Construction of Dataset for Webtoon Creation by Treatment

Kyung-Ho Yu
Chosun university
Gwangju, 61452
Republic of Korea
infinite_gh@naver.com.com

Hyoung-Ju Kim
Chosun university
Gwangju, 61452
Republic of Korea
snowlisakim@gmail.com

Jeongin Kim
Chosun university
Gwangju, 61452
Republic of Korea
jungingim@gmail.com

Pan-Koo Kim
Chosun university
Gwangju, 61452
Republic of Korea
pkkim@chosun.ac.kr

ABSTRACT

With the development of Generative Adversarial Network technology, more and more realistic images have been created. The applications of the generative models have been extended to various domains such as advertising and entertainment industries, and is expanding to the realm of art. In the production of webtoons, GAN can be used to assist an artist in the story planning stage. For example, a story idea (called treatment) can be input into text-to-image GAN, which can generate example webtoon before the actual production. This will give an artist the freedom to try many different ideas without any additional costs or time. However, the main hurdle in adopting such an approach is the unavailability of a representative dataset. Therefore, in this work, we aim to generate toon datasets based on text-to-image GAN models that can be used for in treatment-webtoon production. text-to-image GAN models. In the experiments, we have used the pre-trained CartoonGAN model and built a treatment-toon datasets based on input text caption, which consists of a total 2,465,740 images.

With the development of Generative Adversarial Network technology, more and more realistic images have been created. The applications of the generative models have been extended to various domains such as advertising and entertainment industries, and is expanding to the realm of art. In the production of webtoons, GAN can be used to assist an artist in the story planning stage. For example, a story idea (called treatment) can be input into text-to-image GAN, which can generate example webtoon before the actual production. This will give an artist the freedom to try many different ideas without any additional costs or time. However, the main hurdle in adopting such an approach is the unavailability of a representative dataset. Therefore, in this work, we aim to generate toon datasets based on text-to-image GAN models that can be used in treatment-webtoon production. In the experiments, we have used the pre-trained CartoonGAN model and built a treatment-toon datasets based on input text caption, which consists of a total of 2,465,740 images.

CCS CONCEPTS

• Computing methodologies → Artificial intelligence → Natura l language processing

KEYWORDS

Text-to-image, deep learning, cartoon

1 INTRODUCTION

Generative Adversarial Network (GAN) [ref] is a machine learning framework in which two models namely, a generator and discriminator compete against each other to generate realistic images. Since the announcement of GAN in 2014, research using GAN has been actively conducted, such as computers generating images themselves and converting existing pictures into other paintings' styles. The study of generating images using generative models such as GAN is called 'neural rendering' [1]. Recently, its applications have been extended to advertising and entertainment industries. For example, advertisers generate human images and use them in advertisements, users create avatars to use them on online broadcasting platforms, and create and use emoticons similar to their appearances. In addition, deep learning-based generation models are used in many ways not only in industries such as advertising and entertainment, but also for general purposes.. Such deep learning-based generation model

development is changing people's perception that the creative process is no longer the unique area of professional artists.

Webtoon is a combination of web and cartoon, which is serialized and distributed over the Internet. The creation process of webtoons is divided into story planning, conti/sketch, line drawing, coloring and contrast, and background drawing. The manual production of such webtoons, rather than computer-aided, is expensive and takes a lot of time. . Therefore, to minimize the humans involvement during the production, artificial intelligence software has recently been developed that automatically colors lines in the drawings.

Besides, the applications of GAN in advertising and entertainment industries, they have also been used in the production of art. In the production of webtoons, GAN can be used to assist an artist in the story planning stage. For example, a story idea (called treatment) can be input into text-to-image GAN, which can generate example webtoon before the actual production. This will give an artist the freedom to try many different ideas without any additional costs or time. However, the main hurdle in adopting such an approach is the unavailability of a representative dataset. Therefore, in this work, we aim to generate toon datasets based on text-to-image GAN models that can be used in treatment-webtoon production.

2 RELATED WORK

2.1 Language model

Natural Language Processing (NLP) is a field that studies information analysis, emotion recognition, and individual name recognition by extracting semantic features from text. Language models based on early shallow neural networks learn to predict the surrounding words of window size and perform word embeddings. However, their main disadvantageous is that they cannot represent contextual meaning. Based on this necessity, a bi-lstm-based language model was developed to understand the contextual meaning. Most recently, a Transformer-based language model was developed, which has shown high performance efficiency for various NLP tasks.
The recent language models have a large number of parameters and can learn larger amount of data. For example, language models developed by OpenAI called GPT-3, and by Meta called OPT-175B, both have the same number of parameters that is, 175 billion. These models are available freely and can be fine-tuned according to an application requirements.

2.2 Text-to-image based on Deep learning

Traditional text-to-image involves two steps. First, the important keywords are extracted from the text. Second, the keywords are used to synthesize parts of an image that fit the keywords. The main disadvantage of traditional Text-to-image methods is that its matching images corresponding to the keywords require a lot of human involvement.
On the contrary, text-to-image using deep learning technology allows computers to learn on their own, from the step of extracting key parts of text to the step of generating images. The humans involvement is reduced to designing the models. Recently, in OpenAI, a study was conducted to generate images by learning 250 million pieces of text-image data from VQ-VAE and Transformer [2].

3. Building a Toon Data Set for cartoon generation

We have used MS-COCO datasets as a benchmark to build treatment-webtoon datasets. The existing MS-COCO dataset is in English; however, the present study used the Korean MS-COCO dataset provided by AI-HUB. The Korean MS-COCO dataset consists of 123,287 sheets in total, which are divided training 87,783 sheets and validation 40,504 sheets.

We have used the pre-trained CartoonGAN model to convert MS-COCO images to Toon images. The CartoonGAN model was pre-trained to learn four different toon styles named after their authors such as Hayao, Hosoda, Parika, and Shinkai [3].

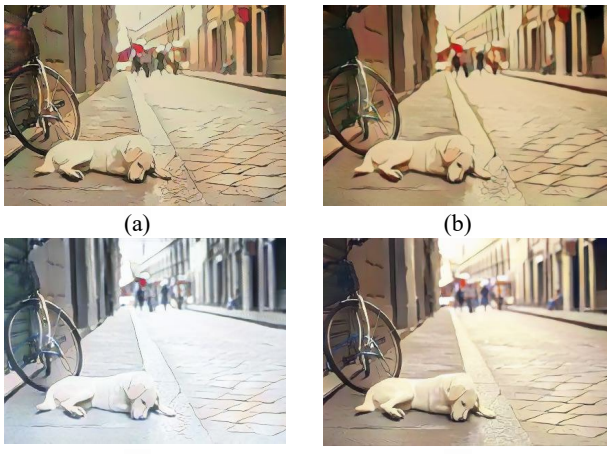

Figure 1. Toon transformation image

Figure 1 shows the results of converting the MS-COCO dataset into toon styles using the pre-trained CartoonGAN. In Figure 1 the image is converted into Hayao (a), Hosoda (b), Parika (c), and Shinkai (d) styles. As a result of the conversion, 493,148 images were generated from the original dataset. In addition, a total treatment-toon dataset consists of 2,465,740 images was constructed because there were five captions per image.

4 CONCLUSIONS AND FUTURE WORK

In this study, the treatment of a webtoon that helps the writers' creative process, was used as an input into the deep learning model to generate a representative dataset for webtoon production. In order to construct the treatment-webtoon dataset, the Korean MS-COCO dataset was used. Images in the dataset were converted into toon images by exploiting the pre-trained CartoonGAN model. The toon images were generated according to four different painting styles. As a result, we construct 2,465,740 treatment-webtoon datasets.

In the future, we will demonstrate the production of webtoons by using the generated Treatment-Toon dataset.

ACKNOWLEDGMENTS

This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1A2C2007091).

REFERENCES

[1] Tewari, Ayush, et al. "State of the art on neural rendering."*Computer Graphics Forum*. Vol. 39. No. 2. 2020.

[2] Ramesh, Aditya, et al. "Hierarchical text-conditional image generation with clip latents."*arXiv preprint arXiv:2204.06125*(2022).

[3] Chen, Yang, Yu-Kun Lai, and Yong-Jin Liu. "Cartoongan: Generative adversarial networks for photo cartoonization." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.