a good object to convert. Art has the power to deeply appeal to people by

relaxing and sometimes straining their hearts. Is it possible to use this power of art to transform the mobile space into an art space?
In this paper, we propose a method to meet such demands by using the latest AI technology, show the concrete method, and show several examples of converting the landscape into art. Section 2 proposes the concept of our proposal including its merit and problem. Section 3 briefly describes CycleGAN, which is one of recent AI technologies called GANs and which play an important role to convert scenery into artistic videos. Section 4 describes our experiment to convert scenery videos shot while driving into artistic videos including several target artworks we have selected. Section 5 shows some of the results we have obtained and describes some discussions. As there are several problems with the obtained results, Section 6 proposes an improvement of our method described in Section 4 and describes the improved results. Finally, Section 7 concludes the paper.

2 Concept

The basic method proposed in this paper is to transform the scenery outside the car into an art-like video in real-time. GANs (Generative Adversarial Networks), which is an AI technology that has been attracting attention recently, is used to convert landscapes into art videos. The details of the method will be described later.

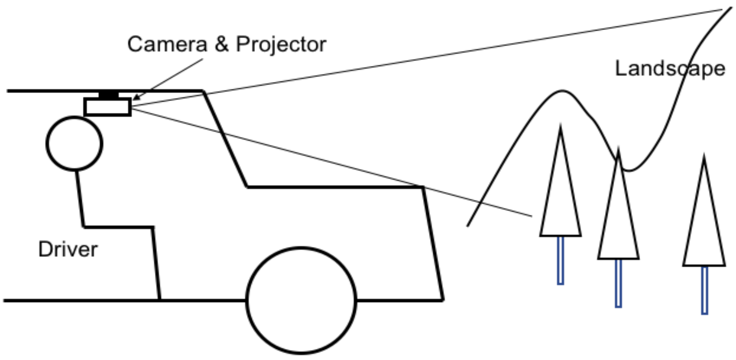

Fig. 1. System structure.

There are levels 1 to 5 for autonomous driving [1][2][3]. Level 5 is for fully autonomous driving, but it takes a considerable amount of time to realize it. For the moment, the realization of levels 2 and 3, in which autonomous driving is restricted in a free-way, etc., is being tried. In this case, a person in charge of driving will be called a driver, because the person needs to intervene in driving if necessary. Fig. 1 shows a system structure that can convert a driver's visible landscape into an art video. This can be applied not only to level 5 which is completely autonomous driving but also to levels 3 and 4 where the driver needs to intervene in driving. However, applying this system to levels 3 and 4 need revision of the Road Traffic Law, so we are currently proposing the system as a concept.

As shown in Fig. 1, a camera and a projector are installed in the vehicle near the driver's head. The video captured by the camera is converted into an art video, which is projected on the windshield by the projector. The windshield is transparent and at the same time has the capability that images can be projected with a projector [4]. The projection on the windshield in this way is an art video of the outside scenery seen from the driver's viewpoint. The features of this method are as follows.
1) A scenery outside a car that a driver sees during driving is converted into an art video. The driver not only appreciates it as an art but also can drive the car in the space expressed in the art video. The driver and other passengers can feel as if they are moving in the art space.
2) Except for the driver, the passengers do not see the converted landscape from their viewpoint. But it is considered that there is a little problem because the difference in viewpoint is not so large.
3) It is possible to select any art style for converting an actual landscape into an art video. For example, it is possible to change the natural scenery outside the car into a Monet style or Cezanne style art video. On the other hand, what kind of effect can be obtained for drivers and passengers by making the outside scenery into an abstract art style, and what kind of abstract art style is suitable for this system are interesting research issues.
4) Further, it is possible to gradually convert the actual landscape into the art video by blending them. In current cars, it is possible to change the air conditioning temperature and air volume at several levels. It is possible to select the blending level, in the same way, depending on the driver or passenger's preference.

3 CycleGAN for Art Style Transfer

The task of converting scenery outside the car into an art video is connected to style transfer topics in AI. There are numerous methods in Deep Learning/Machine Learning [5] dealing with the problem of art style transfer. In this research, we use the celebrated method named CycleGAN, a variation of GANs.

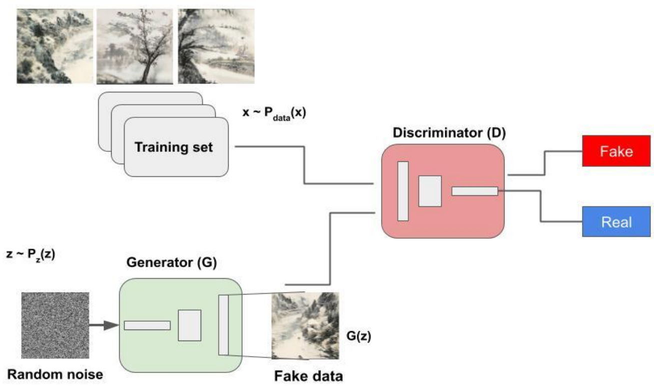

Fig. 2. The basic configuration of GANs

GANs (Generative Adversarial Networks) is a Deep Learning structure including both generative models and discriminative models in a deep neural network [6]. The name "generative" and "discriminative" came from the fact that generative models can generate new data instances while discriminative models can discriminate between different categories of data. In art style transfer, we would expect a generative model to transfer a photo or video frame into a specific style. The advantage of GANs is their efficiency while not requiring large amounts of training data.
The architecture of GANs is represented in Fig. 2. In GANs, generator G learns to generate data from random noise while discriminator network D tries to identify the generated data whether it is real or fake. The training process can be interpreted as a zero-sum game between G and D. The training process on G tries to maximize the probability of the generated data to lie on the distribution of target sets and the training process on D tries to minimize it. This minimax mechanism helps the networks to converge even with a relatively small number of training data. By modifying the basic configuration, a large number of GANs variation has been developed.
Among the variations of GANs, CycleGAN is an elegant method to study the style-to-style level of image transformation [7]. We consider the generative models on CycleGAN as the main tool to transfer landscape into artworks. The architecture of a CycleGAN network is illustrated in Fig. 3. We add an inverse transformation G_{BA} of the generator network G_{AB}, which has the data of domain A as input to transform them into elements of domain B. We also use two discriminators D_A and D_B for the domains A and B, respectively. We would measure the difference between A and $\hat{A}$ (the reconstruction data in A by applying G_{AB} then G_{BA}) and the error caused by the difference between B and the domain given by applying G_{AB} to A. The training process would minimize the sum of these two errors. The data generated by G_{AB} and G_{BA} provides the mutual transformation between the two domains.

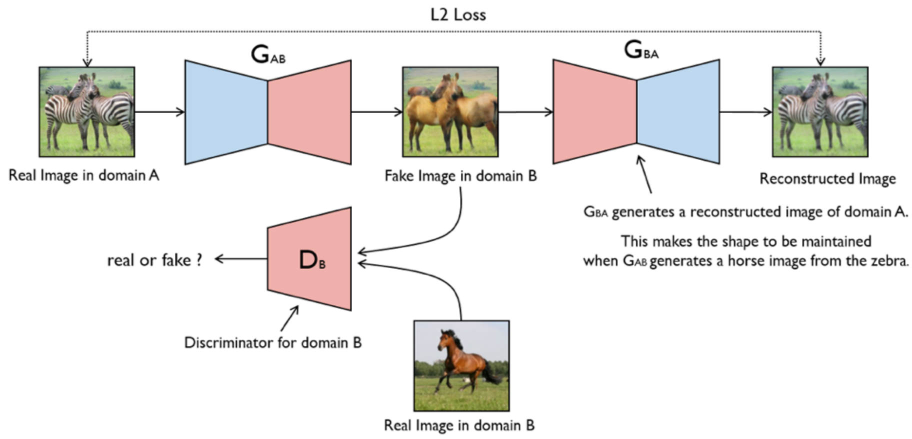

Fig. 3. The basic configuration of CycleGAN ([6])

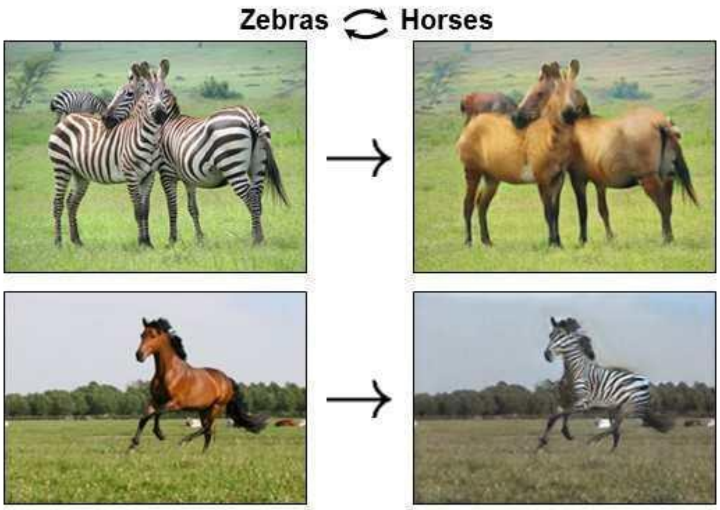

Fig. 4. Horses-Zebras transfer (Image source [7])

The most important difference between generators of GANs and CycleGAN is that GANs learn to generate data to fit in a target set while CycleGAN learns the set-to-set level of transformation. Because of this, CycleGAN could be used to establish mutual conversion between these two groups of images - the key point to do the task of style transfer. Fig. 4 shows how horses are converted into zebras and vice versa. The development of CycleGAN has opened a new way to transform regular images into artworks that have specific art styles, because of the set-to-set level of style transfer instead of image-to-image level transfer. Therefore, we can perform flexible transformations when the outside scenery changes at high speed.

4 Experiment

We used the datasets shown below and make experiments of mutual transformation via CycleGAN between A and B1, B2, B3, B4 respectively. After finishing the training phase, we applied the generator of each experiment to convert landscape video taken in a driving car in real-time (frame by frame).

Dataset A: Road-scene photos mixed by self-taken data and Nuscenes Dataset from APTIV (partly).
Dataset B1: Kandinsky abstract artworks in Wiki Art [8].
Dataset B2: Naoko Tosa's Sound of Ikebana [9][10].
Dataset B3: Sansui artworks in CS231N project "Chinese Painting Generation Using Generative Adversarial Networks" at Stanford University [11].
Dataset B4: Ikebana photos in Flickr.

Fig. 5. Examples of Kandinsky artworks.

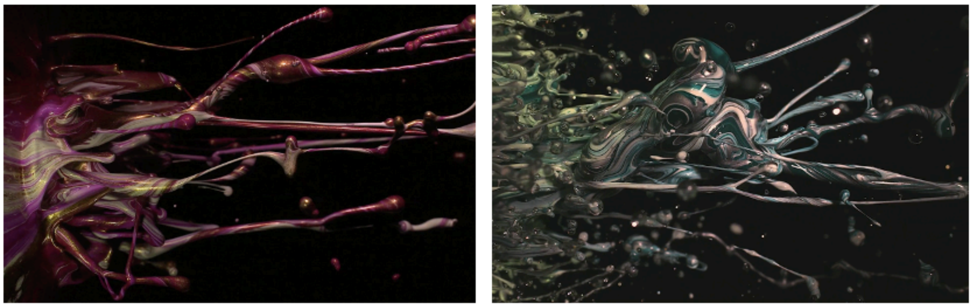

Fig. 6. Examples of Sound of Ikebana images.

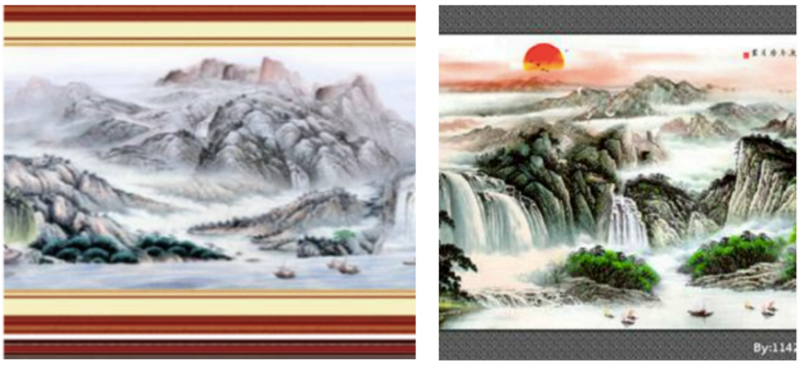

Fig. 7. Examples of Sansui Artworks.

We choose these datasets to transfer the outside scenery into Western abstract art (A-B1), Japanese abstract art (A-B2), traditional Oriental arts (A-B3, A-B4). Fig. 5 shows several of the Kandinsky artworks. Fig. 6 shows several images of the Sound of Ikebana. Fig. 7 shows several Sansui artworks. Fig. 8 shows several Ikebana photos.

Fig. 8. Examples of Ikebana photos.

The Sound of Ikebana was created by Naoko Tosa, one of the authors. It is a video artwork that was created by giving sound vibration to color paints, letting them jump up, and shooting the created form by a high-speed camera. She was inspired by watching various fluid dynamics based phenomena such as “milk crown” and after various experiments succeeded in creating beautiful forms. Although created forms are abstract, many people indicated that they feel Japanese beauty in these abstract forms [9]. Therefore, we think that the Sound of Ikebana would be a good representation of Japanese abstract art.

5 Results and Discussion

After applying the style transfer functions generated by CycleGAN, we get the transformation as shown in Fig. 9. For the original scenes, two different types of scenes were selected; one is a countryside scene, and another is a cityscape scene.

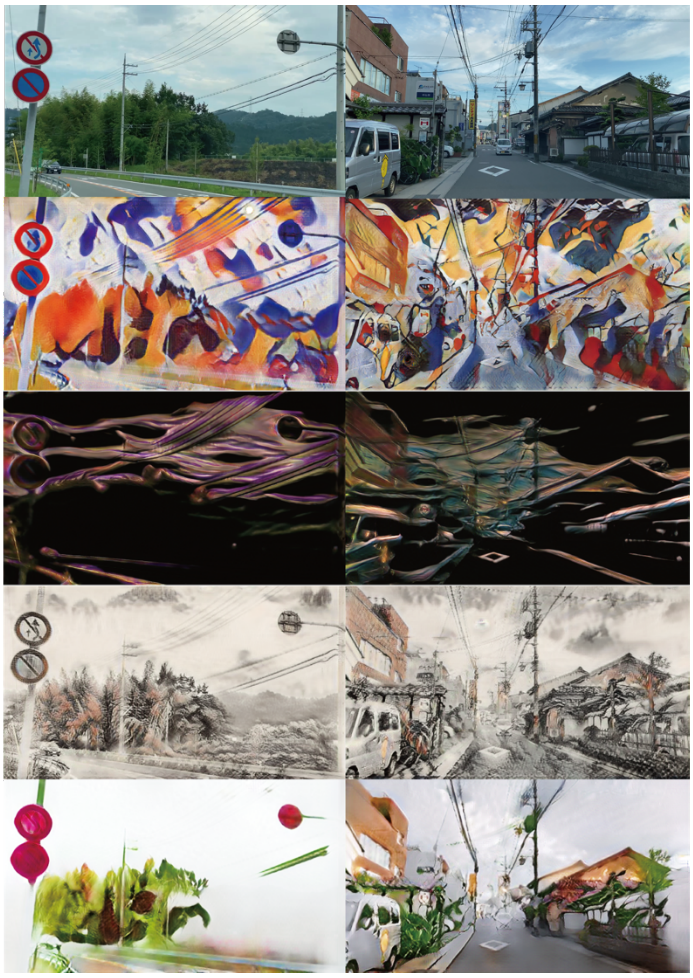

Fig. 9. Transformation of actual scenes into art style scenes (From top to bottom: original scene (A), Kandinsky style (B1), Sound of Ikebana style (B2), Sansui style (B3), Ikebana style (B4))

In our eye-checked evaluation, the transformation A-B1 and A-B2 give the most qualified results in the cityscape scenes while A-B3 and A-B4 give the best results in the countryside scenes. We consider the reason as abstract art is suitable to represent artificial objects such as buildings, traffic lights, etc., and Oriental traditional art was inspired by a natural landscape.

The A to B1, B2, B3, B4 transformation opened a new way to create original video art. CycleGAN is used to work with images of relatively similar sizes, themes, and categories such as horses and zebras, Monet paintings and landscape photos, winter scenes, and summer scenes. But if we perform some "unusual transformation" between relatively different domains of objects (for example between cityscapes and flowers), we could create original art of a high-abstract level.

In the next research, we would improve some limits of the current work. Firstly, the direction of the original training data is vertical or horizontal on a 2D frame but the direction of the front scene while driving is in the last dimension of the 3D frame. Another problem is that sometimes the transition between frames is not smooth enough, we present an improvement for this problem in the next section.

6 DriveGAN: Improvement of the Method

To obtain smooth transitions between frames in the generated video, we try to include frame loss into the total loss of the CycleGAN. The goal is to minimize the difference between the frame transitions in the real video and the generated video. In our experiment, a video captured by a car's camera is extracted to frames that form the source image set (A). The target art image set (B) is described in Section 4.
In the original CycleGAN, the full objective function is defined as follows:

$$\begin{aligned}\mathcal{L}(G_{AB}, G_{BA}, D_A, D_B) = {} & \mathcal{L}_{GAN}(G_{AB}, D_B, A, B) \\ & + \mathcal{L}_{GAN}(G_{BA}, D_A, B, A) \\ & + \lambda\mathcal{L}_{cyc}(G_{AB}, G_{BA}) \end{aligned} \tag{1}$$

where $\mathcal{L}_{GAN}(G_{AB}, D_B, A, B)$ is the adversarial loss for the mapping function (the generator) $G_{AB}: A \to B$ and its discriminator D_B. $\mathcal{L}_{GAN}(G_{BA}, D_A, B, A)$ is the adversarial loss for the mapping function (the generator) $G_{BA}: B \to A$ and its discriminator D_A. $\mathcal{L}_{cyc}(G_{AB}, G_{BA})$ is the cycle consistency loss that captures the forward and backward cycle consistencies: $a \to G_{AB}(a) \to G_{BA}(G_{AB}(a)) \approx a$ and $b \to G_{BA}(b) \to G_{AB}(G_{BA}(b)) \approx b$.

To define the frame loss, we denote the input frame sequence (or source image set) as $A = \{f_{m+it}\}_{i=0}^{N-1}$ where m is the index of the first frame that we want to include in the source images, t is the frame step. t is also a hyperparameter to tune when training our GAN. The frame loss between two sequences of frames is defined as follows:

$$\mathcal{L}_{FRAME}(G_{AB}, G_{BA}) = \mathbb{E}_{a\sim p_A(a)}\left[\left\| \begin{matrix}(\mathrm{G}_{\mathrm{BA}}\left(G_{AB}(f_{m+(i+1)t})\right) - \mathrm{G}_{\mathrm{BA}}(G_{AB}(f_{m+it}))) \\ - (f_{m+(i+1)t} - f_{m+it})\end{matrix} \right\|_1\right]$$

We add this frame loss to the equation (1), the full objective function of our proposed DriveGAN becomes:

$$\begin{aligned}\mathcal{L}(G_{AB}, G_{BA}, D_A, D_B) = {} & \mathcal{L}_{GAN}(G_{AB}, D_B, A, B) \\ & + \mathcal{L}_{GAN}(G_{BA}, D_A, B, A) \\ & + \lambda\mathcal{L}_{cyc}(G_{AB}, G_{BA}) \\ & + \gamma\mathcal{L}_{FRAME}(G_{AB}, G_{BA}) \end{aligned} \tag{2}$$

where γ control the relative importance of the frame loss. The training process would also try to minimize the frame loss to ensure the differences between consecutive frames after transforming would be small if the difference of consecutive frames in the original scenes is small. We replace the training data of A by a collection of video taken by camera to generate sequences of consecutive frames.
Figure. 10 shows the result of the transformation of several continuous landscape images into Kandinsky style images and Sound of Ikebana style images. Based on eye-checked evaluation, the created videos look smooth and could be used as contents that is a transformation of outside landscape into artistic videos.

Fig. 10. Transformation of continuous frame images into art style images (From top to bottom: original scene (A), Kandinsky style (B1), Sound of Ikebana style (B2)

7 Conclusion

A method of transforming outside sceneries seen from the inside of a car into artistic video is proposed. For the transformation, we adopted CycleGAN to transform a set of landscape images into a set of art images, as CycleGAN, one of the latest AI technologies, has the capability of mutual transformation between two image sets.
For source images, we collected various road images (A). As target art images we selected four types of art images; Kandinsky art images (B1), Sound of Ikebana art images created by one of the authors, Naoko Tosa (B2), Sansui art images (B3), and Ikebana photos (B4). By using these training sets and CycleGAN the transformation functions from A to B1, B2, B3, B4 were achieved. Then each frame of several videos taken from a driving car was transformed into each of B1, B2, B3, and B4 styles. And some discussions were carried out regarding the obtained results.
As a next step, we tried to improve the original CycleGAN to delete a flicker observed in the generated video. By including frame loss into the total loss function and by

minimizing the frame loss during the training process, we succeeded in obtaining a smooth transition between frames in the generated video.
For future research, we will carry out a user study to evaluate the feasibility of the proposed method. Also, transformation into several other art styles including anime-style [12] will be carried out.

Reference

1. Self-driving car. Available online: https://en.wikipedia.org/wiki/Self-driving_car (accessed on 10 January 2021),
2. Jesse Levinson, et al., *Towards Fully Autonomous Driving: Systems and Algorithms*. 2011 IEEE Intelligent Vehicles Symposiums, pp.163-168 (2011).
3. Laura Garcia Cuenca, et al., *Machine Learning Techniques for Undertaking Roundabouts in Autonomous Driving*, MDPI Sensors 2019, 19, 2286 (2019).
4. Glass on Web. Available online: https://www.glassonweb.com/news/agc-releases-infover-retm-mirror-enabling-unique-information-display-mirror (accessed on 10 January 2021).
5. Jorn D. Kelleher, *Deep Learning,* MIT Press (2019).
6. Antonia Creswell, et al., *Generative Adversarial Networks: An Overview*, IEEE Signal Processing Magazine, Vol.35, No.1, pp.53-65 (Jan. 2018).
7. Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, *Unpaired image-to-image Translation using Cycle-Consistent Adversarial Networks*, The IEEE International Conference on Computer Vision (ICCV), pp.223-2232 (2017).
8. Wikiart: Wassily Kandinsky. Available online: https://www.wikiart.org/en/wassily-kandinsky (accessed on 10 January 2021)
9. Naoko Tosa, Yunian Pang, Qin Yang, Ryohei Nakatsu, *Pursuit and Expression of Japanese Beauty Using Technology*, Special Issue “The Machine as Artist (for the 21st Century),” Arts journal, MDPI, Vol.8, No.1, 38 (2019).
10. Naoko Tosa, Pang Yunian, Liang Zhao, Ryohei Nakatsu, *Genesis: New Media Art Created as a Visualization of Fluid Dynamics*, Entertainment Computing – ICEC2017, LNCS 10507, Springer, pp.3-13 (2017).
11. Guanyang Wang, Ying Chen, Yuan Chen, *Chinese Painting Generation Using Generative Adversarial Networks*. Available online: http://cs231n.stanford.edu/reports/2017/pdfs/311.pdf
12. Y. Chen, Y. Lai, and Y. Liu, *CartoonGAN: Generative Adversarial Networks for Photo Cartoonization*, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9465-9474 (2018).