Research on Deep Learning for Artificially Controllable Image Synthesis

ZHANG, Zhiqiang

(開始ページ / Start Page)
1
(終了ページ / End Page)
114
(発行年 / Year)
2023-09-15
(学位授与番号 / Degree Number)
32675甲第586号
(学位授与年月日 / Date of Granted)
2023-09-15
(学位名 / Degree Name)
博士(工学)
(学位授与機関 / Degree Grantor)
法政大学 (Hosei University)
(URL)
https://doi.org/10.15002/00030035

博士学位論文
論文内容の要旨および審査結果の要旨

論文題目	Research on Deep Learning for Artificially Controllable Image Synthesis
氏名	ZHANG　Zhiqiang
学位の種類	博士（工学）
学位番号	第 586 号
学位授与年月日	2023 年 9 月 15 日
学位授与の要件	法政大学学位規則第 5 条第 1 項第 1 号該当者（甲）
論文審査委員	主　査　周　金佳　　准教授
	副　査　尾川　浩一　　教授
	副　査　彌富　仁　　教授

１．論文内容の要旨

In recent years, text-to-image synthesis (T2I) has been proposed and gained extensive attention. T2I generates corresponding images through simple, intuitive, and easy-to-enter text information. Due to the text information conforming to people's input habits, this method can realize the artificially controllable image synthesis effect to a certain extent. Nevertheless, T2I still faces the following challenges: 1) The quality of image synthesis needs to be further improved. Quality refers to the accuracy of the visually significant attributes of the generated image such as the color accuracy, exposure accuracy, contrast, and so on. The current T2I methods still produce low image quality, so the overall quality needs to be improved. 2) The controllability of the image synthesis process needs to be further improved. Controllability refers to how much of the generated images can be controlled by the users. By using text information, the current T2I methods can only control the basic content of the synthesized object but cannot control the shape, size, and position information of the synthesized object, so the overall controllability is insufficient. 3) The overall practicability of the image synthesis method needs to be further improved. Practicality is reflected in the application degree of the synthetic method. The current T2I methods can synthesize the corresponding image based on the input text, but it cannot continue to input new text to modify the content of the generated image, which makes the overall practicability of the current method insufficient.

Facing the above challenges, this research is committed to realizing the artificial controllable image synthesis method in the whole process, which is divided into three

parts: 1) Developing better T2I methods to achieve better image quality; 2) Developing controllable image synthesis methods to improve the controllability of the synthesis process; 3) Based on the first and second parts, introduce the image manipulation method to achieve controllable image synthesis and manipulation with high quality, thereby further improving the practicability of the synthesis method. This thesis includes the following chapters.

Chapter1 [Introduction] describes the problems of existing T2I methods, gives the solutions, and summarizes the contribution of this dissertation.

Chapter2 [High Quality Oriented Image Synthesis Methods] describes the proposed methods for high-quality image synthesis that addresses the quality deficiencies of existing methods. Specifically, this dissertation proposes three methods to achieve higher-quality image synthesis results. The basic idea of the first method is to synthesize simple contour information at first, then synthesize foreground content, and then synthesize the final image result; The basic idea of the second method is first to synthesize the foreground content based on the text information, and then synthesize the final image result based on the synthesized foreground and the input text information; The basic idea of the third method is to introduce additional image discrimination types into the discriminator to improve its discriminative ability, and then better discriminant is fed back to the generator to improve the quality of the synthesis result. Extensive experimental results have proved that the three methods proposed above all achieve higher-quality image synthesis results.

Chapter3 [High Controllability Oriented Image Synthesis Methods] presents the proposed methods for high-controllability image synthesis. This dissertation first proposes a text- based and contour-based image synthesis method, where the text can determine the basic synthetic content, and the contour can determine the shape, size and position information of the synthesized object. Since both text and contour information can be entered manually, this method achieves better controllability. In addition, based on the above idea, in order to further improve the quality of synthesized images, this dissertation designs a more complex network structure (including attention mechanism) to improve the quality of synthesis. Experimental results demonstrate that the proposed second network structure achieves better controllable and higher-quality

image synthesis results.

Chapter4 [High Practicality Oriented Image Synthesis Methods] presents the proposed text-guided image synthesis method and the resulting two schemes with high practicality. First, this dissertation proposes a text-guided image manipulation method with improved performance, which is able to modify the content of images using textual information. After that, the proposed text-guided image manipulation method with the methods proposed in Chapter2 and Chapter3 are combined to achieve two high-practicability schemes. The first scheme is text-guided image synthesis and manipulation, which allows the input text manually to synthesize the corresponding image, and then continue to input new text manually to modify the content of the synthesized image. The second shceme is text-guided controllable image synthesis and manipulation, which allows input text and simple contour information to synthesize the corresponding image result, and then can continue to input new text to modify the content of the previously synthesized image. From the experimental results, these two schemes have achieved good practicability. In contrast, the second scheme has better human controllability and practicability.

Chapter5 [Conclusion] concludes the contributions of this dissertation and describes the future plans.

As mentioned above, this dissertation proposes several deep learning-based techniques for enhancing the quality, controllability, and practicality of the generated image. By combining the proposed techniques, a high-quality, controllable, and practical image synthesis application is achieved.

２．審査結果の要旨

Using text to synthesize the corresponding images has gained widespread attention in computer vision research in recent years. This thesis presents deep learning-based algorithms to address quality, controllability, and practicality problems in current research on text-to-image synthesis. The novelty and effectiveness of this dissertation were confirmed in the following points.

1. In terms of improving the quality of image synthesis, three deep learning-based algorithms are proposed. Different from the existing algorithms, the first two proposed algorithms achieve better image quality by refining the image generation process.

Specifically, the first algorithm achieves high-quality results by generating the contour image, then the foreground image, and then the image with the background. The second algorithm improves the quality of the result by first generating the foreground image and then the image with the background. The proposed third algorithm is to improve the performance of the generator by improving the performance of the discriminator to finally achieve high-quality image results. The above three algorithms achieve high-quality image synthesis results.

2. In terms of improving the controllability of image synthesis, a text and contour based image synthesis method is proposed. In the proposed method, text is used to determine the basic composite content, while contour is used to determine the shape, size, and position information. Different from existing image synthesis methods, this idea creates a highly controllable image synthesis paradigm. Besides, based on the proposed paradigm, the thesis also proposes an effective network structure capable of synthesizing higher image quality, thus finally realizing a highly controllable and high-quality image synthesis algorithm.

3. In terms of improving the practicality of image synthesis, a deep learning-based text-guided image synthesis algorithm is proposed to achieve the effect of using text to modify image content. Afterward, the thesis combines all the proposed algorithms to form the image synthesis scheme with high quality, high controllability, and high practicability.

Based on all of these, this examination committee is unanimous that the submitted doctoral thesis is fully qualified as a Doctor of Philosophy (Engineering).