2022年　4月　28日

2021年度若手研究者共同研究プロジェクト実施報告書

法政大学総長　殿

以下のとおり研究実施報告書を提出します。

基本情報	研究課題名：Deep learning for creating images from text description
	研究代表者　氏名：ZHANG Zhiqiang (張 志強)
	（在籍者）研究科・専攻・学年：理工学 研究科　応用情報工学 専攻博士後期課程　２年生在籍 （修了者）所属・職種：
	指導教員（所属・職・氏名）：理工学研究科・准教授・周　金佳 （※在籍者のみ記入）
	共同研究者（所属・職・氏名）：理工学研究科・学生・傅 晨 （※指導教員と同人の場合は記入不要）
	その他　研究分担者：なし
	研究期間：　２０　21年度　～　２０　23年度（※研究修了年度を記載）

年間の研究実施概要

※研究計画の進捗状況を中心に今年度の研究実施状況を記載してください。

According to the research plan, text-to-image synthesis, customizable image synthesis, multi-stage multi-text image synthesis, and multi-stage multi-text customizable image synthesis have been realized this year. The specific implementation situation of each aspect is as follows:

- **Text-to-Image Synthesis**

We propose two improvements in the text-to-image synthesis research: 1) multi-class discriminant method; 2) a method of starting synthesis from the foreground.

The idea of the multi-class discrimination method is to improve its discriminative ability by introducing more discriminant types into the discriminator. According to the adversarial character of GAN, the improvement of the discriminative ability can promote the generator's generation ability to achieve better image synthesis. The subjective results are shown in Fig. 1.

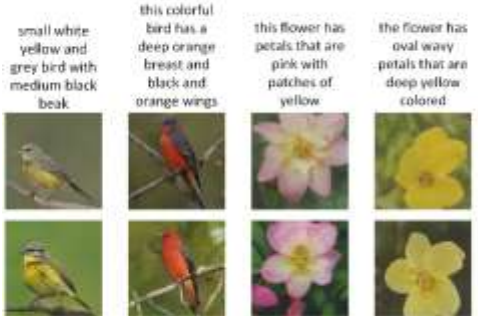
small white yellow and grey bird with medium black beak
this colorful bird has a deep orange breast and black and orange wings
this flower has petals that are pink with patches of yellow
the flower has oval wavy petals that are deep yellow colored

Fig.1. Based on the input text, our method can synthesize multiple high-quality image results.

For the method of synthesizing from the foreground, the specific implementation process is first to synthesize the corresponding foreground content based on the input text and then synthesize the final image based on the synthesized foreground and input text. This synthesis process is divided into two synthesis methods. One is to synthesize the foreground in the first and second stages, and the image with background information is synthesized in the third stage; the other is to synthesize the foreground in the first stage, and the image with background information is synthesized in the second and third stages.

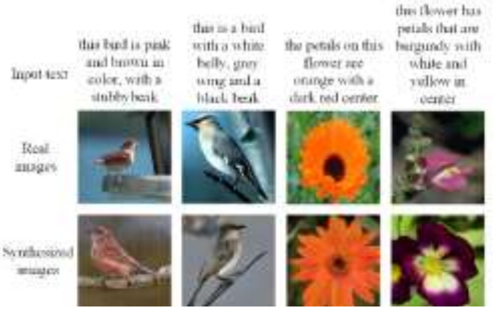
Input text
this bird is pink and brown in color, with a stubby beak
this is a bird with a white belly, grey wing and a black beak
the petals on this flower are orange with a dark red center
this flower has petals that are burgundy with white and yellow in center
Real images
Synthesized images

Fig.2. The comparison between our results and the real images corresponding to the input text.

年間の研究実施概要

- **Customizable Image Synthesis**

The basic idea of customizable image synthesis is to use text and contour information to synthesize corresponding images. The text can determine the basic content of the synthesis, and the contour can determine the shape, size, and position of the synthesized object. On the one hand, using text and contour information to synthesize images achieves better control effects. On the other hand, both text and contour can be manually input, which makes the synthesis method more interactive and practical. The results of the customizable synthesis are shown in Fig. 3.

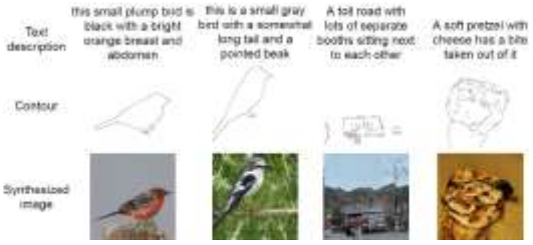
Text description
this small plump bird is black with a bright orange breast and abdomen
this is a small gray bird with a somewhat long tail and a pointed beak
A toll road with lots of separate booths sitting next to each other
A soft pretzel with cheese has a bite taken out of it
Contour
Synthesized image

Fig.3. The results of customizable image synthesis.

- **Multi-stage multi-text image synthesis**

Multi-stage multi-text image synthesis means that for the image synthesized based on text, it can continue to use text to modify the image content. The core of this method is to use text to modify the generated image content. To achieve this goal, we refer the idea of text-to-image synthesis and realize text-guided image manipulation through multi-stage synthesis. We introduce sentence-aware and word-aware in the network structure to improve the image manipulation effect. The comparison results between our method and existing text-guided image manipulation methods are shown in Fig. 4. The figure shows that the manipulation effect of our method is best.

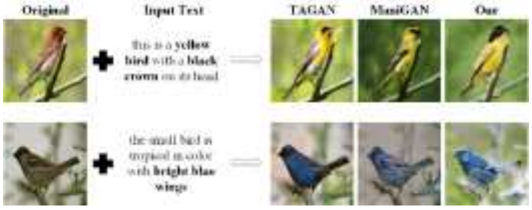
Original
Input Text
TAGAN
ManiGAN
Our
this is a yellow bird with a black crown on its head
the small bird is tropical in color with bright blue wings

Fig.4. The comparison results of our method with existing text-guided image manipulation methods.

After combining our proposed text-guided image manipulation and text-to-image synthesis methods, a multi-stage multi-text image synthesis method is formed. The basic result is shown in Fig. 5.

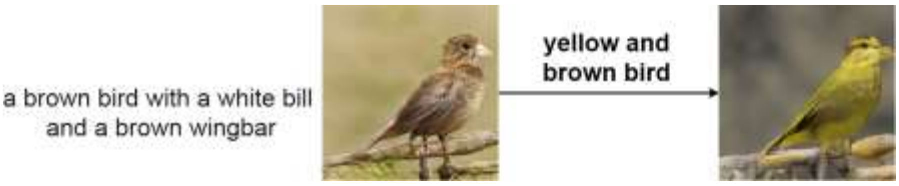

Fig. 5. Based on the text, the corresponding image can be synthesized, and then it can continue to enter the text to modify the content of the generated image.

- **Multi-stage multi-text customizable image synthesis**

We combine text-guided image manipulation with customizable image synthesis to achieve multi-stage multi-text customizable image synthesis. This approach has extremely high controllability. It allows people to input the text and contour to synthesize the corresponding image and then continue to input new text to modify the local content of the generated image. Fig. 6 shows the results of multi-stage multi-text customizable image synthesis. The result is first synthesized based on text and contour, and then can continue to input new text to modify the generated image content.

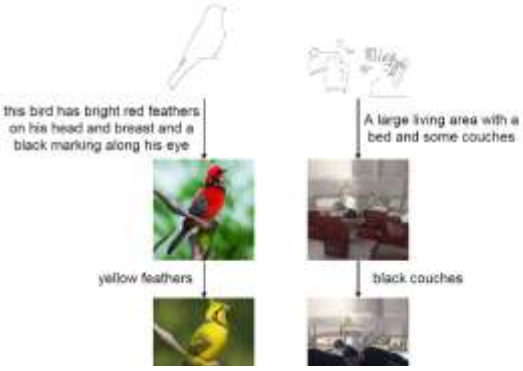
this bird has bright red feathers on his head and breast and a black marking along his eye
A large living area with a bed and some couches
yellow feathers
black couches

Fig.6. The results of the multi-stage multi-text customizable image synthesis.

Summary of the current situation

Overall, we have accomplished the fundamental goal of this research, which is to realize a multi-stage and multi-text synthesis method that can be artificially controllable and highly flexible. At the same time, in terms of quantitative results, the established 15% improvement target has been basically completed.

Existing problems and future research plans

There are two main problems at present, one is that the image results synthesized by the current methods still have room for further improvement; the other is that the current method performs generally in complex image synthesis. For the above problems, we will further optimize the network model of synthetic images in the future to achieve better image synthesis quality and higher quality complex image synthesis results. Besides, we will reduce the weight of the model to improve the applicability of this research.

研究業績

成果発表（学会・論文・研究会等）

学会・論文・研究会等の別	タイトル	発行または発表年月
International Conference on Image Processing (ICIP)	Text to Image Synthesis with Erudite Generative Adversarial Networks	September, 2021
International Conference on Multimedia and Expo (ICME)	Text-guided Image Manipulation based on Sentence-aware and Word-aware Network	Accepted, 2022

その他（アピールすることがあればご記入ください。）