ArtTension: Artistically extended generation based on single-image learning

Jung-Jae Yu
Creative Content Research
Division, ETRI
jungjae@etri.re.kr

Juwon Lee
Creative Content Research
Division, ETRI
zacurr@etri.re.kr

Junmo Kim
Electrical Engineering
KAIST
junmo.kim@kaist.ac.kr

Abstract

To enhance efficiency in creative activities that utilize existing art paintings, we propose an artistic outpainting method based on single-image learning. The proposed method expands the image background by learning only the input image, without pre-learning large-capacity images. When recreating existing artworks by transforming them, the simple task of expanding the background is automatically solved, allowing the artist to focus on more creative work. The proposed method can be used in AI collaboration or immersive media art exhibitions using existing paintings.

1 Introduction: Artistic outpainting

Is it possible to automatically extend the background of paintings whose painter has a distinct style and few paintings? In this study, we propose an artistic outpainting that uses deep neural network to do this. Outpainting refers to a technology that automatically generates the outer region of an input image using machine learning. Most outpainting studies are based on a large number of training images [1], [2], [3]. This learning-based generation technology assumes that the learned and processed images belong to the same domain. However, for paintings, this method frequently cannot be used because a large number of similar styled images cannot be collected. Recently, SinGAN [4], a method for generating images of the same style by learning only one input image, has received considerable attention. However, SinGAN is based on random sampling of a complete image within the learned image distribution; therefore, it is difficult to use it for outpainting. To solve this problem, we modified the method of calculating the adversarial and reconstruction losses in the learning stage, and extended the input image horizontally by inputting a combined noise vector as a seed for image generation in the inference stage. Please refer to the supplementary material for details regarding these modifications. If there is a main foreground object in the picture, consistent outpainting causes a problem in which the components of the foreground object are mixed with the background components in the extended area. To solve this problem, only the image components of the background are learned using a mask prepared in advance. In the inference step, the background image is generated with extension, and the excluded foreground object is recombined at the original position in the generated image to complete outpainting. We call the proposed method ArtTension(Artistic Extension), and we demonstrate through various experiments that ArtTension can be used for both photos and art paintings. The proposed method can be used in an AI collaboration exhibition that displays existing artworks by reprocessing them, or an immersive media art exhibition that displays paintings in various ratios on a large screen.

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

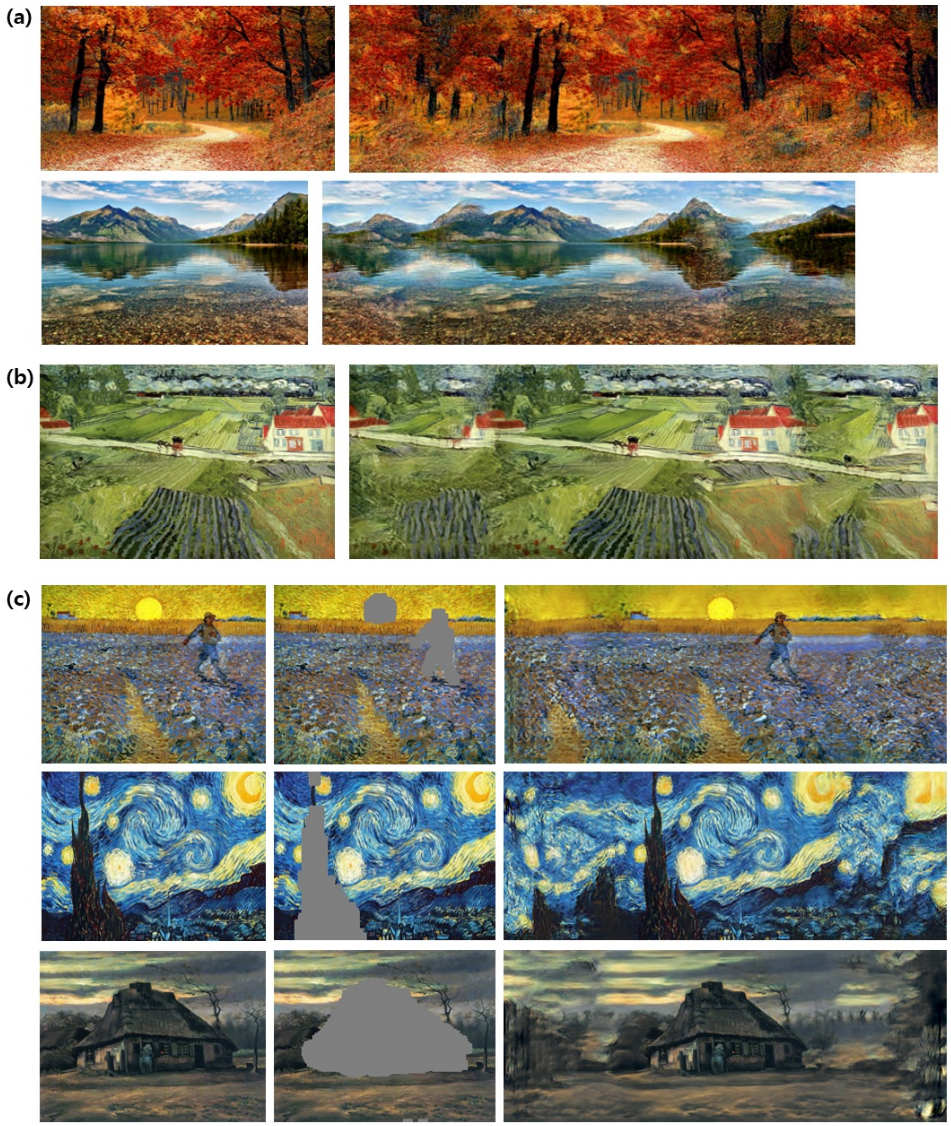

Figure 1: Results of applying ArtTension to photographs and art paintings: input and results for (a) photographs and (b), (c) art paintings. (c) also shows the masks of the foreground objects.

2 Interface and Experimental Results

Outpainting using ArtTension consists of the following 4 steps. The first step is to prepare a mask that includes the main foreground object in the image to be processed. It can be created in a few seconds using the selection function in Adobe Photoshop, or a rough area can be manually created to contain the foreground object. In the second step, a generative network is trained on the image to be processed together with the prepared mask. In the third step, the learned network is used to generate an extended background image. The last step is synthesizing the foreground object back to its original position using the mask. Fig. 1 shows the results of applying ArtTension to art paintings and photographs. The example pictures show the case where a distant landscape image is an input with no mask, and the case where a mask is applied to the foreground object. Each image is created such that the horizontal axis is up to 500 pixels. The training time is an average of 2 hours for each image with a Titan V 1GPU, and the generation time in the inference stage is less than 0.1 second. Compared to manual drawing, our method can increase the automation rate in creative activities using existing artistic paintings.

3 Ethics

Deep neural networks, which have recently been in the spotlight of machine learning, depend on training data. A network that has learned considerable data can have high performance in terms of function, but it may have preconceived notions and biases about gender, race, and culture depending on the learned data. To prevent such problems from occurring, it is necessary to manage and monitor the training data so that it is properly configured. However, human review of massive data incurs a lot of time and cost, and if another automation technology is used for this task, the same problem may arise. Therefore, some of these ethical problems can be solved if neural networks can learn from small-scale learning data that can be appropriately managed by human ethical standards. This makes the proposed ArtTension significant as it learns only the input image to perform outpainting.

Acknowledgments and Disclosure of Funding

This research is suported by Ministry of Culture, Sports and Tourism(Korean government) and Korea Creative Content Agency (Project Number:R2021040122, Development of artificial intelligence-based photo/painting harmonization and content expansion technology).

References

[1] Sabini, Mark, and Gili Rusak. "Painting outside the box: Image outpainting with gans." arXiv preprint arXiv:1808.08483 (2018).

[2] Van Hoorick, Basile. "Image outpainting and harmonization using generative adversarial networks." arXiv preprint arXiv:1912.10960 (2019).

[3] Yang, Zongxin, et al. "Very long natural scenery image prediction by outpainting." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[4] Shaham, Tamar Rott, Tali Dekel, and Tomer Michaeli. "Singan: Learning a generative model from a single natural image." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

A Supplementary Material

A.1 Method

The method proposed in this paper is divided into a training stage, as shown in Fig. 2, and an inference stage, as shown in Fig. 3). The training stage consists of generation, adversarial loss, and reconstruction loss steps, as shown in Fig. 2). In the description of training and inference below, the mask of the foreground object is not considered for convenience, and it is only necessary to consider that the mask is multiplied by the input and generated images in each process.

In the training stage, as shown in Fig. 2, the generation step includes an image generator (multi-scale layer network [4]), G_{Img}, and a noise vector generator, G_z, which generates three types of noise vectors: z_{random}, z_{fixed1}, and z_{fixed2}. z_{random} is generated with a different value for every trial of the training stage, and for z_{fixed1} and z_{fixed2}, the values are stored in the noise vector set, $\{z_{fixed1,2}\}$, when they are first generated in each layer. The stored values are loaded and used again when passing through the same layer in the subsequent iterations. In the training stage, the image generator, G_{Img}, generates extended images for each trial by inputting z_{random}, z_{fixed1}, and z_{fixed2}. In this case, the generated image may have a magnification of one to three times that of the input image in the horizontal direction; for convenience, the proposed method is described by fixing it to two times.

In the adversarial loss step shown in Fig. 2, a discriminator network D is trained for the generated image G(z_{random}) among the three images generated in the previous generation step. The G_{Img} in the generation step is trained to minimize this loss, such that an image of the same style as the input image is generated. In the reconstruction loss step of Fig. 2, $LossR_{center}$ is calculated from G(z_{fixed1}), and $LossR_{side}$ is calculated from G(z_{fixed2}). When G_{Img} is trained to minimize $LossR_{center}$, it is trained to generate the same image as the input image in the central region of the generated image with z_{fixed1}, while stabilizing the generation result. When G_{Img} is trained

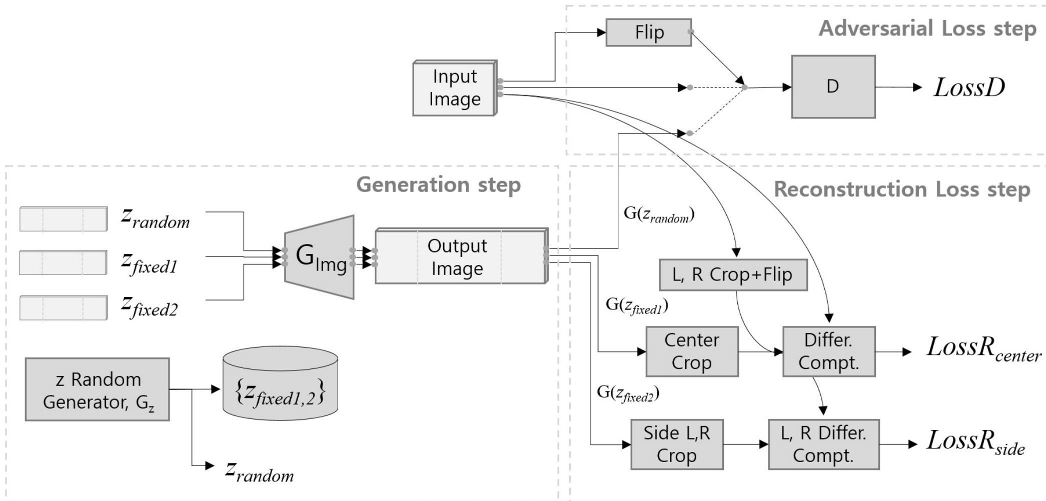

Figure 2: Overall process of training stage of ArtTension

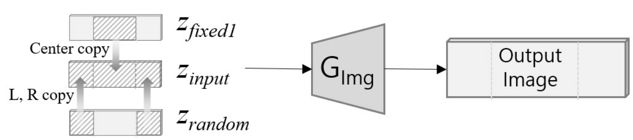

Figure 3: Inference stage of ArtTension

to minimize $LossR_{side}$, the image generated in the left and right peripheral regions is smoothly connected to the left and right ends of the input image.

In the inference stage shown in Fig. 3, the image is extended in the left and right directions using the image generator G_{Img} and z_{fixed1}, which is part of the stored noise vector set. For each trial in the inference stage, z_{input} combined with z_{fixed1} for the central region and z_{random} for the left and right regions are input to G_{Img}, and each image is expanded differently according to the z_{random} in the left and right peripheral regions.

A.2 Results

Above, each image was shown one by one when it was extended in the horizontal direction by applying ArtTension. Here, we show the results of various extensions by applying ArtTension to each image. Fig. 4 and Fig. 5 show various results of extensions when ArtTension is applied to photographic and art painting landscape images, respectively. Fig. 6 shows the results of outpainting by applying a mask to a relatively small foreground object in a painting image. Fig. 7 shows the results of outpainting by applying a mask to a relatively large foreground object in a painting image and photographic image. As each result shows, using ArtTension, it is possible to acquire extended images in various forms with a feeling similar to the style of the input image. Artists can use these results as a background to show their creativity and modify them in a new form or add new elements.
‘

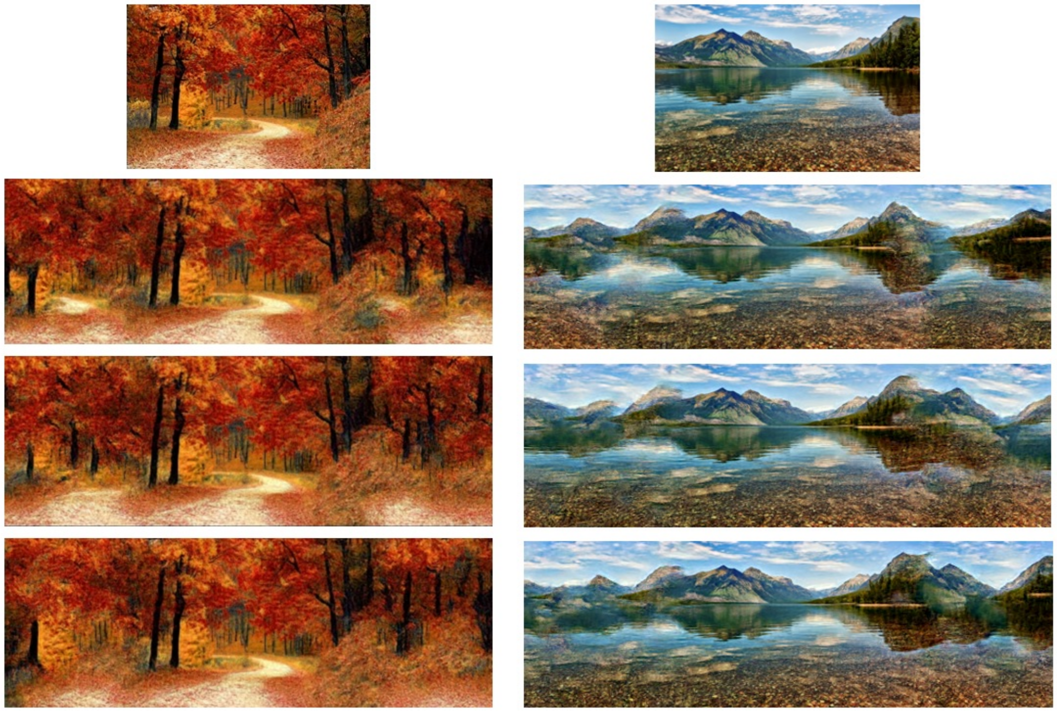

Figure 4: Results of applying ArtTension to photographic images.

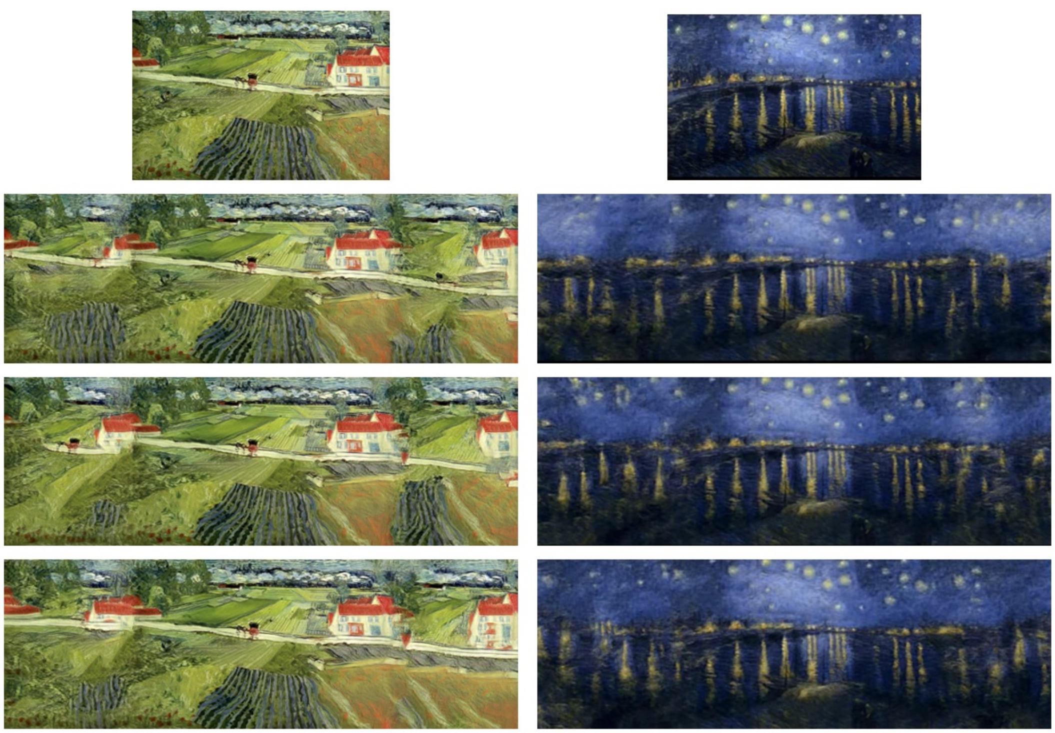

Figure 5: Results of applying ArtTension to art paintings.

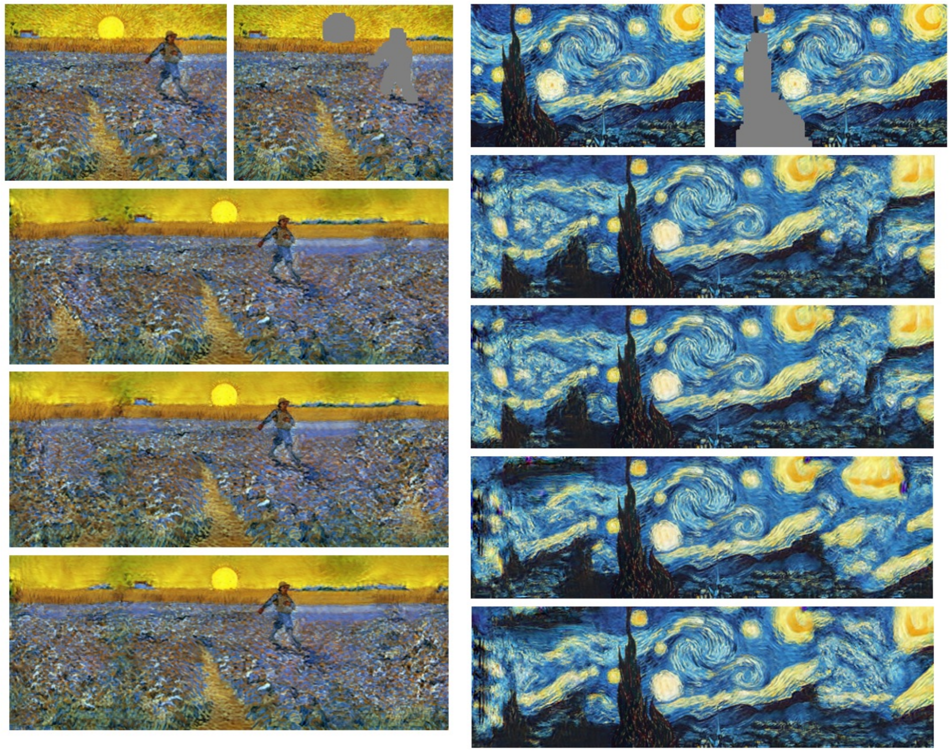

Figure 6: Results of applying ArtTension to art paintings with foreground objects.

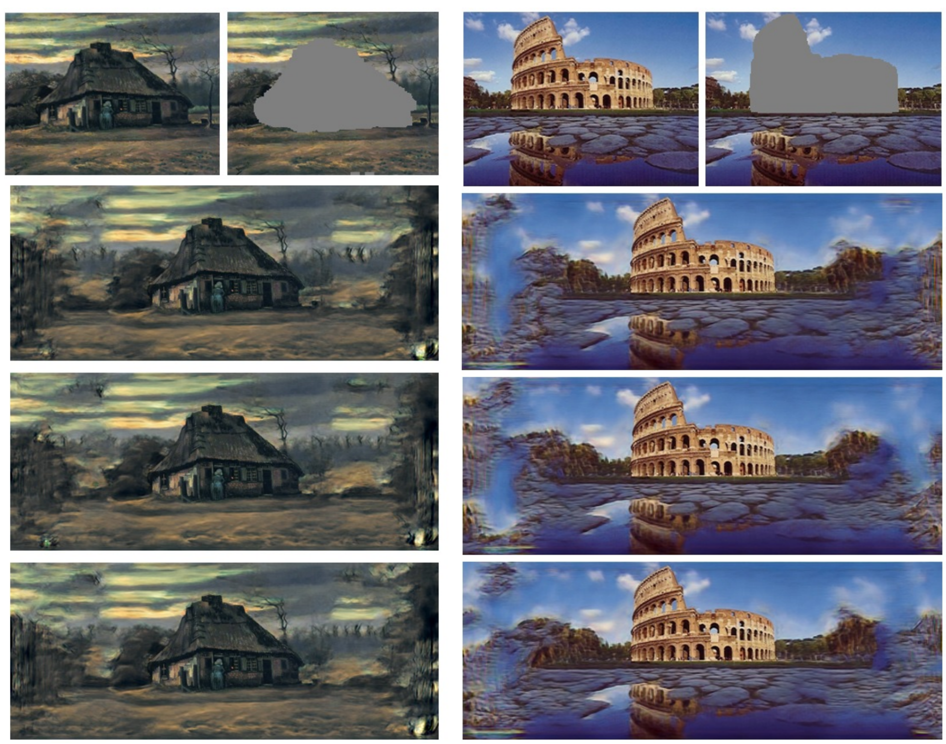

Figure 7: Results of applying ArtTension to an art painting and a photographic image with large foreground objects.