The Effect of Edge Information in Stable Diffusion Applied to Image Coding

Hiroshi Watanabe
Waseda University

Luoxu Jin
Waseda University

Taiga Hayami
Waseda University

Takeshi Chujoh
Sharp Corporation

Yukinobu Yasugi
Sharp Corporation

Sujun Hong
Sharp Corporation

Zheming Fan
Sharp Corporation

Tomohiro Ikai
Sharp Corporation

Abstract

We focus on image coding schemes whose primary purpose is recognition in computer vision rather than consumers viewing. The input image is decomposed into prompt, hyperparameter, and edge information by Stable Diffusion. For decoding, we use the same diffusion model to sequentially remove noise determined by the sampling method and the initial value of random variables. The quality of the generated image depends on the fineness of the edge information. In this study, we investigate the effect of the amount of edge information on the quality of the decoded image.

Keywords: image coding, diffusion model, stable diffusion, variational autoencoders, text-to-image, canny edge detector

1. Introduction

Image coding using text-to-image diffusion model has been actively studied. One of the interesting functions included in Stable Diffusion is image generation with edge constraints. Using the shape information of objects in the image as control information, it is possible to generate a level of encoded images that can be used for recognition.

Originally, image coding using edge information 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 suitable in terms of rate-distortion competition but may be suitable for machine recognition.

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, generated object shapes are sometimes different from an original image. A plug-in extension to Stable Diffusion was released for controlling object shape in 2023. 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. Recently, other approaches using soft-edge in text-to-image coding have been studied [10],[11].

We propose a new image coding scheme, which generates an image from text with edge information. The target is semantic compression like Video Coding for Machines (VCM) 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 and Decoding 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 edge detector can be used for extracting edges 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.

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 as shown in Fig. 2.

3. Coding Control

An object shape reproducibility in Stable Diffusion depends on the amount of edge information. Canny edge detector, used in ControlNet, includes noise reduction, Sobel filter, non-maxim suppression, and thresholding operations. We can control edges by setting high side threshold (Ht) for primary edges and low side threshold (Lt) for additional details.

The basic quality of the coded image can be maximized by setting the criterion *J* regarding hyper parameters, such as, the type of random variables, seeds, prompts, classifier-free guidance scale (CFG) and other control parameters

$$J = \min_{index,\ seed,\ prompt,\ \dots} D\,(I\_in,\ I\_gen) \qquad (1)$$

where *I_in*, *I_gen* are input and generated images, and the distortion measure *D(*,*)* can be a combination of SSIM,

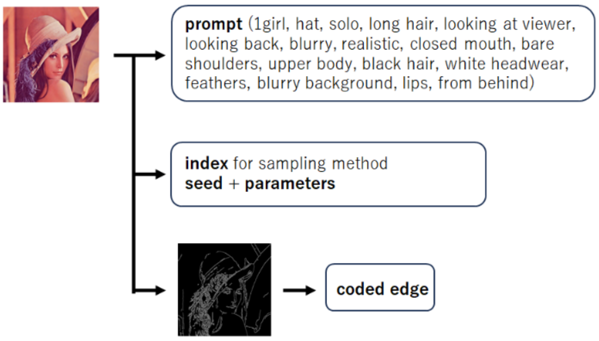
prompt (1girl, hat, solo, long hair, looking at viewer, looking back, blurry, realistic, closed mouth, bare shoulders, upper body, black hair, white headwear, feathers, blurry background, lips, from behind)
index for sampling method
seed + parameters
coded edge

Figure 1: Coding scheme of the proposed method.

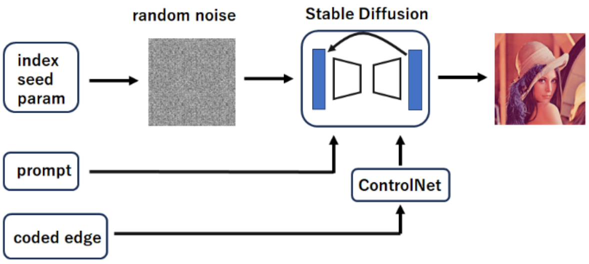

Figure 2: Decoding scheme of the proposed method.

LPIPS, SNR or other human satisfaction measure [12]. Once the distortion measure is defined, we pick up the one which maximize Eq. (1). Depends on the selected parameters, the quality of the coded image drastically changes.

4. Experiment

We experiment to observe the effect of edges on the coded image quality provided by Stable Diffusion. In the experiment, we use Lenna, 512x512 pels, 8bit monochrome image. First, prompt is derived from the image as shown in Fig. 1. Examples of edge images and corresponding coded images are shown in Fig. 3.

We use the following conditions. Steps: 20, Sampler: Euler a, CFG scale: 1, Seed: 3515612223, Size: 512x512, Model: v1-5-pruned-emaonly, ControlNet: "Module: canny, Model: control_canny-fp16 [e3fe7712], Resize Mode: Crop and Resize, Low Vram: False, Processor Res: 512, Version: v1.6.0-2- g4afaaf8a. The quality characteristics for different edge images are shown in Fig. 4. The file size is dominated by edge image coded by JBIG. It varies from about 5 KB to 14KB. The data size or prompts and hyperparameters are about 0.5KB. SSIM and LPIPS are used as evaluation indicators. A weak peak appears at Lt=Ht=100 in the canny edge detector.

From experiments, edges that can represent major object shapes and do not represent detailed textures provide better quality. The original and coded images are shown in Fig. 5. The shapes are similar, but the gray-level areas are different. The overall impression appears to be reasonably well preserved.

5. Conclusion

This paper presents experimental results on how edge information affects image coding using Stable Diffusion. Edges that cover major objects but do not represent detailed textures provide better quality.

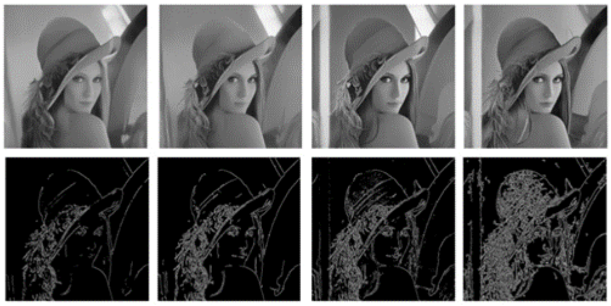

Figure 3: Coded “Lenna” images (upper) and corresponding canny edges (lower). From left, Lt-Ht: 200-200, 100-200, 100-100, 25-100.

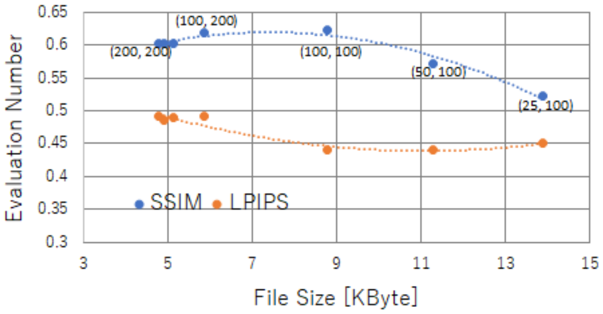

Figure 4: Numerical evaluation results by SSIM(↑) and LPIPS(↓) to generated images using 7 different edge images. Note that PSNR values are almost the same in all cases.

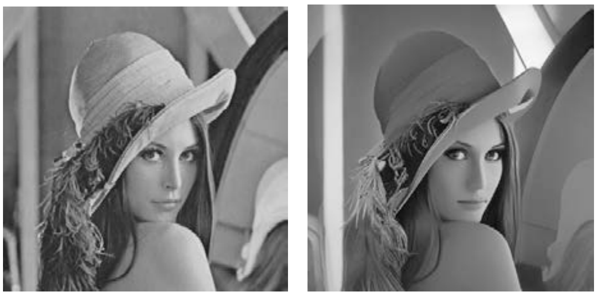

Figure 5: Original Lenna, 512*512 pels (left) and coded image (right) when using Canny edge detector with Lt=Ht=100 (SSIM=0.624, LPIPS=0.44, PSNR=15.05, Size=5.87KB).

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 ISCAS1999, Vol. 4, pp.25-28, May 1999.

[5] R. Rombach, et al.: “High-Resolution Image Synthesis with Latent Diffusion Models”, IEEE/CVF CVPR, pp.10674-10685, Jun. 2022.

[6] L. Theis, et al.: “Lossy Compression with Gaussian Diffusion”, arXiv.2206.08889, Dec. 2022.

[7] R. Feng, et al. “Prompt-ICM: A Unified Framework towards Image Coding for Machines with Task-driven Prompts”, arXiv.2305.02578, May 2023.

[8] B. Kim, et al.: “On Architectural Compression of Text-to-Image Diffusion Models”, arXiv.2305.15798, May 2023.

[9] L. Zhang, et al.: “Adding Conditional Control to Text-to-Image Diffusion Models”, arXiv.2302. 05543, Feb. 2023.

[10] E. Lei, et al.: “Text+Sketch: Image Compression at Ultra Low Rates,” arXiv.2307.01944, Jul. 2023.

[11] M. Careil, et al.: “Towards Image Compression with Perfect Realism at Ultra-Low Bitrates,” arXiv.2310.10325, Oct. 2023.

[12] Tsachy Weissman: “Toward Textual Transform Coding,” arXiv.2305.01857, May 2023.