IkebanaGAN: New GANs Technique for Digital Ikebana Art

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

[1] Kyoto University, Kyoto, Japan
[2] Faculty of Computer Science, Phenikaa University, Ha Noi, 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. In this research, we have carried out various experiments to perform mutual transformation between a domain of Ikebana (Japanese traditional flower arrangement) photos and other domains of images (landscapes, animals, portraits) to create new artworks via a variation of CycleGAN - a GANs technique based on cycle-consistency loss. A pre-trained process on object detection was added to improve the efficiency by avoiding over-transformation.

Keywords: GANs, Cycle GAN, Ikebana, Image Transformation.

1 Introduction

The rapid advance of Deep Learning in recent years raises an interesting question for both computer scientists and artists: "What is the role of AI/Machine Learning/Deep Learning in the future art scene?". For instance, the machine learning technique was used in artwork clustering tasks [1] as well as art evaluation [2]. On the other side, the application of AI, especially Deep Learning in art creation is of interest.

One basic approach of AI toward art is to use the style transfer technique to transform normal photos or sketches into artworks of specific styles. On Deep Learning, style transfer tasks can be performed by applying generative models in GANs (Generative Adversarial Networks [3]). In the training of GANs, generator network G learns to generate new data 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: G tries to maximize the probability of the generated data to lie on the distribution of target sets while D tries to minimize it. GANs training can converge even with a relatively small number of learning data.

A large number of GANs variation has been developed by modifying the basic configuration and performs impressive results on style transfer tasks. Among the variations of GANs, CycleGAN is an elegant method to study the mutual transformation between two sets of data [4]. In comparison to traditional GANs, in CycleGAN an inverse transformation of the generator network has been added to transform data on the target domain back to the input domain. Also, two discriminators are used for the two domains. The training process on CycleGAN tries to minimize the error caused by applying a cycle of forwarding and backward transformation. CycleGAN is flexible and useful for art style transfer because it uses unpaired training sets and set-to-set level transformation to learn the distribution of the target sets, which we could consider as an art style.

Classic examples of CycleGAN and other style transfer techniques were developed by taking the transformation between two sets of data of relatively similar size, with themes or categories such as the transformation between artworks by Monet and landscapes photos, winter and summer landscapes, or horse and zebra photos. So, what would happen if one performs a transfer between two sets of relatively different domains of objects. The authors proposed the idea of "unusual transformation [5]," which achieves a mutual transformation between two sets of different sizes and themes. Several examples were given by transforming portraits and animal photos into Ikebana, the Japanese art of flower arrangement, via CycleGAN. It is impressive that portraits and horse photos turn into Ikebana while one can still recognize the original shape of human faces and horses (Fig. 1). This "unusual transformation" concept would open a new way to create an original art style.

Fig. 1. Transformation of portraits and horse photos into Ikebana by CycleGAN

However, there are some limitations of traditional GANs techniques to perform this unusual transformation task. The experiments with Ikebana in [5] show some failures of CycleGAN to transform photos of complex backgrounds into abstract Ikebana (Fig. 2). In some cases, some photos were over-transformed so that we could not recognize the original shape of the main object. The structure of classic GANs techniques was not designed to learn specific high abstract representation and was difficult to learn an object with various sizes in a collection of photos. In our research, we would improve this limitation by mixing GANs with classic Computer Vision techniques. This idea appeared on CartoonGAN [6] when the authors used edge detection to emphasize the weight of edges to fit with the task of the anime-style transfer. Another interesting example is the Attentive Adversarial Network [7] which uses face recognition to improve the performance of art style transfer for selfie images.

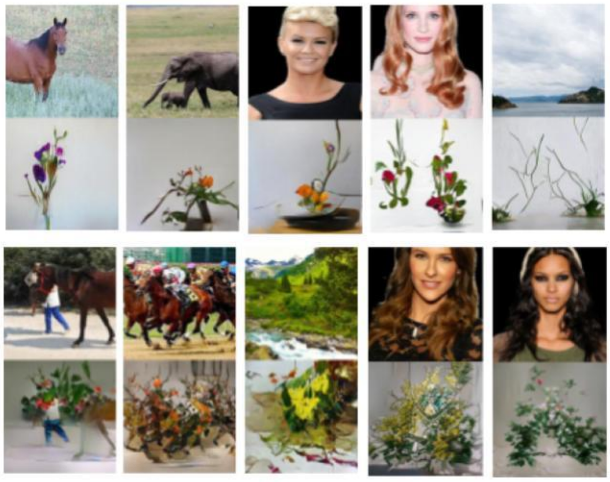

Fig. 2. Several failed transformation in [5]

In our research, we would use pre-trained object recognition and edge detection to overcome the limitation of CycleGAN to improve the transformation of portraits, landscape, and animal photos into Ikebana. The object recognition technique would remove complex background while edge detection would be used to keep the original shape of the main object to avoid over-transformed problems. The adversarial loss function of our proposed method would add an “object edge-promoting loss” to the CycleGAN’s adversarial loss so that the training process would also minimize the loss of the original shape of the main object in input photos.
Our paper is organized as follows: in section 2, we would introduce the basic concept of Ikebana and its connection to modern art. In section 3, we would describe the concept of the “unusual transformation” and the architecture of our IkebanaGAN would be proposed in section4. The experiment results would be shown in section 5 and we would discuss further the obtained research in section 5.

2 Ikebana

Ikebana is one of the most important art forms in Japanese culture. The word "Ikebana" comes from the Japanese words "Ikeru" (means "be living" or "to have a life") and "Hana" (means "flower"). Ikebana is the art of flower arrangement where the flowers are given life under the conceptual arrangements of the artists [8].
Ikebana has a deep root in the Japanese philosophy of art under the strong influence of Zen Buddhism. The tradition of arranging flowers on Buddha from China was brought to Japan in the Heian period (794-1185) by Zen Buddhist monks. In the early stage, Ikebana was just placing flowers in vases under the philosophy of Zen. But Ikebana

then grew to be an important art form along with the development of Zen, it is not just beautifully arranging flowers, but it gives the path to be harmony with nature.
Ikebana has a long history of development and has continued to be a great source of inspiration in modern art. For instance, we note a series of artwork named "Sound of Ikebana" created by one of the authors, Naoko Tosa (Fig. 3) [9]. In this work, she used fluid dynamics to create Ikebana-like forms from different types of water-based solutions. The idea of connecting Ikebana and modern technology inspired the study of developing Ikebana as a digital painting tool in [5]. In this work, the authors use the Deep Learning technique to transfer portrait and animal photos into Ikebana paintings. Our present research continues the work in [5] to improve the quality of Deep Learning-based Ikebana.

Fig. 3. Sound of Ikebana by Naoko Tosa

We emphasize two important elements in Ikebana influenced by Oriental philosophy which would support our idea of the "unusual transformation" - the mutual transfer of relatively different domains of objects. They are the "minimality" and the "flexibility." Under the influence of Zen, "emptiness" plays an essential role in Ikebana. The emptiness appearing in an Ikebana artwork is believed to provide meaning and be harmonic to the whole scene. Moreover, we call Ikebana flexible as the materials can be placed in various shapes and arrangements. We would explain how important these two properties are in our experiment in the next section.

3 Unusual Transformation

3.1 GANs and CycleGAN

As mentioned above, this research is conducted to improve the work in [5]. The fundamental approach of the method is to use the generative models in Deep Learning to transfer photos into Ikebana paintings. Generative models and discriminative models are two kinds of neural networks in Deep Learning. Informally, the goal of generative models is to generate new data instances while discriminative models would

discriminate between different categories of data. Mathematically, the generative model learns the joint probability $p(X, Y)$ of data instances set X and label set Y while the discriminative model learns the conditional probability $p(Y|X)$.
In recent years, GANs (Generative Adversarial Networks) [3] has become a big topic in Deep Learning as their generative model provides a powerful performance on the task of art style transfer with just a relatively small number of training data. The structure of GANs could be described as in Fig. 4 with the basic configuration of two networks, a generator network (G) and a discriminator network (D). The training of GANs is based on a minimax mechanism where network G learns to generate data from random noise while D tries to identify the generated data whether it is real or fake. In mathematics terms, 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. Recently, a large number of GANs variation has been developed by modifying the basic configuration of this minimax mechanism.

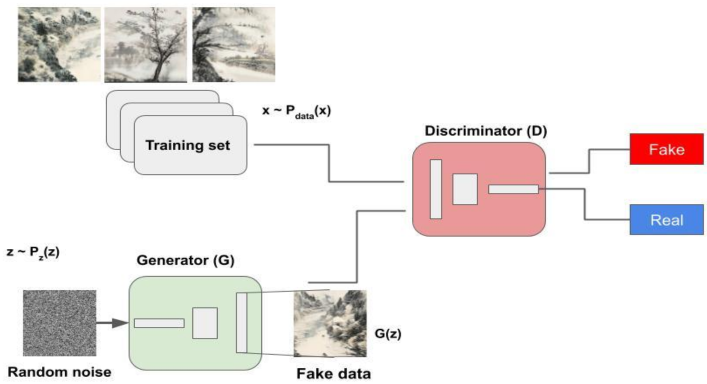

Fig. 4. The basic configuration of GANs

Among the variations of GAN, CycleGAN [4] is an elegant method to study the set-to-set level of mutual transformation between two categories of objects. Its architecture consists of two generators and two discriminators as shown in Fig. 5. Given two image sets A and B, the core goal of CycleGAN is to learn two mappings $G_{AB}: A \rightarrow B$ and $G_{BA}: B \rightarrow A$ given the training samples: $\{a_i\}_{i=1}^{N} \in A$ and $\{b_j\}_{j=1}^{M} \in B$ with the data distributions $a \sim p_A(a)$ and $b \sim p_B(b)$. The two discriminators are D_A and D_B where D_A aims to distinguish between images $\{a\}$ and translated images $\{G_{BA}(b)\}$ and the same analogy applies to D_B. The objective function of CycleGAN contains two types of loss: adversarial losses for matching the generated images to the target images; and cycle consistency loss for preventing the mappings G_{AB} and G_{BA} from contradicting each other.

Adversarial Loss: The adversarial loss applies to both mapping functions.

- For the mapping function $G_{AB}: A \rightarrow B$ and its discriminator D_B:

$$\mathcal{L}_{GAN}(G_{AB}, D_B, A, B) = \mathbb{E}_{b\sim p_B(b)}[\log D_B(b)] + \mathbb{E}_{a\sim p_A(a)}[\log(1 - D_B(G_{AB}(a)))] \quad (1)$$

- For the mapping function $G_{BA}: B \rightarrow A$ and its discriminator D_A:

$$\mathcal{L}_{GAN}(G_{BA}, D_A, B, A) = \mathbb{E}_{\mathrm{a}\sim p_A(a)}[\log D_A(a)] + \mathbb{E}_{b\sim p_B(b)}[\log(1 - D_A(G_{BA}(b)))] \quad (2)$$

Cycle Consistency Loss: For each image, a from domain A, the generated image $\widehat{}$ after applying two transformation G_{AB} and G_{BA} should be similar to a: $a \rightarrow G_{AB}(a) \rightarrow G_{BA}(G_{AB}(a)) \approx a$. This is called forward cycle consistency. Similarly, for the backward path, we have backward cycle consistency: $b \rightarrow G_{BA}(b) \rightarrow G_{AB}(G_{BA}(b)) \approx b$. The cycle consistency loss is a combination of forwarding and backward cycle consistency losses:

$$\mathcal{L}_{cyc}(G_{AB}, G_{BA}) = \mathbb{E}_{a\sim p_A(a)}\left[\left\|\mathrm{G_{BA}}(G_{AB}(a)) - a\right\|_1\right] + \mathbb{E}_{b\sim p_B(b)}\left[\left\|\mathrm{G_{AB}}(G_{BA}(b)) - b\right\|_1\right] \quad (3)$$

The full objective function of CycleGAN is a combination of the adversarial losses and the cycle consistency loss:

$$\mathcal{L}(G_{AB}, G_{BA}, D_A, D_B) = \mathcal{L}_{GAN}(G_{AB}, D_B, A, B) + \mathcal{L}_{GAN}(G_{BA}, D_A, B, A) + \lambda\mathcal{L}_{cyc}(G_{AB}, G_{BA}) \quad (4)$$

where λ is the weight of the cycle consistency loss. In the training phase, the parameters of the networks (G_{AB}, G_{BA}, D_A, and D_B) are estimated by optimizing the full objective function:

$$G_{AB}^*, G_{BA}^* = \arg\min_{G_{AB}, G_{BA}} \max_{D_A, D_B} \mathcal{L}(G_{AB}, G_{BA}, D_A, D_B) \quad (5)$$

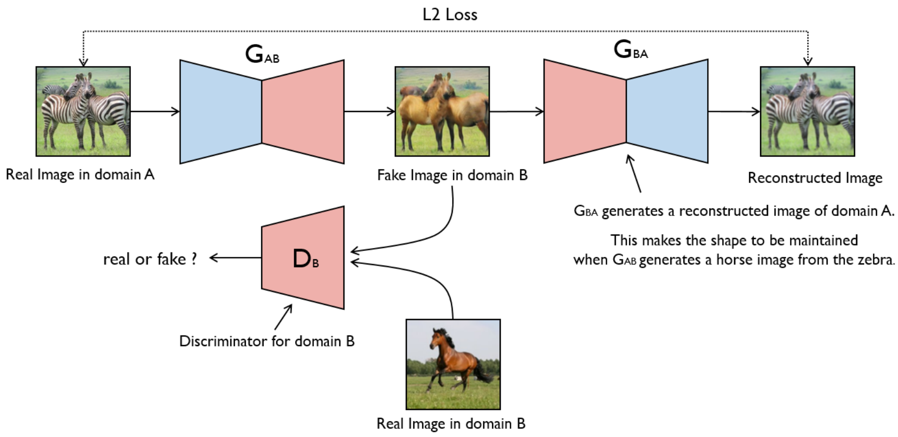

Fig. 5. The basic configuration of CycleGAN ([4])

In general, generative models in CycleGAN learn the set-to-set level of transformation while the original GANs learn to generate data to fit in a target set. Therefore, CycleGAN could be used to establish mutual conversion between these two groups of images such as the art styles of two artists. As in Fig. 4, CycleGAN converts horses into zebras and vice versa. In [5], the authors use CycleGAN to create Ikebana painting via the concept of Unusual Transformation.

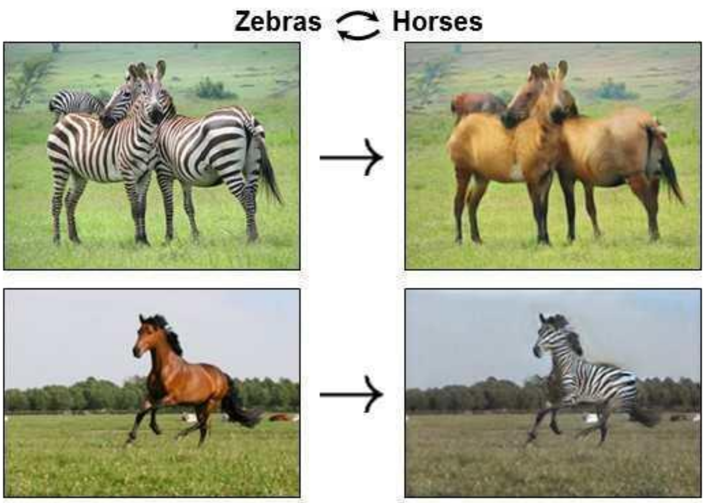

Fig. 6. Horses-Zebras transfer (Image source [3])

3.2 Concept of Unusual Transformation

In classic examples of CycleGAN in [4], the generative models were used to make a mutual transformation between landscape photos and Monet paintings, horses and zebras, winter landscapes, and summer landscapes. The transformation was made between images of relatively similar size, theme, and category. In [5], the authors give the idea of *unusual transformation*, a high-abstracted transformation where CycleGAN was applied to relatively different domains of objects which are difficult to imagine the mutual transformation such as macro and micro-size worlds or between plants and animals.

This concept of unusual transformation is believed to be a key point to create new art. For example in [5], portraits and animals were *unusually transformed* into Ikebana paintings. Another example was made by us in Fig. 7 where portraits are transformed into Sansui paintings.

The unusual transformation is a naturally difficult task with a very low rate of a successful transfer. We suggest a good transformation would include a *painting tool*set of data. The painting toolset would overcome some limitations of Deep Learning transfer including local-based transformation and noise vulnerability. Therefore, we use Ikebana as a good example of a painting toolset because of its minimality and flexibility. Sansui paintings would be another good painting tool as well because emptiness plays an important role in Sansui and natural elements such as rock, stream, mountain would be put in flexible positions.

Fig. 7. Example of unusual transformation: Portrait and Sansui Paintings

4 IkebanaGAN

CycleGAN works well on performing transformation between two sets of data of relatively similar size, themes, or categories such as transformation between art-works by Monet and landscape photos, winter and summer landscapes. In [5], the authors used CycleGAN to perform the usual transformation task between photos with complex backgrounds and abstract Ikebana (Fig. 1, 2). The experiments showed that the original CycleGAN suffers the over-transformation problem, i.e., we could not recognize the original shape of the main object in a photo after transforming.

To circumvent this problem, we try to combine CycleGAN with several computer vision techniques. To keep the original shape of the main objects, we first apply object recognition techniques to remove complex background, then edge detection is used to strengthen the shape of the objects. We define an object edge-promoting loss to enforce the model to keep the original shapes of main objects.

To include object edge-promoting loss to the adversarial loss, from the training images A, we automatically generate a set of images $E = \{e_i\}_{i=1}^{N}Z$ by removing clear edges of the main object in $\{a_i\}_{i=1}^{N}$. In more detail, for each image $a_i \in A$, we apply the following steps: (1) recognize objects in the image by using a pre-trained object detector (e.g., Mobile_Net_SSD), (2) detect edge pixels of objects using Canny edge detector [10], (3) dilate the edge regions, and (4) apply Gaussian smoothing in the dilated edge regions.

In our proposed IkebanaGAN, the goal of discriminator D_A is to maximize the probability of assigning the correct label to $G_{BA}(b)$, the real photos without clear edges of the photos' main objects (i.e., $e_j \in E$) and the real photos (i.e., $a_i \in A$). Therefore, we include object edge-promoting loss to the adversarial loss as follows:

$$\begin{aligned}\mathcal{L}_{GAN}(G_{BA}, D_A, B, A) = {} & \mathbb{E}_{a\sim p_A(a)}[\log D_A(a)] \\ & + \mathbb{E}_{b\sim p_B(b)}[\log(1 - D_A(G_{BA}(b)))] \\ & + \gamma\mathbb{E}_{e\sim p_E(e)}[\log(1 - D_A(G_{BA}(e)))] \qquad (6)\end{aligned}$$

where γ controls the relative importance of the object edge-promoting loss.

5 Experiment

We performed the unusual transformation via IkebanaGAN with the style set A and the object sets B1 and B2 as follows:

- Dataset A: Ikebana photos in Google Image Search
- Dataset B1: Portrait photos in Flickr
- Dataset B2: Kaggle Animal-10 dataset. (https://www.kaggle.com/alessiocorrado99/animals10)

Figures 8a and 8b show several results of A to B1 transformation and Figs. 9a and 9b show several results of A to B2 transformation.

Fig. 8a. Experiment result A-B1: IkebanaGAN and CycleGAN both generate acceptable transformation (the first row is the original photo, the second row is the transformation by CycleGAN, the last row is the transformation by IkebanaGAN)

Fig. 9b. Experiment result A-B1: IkebanaGAN performs better than CycleGAN (the first row is the original photo, the second row is the transformation by CycleGAN, the last row is the transformation by IkebanaGAN)

Fig. 9a. Experiment result A-B2: IkebanaGAN and CycleGAN both generate acceptable transformation (the first row is the original photo, the second row is the transformation by CycleGAN, the last row is the transformation by IkebanaGAN)

Fig. 9b. Experiment result A-B2: IkebanaGAN performs better than CycleGAN (the first row is the original photo, the second row is the transformation by CycleGAN, the last row is the transformation by IkebanaGAN)

6 Discussion and Conclusion

In the examples which IkebanaGAN performs better than CycleGAN, we found that the original shape were well-preserved as in our assumption. IkebanaGAN would improve the successful rate as well as the quality of the unusual transformation of Ikebana and portraits more than animal photos. We consider the reason as the structures of objects in animal photos are more complex than human faces.

As we mentioned before, the unusual transformation is a challenging task because of the different structures of the two data sets. We hope to improve that difficulty by providing some techniques that mixed Deep Learning-based style transfer and classic Computer Vision's object detection. We remark that because of the natural difficulty, the success rate of the transformation is still low.

In the future, we would use another approach by mixing two GANs networks. We would provide a transformation between photos and sketches as well as sketches and Ikebana with the assumption that the sketch structure would remove the difficulty of over-transformation.

References

1. Gultepe E., Conturo T. E., Makrehchi M., Predicting and Grouping Digitized Paintings by Style using Unsupervised Feature Learning. J Cult Herit. 2018 May-Jun;31:13-23. doi: 10.1016/j.culher.2017.11.008. Epub 2017 Dec 20. PMID: 30034259; PMCID: PMC6051702.
2. Hung, M. C., Nakatsu, R., Tosa, N., Kusumi, T., "Learning of Art Style Using AI and Its Evaluation Based on Psychological Experiments," *2020 International Conference on Entertainment Computing* (2020).
3. Creswell, A., et al., "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, Vol.35, No.1, pp.53-65 (2018).
4. Zhu, J, Park, T., Isola, P., Efros, A. A., "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242-2251 (2017).
5. Hung, M. C., Nakatsu, R., Tosa, N., "Developing Japanese Ikebana as a Digital Painting Tool via AI," *2020 International Conference on Entertainment Computing* (2020).
6. Chen, Y., Lai, Y. and Liu, Y., "CartoonGAN: Generative Adversarial Networks for Photo Cartoonization," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.9465-9474 (2018).
7. Li, X., Zhang, W., Shen, T., "Mei, Everyone is a Cartoonist: Selfie Cartoonization with Attentive Adversarial Networks," *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp.652-657 (2019).
8. Sato, S., "The Art of Arranging Flowers: A Complete Guide to Japanese Culture," Harry N. Abrams (1965).
9. Tosa, N., Pang, Y., Yang, Q., Nakatsu, R., "Pursuit and Expression of Japanese Beauty Using Technology," in Frederic Fol Leymarie, Juliette Bessette, G. W. Smith eds., The Machine as Art/The Machine as Artist, MDPI, pp.267-280 (2020)
10. Canny, J., "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence (6)*, pp 679-698 (1986).