Pacific Graphics (2023)
R. Chaine, Z. Deng, and M. H. Kim (Guest Editors)

Supplementary Material-WaveNet: Wave-Aware Image Enhancement

Jiachen Dang†[1,2], Zehao Li[1,2], Yong Zhong[1,2], Lishun Wang[1,2]

[1]Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu, China
[2]University of Chinese Academy of Sciences, Beijing, China

Ablation Study

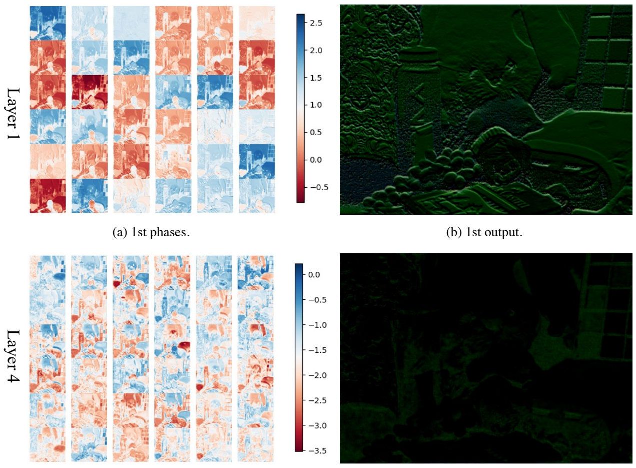
All ablation experiments are performed on the LOL [WWYL18] dataset trained on image patches of size 256×256 for 100 epochs.

Improvements in waves and ASFF: Table 4 shows that CW, SW, and GW, each of them provides **0.5** dB promotion at least. Furthermore, using ASFF to fuse the shallow features and deep features adaptively can provide 0.1dB gain to the waveNet.

Impact of formulating the amplitude and phase estimation: The amplitude and the phase play a significant role in aggregating features and feature representation. There are different methods for formulating amplitude and phase estimation. As shown in Table 5, the identity estimation represents that amplitude and phase are the duplicates of the inputs whose performance is obviously inferior compared with others. Static estimation uses the already-learned parameters to add to the inputs directly. This way cannot provide a good result because of its weakness in processing various inputs. Using the way mentioned in Sec. 3.3 of main paper to extract amplitude and phase information dynamically obtains better results than other methods.

Impact of phase term in different waves: Cosine and sine functions have the same shape but with a $\frac{\pi}{2}$ phase difference. To explore the phase impact in different waves (CW and SW), we observe the effect on the results by adding $\frac{\pi}{2}$ to the phase of the different waves. The results shown in the table 6 show that changing the phase term in SW/CW to transform them into the same type of waveform greatly reduces the model's accuracy. It also indicates the phase information extracted by WTB is different between SW and CW. Therefore, the *sginal-like* feature representation is reasonable and effective and relying on one type of waves (SW or CW) can not provide satisfactory results.

Effectiveness of *wave-like* feature representation: To validate the effectiveness of the proposed *wave-like* feature representation, we replace all periodic functions in Eq. (5) of main paper with GELU [HG16] and LeakyReLU. Table 7 shows that our *wave-like* feature representation obtains **1.5**dB gain and demonstrate the effectiveness of the proposed feature construction mechanism.



(a) 1st phases.

(b) 1st output.

(c) 4th phases.

(d) 4th output.

Figure 1: Visualization of the phase differences between layers. (Zoom in for the best view)

Effectiveness of feature aggregation: Table 8 shows that choosing different kernel sizes of convolution for wave aggregation achieves better results than others with similar parameters and FLOPs. However, using vector-sum to statically aggregate wave features not only improves the computational-consuming but also gets the worst accuracy.

Visualization: For a more intuitive understanding of the information extracted from the phase term, we visualize the phase term of the 1st and 4th layers in Figure 1. As shown in Figure 1 (a) and 1 (c), we visualize the first 36 channels of the phase term. Figure 1 (b) and 1 (d) represent the outputs produced by the phase applied on the input. From a global perspective in Figure 1 (a), we can see that the phase of every Cosine Wave (CW) is close to each other in the same channel. From Figure 1 (b), we can see that the 1st phase terms include **high** frequency information of images and dynamically retain the edge information. In Figure 1 (c), the 4th phase terms more concentrate on aggregating local information. From a global view, the 4th phase terms have a larger phase difference and a wider range of frequencies. From Figure 1 (d), the shape and detail of the input are captured. That is to say, the 4th WTB prefers to capture **low** frequency waves in the feature maps and aggregates local correlations in pixels.

Depth	Encoder			Decoder		
	Input Shape	Output Shape	Layers	Input Shape	Output Shape	Layers
0	H×W×3	H×W×32	Conv3x3_s1_w32 WTB×2_w32	H×W×32	H×W×3	WTB×2_w32 Conv3x3_s1_w3

Table 1: Detailed architectural specifications of WaveNet-T

Depth	Encoder			Decoder		
	Input Shape	Output Shape	Layers	Input Shape	Output Shape	Layers
0	H×W×3	H×W×128	Conv3x3_s1_w128 WTB×2_w128	H×W×128	H×W×3	WTB×2_w128 Conv3x3_s1_w3

Table 2: Detailed architectural specifications of WaveNet-S

Depth	Encoder			Decoder		
	Input Shape	Output Shape	Layers	Input Shape	Output Shape	Layers
0	H×W×3	H/2×W/2×128	Conv3x3_s1_w128 WTB×2_w128 downsampling_w192	H×W×128	H×W×3	WTB×2_w128 ASFF_w128 Conv3x3_s1_w3
1	H/2×W/2×128	H/4×W/4×192	WTB×2_w192 downsampling_w256	H/2×W/2×192	H×W×128	WTB×2_w192 upsampling_w128
2	H/4×W/4×192	H/4×W/4×256	WTB×4_w256	H/4×W/4×256	H/2×W/2×192	Conv3x3_s1_w256 WTB×4_w256 upsampling_w192

Table 3: Detailed architectural specifications of WaveNet-B

Table 4: Impact of each wave and ASFF.

Baseline	√				
+Cosine Wave (CW)		√	√	√	√
+Sine Wave (SW)			√	√	√
+Gating Wave (GW)				√	√
+ASFF					√
PSNR (dB)↑	20.04	20.86	21.28	21.76	**21.87**

Table 5: Formulation of amplitude and phase estimation.

Size	PSNR (dB)↑	SSIM↑	Param (M)↓	FLOPs(G)↓
Identity	21.19	0.835	**7.76**	**82.5**
Static	21.55	0.838	13.5	158.3
Dynamic	**21.87**	**0.841**	14.4	162

Table 6: Impact of phase changing. * indicates adding $\frac{\pi}{2}$ to the phase.

Waves	PSNR (dB)↑	SSIM↑	Param (M)↓	FLOPs(G)↓
SW* & CW	20.43	0.818	14.4	162
SW & CW*	20.19	0.807	14.4	162
SW & CW	**21.87**	**0.841**	14.4	162

Table 7: Comparison of different activation function.

Activation Function	PSNR (dB)↑	SSIM ↑
GELU	23.91	0.858
LeakyReLU	24.05	0.855
Wave Form	**25.44**	**0.864**

Configurations

The detailed specifications of WaveNet family are shown in Table 1 to Table 3. Here Conv3x3_s1_w32 means a convolutional layer with 3×3 kernels, stride 1, and 32 channels. The WaveNet family contains three models with different parameters and computational costs by adjusting the depths and widths of architecture specifications, which are denoted as WaveNet-T (in Table 1), WaveNet-S(in Table 2), and WaveNet-B(in Table 3), sequentially. Considering the efficiency and accuracy, we do not use down-Bp sampling operations for the light-weight model WaveNet-T and WaveNet-B to preserve the informative details and reduce extra parameter-consuming in rescaling operations.

Additional Visual Results

We show images enhanced by WaveNet and other competing approaches as qualitative examples in Figure 2 and Figure 3.

Limitations and Discussions

While WaveNet shows surprising performance on image enhancement tasks, there are still many modules that could be optimized.

Proceedings published by Eurographics - The European Association for Computer Graphics.

For example, our model should enhance capturing the long-range interaction of pixels.

Table 8: The size of convolutional kernel for feature aggregation.

Size	Groups	PSNR (dB)↑	Param (M)↓	FLOPs(G)↓
identity	-	21.51	13.9	155
vector-sum	-	20.03	13.9	378
All 1	1	21.72	16.2	182
All 3	C/4	21.68	14.3	161
All 5	C/2	21.65	14.5	163
All 7	C	21.73	14.4	163
3 5 7	C/4 C/2 C	21.87	14.4	162

References

[BPCD11] Bychkovsky V., Paris S., Chan E., Durand F.: Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR 2011* (2011), IEEE, pp. 97–104. 4

[CGZ18] Cai J., Gu S., Zhang L.: Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing 27*, 4 (2018), 2049–2062. 4

[CLG*22] Cui Z., Li K., Gu L., Su S., Gao P., Jiang Z., Qiao Y., Harada T.: You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. In *BMVC* (2022), p. 238. 4

[HG16] Hendrycks D., Gimpel K.: Gaussian error linear units (gelus). *arXiv.org* (jun 2016). arXiv:1606.08415v4. 1

[LXY*21] Liu J., Xu D., Yang W., Fan M., Huang H.: Benchmarking low-light image enhancement and beyond. *International Journal of Computer Vision 129* (2021), 1153–1184. 4

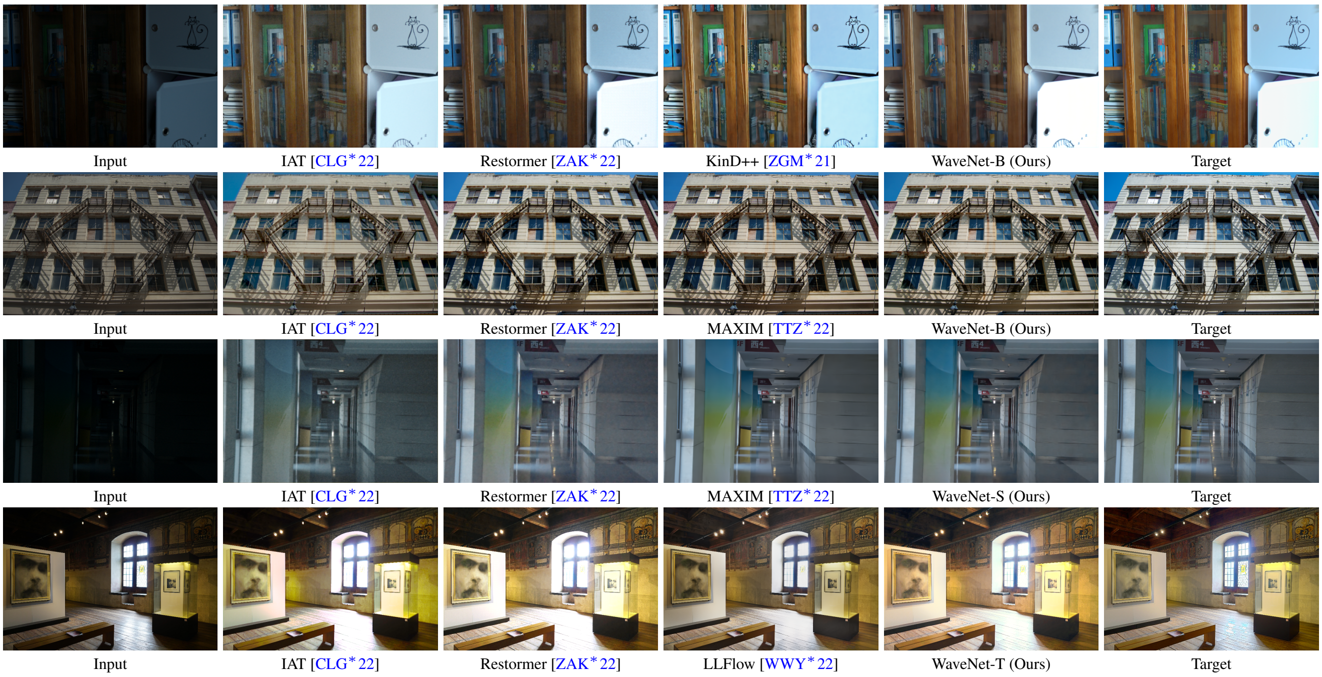
[TTZ*22] Tu Z., Talebi H., Zhang H., Yang F., Milanfar P., Bovik A., Li Y.: Maxim: Multi-axis mlp for image processing. 5769–5780. 4

[WWY*22] Wang Y., Wan R., Yang W., Li H., Chau L.-P., Kot A.: Low-light image enhancement with normalizing flow. 2604–2612. 4

[WWYL18] Wei C., Wang W., Yang W., Liu J.: Deep retinex decomposition for low-light enhancement. *arXiv.org* (aug 2018). arXiv:1808.04560v1. 1, 4

[ZAK*22] Zamir S. W., Arora A., Khan S., Hayat M., Khan F. S., Yang M.-H.: Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022), pp. 5728–5739. 4

[ZGM*21] Zhang Y., Guo X., Ma J., Liu W., Zhang J.: Beyond brightening low-light images. *International Journal of Computer Vision 129* (2021), 1013–1037. 4



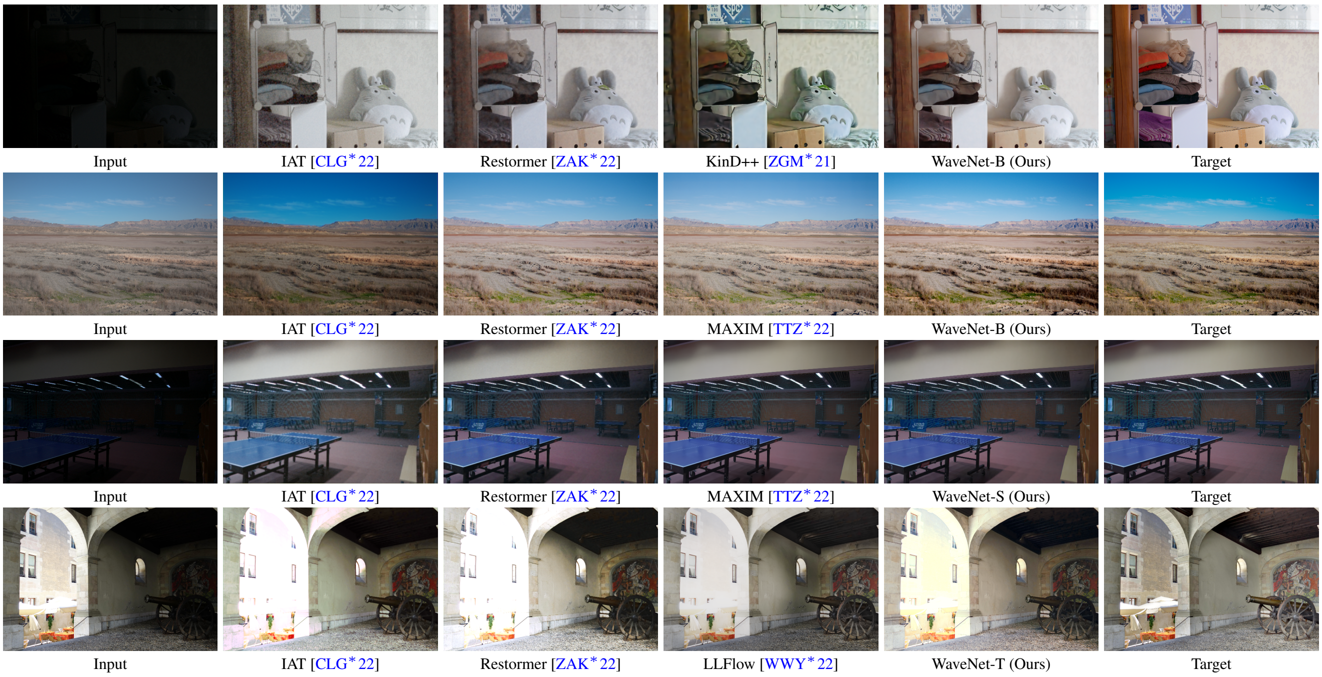
Input | IAT [CLG*22] | Restormer [ZAK*22] | KinD++ [ZGM*21] | WaveNet-B (Ours) | Target

Input | IAT [CLG*22] | Restormer [ZAK*22] | MAXIM [TTZ*22] | WaveNet-B (Ours) | Target

Input | IAT [CLG*22] | Restormer [ZAK*22] | MAXIM [TTZ*22] | WaveNet-S (Ours) | Target

Input | IAT [CLG*22] | Restormer [ZAK*22] | LLFlow [WWY*22] | WaveNet-T (Ours) | Target

Figure 2: Visual comparison with the state-of-the-art methods on LOL [WWYL18] (top row), MIT-Adobe FiveK [BPCD11] (2nd row), VE-LOL [LXY*21] (3rd row) and SICE [CGZ18] (4th row).



Input | IAT [CLG*22] | Restormer [ZAK*22] | KinD++ [ZGM*21] | WaveNet-B (Ours) | Target

Input | IAT [CLG*22] | Restormer [ZAK*22] | MAXIM [TTZ*22] | WaveNet-B (Ours) | Target

Input | IAT [CLG*22] | Restormer [ZAK*22] | MAXIM [TTZ*22] | WaveNet-S (Ours) | Target

Input | IAT [CLG*22] | Restormer [ZAK*22] | LLFlow [WWY*22] | WaveNet-T (Ours) | Target

Figure 3: Visual comparison with the state-of-the-art methods on LOL [WWYL18] (top row), MIT-Adobe FiveK [BPCD11] (2nd row), VE-LOL [LXY*21] (3rd row) and SICE [CGZ18] (4th row).