Appendix of "WaveGAN: Frequency-aware GAN for High-Fidelity Few-shot Image Generation"

Mengping Yang[1,2], Zhe Wang[1,2,⋆], Ziqiu Chi[1,2], and Wenyi Feng[1,2]

[1] Department of Computer Science and Engineering, East China University of Science and Technology, China
[2] Key Laboratory of Smart Manufacturing in Energy Chemical Process, East China University of Science and Technology, China
mengpingyang@mail.ecust.edu.cn wangzhe@ecust.edu.cn
{chiziqiu Y10200096}@mail.ecust.edu.cn

Appendix

This appendix provides the supplementary information that is not elaborated in the main paper: Sec. A provides the implementation details of our model. Sec. B provides quantitative results of our ablation studies. Sec. C shows the efficacy of different High-frequency components. Finally, Sec. D presents the 2D DWT Visualization results of the images generated by our WaveGAN.

A Implementation Details.

Our encoder consists of five convolutional blocks and four wavelet transformation blocks. The five convolutional blocks contain one convolution layer, followed by batch normalization and Leaky-Relu activation. Our decoder is symmetrical with four upsampling blocks and one output convolutional layer. Each upsampling block includes upsample operation followed by one convolutional block. We perform our wavelet transformation after each convolution block in the encoder and employ inverse transformation after each convolution block in the decoder. Our discriminator is the same as LoFGAN with four residual blocks and two fully connected layers.

Adam optimizer is used and we train our model for 100,000 iterations. At the beginning of 50,000 iterations, the learning rates for both the generator and the discriminator are set to 1e-4, after 5000 iterations, the learning rates decay linearly to 0. We set $\lambda_{\text{cls}}^{G} = \lambda_{\text{cls}}^{D} = \lambda_{Fre} = 1$, we save the final checkpoint to synthesis images for evaluation. The batchsize is set to 8, and we sample hundreds of K-shot image generation tasks from $\mathbb{C}_s$. Our model is implemented in PyTorch framework and trained on 1 $\times$ NVIDIA GeForce RTX 3090 GPU.

B Quantitative Results of Ablation Studies

Here we provide the quantitative results of our ablation studies in Sec.4.4 of the main paper, demonstrating the effectiveness of each component of the proposed

⋆ Corresponding author

Table 1. Ablation studies of our WaveGAN. We test the efficacy of each components for our two transformation techniques, *i.e.*, WaveGAN-M and WaveGAN-B.

Conditions	Type	Flowers		Animal Faces		VGGFace	
		FID (↓)	LPIPS (↑)	FID (↓)	LPIPS (↑)	FID (↓)	LPIPS (↑)
WaveGAN-M w/o LoF	Fusion	82.18	0.3720	52.64	0.5071	10.96	**0.3822**
WaveGAN-M w/o LL	Fusion	72.35	0.3709	67.53	0.5061	11.34	0.3795
WaveGAN-M w/o HL	Fusion	87.99	0.3783	105.47	0.4981	21.48	0.3017
WaveGAN-M w/o L_1 Loss	Fusion	73.86	0.3767	62.13	0.5017	12.29	0.3041
WaveGAN-M (Ours)	Fusion	63.79	0.3709	50.98	0.5014	8.62	**0.3822**
WaveGAN-B w/o LoF	Fusion	47.37	0.3733	32.35	0.5080	5.35	0.3308
WaveGAN-B w/o LL	Fusion	44.31	0.3803	31.12	0.5013	5.55	0.3223
WaveGAN-B w/o HL	Fusion	85.25	0.3788	108.82	0.5011	20.61	0.3021
WaveGAN-B w/o L_1 Loss	Fusion	45.52	0.3792	31.45	0.5047	5.99	0.3253
WaveGAN-B (Ours)	Fusion	**42.17**	**0.3868**	**30.35**	**0.5076**	**4.96**	0.3255

Table 2. Efficacy of LH, HL, and HH tested on Flowers.

Method	Metric	Only LF	Only HF	Only HH	Full
WaveGAN-M	FID	69.73	69.49	69.60	63.79
WaveGAN-M	LPIPS	0.3706	0.3654	0.3714	0.3709
WaveGAN-B	FID	48.65	43.76	42.75	42.17
WaveGAN-B	LPIPS	0.3774	0.3779	0.3863	0.3868

WaveGAN. We remove each component and keep other settings unchanged to validate the contributions of each components of our WaveGAN, namely 1) the low-frequency skip connection, 2) the high-frequency skip connection, and 3) frequency L_1-loss. Besides, we remove the LoF module to investigate the influence of local fusion on our model. The quantitative results are given in Tab. 1, from which we can observe that each component boosts the synthesis performance. Combining the visualization results in Fig.6 of the main paper further reflects the effectiveness of our proposed method.

C Efficacy of High-frequency Components

We feed the combination of high-frequency components (*i.e.*, LH, HL, HH) to the decoder in the main version of our WaveGAN. Here we test their efficacy by feeding only one of each to the decoder on Flowers dataset, the results are given in Tab. 2. The results indicate that each high-frequency component has almost the same contribution to the generation quality, and fusing all of them (Full) yields the best results.

D 2D DWT Visualization of High-frequency Components

We provide 2D DWT visualization results of generated images in Fig. 1, which further demonstrates that our method is frequency-aware and produces images with informative frequency signals.

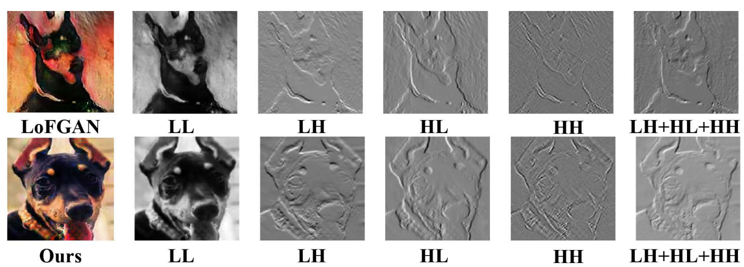

Fig. 1. 2D DWT visualization results of frequency components.