WaveNet: Wave-Aware Image Enhancement

Jiachen Dang\textsuperscript{1,2}*, Zehao Li\textsuperscript{1,2}, Yong Zhong\textsuperscript{1,2}, Lishun Wang\textsuperscript{1,2}

\textsuperscript{1}Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu, China
\textsuperscript{2}University of Chinese Academy of Sciences, Beijing, China

Abstract

As a low-level vision task, image enhancement is widely used in various computer vision applications. Recently, multiple methods combined with CNNs, MLP, Transformer, and the Fourier transform have achieved promising results on image enhancement tasks. However, these methods cannot achieve a balance between accuracy and computational cost. In this paper, we formulate the enhancement into a signal modulation problem and propose the WaveNet architecture, which performs well in various parameters and improves the feature expression using wave-like feature representation. Specifically, to better capture wave-like feature representations, we propose to represent a pixel as a sampled value of a signal function with three wave functions (Cosine Wave (CW), Sine Wave (SW), and Gating Wave (GW)) inspired by the Fourier transform. The amplitude and phase are required to generate the wave-like features. The amplitude term includes the original contents of features, and the phase term modulates the relationship between various inputs and fixed weights. To dynamically obtain the phase and the amplitude, we build the Wave Transform Block (WTB) that adaptively generates the waves and modulates the wave superposition mode. Based on the WTB, we establish an effective architecture WaveNet for image enhancement. Extensive experiments on six real-world datasets show that our model achieves better quantitative and qualitative results than state-of-the-art methods. The source code and pretrained model are available at https://github.com/DeniJsonC/WaveNet.

CCS Concepts

• Computing methodologies → Image processing;

1. Introduction

Challenges for enhancing degraded images exist not only in the real world but also in computer vision tasks. Poor photographing environments, the improper operation of the photographer, or limitations of camera devices often produce the low-quality photos. These degraded images apply harmful visual effects and bad effects on other computer vision tasks. As a low-level computer vision task, image enhancement provides reliable information for downstream visual decisions. Recently, multiple approaches \cite{ZGM21,TTZ22,ZAK20,CWG21} based on CNNs, MLP, and Transformer have been designed for image enhancement. These methods perform well on benchmark datasets, but some may cause unacceptable computational costs and weak versatility.

Different from the method that delicately designs the model to achieve high accuracy without thinking of generalization and computational cost, we aim to explore a robust architecture that can perform well with different efficiency for image enhancement. Besides, we desire to improve the representation way of features for dynamically aggregating them according to semantic contents.

\* Corresponding author:
Mail: dj.chen112@gmail.com (J. Dang)

In this paper, our inspiration comes from signal processing. For an image captured from a digital camera, we simply think that every pixel retains the optical information independent of other pixels. e.g. Each pixel value has recorded a specific light intensity of the captured optical signal. Thus, a digital image that comes from the real world is a discrete digital matrix generated by sampling, quantizing, and encoding a continuous optical signal set. If we can formulate the optical signal recorded by the pixel, we can enhance...
the degraded image according to modulating the signals. Thus, we transform the enhancement task into a signal modulation problem.

How to "present" and "modulate" the signal using learning-based approaches is the key to our work. 1) For "presenting" signals, the Fourier transform theory provides a way of decomposing signals into different frequency sine/cosine waves. Therefore, we describe a pixel as a sample value of a signal with three waves (Cosine Wave (CW), Sine Wave (SW), and Gating Wave (GW)) to realize the wave-like feature representations. For constructing the waves, amplitude and phase are indispensable. The amplitude part measures the maximum intensity of a wave. The phase part represents the initial state of the wave and contains the frequency information. 2) For "modulating" signals, we have to consider not only a single signal modulation but also aggregating the semantic information of signals in a region. Only when we take into account the semantic information in local and non-local, we better modulate the signal. e.g. An individual red pixel provides vague semantic information. We tend to retouch it if the pixel is in an apple, or we tend to remove it as a noise if the pixel is in a banana. Therefore, we need to adaptively modulate these signals with the support of semantic information. Considering above analyses of presenting and modulating waves, we propose the Wave Transform Block (WTB in Figure 3) that allows for efficient and scalable spatial mixing of local and non-local contents and dynamically learns the interaction between waves to enhance the images. Besides, to obtain high-quality results, we use the gating mechanism to control which complementary features should flow forward and allow sub-blocks to focus exclusively on more refined images of attributes. Furthermore, we build the WaveNet (in Figure 2), an effective architecture using WTB. Section 3 discusses the proposed WaveNet in detail.

Figure 1 shows that the proposed WaveNet achieves the best trade-off between accuracy and complexity on various datasets. For example, WaveNet-T obtains the 23.59dB on LOL dataset [WWYL18] with only 80k parameters and outperforms MAXIM [TTZ∗22] by 0.16 dB and IAT [CLG∗22] by 0.21dB. WaveNet-B achieves 25.44dB compared with the previous state-of-the-art method LLFlow [WWY∗22] and obtains 0.44dB gain in PSNR. Besides, WaveNet also achieves strong performance on the VE-LOL [LXY∗21], SICE [CGZ18], MIT-Adobe FiveK [BPCD11] and SID [CCXK18] four image enhancement datasets. For high-level vision tasks, we use face detector DSDF [LWW∗19] to evaluate the face detection performance on various images across various low-light image enhancement methods. Our WaveNet achieves the best results on the DARK FACE dataset [YYR∗19]. Overall, our contribution could be summarized as follows:

- We propose a new way of enhancing feature representation, dubbed wave-like feature representation. We aggregate the features with three waves: Cosine Wave (CW), Sine Wave (Sw), and Gating Wave (Gw).
- We propose the Wave Transform Block (WTB) that is capable of aggregating local and channel information and modeling wave interactions to enhance the original degraded image. We build WaveNet, an effective architecture using WTB for image enhancement.
- Extensive experiments on popular real-world datasets show that our WaveNet achieves SOTA results.

2. Related Work

2.1. Learning-based image enhancement

With the development of deep learning, an amount of research which are learning-based has emerged. Since the ground-breaking methods, LLNet [LAS17] is proposed, learning-based methods have greatly improved. Compared with traditional methods, learning-based methods are more accurate, robust, and faster. These methods used a variety of learning strategies. Most of them used supervised learning, e.g. Retinex-Net [WWWYL18], DeepUPE [WZF∗19], KinD [ZZG19a], LPNet [LLF∗20], DLN [WLSL20], PRIEN [LFH21], and etc. In recent works, MIRNet [ZAK∗20] presented parallel multi-resolution convolution streams for extracting multi-scale features to enhance the degraded images. MAXIM [TTZ∗22] used two MLP blocks to aggregate local and non-local contents for image restoration. IAT [CLG∗22] presented a lightweight network for image enhancement. Nevertheless, these works are unable to achieve promising results on both accuracy and efficiency as the proposed WaveNet.

2.2. Fourier transform

The Fourier transform is proposed to analyze thermal processes and is widely used in various fields due to its excellent performance. In recent years, Fourier transform has been applied in learning-based image processing methods CirCNN [DLW∗17] substantially reduces computational complexity and storage complexity owing to FFT-based fast multiplication. FFC [CJM20] is based on Fourier spectral theory and enables models to have non-local receptive fields. Jae-Han Lee et al. [LHKK18] uses Fourier frequency domain analysis to estimate single-image depth. Fda [YS20] reduces differences between source and target distributions by exchanging low-frequency spectrum. GFNet [RZZ∗21] improves the efficiency by using 2D FFT/IFFT to change the self-attention layer. LaMa [SLM∗22] uses fast Fourier convolution to obtain the bigger receptive field. However, traditional Fourier transform in image processing directly employs the algorithm globally to learn a brute-force relationship between pixels, which may ignore the information effectiveness of different regions. The enhanced results rely on the frequency resolution. Therefore, these works apply the 2D-FFT/2D-IFFT operations in their modules, which are limited by strong hypothesis (signal is smooth), artificial priors, high complexities O(HWClog(HWC)), and the ability to dynamically process various inputs. In contrast, the proposed WaveNet utilizes a learning-based method that decomposes pixels into different waves and outperforms these methods in complexities O(HWC^2), accuracy, and efficiency, which are important for model effectiveness.

3. Method

In this section, we first introduce the recent works about periodic function applications in neural networks briefly. Secondly, we take an overview of WaveNet shown in Figure 2. Then, we discuss the details of our model. Finally, two main blocks of WaveNet are illustrated in detail.

3.1. Preliminaries

Over the past decades, there have been some investigations about periodic non-linearities applications. But so far, no research has

3.2. Pipeline

The mainstream of WaveNet is shown in Figure 2(a). In Figure 2(a), for a degraded image $D \in \mathbb{R}^{H \times W \times 3}$, we first apply a convolutional layer to extract low-level features and expand channels $F_{in} \in \mathbb{R}^{H \times W \times C}$. Then, we use the Wave Transform Blocks (WTB) to transform $F_{in}$ into the signal-like features $\tilde{S}_n \in \mathbb{R}^{H \times W \times C}$. $n$ denotes the $n$-th WTB. Then we add an Adaptively Selective Feature Fusion (ASFF) block to connect the shallow feature $F_{in}$ and signal-like feature $\tilde{S}_n$. Using ASFF aims to enhance the feature presentation and weaken the impact of up-down sampling operations on image enhancement. Finally, we apply a convolutional layer to convert the signal-like features to a residual image $F_{out} \in \mathbb{R}^{H \times W \times 3}$. The enhanced image is obtained as $\hat{E}$. The overall process is summarized as:

$$ F_{out} = W^{out} \ast H_{ASFF}(F_{in}, \tilde{S}_n), $$

$$ \hat{E} = D \oplus F_{out}, \tag{1} $$

where $\oplus$ denotes the element-wise summation, $H_{ASFF}(\cdot)$ denotes the ASFF operation, $W^{out}$ denotes the last convolutional layer with filter size $3 \times 3$.

3.3. Wave Transform Block

As shown in Fig. 2(b), The WTB consists of two parts connected through skip-connection. The first part is a Wave Filter Block (WFB), which represents features in the wave form and allows modeling relationships between waves dynamically. The second part is a standard MLP layer to fuse channel information and enhance the transformation ability. The WTB calculation process can be computed as:

$$ \tilde{S}_n = H_{WFB}(\text{LN}(\tilde{S}_{n-1})) \oplus \tilde{S}_{n-1}, \quad \tilde{S}_0 = \tilde{S}_0, $$

$$ \tilde{S}_n = W_{\text{mlp}}(\text{LN}(\tilde{S}_n)) \oplus \tilde{S}_n, \tag{2} $$

where $\tilde{S}_n \in \mathbb{R}^{H \times W \times C}$ is the output of the first part, $H_{WFB}(\cdot)$ represents the WFB operation. LN(\cdot) stands for the Layer Normalization. $W_{\text{mlp}}$ denotes the Multilayer Perceptron (MLP) with two FC layers and one PReLU layer.

3.4. Wave Filter Block

In this section, we discuss the key component of our WaveNet. First, we recall the Discrete Fourier Transform (DFT) which proposes non-periodic discrete functions within the specified interval can be split into combinations of periodic functions. Its 1D version can be derived by:

$$ X[k] = M^{-1} \sum_{m=0}^{M-1} x[m](\cos(\frac{2\pi}{M} km) - j \sin(\frac{2\pi}{M} km)), \tag{3} $$

where $x[m]$ is a sequence of $M$ complex numbers, $X[k]$ indicates the spectrum at frequency $\frac{2\pi}{M}$, and $j$ represents the imaginary unit. It is clear that the spectrum at any frequency has global information in the frequency domain. Drawing on the idea of 1D-DFT, we adaptively estimate the Fourier coefficients to decompose a pixel into linear combinations of three waves. We take an overview about the formula of signal-like maps $\tilde{S}_n$ as follows:

$$ \tilde{S}_n = \text{Channel-FC}(\overline{GW}_n \oplus \overline{SW}_n \oplus \overline{SW}_n, W^{\text{fc}}) $$

$$ = W^{\text{fc}}(\overline{GW}_n \oplus \overline{SW}_n \oplus \overline{SW}_n), \tag{4} $$
where $W^{C}$ indicates the Channel-FC weights. As shown in Figure 3(b), it indicates that signal-like maps $\hat{S}_{0}$ includes three parts, Cosine Wave maps ($CW_{n}$), Sine Wave maps ($SW_{n}$), and Gating Wave maps ($GW_{n}$), $CW_{n}/SW_{n}/GW_{n} \in \mathbb{R}^{H \times W \times C}$.

1) wave-like representation: Traditional Fourier transform applied to image processing uses the coordinate as input for 2D-FFT calculation and remaps it into the frequency domain. It strongly relies on artificial priors, e.g., the frequency component of basis functions relies on the input resolution and scanning the whole picture. In contrast, we proposed the wave-like representation employing a neural network to transform each pixel into different components of waves for simplicity and efficiency. We represent the intermediate wave-like representation as follows:

$$\overline{CW}_{n} = A_n \otimes \cos(\theta_n), \quad \overline{SW}_{n} = B_n \otimes \sin(\alpha_n), \quad \overline{G}_{n} = \theta_n, \quad \overline{A}_{n}, \overline{B}_{n}, \overline{\theta}_{n}, \overline{A}_{n}, \overline{B}_{n}, \overline{\theta}_{n}, \overline{A}_{n}, \overline{B}_{n}, \overline{\theta}_{n} \in \mathbb{R}^{H \times W \times C},$$

where $\otimes$ denotes the element-wise product. $CW_{n}$ and $SW_{n}$ denote the intermediate waves without wave aggregation. $E_{dc}/E_{ap}/E_{ps}$ represents the set including DC $(A_{n}, B_{n})$, amplitude $(A_{n}, B_{n})$ and phase $(\theta_{n}, \alpha_{n})$ for simple expression. The $A_{n}$ and $B_{n}$ denote the DC component which indicates the original information comes from the previous layer. $A_{n}$ and $B_{n}$ denote the amplitude term that is a real-value feature representing the content of each wave. $\theta_{n}$ and $\alpha_{n}$ denote the phase term that includes the current location of a wave and frequency information. We use the wave-like feature representation to organize in a structured way to extract deeper abstract regularities in the latent space. Due to the periodicity and differentiable invariance, SIREN [SMB+20] also indicates that the periodic activation functions can speed up convergence, and periodicity in the latent space enables smooth interpolations and manipulations between different data points. This structured (wave-like features) representation encourages the model to learn and generate more coherent and realistic images.

2) Constructing wave-like feature maps: To get the wave-like maps in Eq. 5, the DC, amplitude and phase are required. Denote $S_{n-1}$ containing $j$ ($j = 1, 2, \ldots, H \times W$) signals as $S_{n-1} = [s_{n-1}^{1}, s_{n-1}^{2}, \ldots, s_{n-1}^{j}], \ldots, s_{n-1}^{j}$, where each signal $s_{n-1}^{j}$ is a C-dimension vector. For constructing DC/amplitude/phase components, we get them by a plain channel-FC operation, i.e.,

$$E_{dc/ap/ps} = \text{Channel-FC}(\overline{S}_{n-1}, W^{dc/ap/ps}),$$

where $W^{dc/ap/ps}$ is the weight with learnable parameters. There are different strategies for DC, amplitude and phase estimation. The most straightforward strategy represents these components with fixed parameters that can be learned during training. However, this way ignores the diversity of different input images. To dynamically capture the particular attributes according to the input feature, we adopt the Channel-FC in Eq. 6 to capture the particular attributes for simplicity and model performance. There are other constructing methods whose impact on the model performance is empirically investigated in the ablation study.

3) Aggregating wave-like features: The intermediate waves limit the feature structures in the latent space, which concentrates on modeling more general and regular features. However, as we mentioned in Sec. 1, aggregating waves in local and no-local to provide various semantic information is also essential for signal-like feature modulation. The static vector-sum aggregation [FLS11] is popularly employed to calculate wave superposition mode, but it may cause high FLOPs and slow the model speed accordingly. Besides, this static calculation method may not be able to cope with diverse inputs. As the basic operation in CNNs, convolution provides local connectivity and translation equivariance. These properties bring efficiency and generalization to dynamically aggregate waves. To modulate the spatial interactions between different waves, the proposed WaveNet uses different sizes of convolution kernels to gather wave-like features in different sizes of regions dynamically, which is also taken as a learnable linear combination of different waves to fit different local features. Thus, the wave-like dynamical aggregation can be formulated as follows:

$$\overline{CW}_{n}^{agg} = W^{cw}(CW_{n}),$$
$$\overline{SW}_{n}^{agg} = W^{sw}(SW_{n}),$$

where $W^{cw}$ and $W^{sw}$ are both learnable convolutional weights. $CW_{n}^{agg}$ and $SW_{n}^{agg}$ are the superposition waves. In addition, we use
Wave-FC to adaptively linearly combine the basis waves in Eq. 7 to adaptively assign weights to waves and enhance the representation capacity. Like the coefficients in Eq. 3, i.e.,
\[
\text{CW} \otimes \text{SW} = \text{Wave-FC}(\text{I}^\text{ref}_n, \Theta^\text{cw}, \Theta^\text{sw}),
\]
where \(\Theta^\text{cw}\) and \(\Theta^\text{sw}\) are the learnable coefficient maps. \(\text{I}^\text{ref}_n \in \mathbb{R}^{H \times W \times C}\) is the wave aggregation features. As shown in Figure 3(c), we obtain the final wave-like features after Wave-FC operation.

4) Gating Wave: Since the information of CW and SW is periodic, we design the Gating Wave (GW) for handling some non-periodic features. We use the gating mechanism which is activated with Tanh non-linearity to enhance the feature richness of the network. The GWs are formulated as:
\[
\text{GW}_n = \text{Wave-FC}(\text{Gating}(\text{S}_n-1), \text{GW}_n, \Theta^\text{ew}, \Theta^\text{kw}),
\]
where \(\oplus\) denotes the Channel-FC in Eq. 4 for simple expression and \(\oplus\) is the learnable weight to aggregate local information. The GWs retain valid information from the previous layer and provides complementary features to the next layer. The final \(\text{S}_n\) is obtained by Eq. 4. The whole WTB operation indicate that the output \(\text{S}_{n-1}\) from the previous WTB will be modulated by the next WTB.

3.5. Adaptively Selective Feature Fusion

The U-shape methods are proposed to balance the accuracy and computational cost. (which is common in low-level vision tasks because of the high-resolution inputs) However, the U-shape models may lose detail information after many up-down sampling operations. Thus, we design the ASFF to adaptively fuse shallow features and signal-like features. As shown in Figure 4, We use Channel-FC to compress the signal-like maps contents and shallow features, and capture the intra-channel interactions. Furthermore, we employ a Softmax to generate the intra-channel attention maps. Finally, we blend the two branch outputs with attention maps and use the element-wise product to adaptively build the relationship between shallow two kinds of features. In Figure 2, we can see that the output from ASFF goes to a 3 × 3 convolutional layer to transform the signal-like/linear-projection hybrid features maps into the final residual image \(\text{F}_{\text{out}}\). Overall, benefitting from the proposed adaptive wave-like representation, our WaveNet achieves high performance with acceptable computation-consuming.

3.6. Loss Function

Given an output image \(\hat{E}\) and a ground truth image \(E\), we use a signal-based loss function to guide our signal-like feature representation. Specifically, the loss function \(L\) for WaveNet consists of PSNR loss \(L_{\text{psnr}}\), MS-SSIM loss \(L_{\text{ssim}}\) and Edge loss \(L_{\text{edge}}\), i.e.
\[
L = \lambda_1 \cdot L_{\text{psnr}}(\hat{E}, E) + \lambda_2 \cdot L_{\text{ssim}}(\hat{E}, E) + \lambda_3 \cdot L_{\text{edge}}(\hat{E}, E).
\]

4. Experiment

In this section, we test our WaveNet on the popular real-world benchmark datasets for image enhancement and dark face detection. Various ablation studies provided in supplemental material demonstrate performance and effectiveness of wave-like modeling.

4.1. Dataset and Experimental Setup

The proposed WaveNet is tested on six datasets including five low/normal real-captured datasets and one high-level application dataset. The datasets for image enhancement include LOL [WWYL18], VE-LOL [LXY*21], SICE [CGZ18], SID [CCXK18] and MIT-Adobe FiveK [BPCD11]. DARK FACE [YYR*19] is composed of various scenes with faces taken in the dark. Note that we use ‘expert C’ as FiveK ground truth and the Sony subset following the script provided by SID [CCXK18] to transfer the low-light images from RAW to RGB for our training and testing.

Implementation Details: WaveNet is end-to-end trainable and requires no pretraining on large datasets. For data augmentation, we trained the network using random horizontal-vertical flips, rotation, and MixUp [ZCDLP17]. We set the Adam optimizer [KB14] with an initial learning rate of 1×10^{-4}, which is steadily decreased to 10^{-6} with the cosine annealing decay [LH16]. The values of \(\lambda_1/\lambda_2/\lambda_3\) in Eq. 10 are 0.33/0.33/0.1. The proposed WaveNet is trained on a single NVIDIA RTX 3090 using the PyTorch. By varying the width and depth of the model, we build 3 models with different parameters and computational costs, denoted as WaveNet-T, WaveNet-S, and WaveNet-B sequentially.

Metrics: To evaluate the performance of WaveNet, we adopt Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) as the main evaluation metrics, which are classic to measure the difference between output and ground truth.

4.2. Quantitative Evaluation

Note that we obtain these results in Table 1 either from the respective papers or by running the respective public code and full-resolution testing. Low-light Image Enhancement: LOL [WWYL18] and SICE [CCXK18] include extremely dark images with lots of noise. It is challenging to complete the low-light image enhancement task on these datasets. We tested our methods on this dataset, and the results are shown in Table 1. From Table 1, the proposed WaveNet-B achieves 40.44/0.88 dB gain in PSNR over the previous best model LLFlow [WWY*22] and Restormer [ZAK*22] on LOL/SID. WaveNet-T is a lightweight model. It shows promising performance of quality and efficiency. Our WaveNet-T gains better (23.59/22.57dB) result with only 80k parameters compared to the currently popular lightweight method IAT [CLG*22] and image restoration model MAXIM [TTZ*22]. WaveNet-S provides a balance of accuracy and efficiency, which
achieves competitive results on all datasets.

**Image Enhancement**: MIT-Abode FiveK [BPCD11] is a real-world image enhancement dataset containing a wide variety of scenes for training and testing. As shown in Table 1, our WaveNet achieves the best and second-bests scores and surpasses all the baselines. Our WaveNet-B achieves 0.70dB gain in PSNR at most compared to the MAXIM [TTZ*22].

**Cross-dataset Evaluation**: To evaluate our method’s generality and effectiveness in real no-reference degraded images, we take the cross-dataset validation method. e.g. Our method is trained on the LOL dataset, and the model is directly tested on the testing set of VE-LOL (Real) and SICE. From the results in Table 1 we can see that our method significantly outperforms other trained methods in all metrics and provides a 2.05/1.08dB promotion on VE-LOL/SICE compared with previous SOTA methods. From another perspective, our WaveNet can provide a more stable and general result without retraining.

**Efficiency Analysis**: It is worth noting that efficiency term are tested on a 256×256 image. Compared with IAT [CLG*20], which is also a lightweight model, WaveNet-T with 12% fewer parameters than IAT [CLG*20] achieves better PSNR and SSIM scores on all datasets. Compared with the previous SOTA LLFlow [WWW*22], WaveNet-T has 220× fewer parameters and 64× fewer FLOPs but achieves its 94% accuracy on LOL [WWW*18] dataset. Compared to Restormer [ZAK*22], whose FLOPs is close to our WaveNet-S, our WaveNet-S has 18.6× fewer parameters and faster than it. Besides, Our WaveNet-S provides better generality in cross-dataset validation. With acceptable computational costs, WaveNet-B achieves the best results on all datasets in Table 1. Overall, our WaveNet achieves the best trade-off on accuracy and efficiency. The superiority of WaveNet implies that the proposed module WTB has a large potential and modulating the feature representation.

### 4.3. Qualitative Evaluation

The qualitative comparisons of images are given in Figure 5. The higher PSNR and SSIM values in Table 1 indicate that our methods can better restore the color and better reserve details. From the Figure 5 we can see that our WaveNet keeps more fine details, more natural color consistency, higher contrast, and less noise compared with other methods. Specifically, it can be observed in cross-dataset validation visual results (4th and 5th row) in Figure 5 that Our WaveNet not only suppresses overexposure in regions of high brightness but also enhances natural colors and details in regions of shadows. Hence, compared with other methods, the proposed WaveNet use waveform feature representations achieving excellent performance in generality and qualitative evaluation.

### 4.4. High-level Vision Evaluation:

To validate the performance of our WaveNet in high-level vision tasks, we use the DARK FACE dataset [YYR*19] composed of images with faces taken in the dark. The Dual Shot Face Detector (DSFD) [LWW*19] trained on the WIDER FACE dataset.
[YLLT16] is used as the face detector. We take the last 500 images of DARK FACE [YYR∗19] training set enhanced by different methods and feed them to the DSFD [LWW∗19]. The results shown in Table 2 validate that our WaveNet gains the best score. The bottom row in Figure 5 shows that DSFD detected the most faces using the WaveNet-enhanced images.

5. Conclusion

This paper proposes an architecture for image enhancement tasks, dubbed WaveNet, which aims to extract more spatial details, better color accuracy, and a higher contrast from the original degraded images. Inspired by signal processing, we formulate the enhancement task as a signal modulation problem. We use a learning-based method to adaptively construct and modulate the signals come from digital images. Specifically, we utilize the Fourier transform theory to decompose a pixel into three wave functions (Cosine Wave (CW), Sine Wave (SW), Gating Wave (GW)) for feature characterization. The amplitude and phase are essential for constructing wave-like features. Amplitude is the original real-valued feature, and phase modulates the relationship between the varying inputs and fixed weights in WaveNet. Specifically, we propose the Wave Transform Block (WTB) to dynamically construct the wave-like feature maps and modulate the wave interactions (wave superposition) locally and non-locally. Furthermore, WaveNet enhances various inputs by adaptively adjusting the amplitude and phase of each wave in the constructed signal-like feature maps. The extensive experiments on the benchmark datasets validate the effectiveness of our WaveNet for image enhancement.

Acknowledgement. This work was supported by Science and Technology Service Network Initiative (KFJ-STS-QYZD-2021-21-001), Sichuan Science and Technology Program (2019ZDX0006, 2020YFQ0056), and the Talents by Sichuan provincial Party Committee Organization Department.
References


[Gal88] GALLANT: There exists a neural network that does not make avoidable mistakes. 657–664. 3


[PHV16] PARASCANDOLO G., HUTTUNEN H., VIRTANEN T.: Taming the waves: sine as activation function in deep neural networks, 2016. 3


