Prompt-based Image Coding with Edge Information

Hiroshi Watanabe[†] Luoxu Jin[†] Taiga Hayami[†] Takeshi Chujoh[‡] Tomoko Aono[‡]
Yukinobu Yasugi[‡] Sujun Hong[‡] Zheming Fan[‡] Tomohiro Ikai[‡]

Waseda University[†] Sharp Corporation[‡]

Abstract: This paper targets image coding for recognition that does not require visual reproducibility. The image is generated by extracting prompt from the image. A seed for random variables and prompt are fed into the network with edge information. In this method, the quality of the generated image varies depending on the seed value of the random variable. By searching for a random variable that maximizes the existing evaluation index, the quality of the generated image is improved.

1 Introduction

Three component image decomposition approach has been proposed more than 30 years ago [1]. It decomposes an image into edge, primally component, and residuals. Then, several edge-based image compression approaches have been investigated [2]-[4]. They can provide reasonable coded image quality at very low bitrate. This feature may not be suited in the sense of rate-distortion competition, but it could be suitable for machine recognition.

In recent years, the technology to automatically generate images by AI has been rapidly evolving. One approach is to convert text to image by diffusion model. Stable Diffusion, developed by Ludwig Maximilian University of Munich and released by Stability AI, is a publicly available text-to-image generation tool [5]. Then, applying a diffusion model to image compression has been proposed [6]-[8]. However, a generated object shapes are sometimes different from an original image.

In 2023, a plug-in extension to Stable Diffusion was released for controlling object shape. The extension named “ControlNet” can incorporate edge, depth, or skeleton information into Stable Diffusion [9]. In this scheme, edge information can be used to constrain the object shape generated from the text.

We propose a new image coding scheme, which generates an image from text with edge information. The target is semantic compression like VCM (Video Coding for Machines) in MPEG. Prompts that explain an input image, and edge information of objects in the image are extracted. At the decoder, the output image is generated by random variables and prompt with the constrained edge information.

2 Coding Scheme

Coding scheme consists of analysis part of prompt and extraction of edge information shown in Fig.1. To derive prompt, Stable Diffusion has a capability to convert an image to texts. Canny filter can be used for extracting edge in a ControlNet extension of Stable Diffusion. Depends on an initial value (seed) of random variable, results vary drastically. Thus, selection criteria of seed should be defined properly.

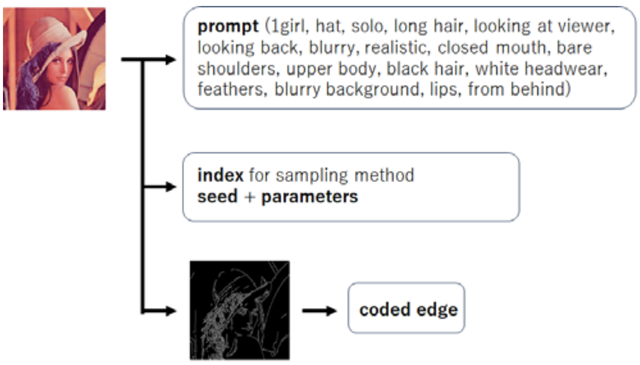


Fig. 1. Coding scheme of the proposed method.

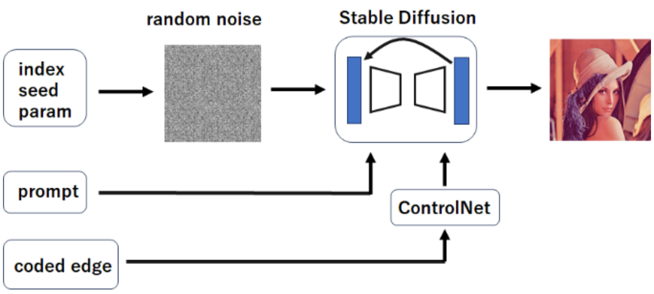


Fig. 2. Decoding scheme of the proposed method.

Decoding process needs index that specify the type of random function, seed value for the function, several control parameters, as well as edge information which is recovered from the coded data.

3 Experimental Results

When Lena is assumed as an input image, the number of characters for prompt is 170, so that 1190bit (170*7bit) is necessary. The dominant amount of the data is canny edge information, which is 4395 Byte as JBIG format for the input 512*512 pel image. Thus, the image compression ratio is around 0.14 bit/pel. To maximize the quality of the coded image, we set the criterion J regarding the type of random variables, seeds, prompts, and other control parameter.

$$J = \min_{index,seed,promt\dots} D(I_{in}, I_{gen}) \qquad (1)$$

where the distortion measure $D(*,*)$ can be a combination of SNR, SSIM, and LPIPS. In the image generation process, the number of batches can be set. By setting the batch number 100, we obtain 100 images. The variation of image quality for SSIM is shown in Fig. 3. The example of generated images is shown in Fig.4 (a), and a typical decoded image is shown in Fig.4 (b).

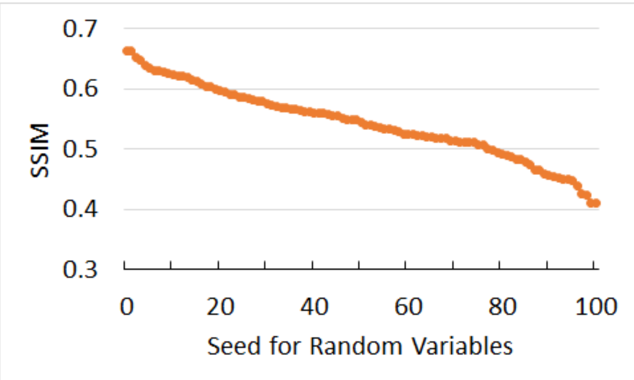
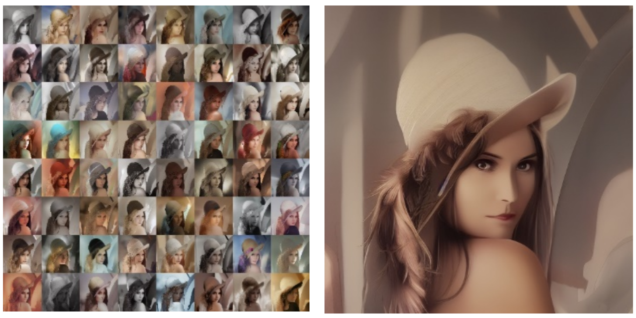


Fig. 3. SSIM variation.



(a) Examples of 64 batches. (b) Typical decoded image.

Fig. 4. The example of generated images.

4 Conclusion

In this paper, we propose a new coding scheme based on prompt and edge information using stable diffusion. Simulation results show that reasonable image quality can be obtained at very low bitrate.

References

[1] Stefan Carlsson: "Sketch Based Coding of Grey Level Images," Signal Processing, No.15, pp.57-83, 1988.

[2] X. Ran and N. Farvadin: "A Perceptually Motivated Three-Component Image Model – Part II: Applications to Image Compression", IEEE Transaction on Image Processing, Vol. 4, No. 4, pp.430-447, Apr. 1995.

[3] Yuji Ito: "An Edge-Oriented Progressive Image Coding", IEEE Transaction on Circuits and Systems for Video Technology, Vol. 6, No. 2, pp.135-142, Apr. 1996.

[4] M. Nakashizuka and H. Kikuchi: "Edge-Based Image Synthesis Model and Its Application to Image Coding", IEEE International Symposium on Circuits and Systems (ISCAS1999), Vol. 4, pp.25-28, May 1999.

[5] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer: "High-Resolution Image Synthesis with Latent Diffusion Models", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.10674-10685, June 2022.

[6] L. Theis, T. Salimans, M. Hoffman and F. Mentzer: "Lossy Compression with Gaussian Diffusion", https://doi.org/10.48550/arXiv.2206.08889, Dec. 2022.

[7] R. Feng, J. Liu, X. Jin, X. Pan, H. Sun and Z. Chen: "Prompt-ICM: A Unified Framework towards Image Coding for Machines with Task-driven Prompts", https://doi.org/10.48550/arXiv.2305.02578, May 2023.

[8] B. Kim, H. Song, T. Castells and S. Choi, "On Architectural Compression of Text-to-Image Diffusion Models", https://doi.org/10.48550/arXiv.2305.15798, May 2023.

[9] L. Zhang, A. Rao, M. Agrawala: "Adding Conditional Control to Text-to-Image Diffusion Models", https://doi.org/10.48550/arXiv.2302.05543, Feb. 2023.

School of Fundamental Science and Engineering, Waseda University, Shillman Hall 401, 3-14-9 Okubo, Shinjuku, Tokyo, 169-0072
Phone: 03–5286–2509, Fax:03–5286–3488
E-mail: hiroshi.watanabe@waseda.jp