From Zero to Detail: Deconstructing Ultra-High-Definition Image Restoration from Progressive Spectral Perspective

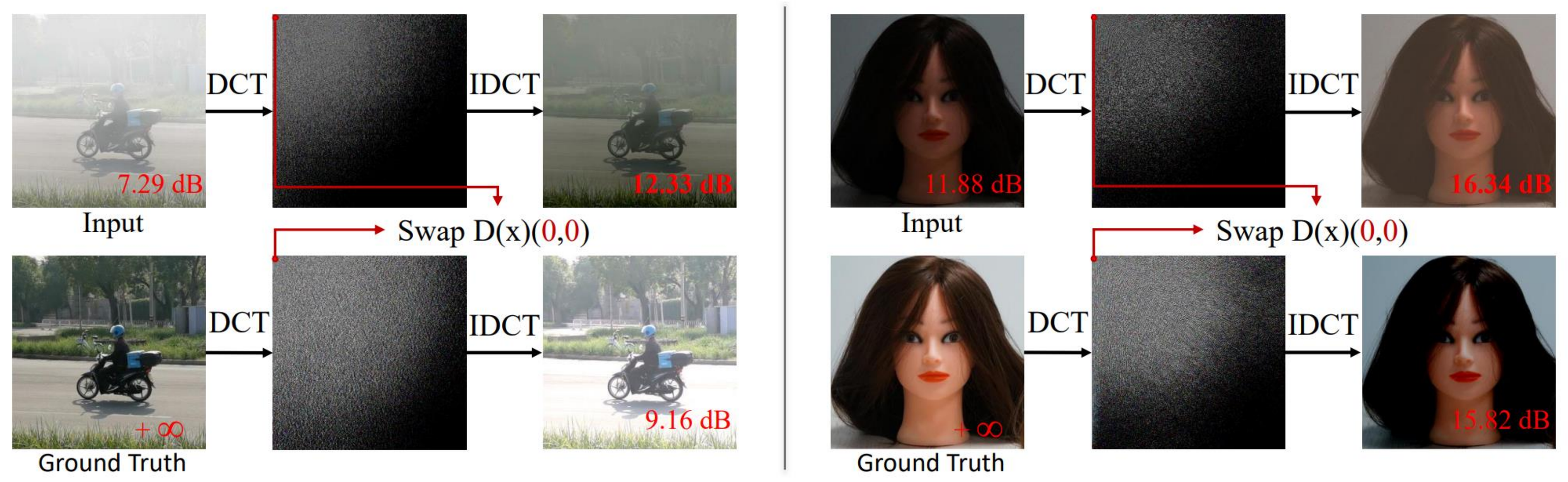
Chen Zhao[1], Zhizhou Chen[1], Yunzhe Xu[1], Enxuan Gu[2], Jian Li[3], Zili Yi[1], Qian Wang[4], Jian Yang[1], Ying Tai[1]
Nanjing University[1], Dalian University of Technology[2], Tencent Youtu[3], China Mobile Institute[4]

CVPR Nashville JUNE 11-15, 2025

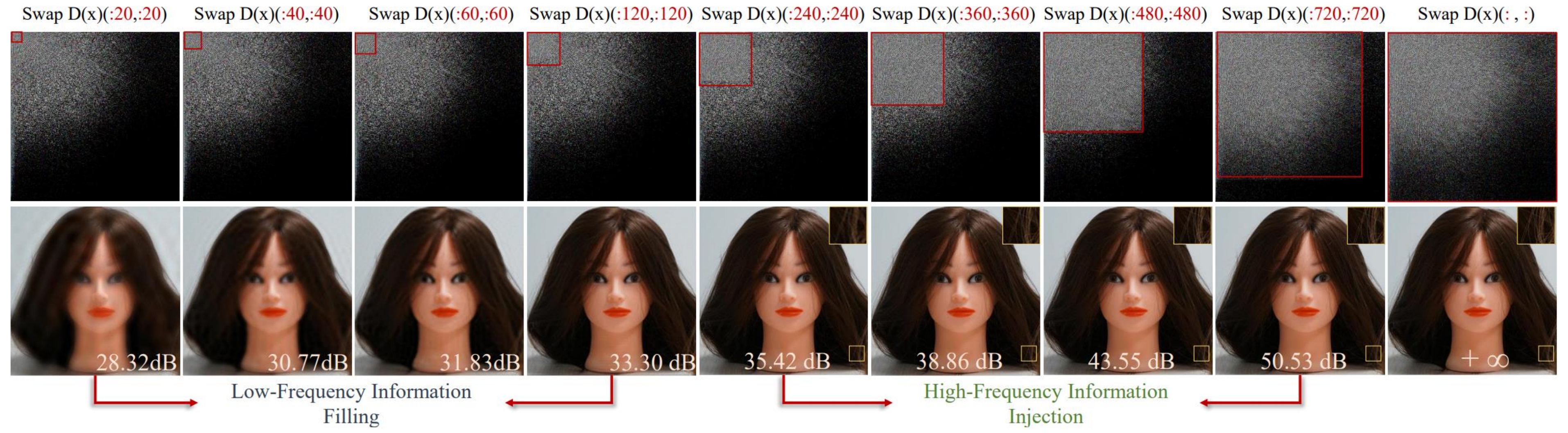
Abstract

UHD image restoration faces significant challenges due to its high resolution. To cope with these challenges, we analyze the restoration process in depth through a progressive spectral perspective, and deconstruct the UHD restoration problem into three progressive stages. Building on this insight, we propose a novel framework, ERR, which comprises three collaborative sub-networks: ZFE, LFR, and HFR.

Motivation



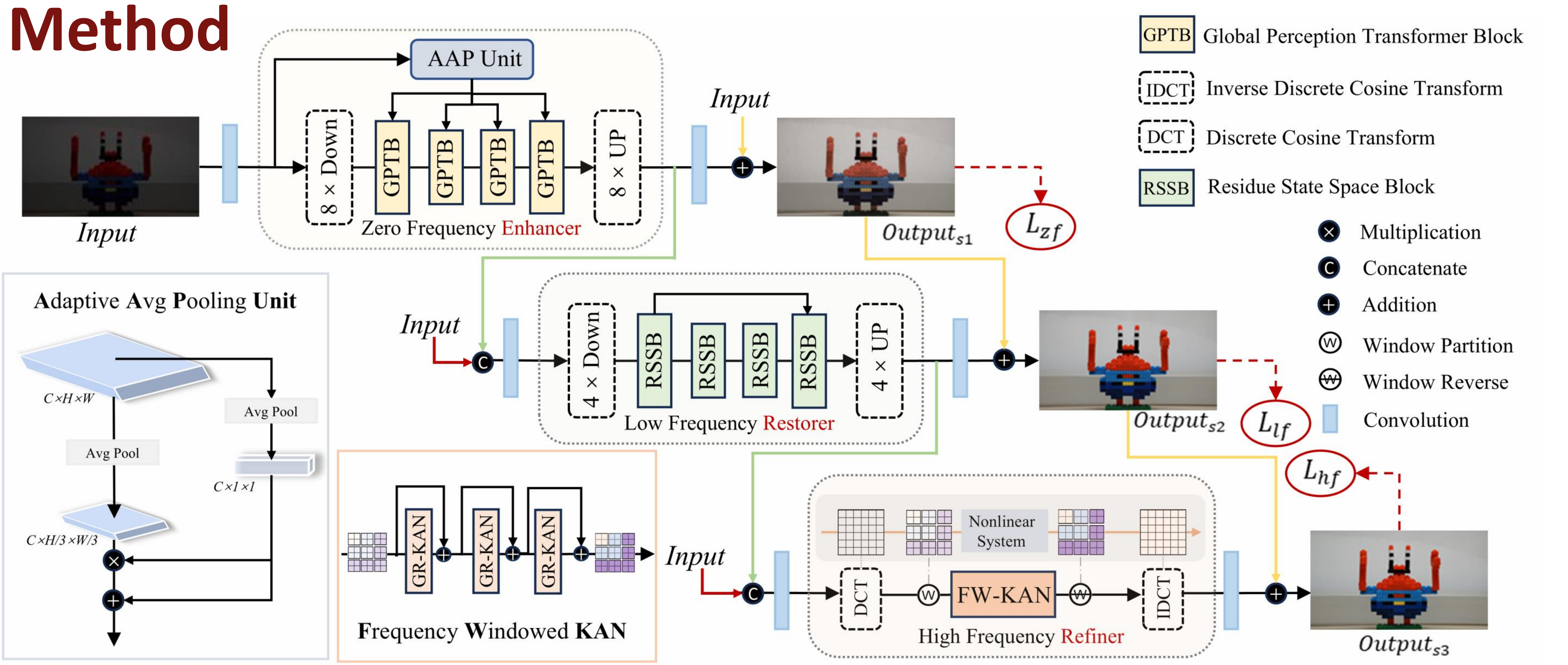
(a) We exchange the information at the (0,0) position in the DCT spectrum, which represents the global information.



(b) The low-frequency filling restores the coarse-grained content, while the high-frequency injection refines the fine-grained textures.

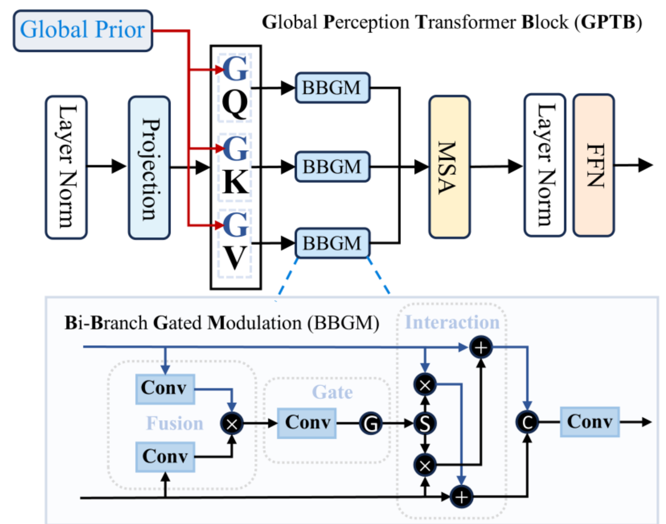
Our core motivation. Based on the observations in (a) and (b), we deconstruct the complex UHD restoration problem into three progressive stages: **zero-frequency enhancement**, **low-frequency restoration**, and **high-frequency refinement**.

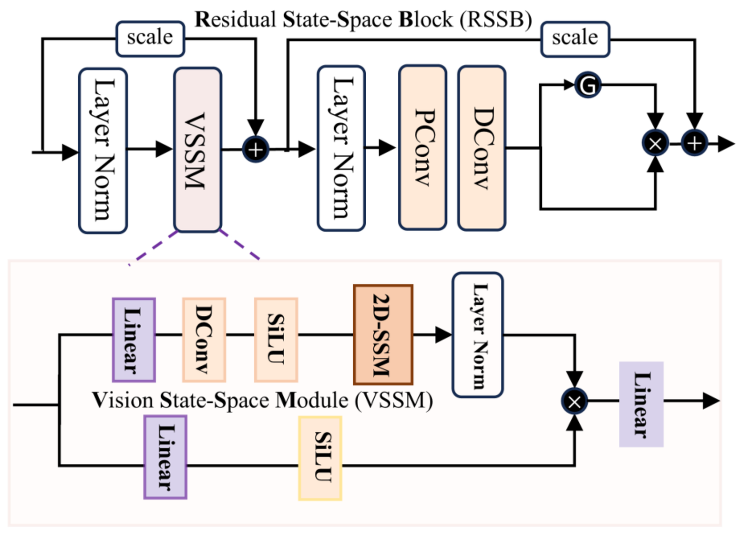
Method



The framework of ERR integrates three collaborative sub-networks: the zero-frequency **enhancer** (ZFE), the low-frequency **restorer** (LFR), and the high-frequency **refiner** (HFR).

Detailed Design





$$\mathcal{L}_{zf} = ||\mathcal{D}(O_{s1})(0,0) - \mathcal{D}(GT)(0,0)||_1.$$

$$\mathcal{L}_{lf} = \sum_{i=0}^{k}\sum_{j=0}^{k} ||\mathcal{D}(O_{s2})(i,j) - \mathcal{D}(GT)(i,j)||_1, \quad \text{where } (i,j) \neq (0,0).$$

$$\mathcal{L}_{hf} = \sum_{i=0}^{N}\sum_{j=0}^{N} ||\mathcal{D}(O_{s3})(i,j) - \mathcal{D}(GT)(i,j)||_1, \quad \text{where } (i \geq k) \vee (j \geq k),$$

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \mathcal{L}_{zf} + \mathcal{L}_{lf} + \mathcal{L}_{hf}.$$

a) **ZFE** is designed to learn global mappings in low-resolution space.

b) **LFR** is designed to recover coarse-grained information in the medium-resolution space.

c) We employ three frequency-domain loss to constrain each.

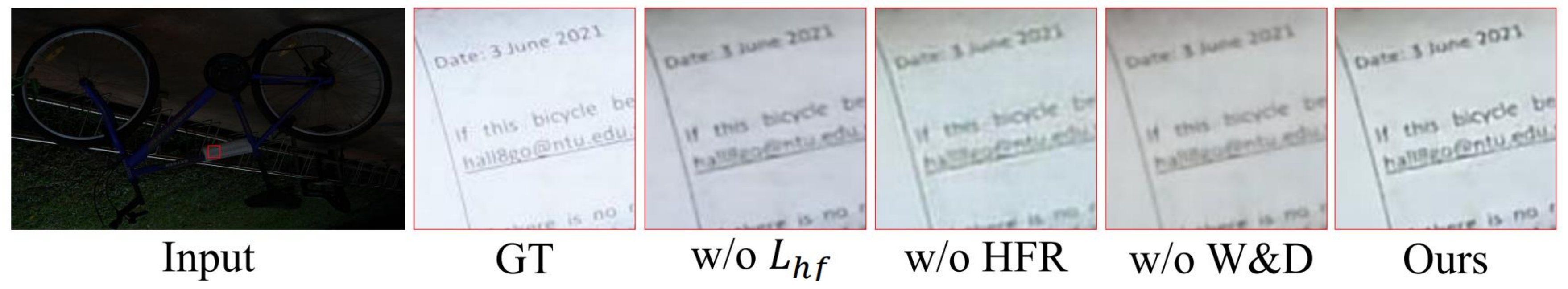
d) The motivation for KAN in HFR. The table below verifies that nonlinear functions primarily contribute to high-frequency enhancement.

Method	Linear system		UHDformer		Difference	
Frequency	Low	High	Low	High	Low	High
PSNR ↑	28.70	34.68	27.92	36.82	-0.78	**2.14**
SSIM ↑	0.9873	0.8539	0.9820	0.9081	-0.0053	**0.0542**

Model	MLP-6	MLP-12	MLP-24	KAN	FW-KAN
PSNR ↑	26.03	26.58	26.97	20.48	**27.57**
SSIM ↑	0.9294	0.9297	0.9311	0.8836	**0.9326**
Parameter↓	30.73K	35.63K	45.42K	27.57K	**27.47K**

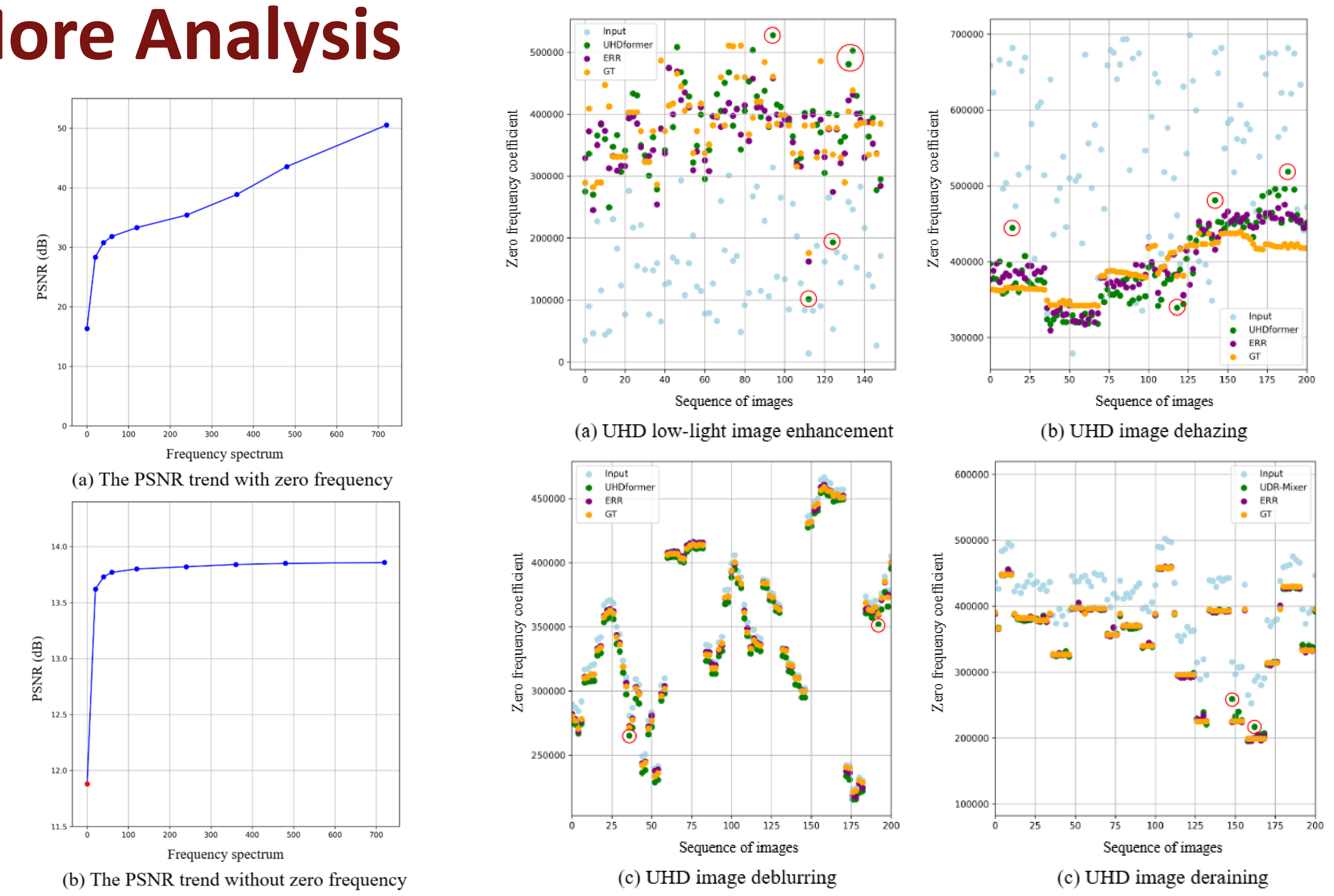
Ablation For ZFE

Ablation For HFR



Results

More Analysis



(a) The PSNR trend with zero frequency

(b) The PSNR trend without zero frequency

(a) UHD low-light image enhancement

(b) UHD image dehazing

(c) UHD image deblurring

(c) UHD image deraining