Spatial-Frequency Fusion for Arbitrary-Scale Ultra-High-Definition Image Super-Resolution

Cuixin Yang, Jun Xiao, and Kin-Man Lam
Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University
Hong Kong, China

***Abstract*—Ultra-high-definition (UHD) image super-resolution (SR) has attracted increasing attention due to the popularity of modern devices, such as smartphones, which support the capture of UHD images, e.g. 4K and 8K images. However, existing UHD SR methods process the image in the spatial domain only. This limitation hinders their ability to effectively utilize the rich details and fine-grained textures in local areas of UHD images. To address this issue, our proposed method comprehensively exploits the global and local features of UHD images by combining spatial and frequency features. Additionally, previous UHD image SR methods can only handle a fixed scaling factor, but real-world applications very often require upscaling low-resolution images with different scales. Therefore, we employ an arbitrary-scale strategy in the SR training process, enabling super-resolution of UHD images at any scale with a single trained model. Experimental results demonstrate the effectiveness and superiority of our proposed method.**

***Index Terms*—Super-resolution, ultra-high-definition, frequency domain, arbitrary-scale**

I. Introduction

Typical image super-resolution (SR) network architectures are capable of generating excellent global representations of images. However, these networks are primarily designed for high-definition (HD)/standard-definition (SD) images, rather than UHD images. Extracting discriminative features from the frequency domain is crucial for UHD images, as it allows for better exploration of finer-grained local details and complementing global features. To address this issue, we propose to incorporate a frequency module (FM) into a representative SR backbone, thereby enhancing the overall representativeness. Moreover, we adopt an arbitrary-scale super-resolution strategy, enabling super-resolution of UHD images with various scales, rather than being limited to a fixed scale under a trained SR model.

II. Methodology

UHD SR aims to reconstruct a high-resolution (HR) image from its corresponding low-resolution (LR) image. In the proposed method, EDSR [1] is adopted as the backbone for handling spatial features. As shown in Fig. 1, the LR input I^{LR} is firstly convolved to generate feature maps f. These feature maps f are then passed into N spatial-frequency blocks. Specifically, in the FM, we decompose the feature maps into four sub-bands using discrete wavelet transform (DWT). The four sub-bands are then divided into two groups, which are respectively input to two branches: a low frequency branch and a high frequency branch. The low-frequency branch consists of only the LL band, while the high-frequency branch takes the other three sub-bands (LH, HL, HH) as input. The outputs from the low frequency and high frequency branches are transformed back to the spatial domain using the inverse discrete wavelet transform (IDWT). The transformed features are then concatenated with the spatial features from the residual block of EDSR. Finally, after N spatial-frequency blocks, the generated features are fed into the arbitrary-scale module to reconstruct an UHD HR image.

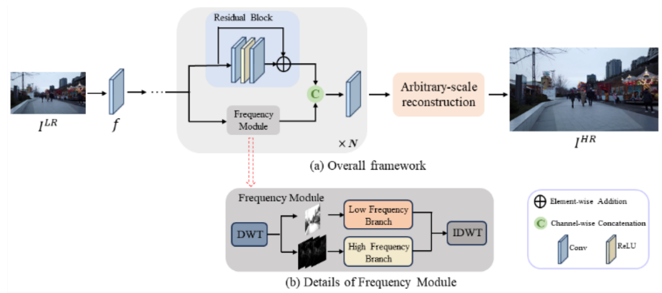

Fig. 1. (a) Overall framework of the proposed method. (b) Structure of the frequency module (FM).

III. Experimental Results

TABLE I

QUANTITATIVE COMPARISON OF DIFFERENT METHODS ON THE UHDSR TESTING DATASETS (PSNR/dB)

	UHDSR-4K			DIV8K		
Scaling Factor	×6	×8	×12	×6	×8	×12
CLIT [2]	31.87	29.46	26.88	30.65	29.24	27.58
CLIT+FM	**31.92**	**29.49**	**26.92**	**30.70**	**29.28**	**27.62**
LTE [3]	31.95	29.51	26.94	30.69	29.30	27.65
LTE+FM	**31.99**	**29.55**	**26.97**	**30.74**	**29.33**	**27.67**

To demonstrate the effectiveness and superiority of the proposed method, we conducted two sets of experiments based on two arbitrary-scale SR baselines (CLIT [2] and LTE [3]). Table I presents the quantitative comparison of the baselines and the corresponding architectures with FM on the public UHDSR-4K and DIV8K testing datasets. The challenging large scaling factors, e.g., ×6, ×8, ×12, are considered in our experiments. Notably, SR on UHD images with different scaling factors is performed using the same trained model. The results demonstrate that the proposed method outperforms the original backbones.

IV. References

[1] Bee Lim, et al. “Enhanced deep residual networks for single image super-resolution,” in CVPRW, 2017, pp. 136–144.
[2] Hao-Wei Chen, et al. “Cascaded local implicit transformer for arbitrary-scale super-resolution,” in CVPR, 2023, pp. 18257–18267.
[3] Jaewon Lee and Kyong Hwan Jin, “Local texture estimator for implicit representation function,” in CVPR, 2022, pp. 1929–1938.