

2022 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（学生証番号） : 理工学・応用情報工学・博士後期課程 3 年生 (20R9805)
	【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 :
年間の研究実施概要	研究期間 : 2021 年度 ~ 2023 年度（※研究終了年度を記載）
	<p>※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。</p> <p>According to the research plan, text-to-image synthesis, customizable image synthesis, text-guided image manipulation and text-guided controllable image synthesis and manipulation have been realized this year. The specific implementation situation of each aspect is as follows:</p> <p>● Text-to-Image Synthesis</p> <p>We propose one improvement in the text-to-image synthesis research: a method of starting synthesis from the foreground.</p> <p>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. The corresponding results are shown in Fig. 1.</p> <div style="text-align: center;"> </div> <p>Fig.1. The comparison between our results and the real images corresponding to the input text.</p>

● 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. 2.

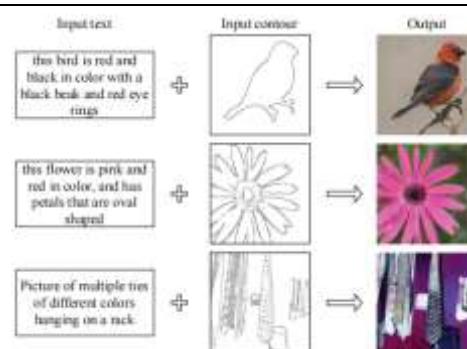


Fig.2. The results of customizable image synthesis.

● Text-guided Image Manipulation

In our target, we want to achieve multi-stage multi-text image synthesis. It means that for the image synthesized based on text, it can continue to use text to modify the image content. The core of this idea 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. 3. The figure shows that the manipulation effect of our method is best.

After combining our proposed text-guided image manipulation and text-to-image synthesis methods, a text-guided controllable image synthesis and manipulation is formed. The basic result is shown in Fig. 4.

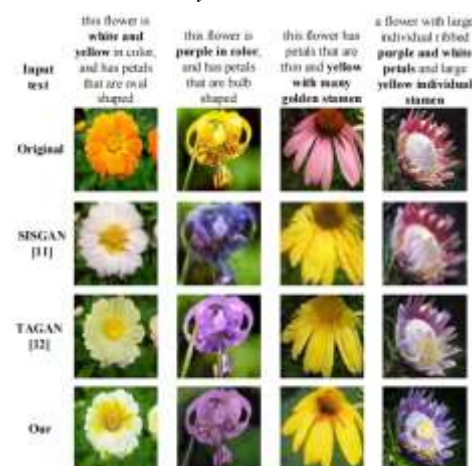


Fig.3. The comparison results of our method with existing text-guided image manipulation methods.

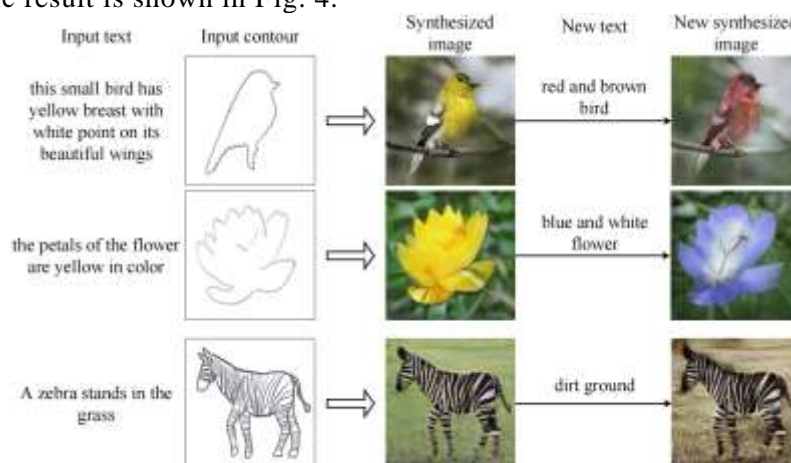


Fig. 4. The results of text-guided controllable image synthesis and manipulation.

● Text-guided controllable image synthesis and manipulation

We combine text-guided image manipulation with customizable image synthesis to achieve text-guided controllable image synthesis and manipulation. 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. In Fig. 4, it can be found that the result is first synthesized based on text and contour, and then can continue to input new text to modify the generated image content.

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. Compared to last year, we have further improved the quality of synthesis results, especially for complex image synthesis. In Fig. 1, 3, and 5, we can clearly find that the complex images synthesized by our proposed method have satisfactory quality.

Existing problems and future research plans

There still have one problem at present, that is, the network parameters are still large. For this problem, we will reduce the parameters 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	July, 2022
	Information Sciences	Text-to-image synthesis: Starting composite from the foreground content	August, 2022
	Applied Sciences	Text-Guided Customizable Image Synthesis and Manipulation	October, 2022
その他（アピールすることがあればご記入ください。）			