Foveation Improves Payload Capacity in Steganography

Anonymous Author(s)

Abstract

Steganography finds its use in visual medium such as providing metadata and watermarking. With support of efficient latent representations and foveated rendering, we trained models that improve existing capacity limits from 100 to 500 bits, while achieving better accuracy of up to 1 failure bit out of 2000, at 200K test bits. Finally, we achieve a comparable visual quality of 31.47 dB PSNR and 0.13 LPIPS, showing the effectiveness of novel perceptual design in creating multi-modal latent representations in steganography.

CCS Concepts

• **Computing methodologies** → **Image representations**; **Perception**; *Learning latent representations*; *Image processing*; Reconstruction; • **Computer systems organization** → Redundancy; • **Information systems** → Multimedia content creation.

Keywords

Steganography, Perception, Latent Representations

ACM Reference Format:

Anonymous Author(s). 2025. Foveation Improves Payload Capacity in Steganography. In *Proceedings of SIGGRAPH Asia 2025 Posters (SA Posters' 25).* ACM, New York, NY, USA, 2 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Steganography concerns about hiding data in another medium [Wang et al. 2023]. More specifically, our work studies embedding information in images. Steganography is therefore useful in conveying multi-modal information such as labels, scene descriptions, or copyright marks. Growing number of AI-generated content and introduction of AR/VR systems increases its importance further by broadening the application scope [Rezaei et al. 2024].

Our work leverages latent representations [Yılmaz et al. 2024] and a foveated rendering loss [Walton et al. 2022] to increase payload capacity in steganography. With only 2000 training images, we achieve bit accuracy of 99.99% for 40K test set bits. Concerning state of the art in latent methods [Bui et al. 2023], our approach increases the payload capacity from 100 bits up to 500 bits with up to 100% recovery under non-distortion condition, 27.56 db PSNR and 0.26 LPIPS. Our final contribution is the introduction of Metameric Foveated Rendering loss in steganography, which noticeably improves all visual metrics and quality with respect to classic L2 loss.

SA Posters' 25, Hong Kong, China

ACM ISBN 978-1-4503-XXXX-X/2025/06
https://doi.org/XXXXXXX.XXXXXXX

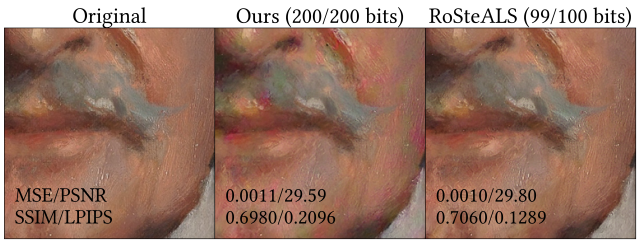

Figure 1: Visual example of stego and payload recovery using our proposed foveated steganography and RoSteALS (Source: MetFaces [Karras et al. 2020]).

2 Method

Considering a message payload $P \in \{0,1\}^k$ consisting of k bits and a input image (cover) $I \in \mathbb{R}^{h\times w\times c}$, find two functions H and R, to produce an output image (stego), $H(I,P) = I' \in \mathbb{R}^{h\times w\times c}$, and output payload, $R(I') = P' \in \{0,1\}^k$. The aim is to reduce the distortion between I and I' while maximizing the accuracy between P and P'.

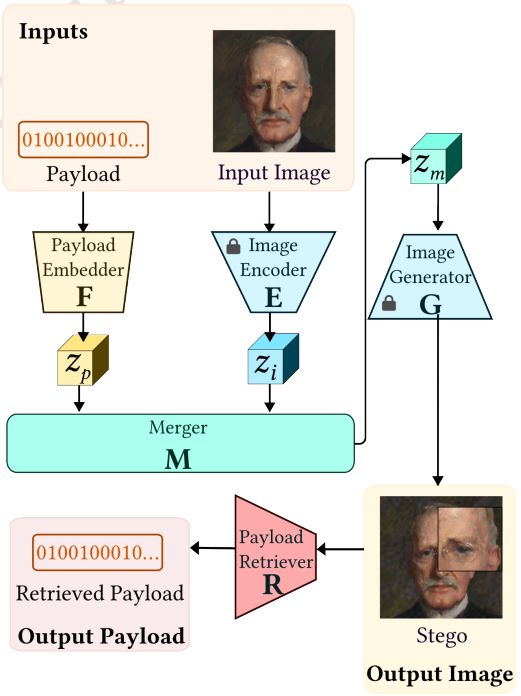

Figure 2: Our proposed foveated steganography approach (Source: MetFaces [Karras et al. 2020]).

Our framework approaches this problem as depicted in Figure 2. In the hiding stage, a frozen image encoder, E, transforms input image into a latent representation, $E(I) = Z_i$. Payload embedder, F, creates also a learned representation, $P(I) = Z_p$. Together,

Table 1: Table of results. Baseline is trained on MSE, opposed to Metameric-Trained which replaces it with Metameric Loss.

Experiment	Resolution	Capacity	Bit Accuracy	MSE	PSNR	SSIM	LPIPS	Metameric Loss
RoSteALS	256	100	0.9942	0.0009	32.16	0.8971	0.0780	0.0016
Baseline	256	100	1	0.0015	28.90	0.8833	0.1621	0.0072
Baseline	256	200	1	0.0018	28.37	0.8681	0.2047	0.0072
Baseline	256	500	0.9997	0.0021	27.56	0.8348	0.2613	0.0077
Metameric-Trained	256	200	0.9998	0.0010	31.47	0.8871	0.1288	0.0017
Baseline	128	500	0.4997	0.0007	33.11	0.9061	0.0570	0.0023
Metameric (still training)	512	2000	0.9097	0.0070	-	-	-	0.0280

they are manipulated by the merger, M, producing a merged latent, $M(Z_i, Z_p) = Z_m$, which a frozen image generator, G, uses to reconstruct the output image, $G(Z_m) = I'$. Finally, a payload retriever, R, extracts the output payload, $R(I') = P'$. The loss function is defined as the combination of payload and image quality losses, being BCE and Metameric Foveated Rendering [Walton et al. 2022] (defaulted to center) losses respectively. Formally, $\mathcal{L}_{total} = \mathcal{L}_{payload} + \lambda_i \cdot \mathcal{L}_{image} = \text{BCE}(P, P') + \lambda_i \cdot (\text{MetamericLoss}(I, I'))$, where λ_i controls the trade-off between the two losses.

The dataset is a balanced mixture of 2000 training, 400 validation, and 400 test images from MetFaces [Karras et al. 2020] and CLIC datasets [Toderici et al. 2020]. For preprocessing, images are randomly cropped and padded to the size of input and normalized as autoencoder requires. Notably, this dataset is much smaller than typical datasets used for the same purpose, but is found sufficient to learn performing 100-bits steganography, within controlled computing resources, about two hours on a single RTX 4090 GPU.

3 Results and Discussion

The frozen pair of image encoder and image generator to create a high-quality latent representation, is the F4-with-attention version autoencoder from LDM VQGAN series [Rombach et al. 2022]. After evaluating empirically, we found its high reconstruction quality is suitable for the embedding process. Compared to other backbones, this one converges slower at payload embedding, but achieves better image quality in the end. Keeping payload embedder as a fully connected layers is sufficient to encode the information after experimentation. For merger, the best performing architecture is adding two convolutional layer sandwiching the sum of image and payload latent, to soften the transition. Finally, ResNet50 was used as payload retriever as a popular and well-studied architecture.

Apart from common metrics, we also report Metameric Loss, which is a perceptual criterion akin to foveated gaze. Modeling the human visual system, this loss is more forgiving of visual distortions in the periphery and more harsh in the fovea.

The main results are shown in Table 1. Our baseline is trained in multiple resolutions and payload capacities, including the native resolution of RoSteALS [Bui et al. 2023], the benchmarked method. The baseline in the smallest setting achieves a bit accuracy of 99.99%, failing to decode only 4 out of 40K test bits. Noticeably, we achieve 100% recovery in the same setting as the benchmark, while other settings also all exceed 99.95% compared to benchmark failing to reach 99.5%. Nevertheless, RoSteALS has better perceptual image quality, although it could be explained with lack of incorporation of LPIPS and finer-grained optimization in the training.

Another important observation is that Metameric Loss consistently and significantly improves the quality of the reconstructed images while keeping same level of bit accuracy. Compared to the benchmark, Figure 1 shows an example of resulting stego and recovered payload. This shows effectiveness of this visual technique in enhancing perceptual fidelity of images.

Despite successfully unlocking higher message length, we notice tangible limits of payload capacity, such as failing to learn 500-bit payload at 128 resolution. Resolution bounds the upper payload capacity under similar perceptual fidelity of images, and we hope to enhance this by introducing gaze as a new parameter. This work provides a light-weighted, human-centered, latent-based steganography framework which boosts payload capacity and accuracy while maintaining image quality. By satiating the need of large capacity in transmitting messages, we step towards practical applications of steganography in real-world scenarios.

References

Tu Bui, Shruti Agarwal, Ning Yu, and John Collomosse. 2023. Rosteals: Robust steganography using autoencoder latent space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 933–942.

Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. 2020. Training generative adversarial networks with limited data. *Advances in neural information processing systems* 33 (2020), 12104–12114.

Ahmad Rezaei, Mohammad Akbari, Saeed Ranjbar Alvar, Arezou Fatemi, and Yong Zhang. 2024. Lawa: Using latent space for in-generation image watermarking. In *European Conference on Computer Vision*. Springer, 118–136.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

George Toderici, Wenzhe Shi, Radu Timofte, Lucas Theis, Johannes Balle, Eirikur Agustsson, Nick Johnston, and Fabian Mentzer. 2024, 2022, 2021, 2020. Workshop and challenge on learned image compression (CLIC2020 to CLIC2024). In *CVPR*.

David R Walton, Koray Kavaklı, Rafael Kuffner Dos Anjos, David Swapp, Tim Weyrich, Hakan Urey, Anthony Steed, Tobias Ritschel, and Kaan Akşit. 2022. Metameric varifocal holograms. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 746–755.

Zihan Wang, Olivia Byrnes, Hu Wang, Ruoxi Sun, Congbo Ma, Huaming Chen, Qi Wu, and Minhui Xue. 2023. Data hiding with deep learning: A survey unifying digital watermarking and steganography. *IEEE Transactions on Computational Social Systems* 10, 6 (2023), 2985–2999.

Doğa Yılmaz, Towaki Takikawa, Duygu Ceylan, and Kaan Akşit. 2024. Learned Single-Pass Multitasking Perceptual Graphics for Immersive Displays. *arXiv preprint arXiv:2408.07836* (2024).