

# FourLLIE: Boosting Low-Light Image Enhancement by Fourier Frequency Information

Chenxi Wang

Sun Yat-sen University

Shenzhen, Guangdong, China

wangchx67@mail2.sysu.edu.cn

Hongjun Wu

Sun Yat-sen University

Shenzhen, Guangdong, China

wuhj33@mail2.sysu.edu.cn

Zhi Jin\*

Sun Yat-sen University

Shenzhen, Guangdong, China

jinz26@mail2.sysu.edu.cn

## ABSTRACT

Recently, Fourier frequency information has attracted much attention in Low-Light Image Enhancement (LLIE). Some researchers noticed that, in the Fourier space, the lightness degradation mainly exists in the amplitude component and the rest exists in the phase component. By incorporating both the Fourier frequency and the spatial information, these researchers proposed remarkable solutions for LLIE. In this work, we further explore the positive correlation between the magnitude of amplitude and the magnitude of lightness, which can be effectively leveraged to improve the lightness of low-light images in the Fourier space. Moreover, we find that the Fourier transform can extract the global information of the image, and does not introduce massive neural network parameters like Multi-Layer Perceptrons (MLPs) or Transformer. To this end, a two-stage Fourier-based LLIE network (FourLLIE) is proposed. In the first stage, we improve the lightness of low-light images by estimating the amplitude transform map in the Fourier space. In the second stage, we introduce the Signal-to-Noise-Ratio (SNR) map to provide the prior for integrating the global Fourier frequency and the local spatial information, which recovers image details in the spatial space. With this ingenious design, FourLLIE outperforms the existing state-of-the-art (SOTA) LLIE methods on four representative datasets while maintaining good model efficiency. Notably, compared with a recent Transformer-based SOTA method SNR-Aware, FourLLIE reaches superior performance with only 0.31% parameters. Code is available at <https://github.com/wangchx67/FourLLIE>.

## CCS CONCEPTS

- Computing methodologies → Computer vision; Image processing.

## KEYWORDS

Low-light image enhancement, Fourier frequency information, Amplitude transform map, Signal-to-noise-ratio map,

## ACM Reference Format:

Chenxi Wang, Hongjun Wu, and Zhi Jin. 2023. FourLLIE: Boosting Low-Light Image Enhancement by Fourier Frequency Information. In *Proceedings*

\*Corresponding author.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*MM '23, October 29–November 3, 2023, Ottawa, ON, Canada.*

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0108-5/23/10.

<https://doi.org/10.1145/3581783.3611909>

*of the 31st ACM International Conference on Multimedia (MM '23), October 29–November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3581783.3611909>

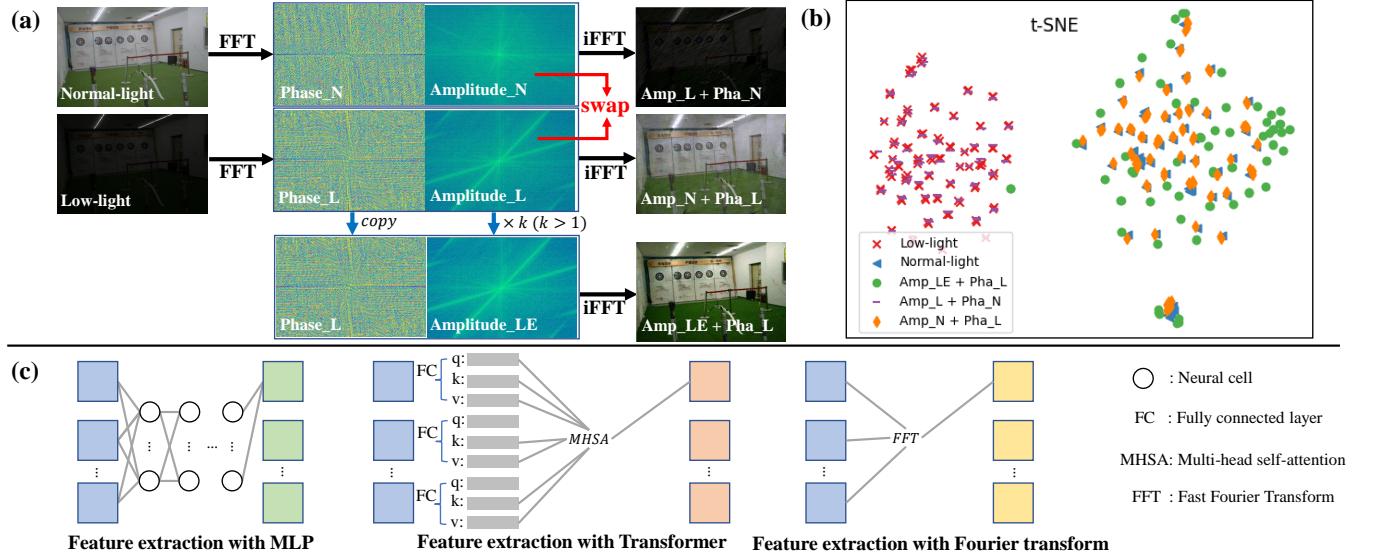
## 1 INTRODUCTION

The fast development of imaging devices eases image capture during our daily life. However, images obtained under low-light conditions easily suffer from unpleasing visibility, which may be unfriendly to users and some high-level computer vision tasks (e.g., action recognition [32], face detection [46], and object detection [35]). Thus, Low-Light Image Enhancement (LLIE), which aims to recover hidden information and improve the quality of low-light images, is posed as an active research field in computer vision.

There are many impressive LLIE methods have been proposed with great performance. Generally, they can be roughly divided into non-learning based methods [5, 6, 12, 20, 33, 34, 40] and learning based methods [9, 17, 23, 26, 27, 29, 42, 43, 54]. However, as mentioned in [14], most of them are based on the spatial information and rarely consider the Fourier frequency information, which has been proved to be effective for improving the image quality.

Recently, some methods [14, 24] explore the Fourier frequency information for LLIE. They find that, in the Fourier space, the most lightness representation is concentrated in the amplitude component while the phase component contains the lightness-irrelevant information (e.g., structure [14] or noise [24]). Based on this observation, they integrate both the Fourier frequency and spatial information into neural networks and achieve impressive results.

Inspired by previous Fourier-based works [14, 24], we further explore the properties of the Fourier frequency information for LLIE. Our motivations are shown in Fig. 1. Firstly, given two images with the same context but different light conditions (i.e., low-light and normal-light), we swap their amplitude components and combine them with corresponding phase components in the Fourier space. The recombined results show that the light conditions are swapped following the amplitude swapping (see the top two rows of Fig. 1 (a)). This phenomenon indicates **the amplitude component represents the lightness of an image**. Then, by only enlarging the magnitude of the amplitude component of the low-light image and keeping the phase component, we find that the low-light image is brightened (see the bottom row of Fig. 1 (a)) and has closer distributions to the normal-light image (see Fig. 1 (b)). From this, we conclude that **the lightness of low-light images can be improved by enlarging the magnitude of its amplitude component in the Fourier space**. Moreover, we notice that both Transformer [38], MLP, and Fourier transform [2] can extract global information (see Fig. 1 (c)), while feature extraction with



**Figure 1: Our motivations.** (a) The amplitude components of low-light and normal-light images with the same context are swapped to get recombined results  $\text{Amp\_L+Pha\_N}$  and  $\text{Amp\_N+Pha\_L}$  (the top and middle rows). The amplitude component of the low-light image is enlarged by a constant value  $k$  to get  $\text{Amp\_LE+Pha\_L}$  (the bottom row). From visual results we can infer that the amplitude component reflects the lightness representation of an image and the magnitudes of amplitude represent the magnitudes of lightness. (b) The t-SNE [37] embeddings of low-light, normal-light,  $\text{Amp\_LE+Pha\_L}$ ,  $\text{Amp\_L+Pha\_N}$  and  $\text{Amp\_N+Pha\_L}$ . The distributions of  $\text{Amp\_LE+Pha\_L}$  and  $\text{Amp\_N+Pha\_L}$  are similar to normal-light, while the distribution of  $\text{Amp\_L+Pha\_N}$  is similar to low-light. (c) The features extracted by Transformer [38], MLP, and Fourier transform [2] are global information. However, the feature extracted by the Fourier transform does not introduce massive parameters of neural networks.

the Fourier transform is more efficient since it **does not introduce massive parameters of neural networks**.

Based on the above observations, in this work, we propose a new LLIE method (called FourLLIE) by leveraging the Fourier frequency information. FourLLIE accomplishes enhancement through two stages: frequency stage and spatial stage. In the frequency stage, FourLLIE achieves lightness improvement by estimating the transform map of the amplitude component. This process is similar to illumination map estimation methods [26, 29, 39] based on Retinex theory [20], while it is implemented in the Fourier space. Then, we introduce the Signal-to-Noise-Ratio (SNR) map into the spatial stage for further details refinement. Generally, the image regions with different SNR values indicate different amount of information and noise. The regions with higher SNR values are better improved with local information, oppositely, they are better improved with global information [45]. Based on this property and the global property of the Fourier frequency information, we propose to recover image regions with higher SNR values by the local spatial information and image regions with lower SNR values by the Fourier frequency information. In this way, a low-light image is first brightened in the frequency stage and then refined in the spatial stage.

Compared with the existing Fourier-based image enhancement methods [14, 24], we go step further to explore the property that the magnitude of amplitude is positively correlated with the magnitude of lightness, and estimate a transform map to improve the magnitude of the amplitude component. Experiments in Sec. 3.2 demonstrate the effectiveness of the proposed design. Besides, we

introduce the SNR map to well utilize the global property of the Fourier frequency information. Compared with SNR-Aware [45], which adopts Transformer [38] as the global information extractor to process low SNR regions, we adopt the Fourier frequency information to represent the global information and remarkably reduce model complexities (see experiments in Sec. 3.3). To sum up, our contributions are three folds:

1) We explore the positive correlation between the magnitudes of amplitude component in the Fourier space and lightness in the spatial space. By utilizing this property, we provide an effective and reasonable way to achieve lightness improvement in the Fourier space.

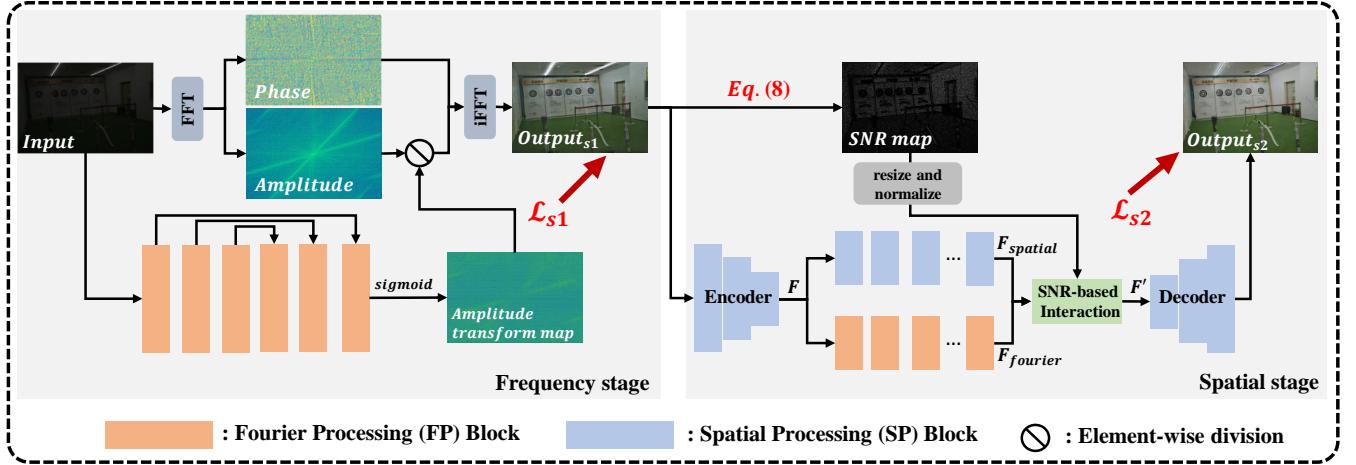
2) We introduce the SNR map to integrate the Fourier frequency information and spatial information. It takes full advantage of the global property of the Fourier frequency information.

3) We conduct extensive experiments on four commonly used LLIE datasets, demonstrating proposed method outperforms existing SOTA methods while preserving good model efficiency.

## 2 RELATED WORK

### 2.1 Low-Light Image Enhancement

Low-Light Image Enhancement (LLIE) aims to improve the quality of low-light images. Before the development of deep learning, many non-learning based LLIE methods, which include HE-based [33] and Retinex-based [20, 34], are proposed. However, these methods may hardly handle noise and color well. Recently, with the rapid



**Figure 2: Overall architecture of the proposed method.** The input image is first processed in the frequency stage to improve the lightness. Then, an SNR-based spatial stage is designed to further recover the details. Note that, in the spatial stage, the SP blocks in encoder and decoder modules are followed by down-sampling and up-sampling operations.

development of deep learning, a rich set of learning based architectures [9, 17–19, 23, 25–27, 29, 31, 42, 45, 54, 55] are proposed. Lore *et al.* [27] first introduced Convolutional Neural Networks (CNNs) to LLIE. Gharbi *et al.* [8] accomplished enhancement by the bilateral up-sampling network. Zamir *et al.* [51] proposed more advanced CNN architecture, which includes attention mechanisms and multi-scale designs, to achieve robust LLIE. Xu *et al.* [45] introduced Transformer [38] and SNR map to LLIE. Besides, another line of LLIE methods tries to combine Retinex theory and deep learning. Wei *et al.* [42] and Zhang *et al.* [54] achieved enhancement by decomposing a low-light image to reflectance and illumination components. Methods in [26, 29, 39] estimated the illumination map of low-light images. However, the above methods may pay less attention to the Fourier frequency information, which is effective for LLIE.

Recently, researchers explore the Fourier frequency information for LLIE. Huang *et al.* [14] pointed out amplitude component can reflect lightness representation of under-/over-exposure images. Li *et al.* [24] found lightness and noise can be decomposed in the Fourier space. Based on these observations, they reached remarkable performance by integrating both the Fourier frequency information and spatial information. However, these works may hardly explore the relationship between the amplitude components of low-/normal- light images and well utilize the global property of the Fourier frequency information.

## 2.2 Fourier Frequency Information

Fourier frequency information has been demonstrated as the effective representation in many computer vision areas [7, 10, 14, 16, 24, 44, 49, 50, 56–58]. For example, Xu *et al.* [44] developed a Fourier-based data augmentation for domain generation. Fuoli *et al.* [7] adopted Fourier-based loss to help restore the high-frequency information in image super-resolution. Yu *et al.* [50] applied the Fourier frequency information to the image dehazing and methods

in [56, 57] utilized it for pan-sharpening. Zhou *et al.* [58] designed a Fourier-based up-sampling and can improve many computer vision tasks in a plug-and-play way. Besides, Huang *et al.* [14] and Li *et al.* [24] developed the Fourier-based LLIE algorithms, while they still have certain limitations as analyzed before. The above advances provide wide applications of the Fourier frequency information and motivate us to further explore its properties in LLIE.

## 3 METHOD

### 3.1 Fourier Frequency Information

Firstly, we briefly introduce the Fourier frequency information. Given an input image  $x$ , whose shape is  $H \times W$ , the transform function  $\mathcal{F}$  which converts  $x$  to the Fourier space  $X$  can be represented as follow:

$$\mathcal{F}(x)(u, v) = X(u, v) = \frac{1}{\sqrt{HW}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} x(h, w) e^{-j2\pi(\frac{h}{H}u + \frac{w}{W}v)} \quad (1)$$

where  $h, w$  are the coordinates in the spatial space and  $u, v$  are the coordinates in the Fourier space,  $j$  is the imaginary unit, the inverse process of  $\mathcal{F}$  is denoted as  $\mathcal{F}^{-1}$ ,  $X(u, v)$  consists of complex values and can be represented by:

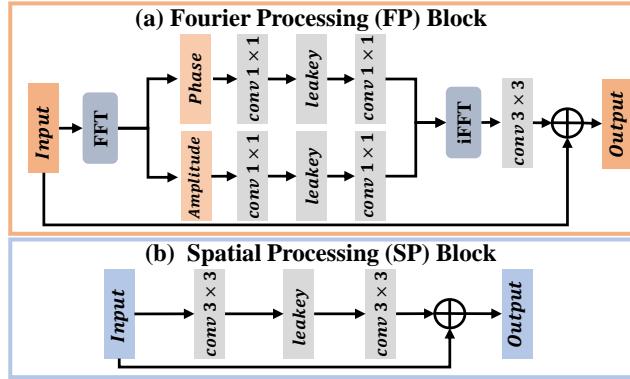
$$X(u, v) = R(X(u, v)) + jI(X(u, v)) \quad (2)$$

where  $R(X(u, v))$  and  $I(X(u, v))$  are the real and imaginary parts of  $X(u, v)$ , respectively.

Besides,  $X(u, v)$  in the Fourier space can be represented by an amplitude component  $\mathcal{A}(X(u, v))$  and a phase component  $\mathcal{P}(X(u, v))$  as follows:

$$\mathcal{A}(X(u, v)) = \sqrt{R^2(X(u, v)) + I^2(X(u, v))} \quad (3)$$

$$\mathcal{P}(X(u, v)) = \arctan[\frac{I(X(u, v))}{R(X(u, v))}] \quad (4)$$



**Figure 3: The illustration of Fourier Processing (FP) block and Spatial Processing (SP) block.**

where the  $R(X(u, v))$  and  $I(X(u, v))$  can also be obtained by:

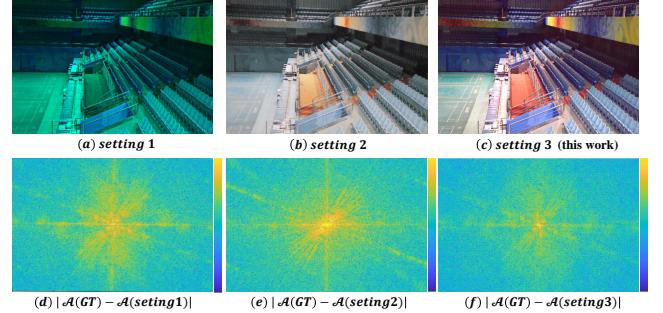
$$\begin{aligned} R(X(u, v)) &= \mathcal{A}(X(u, v)) \times \cos(\mathcal{P}(X(u, v))) \\ I(X(u, v)) &= \mathcal{A}(X(u, v)) \times \sin(\mathcal{P}(X(u, v))) \end{aligned} \quad (5)$$

According to previous Fourier-based methods [14, 24] and Fig. 1, we conclude that: 1) *the lightness of the low-light image can be improved by enlarging the magnitude of amplitude component in the Fourier space.* 2) *the Fourier transform can extract global information and does not introduce massive parameters of neural networks.* Based on the above two conclusions, we propose to enhance low-light images in two stages: frequency stage and spatial stage. The frequency stage improves the lightness of the low-light image by estimating the amplitude transform map to improve the magnitude of the amplitude component in the Fourier space. The spatial stage utilizes the global properties of the Fourier frequency information and SNR map to further recover the details. The detailed implementations are in the following parts.

### 3.2 Frequency Stage in FourLLIE

The enhancement in the Frequency stage is shown in Fig. 2. The input image is fed into six Fourier Processing (FP) blocks with skip connections to estimate the amplitude transform map. FP block is designed to extract the Fourier frequency features (as shown in Fig. 3 (a)). The input of each FP block is first transformed to the Fourier space to get the amplitude and phase components. Then, for each component, two  $1 \times 1$  convolutional layers with a LeakyReLU activation are applied to extract features. Finally, these two components are transformed back to the spatial space followed by a  $3 \times 3$  convolutional layer and the residual of input. The final *sigmoid* activation limits the transform map within  $(0, 1)$ . Therefore, the result of the frequency stage  $Output_{s1}$  is obtained by the division between the amplitude component of the input image and estimated transform map. This process can be expressed as:

$$\begin{aligned} A_{in} &= \mathcal{A}(\mathcal{F}(input)), \quad P_{in} = \mathcal{P}(\mathcal{F}(input)) \\ A_{out} &= A_{in}/(M + \epsilon) \\ R_{out} &= A_{out} \times \cos(P_{in}) \quad I_{out} = A_{out} \times \sin(P_{in}) \\ Output_{s1} &= \mathcal{F}^{-1}(R_{out} + jI_{out}) \end{aligned} \quad (6)$$



**Figure 4: Ablation study for amplitude prediction ways.** GT represents the ground truth image and  $\mathcal{A}(\cdot)$  represents to extract amplitude component in the Fourier space. (a), (b), and (c) present the visual results of different settings. (d), (e), and (f) present the error maps of amplitude components between different settings and GT. It can be seen that the setting of this work (setting 3) reaches the best results no matter in the spatial space or Fourier space.

where  $A_{in}$  and  $P_{in}$  are amplitude and phase components of the input image, respectively,  $A_{out}$  is the output amplitude component,  $R_{out}$  and  $I_{out}$  represent the real and imaginary parts of  $Output_{s1}$  in the Fourier space, respectively,  $M$  is estimated amplitude transform map and  $\epsilon$  is set to  $1.0 \times 10^{-8}$  for avoiding zero-division.

In this way, the lightness of the input low-light image is improved with the enlargement of its amplitude. The loss involved in the frequency stage  $\mathcal{L}_{s1}$  is expressed as:

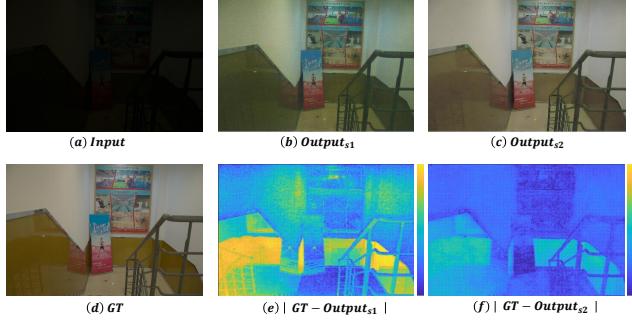
$$\mathcal{L}_{s1} = \|\mathcal{A}(\mathcal{F}(Output_{s1})) - \mathcal{A}(\mathcal{F}(GT))\|_2 \quad (7)$$

where GT is the ground truth image.

Compared with FECNet [14], which also improves the lightness by applying a constrain on the amplitude component, we utilize the positive correlation (an amplitude transform map within  $(0, 1)$ ) between the magnitudes of the amplitude and lightness. It makes enhancement in the Fourier space more reasonable and robust for LLIE.

To verify the effectiveness of estimating the amplitude transform map, we conduct an experiment with three different settings: 1) estimating the output amplitude directly and applying a constraint on the estimated amplitude component, 2) estimating the output image in the spatial space and applying a constraint on the amplitude component by transforming the obtained image to the Fourier space (setting of FECNet [14]), 3) estimating the amplitude transform map and applying a constraint on the amplitude component transformed by estimated transform map (setting of this work). The detailed experiment settings can be found in the supplementary material. We denote them as setting 1, setting 2, and setting 3 for simplicity.

As shown in Fig. 4 (a) and (d), since the amplitude component is very complex and does not have structural property or other related priors like images in the spatial space, it is hard to learn a well enhanced amplitude directly by a neural network. As for setting 2, it easily leads to detail losses (e.g., color) for low-light images since it neglects the relationship between amplitude components



**Figure 5: Visualization of outputs in different stages. (b) and (c) are the outputs of the frequency stage and the spatial stage, respectively. (e) and (f) present the error maps between the outputs of different stages and GT. It can be seen that the frequency stage improves the lightness and the spatial stage recover the details.**

of low-light and normal-light images (see Fig. 4 (b) and (e)). However, when predicting the amplitude component by estimating the transform map like setting 3, the details of the image and structure of amplitude can be recovered better as shown in Fig. 4 (c) and (f).

### 3.3 Spatial Stage in FourLLIE

The lightness of the low-light image can be improved well in the frequency stage, however, it still suffers some detail degradation (as shown in Fig. 5 (b) and (e)). Meanwhile, based on the property of SNR map, regions of an image with lower SNR values need long-range (global) operations to restore, while the regions with higher SNR values prefer short-range (local) operations. In this work, we further extend that the regions with lower SNR values tend to be processed in the Fourier space, while the regions with higher SNR values prefer to be processed in the spatial space. Therefore, we propose to integrate the global Fourier frequency information and local spatial information based on the SNR map and recover details. The enhancement in the spatial stage is designed as shown in Fig. 2. Given the output of the frequency stage  $Output_{s1}$ , we first compute the SNR map  $S$  followed [45] as:

$$\hat{I}_g = blur(I_g), \quad N = abs(I_g - \hat{I}_g), \quad S = \hat{I}_g/N \quad (8)$$

where  $I_g$  is the gray-scale version of  $Output_{s1}$ ,  $blur$  is a Gaussian blur with kernel size of 5,  $abs(\cdot)$  represents to compute the absolute value,  $\hat{I}_g$  represents the noise-free version of  $I_g$ ,  $N$  represents the noise component. Then, we feed the  $Output_{s1}$  into an encoder to extract the features  $F$ . The features  $F$  then separately go through Fourier Processing (FP) blocks and Spatial Processing (SP) blocks to produce the global features  $F_{fourier}$  and the local features  $F_{spatial}$ , respectively. As shown in Fig. 3, FP blocks process features in the Fourier space and can extract global features, while SP blocks process features in the spatial space and can extract local features. Next, the  $F_{fourier}$  and  $F_{spatial}$  are combined based on SNR map  $S$  as:

$$F' = F_{spatial} \times S + F_{fourier} \times (1 - S) \quad (9)$$

where  $F'$  is the output features. Note that  $S$  is normalized to  $[0, 1]$  and resized to the corresponding shape as  $F_{spatial}$  and  $F_{fourier}$ . The result of the spatial stage  $Output_{s2}$  is produced by feeding the

**Table 1: Effectiveness of the Fourier frequency information.** We replace Transformer blocks with FP blocks in SNR-Aware [45]. The best results are boldfaced and the second-best ones are underlined. Note that “nc” represents the number of channels of the latent features.

Methods	nc	PSNR	SSIM	#Para
SNR-Aware-Transformer	64	<b>20.67</b>	<b>0.5910</b>	39.12M
	32	20.45	<u>0.5892</u>	12.94M
	16	20.02	0.5808	4.81M
SNR-Aware-Fourier	64	20.62	<b>0.5910</b>	1.46M
	32	<u>20.63</u>	0.5868	<u>0.37M</u>
	16	20.61	0.5886	<b>0.09M</b>

output features  $F'$  into the decoder. As shown in Fig. 5 (c) and (f), the noise contained in  $Output_{s1}$  is removed after the spatial stage. The loss involved in the spatial stage  $\mathcal{L}_{s2}$  can be expressed as:

$$\mathcal{L}_{s2} = ||Output_{s2} - GT||_2 + \alpha ||\phi(Output_{s2}) - \phi(GT)||_2 \quad (10)$$

where  $\phi$  is a pre-trained VGG [36] network,  $\alpha$  is a weight factor and set to 0.1 empirically.

Compared with previous Fourier-based LLIE methods [14, 24], which accomplish spatial-frequency interaction by features concatenation, FourLLIE introduces the SNR map as prior and well utilizes the global property of the Fourier frequency information. Compared with SNR-Aware [45], which uses Transformer [38] to extract global information, Fourier-based processing can reduce model complexity significantly.

To verify the effectiveness of using Fourier frequency information, we conduct an experiment by replacing the Transformer blocks in SNR-Aware [45] with FP blocks and training it on LSRW-Huawei [13] dataset. We denote SNR-Aware with Transformer blocks as SNR-Aware-Transformer and SNR-Aware with FP blocks as SNR-Aware-Fourier. As shown in Table 1, SNR-Aware-Fourier reaches competitive performance with only about 1/20 parameters of SNR-Aware-Transformer. Besides, we find SNR-Aware-Fourier is not sensitive to channels of the latent features. SNR-Aware-Fourier with only 16 channels of the latent features still can perform well on LSRW-Huawei. Nevertheless, the reduction of channels for SNR-Aware-Transformer may lead to performance degradation. We infer that it benefits from the inherently global nature of the Fourier frequency information, which does not rely on massive parameters of the neural network.

Finally, the overall loss  $\mathcal{L}_{total}$  of FourLLIE can be expressed as :

$$\mathcal{L}_{total} = \mathcal{L}_{s2} + \lambda \mathcal{L}_{s1} \quad (11)$$

where  $\lambda$  is a weight factor set to 0.01 empirically.

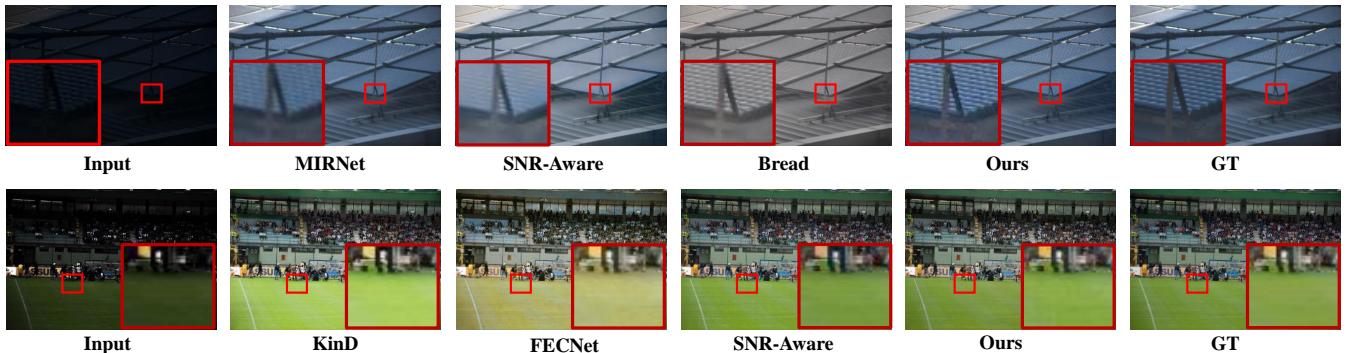
## 4 EXPERIMENT

### 4.1 Datasets and Implementation Details

We choose four widely used LLIE datasets for evaluating FourLLIE, including LOL-Real [48], LOL-Synthetic [48], LSRW-Huawei [13], and LSRW-Nikon [13]. LOL-Real is captured in real scenes by changing exposure time and ISO. It contains 689 low-/normal-

**Table 2: Quantitative comparison on the LOL-Real [48], LOL-Synthetic [48], LSRW-Huawei [13], and LSRW-Nikon [13]. The best results are boldfaced and the second-best ones are underlined.**

Methods	LOL-Real [48]			LOL-Synthetic [48]			LSRW-Huawei [13]			LSRW-Nikon [13]			#Param (M)
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	
LIME [12]	15.24	0.4190	0.2203	16.88	0.7578	0.1041	17.00	0.3816	0.2069	13.53	0.3321	0.1272	-
MF [5]	18.72	0.5089	0.2401	17.50	0.7737	0.1075	18.26	0.4283	0.2153	15.44	0.3997	<u>0.1269</u>	-
NPE [40]	17.33	0.4642	0.2359	16.60	0.7781	0.1079	17.08	0.3905	0.2303	14.86	0.3738	0.1464	-
SRIE [6]	14.45	0.5240	0.2160	14.50	0.6640	0.1484	13.42	0.4282	0.2166	13.26	0.3963	0.1396	-
DRD [42]	16.08	0.6555	0.2364	18.28	0.7737	0.1470	18.23	0.5220	0.1926	15.18	0.3809	0.1690	0.86
Kind [54]	20.01	0.8412	0.0813	22.62	0.9041	0.0515	16.58	0.5690	0.2259	11.52	0.3827	0.1860	8.02
Kind++ [53]	20.59	0.8294	0.0875	21.17	0.8814	0.0678	15.43	0.5695	0.2366	14.79	0.4749	0.2111	8.27
MIRNet [51]	<u>22.11</u>	0.7942	0.1448	22.52	0.8997	0.0568	19.98	0.6085	0.2154	17.10	<u>0.5022</u>	0.2170	31.79
SGM [48]	20.06	0.8158	<u>0.0727</u>	22.05	0.9089	0.4841	18.85	0.5991	0.2492	15.73	0.4971	0.2234	2.31
FECNet [14]	20.67	0.7952	0.0995	22.57	0.8938	0.0699	<u>21.09</u>	0.6119	0.2341	17.06	0.4999	0.2192	<u>0.15</u>
HDMNet [25]	18.55	0.7132	0.1717	20.54	0.8539	0.0690	20.81	0.6071	0.2375	16.65	0.4870	0.2157	2.32
SNR-Aware [45]	21.48	<b>0.8478</b>	0.0740	<u>24.13</u>	<b>0.9269</b>	<b>0.0318</b>	20.67	0.5910	0.1923	<u>17.54</u>	0.4822	<b>0.0982</b>	39.12
Bread [11]	20.83	0.8217	0.0949	17.63	0.8376	0.0681	19.20	<u>0.6179</u>	0.2203	14.70	0.4867	0.1766	2.12
Ours	<b>22.34</b>	<u>0.8468</u>	<b>0.0511</b>	<b>24.65</b>	<u>0.9192</u>	<u>0.0389</u>	<b>21.30</b>	<b>0.6220</b>	<b>0.1719</b>	<b>17.82</b>	<b>0.5036</b>	0.2150	<b>0.12</b>



**Figure 6: Qualitative comparison in LOL-Real [48] (the first row) and LOL-Synthetic [48] (the second row). It can be seen that the proposed method reaches the best visual results. Note that we compare the proposed method with three benchmarks with the best PSNR values on the corresponding datasets.**

light image pairs for training and 100 low-/normal- light image pairs for testing. It is worth noting that LOL-Real is the extended version of LOL [42], which contains 485 training image pairs and 15 testing image pairs. Considering the image pairs from LOL-Real and LOL are overlapped, we only evaluate in LOL-Real, since it is more diverse. LOL-Synthetic is synthesized from raw images by analyzing the distribution of luminance channels of low-/normal-light images. It contains 900 low-/normal- light image pairs for training and 100 low-/normal- light image pairs for testing. LSRW-Huawei and LSRW-Nikon are captured in real scenes like LOL-Real but with different devices. LSRW-Huawei is collected by a Huawei P40 Pro and LSRW-Nikon is collected by a Nikon D7500. LSRW-Huawei contains 3150 training image pairs and 20 testing image pairs. LSRW-Nikon contains 2450 training image pairs and 30 testing image pairs. Besides, we also evaluate FourLLIE on five unpaired datasets DICM [22] (64 images), LIME [12] (10 images), MEF [28] (17 images), NPE [40] (85 images), and VV<sup>1</sup> (24 images).

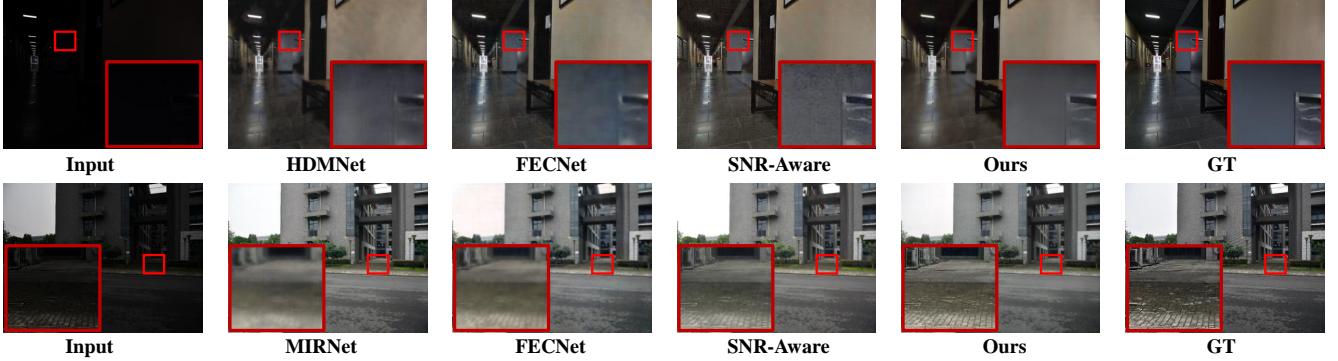
We implement FourLLIE in PyTorch and train it on an NVIDIA 3090 GPU for 0.5 days. The learning rate is initialized to  $4.0 \times 10^{-4}$  and a multi-step scheduler is adopted. We choose Adam [21] with momentum 0.9 as the optimizer. During the training, the input image is cropped to  $384 \times 384$ , and random rotation and flip augmentations are adopted. The batch size is set to 4 and the total training iterations are set to  $2.0 \times 10^5$ .

## 4.2 Comparison with State-Of-The-Arts

In this paper, we compare the proposed method with thirteen state-of-the-art LLIE methods, including traditional methods LIME [12], MF [5], NPE [40], SRIE [6] and deep learning-based methods DRD citewei2018deep, Kind [54], Kind++ [53], MIRNet [51], SGM [48], FECNet [14], HDMNet [25], SNR-Aware [45], Bread [11]. Note that all deep learning-based methods are trained on the same datasets with respective public codes.

**Quantitative comparison.** We adopt Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) [41], and Learned Perceptual Image Patch Similarity (LPIPS) [52] as the evaluation

<sup>1</sup><https://sites.google.com/site/vonikakis/datasets>



**Figure 7: Qualitative comparison in LSRW-Huawei [13] (the first row) and LSRW-Nikon [13] (the second row). It can be seen that the proposed method reaches the best visual results. Note that we compare the proposed method with three benchmarks with the best PSNR values on the corresponding datasets.**

**Table 3: NIQE scores on DICM, LIME, MEF, NPE, and VV datasets. The best results are boldfaced and the second-best ones are underlined. Note that “AVG” represents the average values of the NIQE score on five datasets. All methods are pre-trained on LSRW-Huawei [13].**

Methods	LIME	VV	DICM	NPE	MEF	AVG
KinD	4.772	3.835	3.614	4.175	4.819	4.194
MIRNet	6.453	4.735	4.042	5.235	5.504	5.101
SGM	5.451	4.884	4.733	5.208	5.754	5.279
FECNet	6.041	3.346	4.139	4.500	4.707	4.336
HDMNet	6.403	4.462	4.773	5.108	5.993	5.056
SNR-Aware	<u>4.618</u>	<b>2.207</b>	<b>3.227</b>	<u>3.975</u>	<u>4.589</u>	<b>3.887</b>
Bread	4.717	3.304	4.179	4.160	5.369	4.194
Ours	<b>4.402</b>	<u>3.168</u>	<u>3.374</u>	<b>3.909</b>	<b>4.362</b>	<u>3.907</u>

matrices. LPIPS measures the distance between two images in high-level features. Generally, the higher PSNR, higher SSIM, and lower LPIPS represent that two images are more similar.

As shown in Table 2, compared with recent methods, the proposed method reaches the best results in most of the cases, for the rest it almost achieves the second-best. Notably, compared with a recent Transformer based method SNR-Aware [45], the proposed method achieves generally superior performance with only 0.03 % parameters due to the efficient global representation of the Fourier frequency information. Compared with a recent Fourier-based method FECNet [14], the proposed method overall outperforms it with less parameters benefited from further use of the Fourier frequency information. Besides, our method also has competitive performance with UHDFour [24], which has 21.78dB PSNR and 0.87 SSIM on LOL-Real [48] but with 17.54M parameters.

Besides, we measure the naturalness image quality evaluator (NIQE) score on five unpaired datasets. The images with the lower NIQE scores represent the higher naturalness image quality. Table 3 presents the results of the NIQE evaluation. It can be seen that the proposed method outperforms most existing LLIE methods. Although SNR-Aware [45] reaches competitive performance in the NIQE evaluation, it has much more parameters than the proposed method (see Table 2).

**Table 4: Quantitative comparison on the SICE [3]. The best results are boldfaced and the second-best ones are underlined. Note that the results of other methods are collected from [14].**

Methods	Under		Over		Average	
	PNSR	SSIM	PNSR	SSIM	PNSR	SSIM
DRD [42]	12.94	0.5157	12.87	0.5252	12.90	0.5212
DPED [15]	16.83	0.6133	7.99	0.4300	12.41	0.5217
DRBN [47]	17.96	<u>0.6767</u>	17.33	0.6828	17.65	0.6798
SID [4]	19.51	0.6635	16.76	0.6444	18.15	0.6540
MSEC [1]	19.62	0.6512	17.59	0.6560	18.58	0.6536
CMEC [30]	17.68	0.6592	18.17	0.6811	17.93	0.6702
FECNet [14]	<b>22.01</b>	0.6737	<u>19.91</u>	<u>0.6960</u>	<b>20.96</b>	<u>0.6849</u>
Ours	<u>21.36</u>	<b>0.6793</b>	<b>20.14</b>	<b>0.7143</b>	<u>20.75</u>	<b>0.6968</b>

**Qualitative comparison.** We first present the qualitative comparison of LOL-Real and LOL-Synthetic in Fig. 6, where the first row is the results of LOL-Real and the second row is the results of LOL-Synthetic. We select three methods with the best PSNR values. It can be seen that the proposed method reaches the best visual results both in detail and lightness. Even for some complex textures, the proposed method can recover well (see the first row of Fig. 6).

Fig. 7 presents the comparison results on LSRW-Huawei (first row) and LSRW-Nikon (second row). It can be seen that the results of the proposed method have fewer noise and clearer details than other methods.

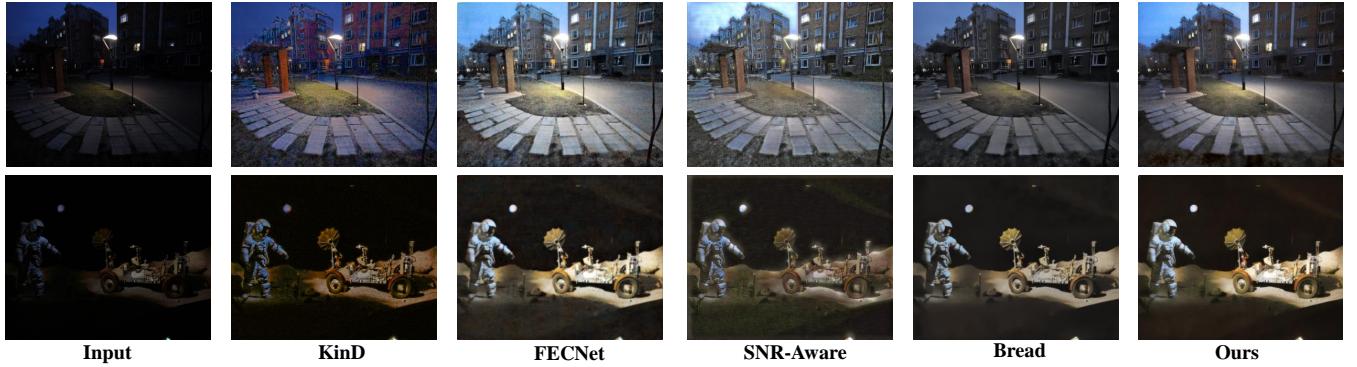
Besides, we also present the visual comparison on unpaired datasets. Fig. 8 presents the results of DICM (first row) and LIME (second row). We can see that the proposed method has best visual results, which demonstrates the decent generalization of the proposed method. The rest visual results can be found in supplementary material.

### 4.3 Extension on Exposure Correction.

In this work, we find the lightness of low-light images can be improved by enlarging the magnitude of the amplitude component in the Fourier space. Alternatively, the lightness also can be depressed by reducing the magnitude of amplitude. Therefore, we extend FourLLIE to the exposure correction task and evaluate it in the

**Table 5: Ablation study. The best results are boldfaced and the second-best ones are underlined.**

Methods	LOL-Real [48]			LOL-Synthetic [48]			LSRW-Huawei [13]			LSRW-Nikon [13]			#Param (M)
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	
Ours w/o F	21.01	<u>0.8398</u>	<u>0.0571</u>	<u>24.14</u>	<b>0.9196</b>	<b>0.0297</b>	20.61	0.5986	0.1928	17.24	0.4866	0.1898	<u>0.09</u>
Ours w/o S	20.20	0.7128	0.1199	17.70	0.7156	0.1423	17.25	0.5358	0.1976	15.17	0.4328	0.1714	<b>0.03</b>
Ours w/o SNR	20.52	0.8291	0.0638	22.39	0.9009	0.1470	20.61	0.6020	0.2012	17.19	0.4889	<b>0.1269</b>	0.12
Ours w/o $\mathcal{L}_{s1}$	<u>21.29</u>	0.8282	0.0634	22.94	0.9036	0.0443	<u>20.96</u>	0.6050	0.2303	16.93	0.4761	<u>0.1464</u>	0.12
Ours w/o $\mathcal{L}_{vgg}$	19.61	0.7874	0.1050	21.54	0.8795	0.0501	20.70	<u>0.6151</u>	<u>0.1745</u>	<u>17.27</u>	<u>0.5020</u>	0.2152	0.12
Ours	<b>22.34</b>	<b>0.8468</b>	<b>0.0511</b>	<b>24.65</b>	0.9192	<u>0.0389</u>	<b>21.30</b>	<b>0.6220</b>	<b>0.1719</b>	<b>17.82</b>	<b>0.5036</b>	0.2150	0.12



**Figure 8: Qualitative comparison in unpaired datasets (the first row is DICM [22] and the second row is LIME [12]). It can be seen that the proposed method reaches the best visual results. Note that we compare the proposed method with four benchmarks with the best average NIQE values in Table 3.**

SICE [3] dataset, which contains 512 over-/normal- exposure image pairs and 512 under-/normal- exposure pairs for training, and 60 image pairs for testing. Note that the frequency stage should estimate an additional amplitude transform map for lightness depression (details can be seen in supplementary materials). The results in Table 4 show that the proposed method outperforms existing methods in PSNR and SSIM. It demonstrates the potential of the proposed method in exposure correction.

#### 4.4 Ablation Study

FourLLIE utilizes the Fourier frequency information to enhance the low-light images through two stages: frequency stage and spatial stage. In this subsection, we conduct the ablation study with five different settings to verify the effectiveness of the proposed designs and adopted loss functions. 1) “Ours w/o F” removes the frequency stage. 2) “Ours w/o S” removes the spatial stage. 3) “Ours w/o SNR” removes SNR-based frequency and spatial interaction and replaces it with the feature concatenation. 4) “Ours w/o  $\mathcal{L}_{s1}$ ” removes the  $\mathcal{L}_{s1}$ . 5) “Ours w/o  $\mathcal{L}_{vgg}$ ” removes the perceptual loss  $\mathcal{L}_{vgg}$  of  $\mathcal{L}_{s2}$ . Table 5 presents the results on both four datasets. Compared with all ablation settings, our full setting almost reaches the best PNSR, SSIM, and LPIPS. It demonstrates the effectiveness of the proposed designs and adopted loss functions.

## 5 CONCLUSION

Inspired by previous Fourier-based LLIE methods, in this work, we further explore the properties of the Fourier frequency information and propose a new Fourier-based LLIE method. We first analyze the relationship between the amplitude component and

lightness and conclude that lightness can be improved by enlarging the magnitude of the amplitude component. Then, we find the Fourier frequency information has nice global properties and is efficient to extract, since it does not introduce massive parameters of neural networks. Based on the above observations, we design a two-stage architecture FourLLIE, which first estimates the amplitude transform map to improve the lightness in the frequency stage. Then, the SNR map is introduced to accomplish frequency and spatial interaction in the spatial stage and to recover the detail. Benefiting from the effectiveness of Fourier frequency information, FourLLIE outperforms existing SOTA LLIE methods with a lightweight architecture.

Recently, many related works have demonstrated the enormous potential of the Fourier frequency information in low-/high- level computer vision tasks. In the future, we will explore more properties of the Fourier frequency information to make the proposed method adaptive for more diverse degradation.

## ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant No. 62071500. Supported by Sino-German Mobility Programme M-0421.

## REFERENCES

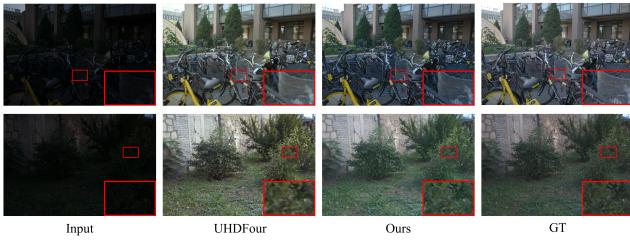
- [1] Mahmoud Afifi, Konstantinos G. Derpanis, Bjorn Ommer, and Michael S. Brown. 2021. Learning Multi-Scale Photo Exposure Correction. In *IEEE Conf. Comput. Vis. Pattern Recog.* 9157–9167.
- [2] E Oran Brigham and RE Morrow. 1967. The fast Fourier transform. *IEEE spectrum* 4, 12 (1967), 63–70.
- [3] Jianrui Cai, Shuhang Gu, and Lei Zhang. 2018. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Trans. Image Process.* 27, 4

- (2018), 2049–2062.
- [4] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. 2018. Learning to see in the dark. In *IEEE Conf. Comput. Vis. Pattern Recog.* 3291–3300.
  - [5] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. 2016. A fusion-based enhancing method for weakly illuminated images. *Signal Processing* 129 (2016), 82–96.
  - [6] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. 2016. A weighted variational model for simultaneous reflectance and illumination estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.* 2782–2790.
  - [7] Dario Fuoli, Luc Van Gool, and Radu Timofte. 2021. Fourier space losses for efficient perceptual image super-resolution. In *IEEE Conf. Comput. Vis. Pattern Recog.* 2360–2369.
  - [8] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédé Durand. 2017. Deep bilateral learning for real-time image enhancement. *ACM Trans. Graph.* 36, 4 (2017), 1–12.
  - [9] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *IEEE Conf. Comput. Vis. Pattern Recog.* 1780–1789.
  - [10] Xin Guo, Xueyang Fu, Man Zhou, Zhen Huang, Jialun Peng, and Zheng-Jun Zha. 2022. Exploring fourier prior for single image rain removal. In *IJCAI*. 935–941.
  - [11] Xiaojie Guo and Qiming Hu. 2023. Low-light image enhancement via breaking down the darkness. *Int. J. Comput. Vis.* 131, 1 (2023), 48–66.
  - [12] Xiaojie Guo, Yu Li, and Haibin Ling. 2016. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* 26, 2 (2016), 982–993.
  - [13] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. 2023. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation* 90 (2023), 103712.
  - [14] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. 2022. Deep Fourier-Based Exposure Correction Network with Spatial-Frequency Interaction. In *Eur. Conf. Comput. Vis.* Springer, 163–180.
  - [15] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhooye, and Luc Van Gool. 2017. Dslr-quality photos on mobile devices with deep convolutional networks. In *Int. Conf. Comput. Vis.* 3277–3285.
  - [16] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. 2021. Focal frequency loss for image reconstruction and synthesis. In *Int. Conf. Comput. Vis.* 13919–13929.
  - [17] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. 2021. Enlightengan: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.* 30 (2021), 2340–2349.
  - [18] Zhi Jin, Muhammad Zafar Iqbal, Dmytro Bobkov, Wenbin Zou, Xia Li, and Eckehard Steinbach. 2019. A flexible deep CNN framework for image restoration. *IEEE Trans. Multimedia* 22, 4 (2019), 1055–1068.
  - [19] Zhi Jin, Muhammad Zafar Iqbal, Wenbin Zou, Xia Li, and Eckehard Steinbach. 2020. Dual-stream multi-path recursive residual network for JPEG image compression artifacts reduction. *IEEE Trans. Circuit Syst. Video Technol.* 31, 2 (2020), 467–479.
  - [20] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. 1997. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* 6, 7 (1997), 965–976.
  - [21] Diederik P Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *Int. Conf. Learn. Represent.*
  - [22] Chulwoo Lee, Chul Lee, and Chang-Su Kim. 2012. Contrast enhancement based on layered difference representation. In *IEEE Int. Conf. Image Process.* IEEE.
  - [23] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. 2021. Low-light image and video enhancement using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 12 (2021), 9396–9416.
  - [24] Chongyi Li, Chun-Le Guo, Man Zhou, Zhenxin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. 2023. Embedding Fourier for Ultra-High-Definition Low-Light Image Enhancement. *arXiv preprint arXiv:2302.11831* (2023).
  - [25] Yudong Liang, Bin Wang, Wenqi Ren, Jiaying Liu, Wenjian Wang, and Wangmeng Zuo. 2022. Learning hierarchical dynamics with spatial adjacency for image enhancement. In *ACM Int. Conf. Multimedia*. 2767–2776.
  - [26] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. 2021. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *IEEE Conf. Comput. Vis. Pattern Recog.* 10561–10570.
  - [27] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. 2017. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition* 61 (2017), 650–662.
  - [28] Kede Ma, Kai Zeng, and Zhou Wang. 2015. Perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. Image Process.* 24, 11 (2015), 3345–3356.
  - [29] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. 2022. Toward fast, flexible, and robust low-light image enhancement. In *IEEE Conf. Comput. Vis. Pattern Recog.* 5637–5646.
  - [30] NE Nsamp, Zhongyun Hu, and Qing Wang. 2018. Learning exposure correction via consistency modeling. In *Brit. Mach. Vis. Conf.* 1–12.
  - [31] Jinwang Pan, Deming Zhai, Yuanchao Bai, Junjun Jiang, Debin Zhao, and Xianming Liu. 2022. ChebyLighter: Optimal Curve Estimation for Low-light Image Enhancement. In *ACM Int. Conf. Multimedia*. 1358–1366.
  - [32] Qingzhe Pan, Zhipu Zhao, Xuemei Xie, Jianan Li, Yuhua Cao, and Guangming Shi. 2021. View-normalized skeleton generation for action recognition. In *ACM Int. Conf. Multimedia*. 1875–1883.
  - [33] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. 1987. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing* 39, 3 (1987), 355–368.
  - [34] Zia-ur Rahman, Daniel J Jobson, and Glenn A Woodell. 2004. Retinex processing for automatic image enhancement. *Journal of Electronic Imaging* 13, 1 (2004), 100–110.
  - [35] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inform. Process. Syst.* 28 (2015).
  - [36] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
  - [37] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008).
  - [38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Adv. Neural Inform. Process. Syst.* 30 (2017).
  - [39] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. 2019. Underexposed photo enhancement using deep illumination estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.* 6849–6857.
  - [40] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.* 22, 9 (2013), 3538–3548.
  - [41] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 4 (2004), 600–612.
  - [42] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2018. Deep retinex decomposition for low-light enhancement. In *Brit. Mach. Vis. Conf.*
  - [43] Hongjun Wu, Haoran Qi, Jingzhou Luo, Yining Li, and Zhi Jin. 2022. A Light-weight Image Entropy-Based Divide-and-Conquer Network for Low-Light Image Enhancement. In *Int. Conf. Multimedia and Expo.* IEEE, 01–06.
  - [44] Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. 2021. A fourier-based framework for domain generalization. In *IEEE Conf. Comput. Vis. Pattern Recog.* 14383–14392.
  - [45] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. 2022. SNR-Aware Low-Light Image Enhancement. In *IEEE Conf. Comput. Vis. Pattern Recog.* 17714–17724.
  - [46] Bowen Yang, Chun Yang, Qi Liu, and Xu-Cheng Yin. 2019. Joint rotation-invariance face detection and alignment with angle-sensitivity cascaded networks. In *ACM Int. Conf. Multimedia*. 1473–1480.
  - [47] Wenhan Yang, Shiqi Wang, Yuning Fang, Yue Wang, and Jiaying Liu. 2020. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *IEEE Conf. Comput. Vis. Pattern Recog.* 3063–3072.
  - [48] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. 2021. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Trans. Image Process.* 30 (2021), 2072–2086.
  - [49] Yanchao Yang, Dong Lao, Ganesh Sundaramoorthy, and Stefano Soatto. 2020. Phase Consistent Ecological Domain Adaptation. In *IEEE Conf. Comput. Vis. Pattern Recog.*
  - [50] Yu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. 2022. Frequency and spatial dual guidance for image dehazing. In *Eur. Conf. Comput. Vis.* Springer, 181–198.
  - [51] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2020. Learning enriched features for real image restoration and enhancement. In *Eur. Conf. Comput. Vis.* Springer, 492–511.
  - [52] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conf. Comput. Vis. Pattern Recog.* 586–595.
  - [53] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. 2021. Beyond brightening low-light images. *Int. J. Comput. Vis.* 129, 4 (2021), 1013–1037.
  - [54] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. 2019. Kindling the darkness: A practical low-light image enhancer. In *ACM Int. Conf. Multimedia*. 1632–1640.
  - [55] Naishan Zheng, Jie Huang, Qi Zhu, Man Zhou, Feng Zhao, and Zheng-Jun Zha. 2022. Enhancement by Your Aesthetic: An Intelligible Unsupervised Personalized Enhancer for Low-Light Images. In *ACM Int. Conf. Multimedia*. 6521–6529.
  - [56] Man Zhou, Jie Huang, Chongyi Li, Hu Yu, Keyu Yan, Naishan Zheng, and Feng Zhao. 2022. Adaptively Learning Low-high Frequency Information Integration for Pan-sharpening. In *ACM Int. Conf. Multimedia*. 3375–3384.
  - [57] Man Zhou, Jie Huang, Keyu Yan, Hu Yu, Xueyang Fu, Aiping Liu, Xian Wei, and Feng Zhao. 2022. Spatial-frequency domain information integration for pan-sharpening. In *Eur. Conf. Comput. Vis.* Springer, 274–291.

- [58] Man Zhou, Hu Yu, Jie Huang, Feng Zhao, Jinwei Gu, Chen Change Loy, Deyu Meng, and Chongyi Li. 2022. Deep Fourier Up-Sampling. *Adv. Neural Inform. Process. Syst.* 35 (2022), 22995–23008.

**Table 6: Quantitative comparison with UHDFour [24]. Note that UHDFour (8×) represents UHDFour with 8× down-sampling in the low-resolution network (LRNet) and is suitable for high-resolution images, while UHDFour (2×) represents UHDFour with 2× down-sampling in LRNet and is suitable for low-resolution images.**

Methods	LOL-Real [48]		LOL-Synthetic [48]		LSRW-Huawei [13]		LSRW-Nikon [13]		#Param (M)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
UHDFour(8×)	20.91	0.76	22.70	0.87	20.64	0.58	17.72	0.48	17.54
UHDFour(2×)	<u>21.78</u>	<u>0.87</u>	<u>23.74</u>	<u>0.90</u>	<u>20.88</u>	<u>0.61</u>	<u>17.57</u>	<u>0.51</u>	17.54
Ours	22.34	0.85	24.65	0.92	21.30	0.62	17.82	0.50	0.12



**Figure 9: Visual comparison with UHDFour [24].**

## A DETAILED EXPERIMENT SETTINGS IN SECTION 3.2

In Section 3.2, we conduct an experiment to demonstrate the effectiveness of estimating the amplitude transform map. In this part, we present the detailed implementations of this experiment. As shown in Fig. 10, setting 1 predicts the amplitude component directly (we add a residual connection for faster convergence), this process can be expressed by:

$$\begin{aligned}
 A_{res} &= NN(input) \\
 A_{in} &= \mathcal{A}(\mathcal{F}(input)), \quad P_{in} = \mathcal{P}(\mathcal{F}(input)) \\
 A_{out} &= A_{in} + A_{res} \\
 R_{out} &= A_{out} \times \cos(P_{in}) \quad I_{out} = A_{out} \times \sin(P_{in}) \\
 Output\_1 &= \mathcal{F}^{-1}(R_{out} + I_{out}j)
 \end{aligned} \tag{12}$$

where  $NN$  represents the neural networks in Fig. 10. Then, the constraint of setting 1  $\mathcal{L}_{setting1}$  is expressed as:

$$\mathcal{L}_{setting1} = \|\mathcal{A}(\mathcal{F}(Output\_2)) - \mathcal{A}(\mathcal{F}(GT))\|_2 \tag{13}$$

where GT is the ground truth.

Setting 2 predicts the enhanced image directly (residual connection is also adopted) and employs constrain on the amplitude component of the predicted enhanced image. This process can be expressed by:

$$Output\_2 = NN(input) \tag{14}$$

then, the constraint of setting 1  $\mathcal{L}_{setting1}$  is expressed as:

$$\mathcal{L}_{setting2} = \|\mathcal{A}(\mathcal{F}(Output\_2)) - \mathcal{A}(\mathcal{F}(GT))\|_2 \tag{15}$$

Setting 3 estimates the amplitude transform map based on the observation that the magnitudes of the amplitude component can reflect the magnitudes of the lightness. This process can be expressed by:

$$\begin{aligned}
 M &= sigmoid(NN(input)) \\
 A_{in} &= \mathcal{A}(\mathcal{F}(input)), \quad P_{in} = \mathcal{P}(\mathcal{F}(input)) \\
 A_{out} &= A_{in} / (M + \epsilon) \\
 R_{out} &= A_{out} \times \cos(P_{in}) \quad I_{out} = A_{out} \times \sin(P_{in}) \\
 Output\_3 &= \mathcal{F}^{-1}(R_{out} + I_{out}j)
 \end{aligned} \tag{16}$$

where  $M$  is the amplitude transform map,  $\epsilon$  is set to  $1e-8$  for avoiding zero-division. Then, the constraint of setting 3  $\mathcal{L}_{setting3}$  is expressed as:

$$\mathcal{L}_{setting3} = \|\mathcal{A}(\mathcal{F}(Output\_3)) - \mathcal{A}(\mathcal{F}(GT))\|_2 \tag{17}$$

## B VARIANT FOR EXPOSURE CORRECTION

To accomplish over-exposure depression, FourLLIE needs to be able to reduce the magnitudes of amplitude components. As shown in Fig. 11, we predict two amplitude transform maps, where one is responsible for improving the lightness and another is responsible for depressing the lightness, to accomplish exposure correction.

## C COMPARISON WITH UHDFOUR

We compare the proposed method with a recent Fourier-based LLIE method UHDFour [24]. As shown in Table 6 and Fig. 9, the proposed method reaches overall better performance.

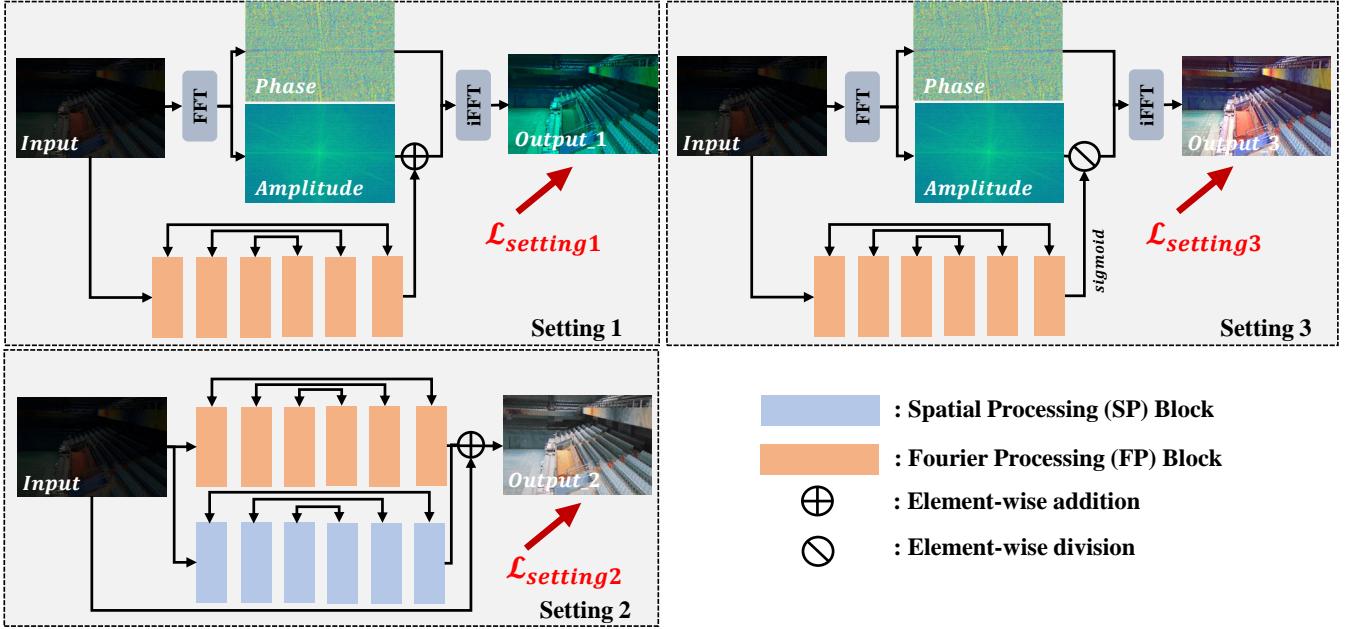


Figure 10: Detailed experiment settings in Section 3.2. Note that considering the input and output of setting 2 are in spatial space, we add a spatial branch for better convergence.

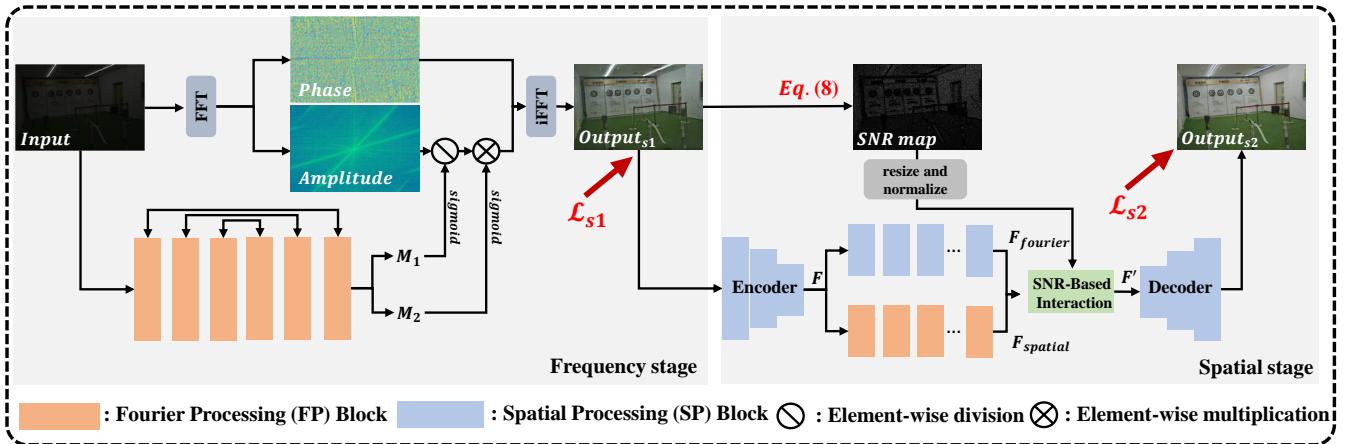


Figure 11: The variant of FourLLIE for exposure correction. It estimates an additional amplitude transform map to depress the magnitudes of amplitude to accomplish the lightness depression.