

Supplementary Materials: GAN-based Symmetric Embedding Costs Adjustment for Enhancing Image Steganographic Security

Anonymous Authors

1 APPENDIX

In the supplementary material, we will present additional results concerning the proposed method, all derived using the baseline steganography HILL with a payload of 0.4 bpp. 1.1 will provide the cover image and its corresponding stego image generated by the proposed method. A.2 will depict the modification maps generated by the proposed method alongside those produced by other relevant methods. Additionally, the modification maps for each sub-image will also be given. A.3 will showcase the loss curves of both the generator and discriminator for different combinations of D_1 and D_2 as well as with different strategies.

1.1 Cover and Stego Image

Unlike other GAN-based image generation tasks, the goal of GAN-based steganography is to embed secret information into images without visual perception. Consequently, the cover image and the stego image are visually indistinguishable. Figure 1 demonstrates the cover image and its corresponding stego image generated by the proposed method, with a modification rate of approximately 10%. From Figure 1, it is difficult to differentiate the cover and stego image with the naked eye. Additionally, they also exhibit high values in SSIM (0.99) and PSNR (57.92 dB), indicating the superior imperceptibility of the proposed method.

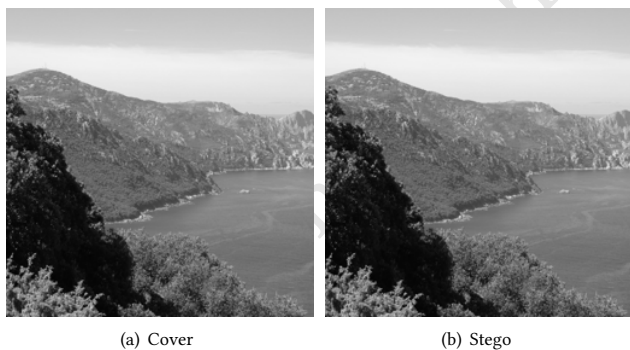


Figure 1: The cover image and the corresponding stego images generated in the proposed method.

1.2 Modification Maps

Figure 2 illustrates the local modification maps of the proposed method with other related methods, including CMD, UGS and ReLOAD. Notably, CMD-HILL demonstrates the most noticeable concentration effect compared to the other methods. This can be attributed to CMD's design principle of synchronizing modification directions. Following CMD, the modifications generated by the proposed method also exhibit a clustering effect. However, these

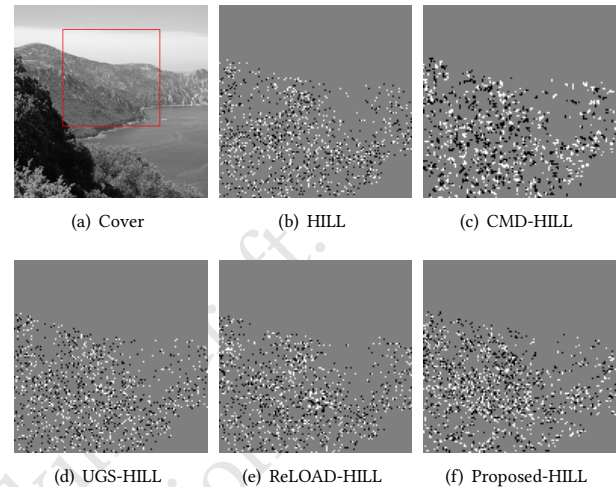


Figure 2: The cover image and the corresponding local modification maps of methods on embedding costs adjustment. In the following modification maps, the white dots represents +1 and the black dots represent -1.

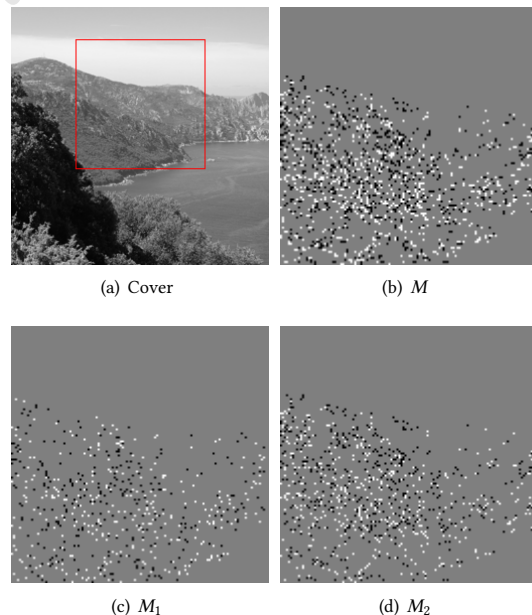


Figure 3: The cover image with the corresponding final modification maps, as well as those in each sub-image in the proposed method.

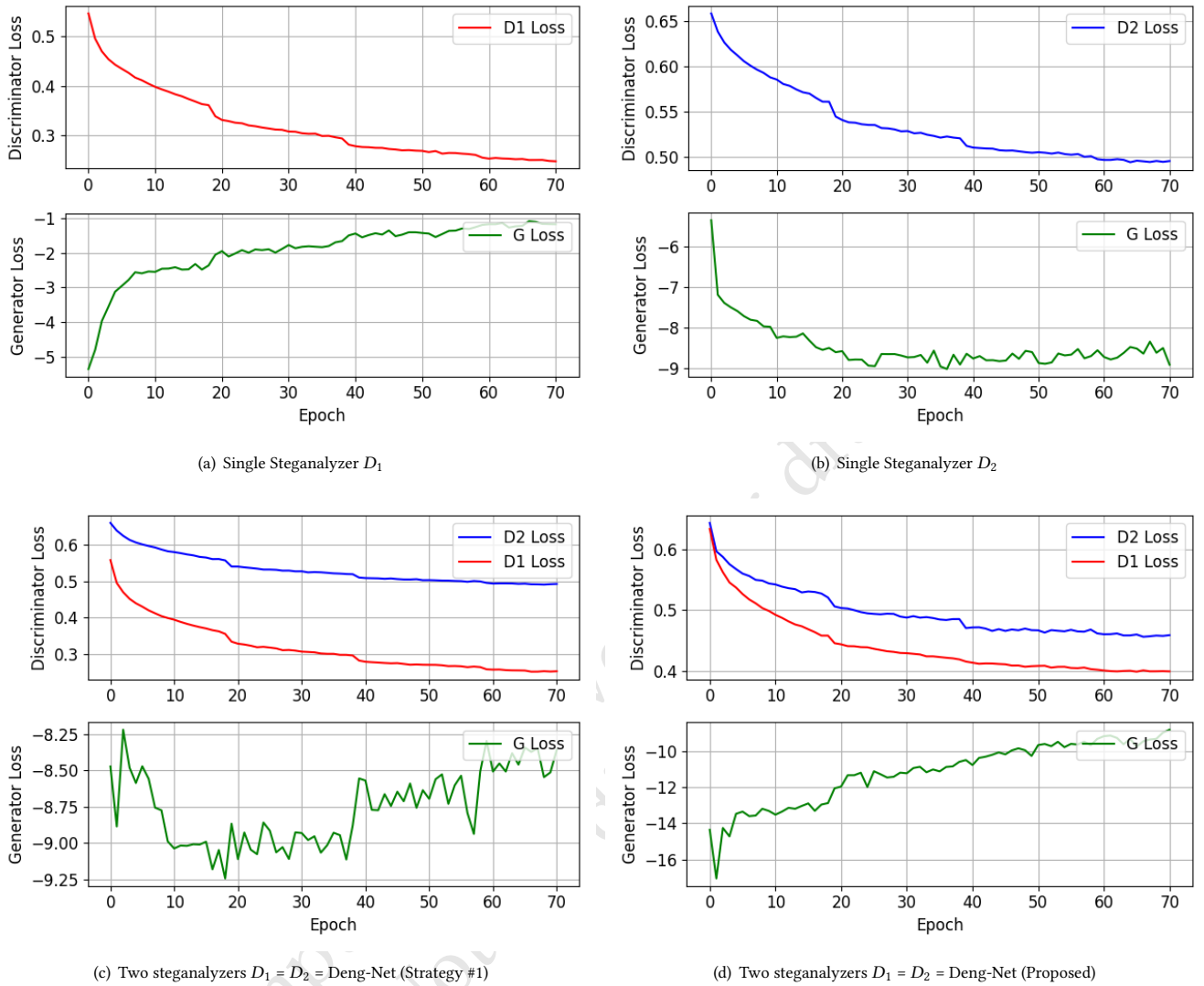


Figure 4: The loss curves of the generator and discriminator with different combinations of D_1 and D_2 .

patterns are learned by the generator under the diverse guidance provided by the discriminator, which distinguishes it from CMD.

Additionally, Figure 3 presents the local final modification maps M as well as those in each sub-image generated during the process of the proposed method. It is evident that most modifications in M_2 are consistent with those in M_1 , resulting in harmonious final modifications M . This can be credited to the inclusion of D_2 , which discerns between the partially embedded stego images Y_1 and Y_2 , capturing the correlations of modifications between neighboring pixels. As a result, it promotes consistent modifications in both sub-images.

1.3 Loss Curves

Figure 4 illustrates the loss curves of the generator and discriminator during the training phase, with the discriminator using a single

D_1 , single D_2 , combining D_1 and D_2 directly (i.e., Strategy #1 in Section ??), as well as combining with the proposed strategy. From Figure 4, three observations can be made:

- As shown in Figure 4(a), with a single D_1 , the loss of the generator increases during the training. This suggests that the optimization of the generator is slower compared to that of D_1 , indicating that D_1 readily detects the output of the generator. Hence, we focus on narrowing the performance gap between them when designing the update strategy.
- As shown in Figure 4(b), when training with a single D_2 , the generator's loss gradually decreases, indicating that a balance between the generator and the steganalyzer D_2 has already been achieved.
- As depicted in Figure 4(c), when combining D_1 and D_2 directly without making any adjustments (i.e., Strategy #1),

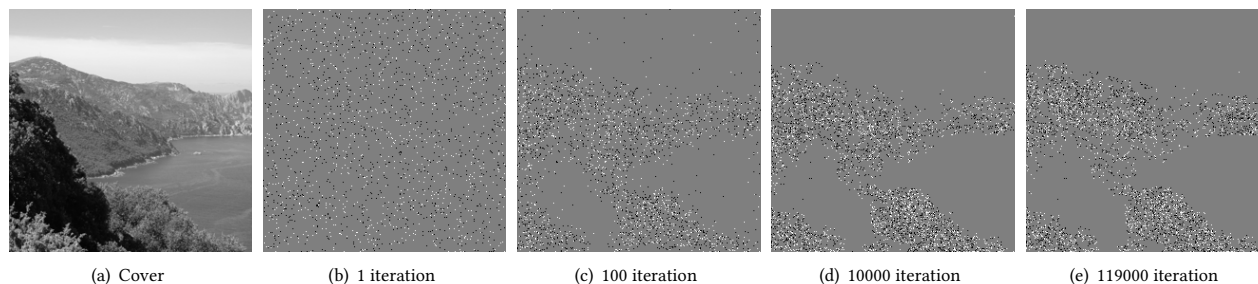


Figure 5: The cover image and the corresponding modification maps at different iterations during the training process of the proposed method.

the generator’s loss initially decreases but then begins to increase, signaling an unstable training process. Conversely, as demonstrated in Figure 4(d), when combined with the proposed update strategy, the generator’s loss shows a consistent upward trend, suggesting that although the overall performance of the discriminator remains superior to that of the generator, the implementation of our strategy ensures stability in the training dynamics. Moreover, as evidenced in [1], it has been observed that even when the losses of the generator increase, the quality of samples improves, indicating no significant correlation between sample quality (i.e., the ability of the generator) and its loss. This underscores

the reason why combining D_1 and D_2 with the proposed update strategy still achieves the highest performance.

Additionally, Figure 5 illustrates the modification maps at different iterations during the training process. At the beginning, the modifications appear to be random. As the training progresses, however, the modifications gradually concentrate in regions with complex textures. This suggests that the generator is being optimized under the guidance of the steganalyzers D_1 and D_2 .

REFERENCES

- [1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*. 214–223.