Prompt-Based Segmentation and Inpainting : A New Approach to Disaster Image Creation

Minji Choi
Information and Communication Engineering
University of Science and Technology, Korea
Daejeon, Republic of Korea
cmj@etri.re.kr

Ru-Bin Won
Information and Communication Engineering
University of Science and Technology, Korea
Daejeon, Republic of Korea
rubrub@etri.re.kr

Ji Hoon Choi
Media Intellectualization Research Lab
ETRI (Electronics and Telecommunications Research Institute)
Daejeon, Republic of Korea
cjh@etri.re.kr

Byungjun Bae
Media Intellectualization Research Lab
ETRI (Electronics and Telecommunications Research Institute)
Daejeon, Republic of Korea
1080i@etri.re.kr

Abstract— **This paper proposes a novel approach to disaster image generation using prompt-based segmentation techniques. By segmenting terrains based on the provided prompt and inputting disaster-related prompts into the segmented area, we explore a method to generate images that reflect disaster scenarios. Moreover, we propose a method to adjust the inpainting mask area according to the severity of the disaster, providing a visual representation that varies with the situation. We acknowledge limitations in areas such as the inpainting mask region and the overall disaster image generation, and suggest directions for further research to overcome these challenges. We emphasize that the methodology of our study contributes significantly to disaster management and information dissemination.**

Keywords—image segmentation, image inpainting, disaster alert, disaster image, image processing, text to image model

I. INTRODUCTION

In effectively communicating and explaining urgent disaster situations to the public, providing images with disaster alerts plays a important role. Especially for groups like children, the elderly, individuals with cognitive impairments, and foreigners, who might find it challenging to quickly recognize disaster warnings, they are relatively more vulnerable to be exposed to disasters. For these populations susceptible to disaster recognition, it's essential to have disaster images that can assist in immediate situation assessment. Furthermore, these disaster images should accurately depict the type of disaster and the terrain where it has occurred.

However, most of the current Text-to-Image models have been trained on extensive datasets, tend to be more proficient in depicting prominent landmarks and are often lacking when it comes to general terrains [1]. Additionally, Text-to-Image models sometimes face challenges when trying to aggregate multiple prompt words into a single coherent image output [2]. Consequently, disaster images generated through conventional image creation models may not accurately reflect the intensity, type, and terrain of disasters occurring domestically, making them less suitable for real-world disaster alerts within the country.

Considering these challenges, we propose applying image inpainting techniques for disaster image generation. The goal is to provide citizens with more detailed and diverse visual information. This paper presents the starting point for such disaster image creation and delves into the specific approaches taken.

II. METHOD

In this paper, we propose a novel approach that integrates image segmentation and inpainting techniques to distinctly represent the intensity of disaster notifications. In alignment with the two levels of disaster alerts recognized in Korea, specifically advisory and warning, we assign different extents of inpainting regions to generate images capable of distinguishing the severity of disasters [3]. Utilizing the provided object prompts, we first segment objects within the image. Following this, we suggest disaster-related prompts to inpaint disaster scenarios within the segmented object areas. By combining these two techniques, we produce effective images that comprehensively capture the intensity, type, and topography of the disaster.

A. Image Segmentation

Image segmentation is an extended field of image classification that identifies objects within an image, discerns their location and contours, and assigns individual labels to each object region [4].

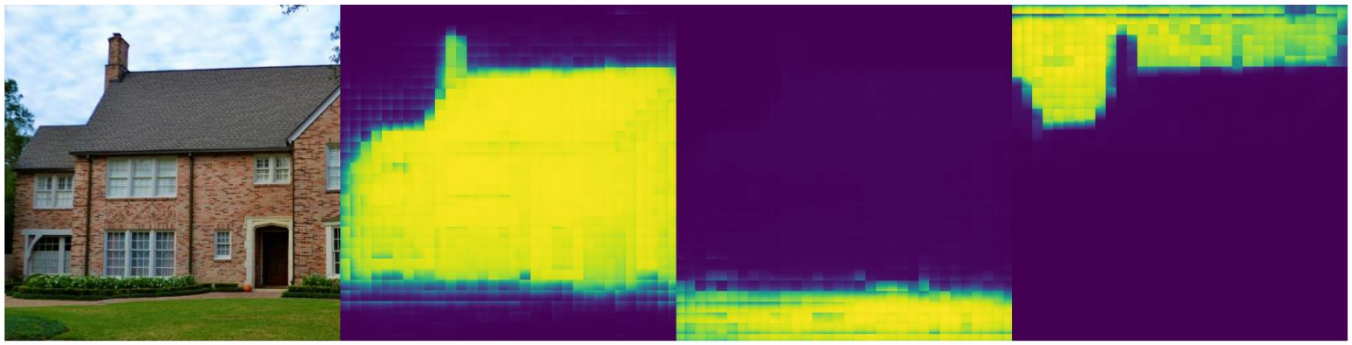

Fig. 1. Segmentation results

In this paper, we apply the CLIPSeg model, which utilizes the Transformer as its image segmentation model, to images [5,6,7,8]. A notable limitation of most image segmentation models is that they operate solely on a fixed list of categories. Thus, when one wishes to label data with a new category, a new model needs to be trained, necessitating significant cost and time. In contrast, CLIPSeg offers a Zero Shot Segmentation model, capable of segmenting almost any kind of object without additional training [9,10]. Moreover, both the query and prompt in CLIPSeg are inputted as CLIP image embeddings. Using these two inputs, CLIPSeg produces a binary segmentation mask.

In the case of the Transformer-based CLIPSeg model, a Transformer-based decoder is trained on top of the CLIP model to achieve zero shot image segmentation. The decoder receives the CLIP representation of the input image and the CLIP representation of the desired object for segmentation. The segmentation results are represented as an attention mask.

B. Image Inpainting

Image inpainting is a technique for regenerating and restoring damaged or missing parts of an image. When users specify unwanted content or areas they wish to modify within an image, the method analyzes the surrounding pixel patterns and textures to naturally restore the specified area [11].

In this paper, we utilize the inpainting technique from Stable Diffusion [12,13]. Inpainting in stable diffusion operates by applying a diffusion process to the image pixels surrounding the damaged area. This process assigns values to the image pixels and determines them differently based on the proximity to the damaged area. By employing the diffusion inpainting technique, we aim to create natural disaster images. In this study, the attention areas represented by CLIPSeg are designated as the inpainting mask areas. If the input disaster alert is a warning, the inpainting mask region is expanded to generate images that further emphasize the severity of the disaster situation.

III. Experiments & Results

All experiments are conducted using PyTorch and Python 3.10.12 in a Google Colab GPU environment. All images are sourced from: https://pixabay.com/. Due to the stable diffusion model being based on a 512x512 image size, images are cropped to fit this dimension before processing.

A. Segmentation & Inpainting Results

Figure 1 shows the results of segmentation by directly inputting three prompts: house, ground, and sky. One can visually confirm that the attention regions are appropriately represented based on the entered prompts. In the case of house segmentation, details as intricate as the chimney are included. Moreover, the sky segmentation not only excludes the house and ground but also distinctly separates and excludes the tree on the left, thereby faithfully representing the 'sky' prompt. These results corroborate the high performance of the CLIPSeg model in prompt-based segmentation.

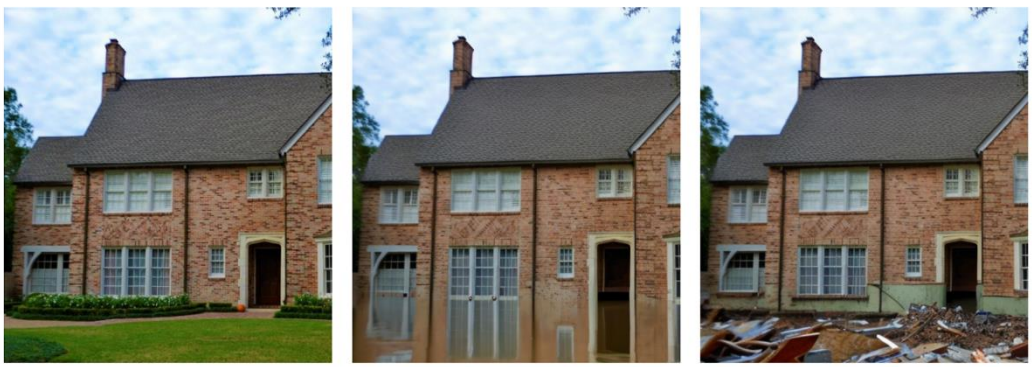

Fig. 2. Disaster inpainting results

Fig. 3. Inpainting area expansion results

Figure 2 shows both the inpainting prompts and their corresponding image inpainting outcomes. In (a), a flood is inpainted into the ground region, resulting in an image where the lower portion appears entirely submerged in water. In (b), by inputting the term "collapse" into the ground area, an image is produced that resembles the aftermath of an earthquake or a landslide.

By inpainting disaster scenarios into segmented areas, a variety of disaster images can be generated in line with the entered prompts and user intentions. Such crafted disaster images maintain the original image's topography while reflecting diverse disaster situations.

B. Mask Area Expanded Results

Disaster alerts in Korea are divided into two categories: advisory and warning. To visually convey these differences in severity, an expansion of the inpainting mask area is proposed. When a disaster warning is entered, the inpainting mask area is expanded to generate a disaster image that emphasizes the severity and importance of the disaster even more.

Figure 3 shows: (a) the result of inpainting in the original segmentation area, and (b) the image inpainted after expanding the inpainting mask area. Although flooding was inpainted, the segmentation area was classified based on the 'ocean' prompt. As a result, aside from the color difference from the original image's sea, there is no significant change that distinctly represents the flood. When the inpainting area was expanded, the sea level rose, making the flooding situation appear much more severe than in the previous image.

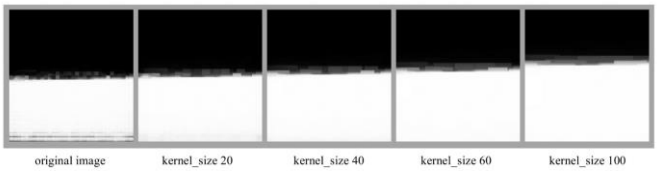

Fig. 4. Masked area for each kernel size

The expansion of the inpainting area is achieved by adjusting the kernel size of the mask. Figure 4 shows the mask area designated for different kernel sizes. As the kernel size number increases, one can observe a gradual enlargement of the inpainting mask area. Based on these results, by expanding the inpainting area through kernel size, we can create disaster images that adjust the severity of the situation depending on the distinction between disaster advisory and disaster warning.

Figure 5 shows the inpainting results for each kernel size. The difference between the original image and image (a) is minimal, excluding the color change in the inpainted area. As the kernel size number grows, there are noticeable changes in the sea level and topography. However, as the kernel size area continues to increase, trees or buildings that were not present originally start to emerge. Starting from the kernel size of 40, the emergence of trees that were not originally present and the deformation of the bridge's form become apparent. Based on these outcomes, it seems inadvisable to input excessively large numbers for the kernel size during the inpainting area expansion process. Further research is required to determine the appropriate kernel size.

Fig. 5. Image inpainting results according to kernel size

IV. Conclusion

In this paper, we propose the creation of disaster images through inpainting by first segmenting terrains using prompt-based segmentation and then entering disaster prompts into those segmented areas. Furthermore, we offer a visual representation suited to the severity of the disaster situation by expanding the inpainting area.

Currently, most Text-to-Image models fail to accurately reflect the specific terrains of particular countries. Additionally, training a model to understand a country's entire topography through additional learning requires significant computational resources, time, and training costs. The inpainting technique proposed in this study overcomes these limitations, suggesting a simple and efficient method for disaster image creation. This method is anticipated to be more accurate and effective for disaster management and information dissemination. Moreover, this inpainting technique isn't limited to just creating disaster images; it can be utilized for other specific purposes. It showcases the potential to be applied to various image creation challenges, paving the way for more extensive future research.

However, there are limitations in the current research on disaster image creation through image inpainting. First, there is the issue of adjusting the size of the mask area. In this study, we suggested expanding the mask area using kernel size. However, reducing the mask area using kernel size is technically complex, underscoring the need for an appropriate method. Secondly, for disasters like heavy rain or snowfall, most or all of the image needs to be inpainted, which is a limitation in the current technology that segments only specific parts through attention.

To overcome these limitations, further research to decrease the inpainting mask area is planned, with the intent of incorporating various disaster scenarios. Additionally, for the creation of comprehensive disaster images like heavy rain or snowfall, we aim to carry out additional research on image style transformation or image synthesis.

Acknowledgment

"This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (2022-0-00083, Development of customized disaster media service platform technology for the vulnerable in disaster information awareness)"

References

[1] M. Choi, RB. Won, JH. Choi, and B. Bae, “An Analysis of Diffusion-Based Disaster image Generation Results for Vulnerable Populations,” in press.

[2] H. Chefer, Y. Alaluf, Y.Vinker, L. Wolf, and D. Cohen-Or, "Attend-and-excite: Attention-based semantic guidance for text-to-image diffusion models," arXiv preprint arXiv:2301.13826, 2023.

[3] https://www.weather.go.kr/w/weather/warning/standard.do

[4] Minaee, Shervin, et al. "Image segmentation using deep learning: A survey," IEEE transactions on pattern analysis and machine intelligence 44.7, 2021: 3523-3542.

[5] Radford, Alec, et al. "Learning transferable visual models from natural language supervision," International conference on machine learning. PMLR, 2021.

[6] Lüddecke, Timo, and Alexander Ecker. "Image segmentation using text and image prompts," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.

[7] https://github.com/timojl/clipseg

[8] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.

[9] P. Li, Y. Wei, and Y. Yang, "Consistent structural relation learning for zero-shot segmentation," Advances in Neural Information Processing Systems 33, 2020: 10317-10327.

[10] Y. Xian, Ch. Lampert, B. Schiele, and Z. Akata, "Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly," IEEE transactions on pattern analysis and machine intelligence 41.9, 2018: 2251-2265.

[11] O. Elharrouss, N. Almaadeed, S. AI-Maadeed, and Y. Akbari, "Image inpainting: A review," Neural Processing Letters 51, 2020: 2007-2028.

[12] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022.

[13] https://huggingface.co/stabilityai/stable-diffusion-2-inpainting