Enhancing Super-Resolution via Frequency Domain Losses

Min Woo Kim and Nam Ik Cho

***Abstract*— Single-image super-resolution (SISR) is a technique that generates a high-resolution (HR) image from a low-resolution (LR) one. In general, the LR image lacks high-frequency details compared to the HR image. Therefore, the main task of SISR is to restore the lost information. To accomplish this, we propose to penalize the training loss not only in the spatial domain but also in the frequency domain. We introduce an adversarial loss for the training patches that are converted to the frequency domain using the Discrete Cosine Transform. Additionally, we incorporate the Wavelet-domain High-Frequency Loss because it emphasizes the high-frequency spectrum. Our experiments show that our approach can improve both the quantitative and qualitative outcomes.**

I. INTRODUCTION

Super-resolution (SR) aims to reconstruct frequencies across different bands, relying on the low-frequency band. However, one of the main limitations is that the objective functions mainly target the low-frequency band within the spatial domain. Therefore, Ledig *et al.*[1] suggested to utilize not only the perceptual loss but also the adversarial loss with a generative adversarial network (GAN). Fuoli *et al.*[2] further advanced this approach by employing losses in the Fourier space. They measured the Euclidean distance between real and fake images and also calculated an adversarial loss. However, the discriminator is constrained to accommodate inputs of a predetermined size, owing to its composition of multilayer perceptron (MLP).

The focus of our paper is on improving the performance of the discriminator by transforming inputs into frequency domain with the Discrete Cosine Transform (DCT), using CNN instead of MLP, and employing the Wavelet domain High-Frequency Loss (WHFL) to enhance high-frequency components [3]. By using MLP, we resolve the issue of image size, and by utilizing WHFL-based loss, we enhance high frequencies. We have applied our method to the baseline[1], and the results show significant improvements in both quantitative and qualitative aspects for all test datasets.

II. METHOD

Fig 1. describes the comprehensive workflow of the proposed method. Initially, the SR network generates an SR image from its LR counterpart. Then, the training patches extracted from the images undergo a transformation into the frequency domain by DCT, and the transformed inputs pass through the discriminator comprised of CNN. Moreover, since the WHFL[3] emphasizes the high-frequency components, we choose the function as the auxiliary loss. As a result, the final loss equation can be written as:

$$L_{total} = L_{spatial} + \lambda_1 L_{WHFL} + \lambda_2 L_{adversarial}, \quad (1)$$

where λ_1, λ_2 are hyper-parameters calibrating the balance among each constituent term in the loss function.

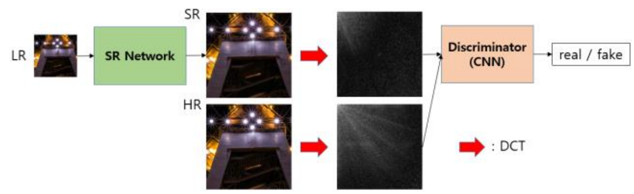

Figure 1. The overall pipeline of the proposed method.

III. EXPERIMENTAL RESULTS

We applied our proposed method in conjunction with the baseline [1]. It's worth noting that we used binary cross-entropy to set up the adversarial loss across all experimental settings and trained the networks on the DIV2K training set. Table 1 demonstrates that our method significantly improves all evaluation metrics. Additionally, Fig. 2 illustrates that our approach mitigates artifacts around edges, leading to enhanced outcomes.

TABLE I. THE QUANTITATIVE RESULTS. (VAL MEANS VALIDATION)

4X upscaling	***BSD100***	***Urban100***	***DIV2K Val***	***Set14***
	PSNR ↑			
Baseline[1]	25.316	22.730	27.490	25.129
Proposed	**25.439**	**22.879**	**27.659**	**25.294**
	LPIPS ↓			
Baseline[1]	0.318	0.310	0.282	0.289
Proposed	**0.315**	**0.306**	**0.278**	**0.281**

Figure 2. An Example of Qualitative Results

REFERENCES

[1] Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[2] Fuoli, Dario, Luc Van Gool, and Radu Timofte. "Fourier space losses for efficient perceptual image super-resolution." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021..

[3] Kim, Min Woo, and Nam Ik Cho. "Whfl: Wavelet-domain high frequency loss for sketch-to-image translation." Proceedings of the IEEE/CVF Winter Conference on applications of computer vision. 2023.

*Research supported by Seoul National University.

M.W. Kim and N. I. Cho. are with the Electrical and Computer Engineering Department, INMC, Seoul National University, Seoul, Korea (e-mail: mwk0614@snu.ac.kr, nicho@snu.ac.kr).