REVITALIZING CHANNEL-DIMENSION FOURIER TRANSFORM FOR IMAGE ENHANCEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Exploring the global representations of Fourier transform for image enhancement has become an alternative and made significant advancements. However, previous works only operate in the spatial dimension, overlooking the potential of the channel dimension that inherently possesses discriminative features. In this work, we propose a fresh perspective, channel-dimension Fourier transform, for image enhancement. Our designs are simple yet effective and comprise three straightforward steps: applying the Fourier transform to the channel dimension to obtain channel-wise Fourier domain features, performing a channel-wise transformation on both its amplitude and phase components, and then reverting back to the spatial domain. Following the above rules, we offer three alternative implementation formats of the channel transform in different operational spaces, performing operations *1) in the global vector with higher orders*; *2) in the global vector with channel groups*; and *3) in the Fourier features derived from spatial-based Fourier transform*. The above core designs, as generic operators, can be seamlessly integrated with enhancement networks, achieving remarkable gains and building efficient models. Through extensive experiments on multiple image enhancement tasks, like low-light image enhancement, exposure correction, SDR2HDR translation, and underwater image enhancement, our designs exhibit consistent performance gains. The code will be publicly available.

1 INTRODUCTION

Image enhancement aims to recover the latent high-visibility image from its given low-visibility version captured under unfavorable lightness conditions such as underexposure. This endeavor greatly promote the application of computer vision techniques. Therefore, quantities of image enhancement methods have been developed to explore effective clues for adjusting components related to brightness. Representative image enhancement tasks encompass low-light image enhancement, exposure correction, SDR2HDR translation, and underwater image enhancement, among others.

Deep-learning-based techniques have witnessed remarkable advancements in image enhancement and exhibited powerful capability in modeling lightness adjustment procedures (Guo et al., 2020b). A line of works customize degradation prior-aware paradigms to explicitly learn the lightness component, like curve-adjustment and Retinex theory based methods (Guo et al., 2020a; Wei et al., 2018). These studies typically divide the learning process into global and local components and may not fully capture the dependencies within the feature space. In addition, another line of research concentrates on roughly designing complex frameworks to implicitly learn lightness enhancement procedures (Xu et al., 2022b; Zheng et al., 2023). However, these approaches have not deeply explored the underlying mechanism of image enhancement or introduced dedicated operations for handling global components, thus constraining their ability to effectively learn lightness adjustment.

Recently, Fourier transform has demonstrated its effectiveness in global information modeling (Yu et al., 2022a; Zhou et al., 2022a; Huang et al., 2022a; Zhou et al., 2022b; 2023b). As a supplement to the aforementioned works, some endeavors have integrated Fourier transform-based operations into image enhancement architectures to adjust global components (zhou et al., 2022; Li et al., 2023; Zhou et al., 2023a). By employing Fourier transform on the spatial dimension, which yields global statistical information for each channel, this operation improves the distinguishability of various global representations and streamlines their learning process. Despite its effectiveness, we argue

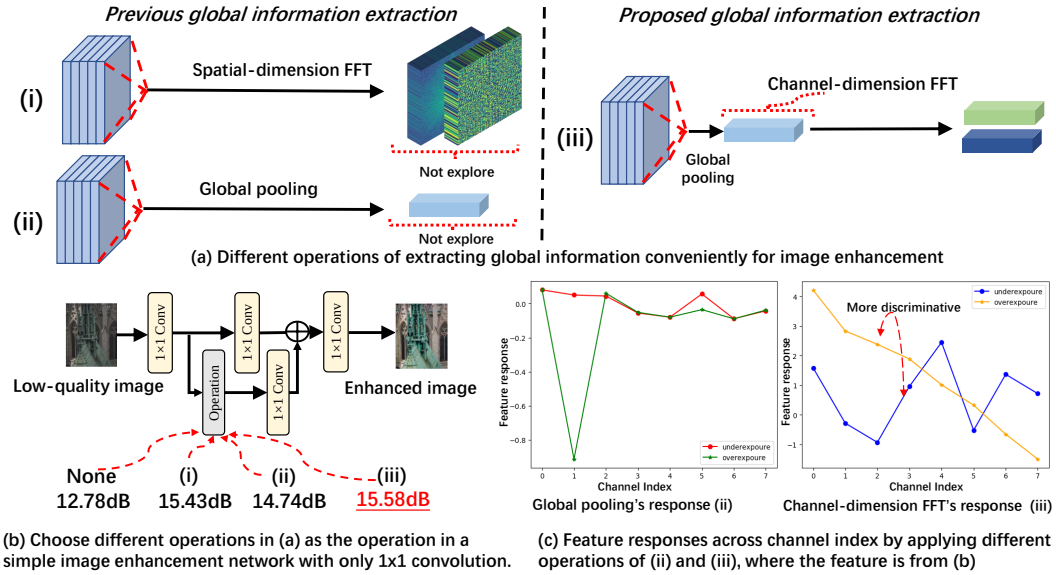


Figure 1: Motivation. (a) presents the different global representation formats for image enhancement involving global stylistic lightness correction. As shown in (b) which performs exposure correction using a simple toy network, when combined with our proposed channel-dimension FFT, it outperforms other formats presented in (a), making it a perfect fit for image enhancement tasks. (c) provides evidence that the channel feature response of channel-dimension FFT is more discriminative across different lighting conditions compared to the response curve of other global formats.

that the discriminability of global representation can also be addressed by modeling the channel distribution based on the style-represented Gram matrix in style transfer. Learning within this space using Fourier transform can enhance the discriminability of global representations and thus contribute to performance improvement, proofed in Appendix.

In this work, we propose a novel Channel-dimension Fourier transform learning (CFTL) mechanism for image enhancement tasks. The primary objective of CFTL is to capture global discriminative representations by modeling channel-dimension discrepancy in an efficient learning space. As illustrated in Figs. 1 and 3, we implement the CFTL with three straightforward steps: (1) Apply the Fourier transform to the global averaged vector in the channel dimension; (2) Derive the amplitude and phase components of the transformed features and perform the channel-wise operation on them; (3) Convert the processed feature to its original space through an inverse Fourier transform in the channel dimension and then add it to the spatially processed features. As depicted in Fig. 1, by employing Fourier transform over channel dimension, it has the strongest ability for fitting lightness adjustment, and the discriminability among different brightness representations is heightened, resulting in an efficient space for channel statistics adjustment.

Following the above rules, we provide several usage formats over different spaces in Fig. 3: (1) Performing operations in the global vector space with its different orders; (2) Performing operations on the global vector with its group of channels; (3) Performing operations on the Fourier features converted by spatial-based Fourier transform. Across various image enhancement tasks, our extensive experiments consistently demonstrate performance enhancements achieved by incorporating CFTL. Furthermore, CFTL also contributes to the development of a lightweight backbone, achieving an elegant balance between effectiveness and efficiency.

Our contributions are summarized as follows: **(1)** We provide a new mechanism for image enhancement through revitalizing Channel-dimension Fourier transform learning for the first time. This mechanism enhances the discriminability of global representation through channel statistics modeling, which further acts as a representative space that efficiently adjusts global information; **(2)** We provide several applicable formats of Channel-dimension Fourier transform learning (CFTL) mechanism, and conducted extensive experiments in various mainstream image enhancement tasks, highlighting its potential in wide-range applications; **(3)** Our proposed CFTL paradigm is compatible with existing image enhancement architectures, leading to performance improvement with negligible computation cost.

2 RELATED WORK

Image Enhancement. Image enhancement tasks aim to improve the quality of low-visibility images by adjusting the global lightness and contrast components (*i.e.,* illumination, color, and dynamic range). Different tasks in image enhancement play different roles in various scenarios. In the low-light image enhancement task, algorithms are tailored to enhance the visibility of images acquired in low-light conditions (Chen et al., 2018; Yang et al., 2020b; Wu et al., 2022; Zhang et al., 2019; 2021a; Guo et al., 2020a; Wei et al., 2018; Zhu et al., 2020; Wang et al., 2019b). For exposure correction, methods concentrate on rectifying images captured under both underexposure and overexposure scenes to normal exposure (Yang et al., 2020b; Afifi et al., 2021; Wang et al., 2022; Huang et al., 2022b). For SDR2HDR translation, this task aims to convert images from a low-dynamic range to a high-dynamic range (Chen et al., 2021a; Dong et al., 2021; Chen et al., 2021c; Kim et al., 2019; He et al., 2020b). While in underwater image enhancement, the contrast and color shift are required to be adjusted (Li et al., 2019b). Since recent approaches in these enhancement tasks are designed with deep neural networks, adjusting the lightness-related components would effectively strengthen the performance.

Fourier transform. Fourier transform aims to analyze the frequency content in signals, which transforms the signal to the domain with global statistical property and thus is employed to perform computer vision tasks. As a classic application, Fourier transform has been developed for domain generalization and adaptation with its global domain information modeling. For instance, Xu (Xu et al., 2021a) developed a Fourier-based data augmentation strategy that derives samples with wide-range styles for domain generation. Lee (Lee et al., 2023) proposed to improve the normalization for domain generalization by recomposing its different components in the Fourier domain. As another application, the Fourier transform mechanism is employed for designing effective backbones with its global information representativity. For example, FFC (Chi et al., 2020) is proposed to process partial features in the Fourier domain that enable the models to have a non-local receptive field. GFNet (Rao et al., 2021) employs FFT/IFFT to derive the Fourier domain features that serve as global filters for effective attention modeling. All the above works depict the representability of Fourier domain features in capturing global spatial statistics.

More recently, based on the above merits of the Fourier transform, this technique has also been introduced to low-level vision tasks as a design choice (Fuoli et al., 2021; Xintian Mao & Wang, 2023; Suvorov et al., 2022; Huang et al., 2022b; Yu et al., 2022b; Li et al., 2023). As an early attempt, Fuoli *et al* (Fuoli et al., 2021) proposed a Fourier transform-based loss to optimize the global high-frequency information of efficient image super-resolution. Later, DeepRFT (Xintian Mao & Wang, 2023) is proposed for image deblurring that captures both low-frequency and high-frequency properties of various blurs with a global-receptive field, and a similar design is also employed for effective image inpainting (Suvorov et al., 2022). FECNet (Huang et al., 2022b) and UHDFour (Li et al., 2023) point out that the amplitude of the Fourier feature decouples the global lightness components and thus is efficacy for image enhancement. Yu *et al* (Yu et al., 2022b) also observes a similar phenomenon in image dehazing that the amplitude reflects the global haze-related information and then designs the corresponding solution.

3 METHOD

In this section, we commence by revisiting the traditional 2D Fourier transform, followed by the introduction of our innovative channel-dimension Fourier transform. Subsequently, we delve into the details of the proposed CFTL. Finally, we illustrate the implementation of the CFTL variants.

3.1 PRELIMINARY OF FOURIER TRANSFORM

As recognized, the Fourier transform is widely used to analyze the frequency representation of images. **Typically**, this operation is independently conducted over the spatial dimension of each individual channel. Given an image $x \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$, the Fourier transform $\mathcal{F}(\cdot)$ converts it to Fourier space as the complex component $\mathcal{F}(x)$, which is expressed as:

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

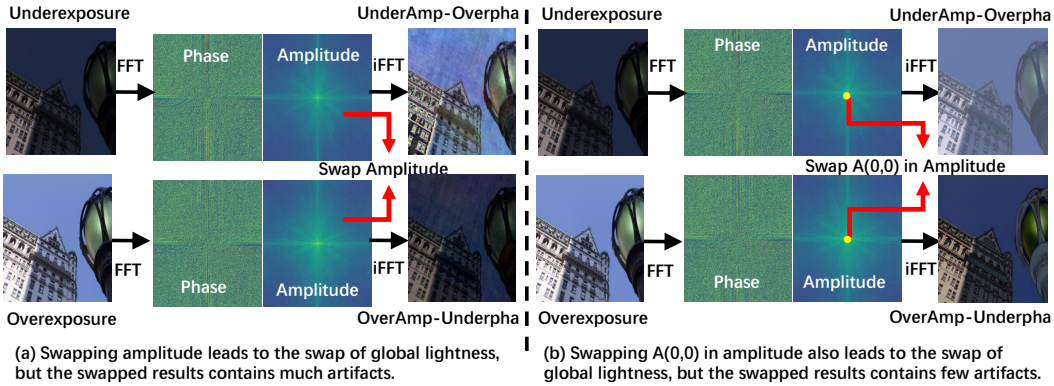


Figure 2: The illustration of swapping amplitude component in the Fourier domain. Both swapping the amplitude and $\mathcal{A}(x)(0,0)$ leads to lightness swapping, while the latter exhibits fewer artifacts.

$\mathcal{F}^{-1}(\cdot)$ defines the inverse Fourier transform accordingly. Both the Fourier transform and its inverse procedure can be efficiently implemented by FFT/IFFT algorithms (Frigo & Johnson, 1998). The amplitude component $\mathcal{A}(x)(u,v)$ and phase component $\mathcal{P}(x)(u,v)$ are expressed as:

$$\begin{aligned}\mathcal{A}(x)(u,v) &= \sqrt{R^2(x)(u,v) + I^2(x)(u,v)}, \\ \mathcal{P}(x)(u,v) &= \arctan[\frac{I(x)(u,v)}{R(x)(u,v)}],\end{aligned} \tag{2}$$

where $R(x)$ and $I(x)$ represent the real and imaginary part respectively. When referring to the amplitude component $\mathcal{A}(x)(u,v)$, it measures the magnitude of the frequency index (u,v) of an image and acts as a statistic indicator of frequency.

Targeting image enhancement, previous works have demonstrated that global information such as lightness is mainly preserved in the amplitude component (Li et al., 2023; Zhou et al., 2023a). However, we argue that the primary characteristic of global information remains conserved within $\mathcal{A}(x)(0,0)$, as shown in Fig. 2), which represents the low-frequency DC component within the original image, while using the whole amplitude may bring artifacts in the swapped result in Fig. 2) (we provide more discussion in the Appendix). With regard to Eq. 2, the globally averaged vector across the 2D dimension is equal to $\mathcal{A}(x)(0,0)$, and we utilize it as the subject of manipulation.

On the other hand, different channels exhibit different properties of spectral information, which also determine the global information of an image when conjunct different channels. A comparable inference can be drawn from style transfer studies, wherein the Gram matrix signifies global style information (Li et al., 2017). This impels us to employ the Fourier transform on the channel dimension to enrich the representation of global information, as detailed below.

3.2 Channel-dimension Fourier transform

We introduce the Channel-dimension Fourier transform by individually applying the Fourier transform along the channel dimension for each spatial position. For each position ($h \in \mathbb{R}^{H-1}, w \in \mathbb{R}^{W-1}$) within the feature tensor $x \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$, denoted as $x(h,w,0:C-1)$ and abbreviated as $y(0:C-1)$, the Fourier transform $\mathcal{F}(\cdot)$ converts it to Fourier space as the complex component $\mathcal{F}(y)$, which is expressed as:

$$\mathcal{F}(y(0:C-1))(z) = \frac{1}{C}\sum_{c=0}^{\mathrm{C}-1} y(c) e^{-j2\pi \frac{c}{\mathrm{C}} z}, \tag{3}$$

Correspondingly, the amplitude component $\mathcal{A}(y(0:C-1))(z)$ and phase component $\mathcal{P}(y(0:C-1))(z)$ of $\mathcal{F}(y(0:C-1))(z)$ are expressed as:

$$\begin{aligned}\mathcal{A}(y(0:C-1))(z) &= \sqrt{R^2(y(0:C-1))(z) + I^2(y(0:C-1))(z)}, \\ \mathcal{P}(y(0:C-1))(z) &= \arctan[\frac{I(y(0:C-1))(z)}{R(y(0:C-1))(z)}].\end{aligned} \tag{4}$$

These operations can also be applied for the global vector $x_g \in \mathbf{R}^{1\times1\times\mathrm{C}}$. In this way, $\mathcal{A}(y)(z)$ and $\mathcal{P}(y)(z)$ signify the magnitude and directional changes in magnitude of various channel frequencies, respectively. Both of these metrics encapsulate global statistics related to channel information.

We provide visualizations to validate the properties of this operation in Fig. 15. It is evident that distinct representations of lightness become more discernible following the channel-based Fourier transform, both in terms of $\mathcal{A}(y)(z)$ and $\mathcal{P}(y)(z)$. This indicates that this operation improves the distinguishability of global information components, with channel statistics adjustments conveniently influencing its properties. Therefore, it is believed that the transformed feature can act as a representative space for global information adaptation.

3.3 Channel-based Fourier transform learning

Based on the above analysis, we introduce the CFTL implementation, shown in Fig. 3, which conducts operations on channel-based Fourier transformed features.

Operation Description. The core construction of the CFTL involves a three-step sequential process: firstly converts the features to a global information representative space, then applies operations on the transformed features, and finally, converts the features back to the original space.

Given the feature $x \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$, the initial step involves transforming it into a global vector $x_g \in \mathbb{R}^{1\times1\times\mathrm{C}}$ using global average pooling as

$$x_g = \frac{1}{\mathrm{HW}} \sum_{h=0}^{\mathrm{H}-1} \sum_{w=0}^{\mathrm{W}-1} x(h, w). \tag{5}$$

Here, x_g equals $\mathcal{A}(x)(0, 0)$ as described above, effectively encapsulating global information. Then, x_g is transformed into channel-dimension Fourier domain using Eq. 3, denoted as $\mathcal{F}(x_g)(z)$.

Secondly, we utilize Eq. 3 to transform the feature $\mathcal{F}(x_g)(z)$ into its amplitude component $\mathcal{A}(x_g)(z)$ and phase component $\mathcal{P}(x_g)(z)$. Instead of acting upon $\mathcal{F}(x_g)(z)$, we advocate performing operations on $\mathcal{A}(x_g)(z)$ and $\mathcal{P}(x_g)(z)$ due to their explicit information significance, elaborated in Sec. 3.1. Conversely, $\mathcal{F}(x_g)(z)$ lacks the requisite discriminative properties, as detailed in the Appendix. Here, we introduce attention-based operations on $\mathcal{A}(x_g)(z)$ and $\mathcal{P}(x_g)(z)$ as

$$\begin{aligned} \mathcal{A}(x_g)(z)' &= \texttt{Seq1}(F_g) \odot \mathcal{A}(F_g)(z), \\ \mathcal{P}(x_g)(z)' &= \texttt{Seq2}(F_g) \odot \mathcal{P}(F_g)(z), \end{aligned} \tag{6}$$

where $\texttt{Seq1}(\cdot)$ and $\texttt{Seq2}(\cdot)$ denote sequences of 1×1 convolutions followed by LeakyReLU activation, as illustrated in Fig. 3. The symbol $\odot$ signifies element-wise multiplication for attentive adjustment. Hence, Eq. 6 signifies the process of modifying the global information encapsulated within the channel statistics $\mathcal{A}(F_g)(z)$ and $\mathcal{P}(F_g)(z)$.

Finally, we convert the processed channel-dimension Fourier domain feature $\mathcal{A}(F_g)(z)'$ and $\mathcal{P}(F_g)(z)'$ to their original space by employing the inverted channel-based Fourier transform as

$$F_g' = \mathcal{F}^{-1}(\mathcal{A}(F_g)(z)', \mathcal{P}(F_g)(z)'), \tag{7}$$

where F_g' is the final processed feature of CFTL. Furthermore, we replicate it to the original resolution of $\mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$ to align with the size of x, as illustrated in Fig. 3.

Integrate CFTL into the backbone. Upon the processed feature F_g', we seamlessly integrate it with the processed original feature x, rendering CFTL compatible with existing backbone architectures. Given the resolution disparity between F_g' and x, we initially expand F_g' by repeating it by $\mathrm{H} \times \mathrm{W}$ times to match the size of x. Subsequently, we integrate F_g' with the processed x. As depicted in Fig. 3, within the CNN-based backbones, x undergoes a local information branch (i.e., convolutional layers) for local information processing, prioritizing the extraction of local representations.

3.4 Variants implementation of CFTL

Following the above rules, we offer three alternative implementation formats of the channel transform in different operational spaces of x_g.

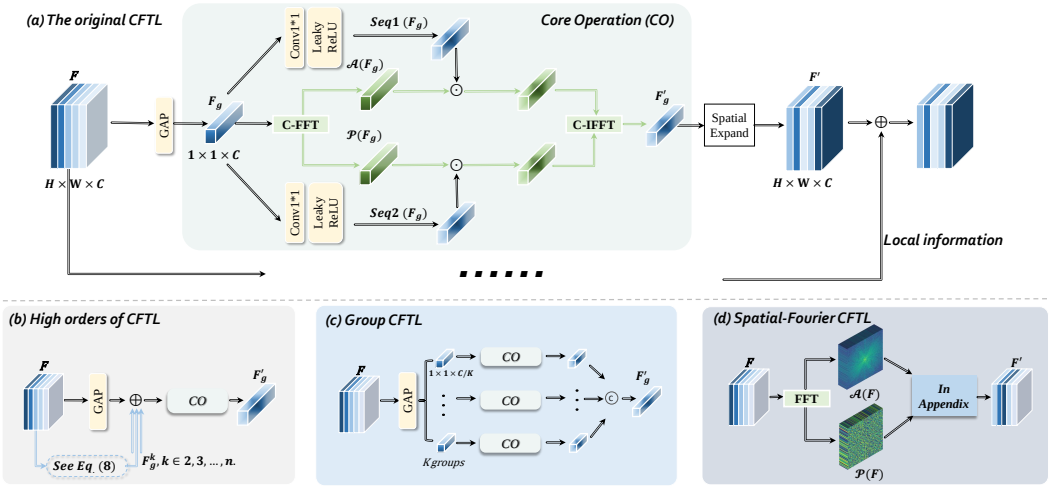


Figure 3: The illustration of the CFTL operation and other CFTL variants for image enhancement

The original CFTL. We construct the original CFTL by setting x_g as the global averaged vector of $x \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$ as shown in Fig. 3, which is plug-and-play for enhancement networks.

High orders of CFTL. Regarding the global average vector x_g in Eq. 5 as the first order of the global information from x, we can introduce more orders to strengthen the representation ability of x_g, which is illustrated in Fig. 3 (b). Specifically, for the k-th order of the global information, we denote it as x_g^k, which is expressed as:

$$x_g^k = \sqrt[k]{\frac{1}{\mathrm{HW}} \sum_{h\in[1,\mathrm{H}],\mathrm{w}\in[1,\mathrm{W}]} (x(h,w) - x_g)^k}, \quad k \in 2, 3, ..., \mathrm{n}. \tag{8}$$

This formula is derived from the color moment (), which is another representation of global information. For instance, when k equals 2, x_g^k equals the standard deviation of x, we set k as 2 defaultly in this format (as ablated in the Appendix). We further introduce x_g^k to x_g as its strengthed version:

$$x_g = x_g + x_g^2 + ...x_g^k, \qquad k \in 2, 3, ..., \mathrm{n}. \tag{9}$$

We follow other operations in CFTL to process x_g and thus regard it as the high orders of CFTL.

Group CFTL. Since the original operations in CFTL are conducted on all channels of $x_g \in \mathbb{R}^{1\times1\times\mathrm{C}}$, we can further process different groups of channel statistics in x_g, which is similar to the group normalization (Wu & He, 2018). Specifically, we divide x_g into K groups along the channel dimension, which is expressed as:

$$x_g = [x_g(1), ..., x_g(k)|k \in 1, ..., \mathrm{C/K}], \tag{10}$$

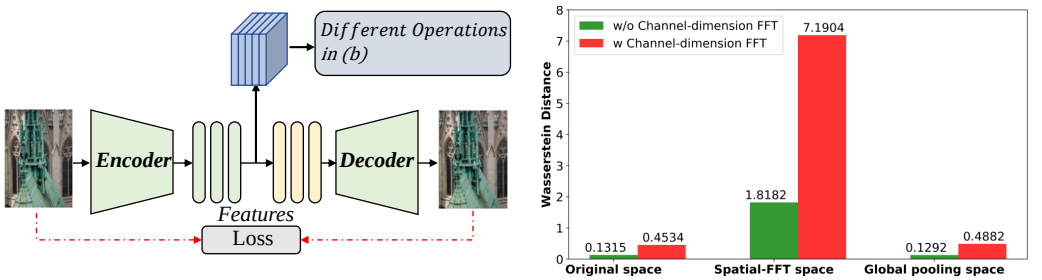
where $x_g(k)$ is the k-th channel group of x_g. As shown in Fig. 3 (c), in the following operations of the Group CFTL, different $x_g(k)$ are processed by the operations in the CFTL respectively with different weights in Eq. 10. Finally, these processed features $x_g(k)'$ are concatenated along the channel dimension, which are expressed as:

$$x_g' = \mathtt{cat}[x_g(1)', ..., x_g(k)'], \tag{11}$$

where $\mathtt{cat}[\cdot]$ denotes the concatenate operation. In this way, the parameters of the Group CFTL are less than the original CFTL as depicted in the Appendix, while keeping competitive performance.

Spatial-Fourier CFTL. All the above variants are based on the global vector x_g, which is a special case of $A(x)(u,v)$. Here, we propose to extend x_g as the amplitude component $A(x)(u,v)$ converted by the spatial Fourier transform in Eq. 4, and we denote it as Spatial-Fourier CFTL. Therefore, $x_g \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$ has the same shape as x, which also exhibits global representation.

As shown in Fig. 3, to process x_g, we follow the rules of the CFTL to process it and obtain the result F_g' (due to unstable training, we discard the few operations, which are detailed in the Appendix



(a) The illustration of self-reconstruction (b) Distribution distance of under-/over-exposure features from (a)

Figure 4: Toy experiment. (a) depicts the self-reconstruction, it extracts the feature that is processed by different operations in Fig. 1. Evidenced by the feature similarity in (b), our proposed channel-dimension FFT enhances the discriminability between under- and over-exposure in different spaces.

shown in Fig. 10). Finally, F_g' and $\mathcal{P}(x)(u,v)$ are converted back to the space of x through the inverted spatially-based Fourier transform as follows:

$$F' = \mathcal{F}^{-1}(F_g', \mathcal{P}(x)(u,v)), \tag{12}$$

where F' is the final result that is integrated into the processed x.

CFTL-Net. We introduce the CFTL-Net, an efficient network architecture depicted in Fig. 11, which combines High-order CFTL and Spatial-Fourier CFTL. This framework adopts an encoder-decoder-based architecture. Further details and discussions are available in the Appendix.

4 Experiment

To demonstrate the efficacy of our proposed CFTL paradigm, we conduct extensive experiments on multiple image enhancement tasks, More results can be found in the Appendix.

4.1 Toy experiments

To showcase the enhanced discriminability of global information through the channel-based Fourier transform, we conducted a toy experiment to convert the image to the feature space (Fig. 4): We constructed an encoder-decoder architecture for image reconstruction, trained on a dataset of 1000 samples sourced from the MIT-FiveK dataset (Bychkovsky et al., 2011). During testing, we used 100 underexposed and 100 overexposed samples from the SICE dataset as inputs, owing to their significant global information variations. We assessed their encoder-decoder features using various operations: spatial domain Fourier transform, global average pooling, and channel-based Fourier Transform on the global vector. Fig. 15 demonstrates that our proposed operation achieves the highest discriminability between underexposed and overexposed samples, as indicated by the maximum distribution distance. Additional results at the sample level are available in the Appendix.

Settings	#Param	Flops (G)	LOL	Huawei	FiveK
DRBN (Baseline) Yang et al. (2020b)	0.532M	39.71	20.73/0.7986	19.93/0.6810	22.11/0.8684
+Pooling attention	0.533M (+0.01M)	39.72 (+0.01)	21.84/0.8176	20.13/0.6838	23.15/0.8702
+Spaial Fourier	0.533M (+0.01M)	39.88 (+0.17)	22.07/0.8355	20.28/0.6844	23.72/0.8735
+Original CFTL	0.534M (+0.02M)	39.73 (+0.02)	**23.71/0.8492**	20.82/0.6933	**24.03**/0.8751
+Group CFTL	0.532M (+0M)	39.71 (+0)	22.98/0.8445	20.80/0.6952	23.93/**0.8768**
+High-order CFTL	0.534M (+0.02M)	39.73 (+0.02)	23.05/0.8457	20.81/0.6930	23.95/0.8755
+Spatial-Fourier CFTL	0.536M (+0.04)	40.27 (+0.56)	22.31/0.8376	**20.89/0.6954**	23.96/0.8755
Restormer (Baseline) (Zamir et al., 2022)	26.10M	563.96	20.49/0.7886	20.02/0.6663	23.13/0.8891
+Pooling attention	26.12M(+0.02M)	569.41(+5.45)	20.95/0.7952	20.34/0.6685	23.45/0.8915
+Spatial Fourier	26.12M(+0.02M)	569.41(+5.45)	21.01/0.8003	20.65/0.6713	23.66/0.8931
+Original CFTL	26.10M(+0M)	563.97(+0.001)	21.34/0.8020	20.70/0.6706	23.87/**0.8959**
+Group CFTL	26.10M(+0.01M)	563.97(+0.01)	**21.49**/0.8008	20.59/0.6715	**23.91**/0.8942
+High-order CFTL	26.10(+0M)	563.97(+0.01)	21.27/**0.8061**	**20.76/0.6733**	23.81/0.8944
+Spatial-Fourier CFTL	26.11(+0.01M)	565.19(+1.23)	21.37/0.8018	20.71/0.6713	23.82/0.8948
CFTL-Net	0.028M	3.64	22.50/0.8139	20.91/06941	24.03/0.8904

Table 1: Quantitative results of low-light image enhancement in terms of PSNR/MS-SSIM.

Settings	#Param	Flops (G)	MSEC	SICE
DRBN (Baseline)	0.532M	39.71	19.52/0.8309	17.88/0.6798
+Pooling attention	0.533M	39.72	22.89/0.8604	20.75/0.7095
+Spaial Fourier	0.533M	39.88	22.94/0.8642	20.94/0.7036
+Original CFTL	0.534M	39.73	**23.34/0.8683**	21.32/**0.7250**
+Group CFTL	0.532M	39.69	23.33/0.8672	21.30/0.7177
+High-order CFTL	0.534M	39.73	23.19/0.8667	**21.64**/ 0.7243
+Spatial-Fourier CFTL	0.536M	40.27	23.04/0.8645	21.33/0.7201
LCDPNet (Baseline)	0.961M	9.40	22.30/0.8552	20.46/0.6843
+Pooling attention	0.962M	9.41	22.41/0.8568	20.57/0.6835
+Spaial Fourier	0.962M	9.43	22.47/0.8561	20.94/0.6946
+Original CFTL	0.962M	9.41	22.68/0.8572	21.25/**0.7063**
+Group CFTL	0.961M	9.40	22.70/**0.8579**	20.96/0.6952
+High-order CFTL	0.962M	9.41	**22.74**/0.8565	**21.38**/0.7024
+Spatial-Fourier CFTL	0.967M	9.48	22.52/0.8563	20.58/0.6865
CFTL-Net	0.028M	3.64	22.88/0.8594	21.24/0.6999

Table 2: Results of exposure correction.

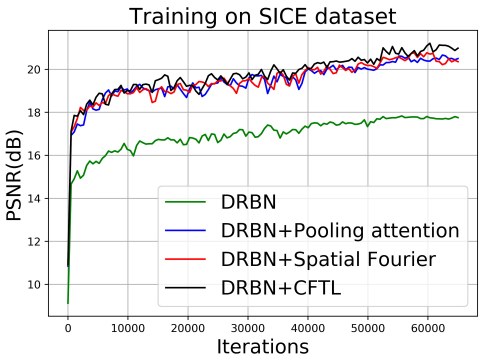


Figure 5: Training on exposure correction.

4.2 Experimental Settings

Low-light image enhancement. Following previous works (Hai et al., 2021; Zhao et al., 2021), we employ three widely used datasets for evaluation, including LOL dataset (Wei et al., 2018), Huawei dataset (Hai et al., 2021) and MIT-FiveK dataset (Bychkovsky et al., 2011). We employ two image enhancement networks, DRBN (Yang et al., 2020b) and Restormer (Zamir et al., 2022) as baselines.

Exposure correction. Following (Huang et al., 2022b), we adopt MSEC dataset (Afifi et al., 2021) and SICE dataset (Cai et al., 2018) for evaluations. The above two architectures, i.e., DRBN (Yang et al., 2020b) and LCDPNet (Wang et al., 2022) are regarded as baselines.

SDR2HDR translation. Following (Chen et al., 2021c), we choose the HDRTV dataset (Chen et al., 2021c) for evaluation. We employ CSRNet (He et al., 2020b) as the baseline in the experiments.

Underwater image enhancement. Following prior works (Li et al., 2019b), we select UIEC^2-Net (Wang et al., 2021) as baseline over UIEB (Li et al., 2019b) dataset for validation.

Comparison operators. We set two comparison operators corresponding to (i) and (ii) in the motivation, which are respectively based on the global pooling and spatial FFT. We name them "Pooling attention" and "Spatial Fourier", which are illustrated in Fig. 19 and the format is refer to CFTL.

4.3 Implementation Details

Since there exist three CFTL formats in Sec. 4.5, we respectively integrate them into the baseline to conduct experiments. For comparison, we perform the experiments of baseline networks, and the integration of the global average pooling and the spatial Fourier transform. Additionally, the CFTL-Net in Sec. 4.5 is also performed in experiments. We train all baselines and their integrated formats following the original settings, and our CFTL-Net until it converges. More implementation details are provided in the Appendix.

4.4 Comparison and Analysis

Quantitative Evaluation. We perform quantitative performance comparison on the mainstream image enhancement tasks in Table 1, Table 2, Table 3, and Table 4, where the best results are highlighted in bold. From the results, it can be observed that all formats of our proposed paradigm elevate the performance of the baselines across the datasets in all tasks, validating the effectiveness of our proposed method. In contrast, applying the pooling attention operation or the Spatial Fourier transform could not surpass our proposed CFTL in most datasets. Moreover, the proposed CFTL-Net also achieves effective performance with efficacy. We also demonstrate that the proposed CFTL helps improve the training performance as shown in Fig. 5. Note that all the above evaluation exhibits the convenience of applying our CFTL while introducing little computation cost.

Qualitative Evaluation. We present the visual results of exposure correction on the SICE dataset due to the limited space. As shown in Fig. 6, the integration of the CFTL leads to a more visually pleasing effect with less lightness and color shift problems compared with the original baseline. We provide more visual results in the supplementary material.

Settings	PSNR/SSIM	#Param/FLops(G)
CSRNet (Baseline)	35.34/0.9625	0.035M/1.58
+Pooling attention	35.67/0.9635	0.045M/1.62
+Spaial Fourier	35.59/0.9647	0.048M/1.63
+Original CFTL	35.93/**0.9712**	0.045M/1.62
+Group CFTL	35.82/0.9703	0.044M/1.62
+High-order CFTL	35.97/0.9694	0.045M/1.62
+Spatial-Fourier CFTL	35.74/0.9654	0.048M/1.62
CFTL-Net	**37.37**/0.9683	0.028M/3.64

Table 3: Results of SDR2HDR translation.

Settings	PSNR/SSIM	#Param/FLops(G)
UIEC^2-Net (Baseline)	21.39/0.8957	0.53M/104.25
+Pooling attention	21.46/0 .8996	0.58M/113.93
+Spaial Fourier	21.63/0.9012	0.58M/113.93
+Original CFTL	21.81/0.9005	0.54M/104.27
+Group CFTL	22.20/**0.9042**	0.54M/104.27
+High-order CFTL	**22.33**/0.9023	0.54M/104.27
+Spatial-Fourier CFTL	22.26/0.9030	0.55M/106.43
CFTL-Net	22.27/0.9034	0.028M/3.64

Table 4: Results of underwater image enhancement.

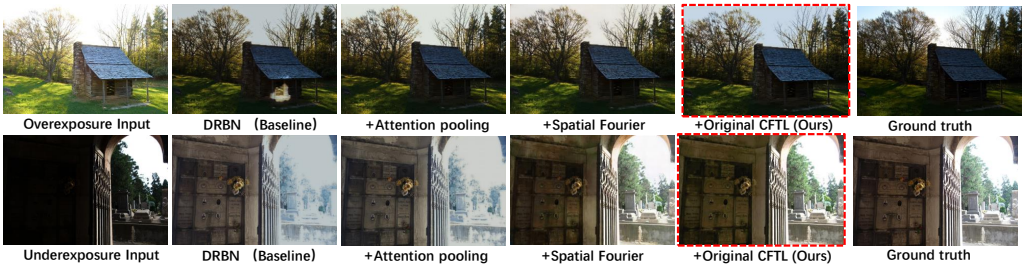


Figure 6: The visualization results on the SICE dataset for exposure correction.

Configurations	PSNR	SSIM
Baseline (DRBN)	17.88	0.6798
+CFTL	**21.32**	**0.7250**
+CFTL w/o global pooling	18.24	0.6853
+CFTL w/o channel-based ifft	21.05	0.7228
+CFTL w/o processing amplitude	21.21	0.7194
+CFTL w/o processing phase	20.94	0.7201

Table 5: Ablation studies for the configuration of CFTL on the SICE dataset.

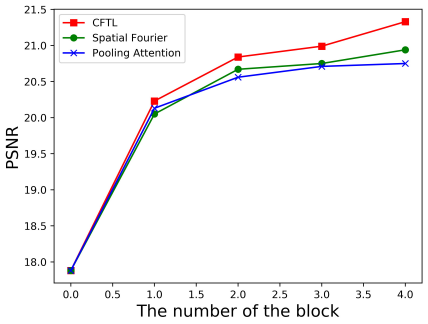


Figure 7: Ablations of investigating CFTL number on the SICE dataset.

4.5 Ablation Studies

We conduct ablation studies in the exposure correction task in terms of the CFTL design and the number of CFTL in DRBN baseline. More ablated studies can be found in the Appendix.

Investigate the design of CFTL. To explore the design of CFTL, we perform experiments by setting the CFTL with different configurations. The quantitative results are presented in Table 5. As can be seen, introducing the global pooling in the CFTL would lead to significant performance improvement. Meanwhile, converting the feature back to the original space with channel-dimension iFFT also works well. Note that conducting the operation on either the amplitude or phase components would lead to sub-optimal results. All results depict the reasonableness of our method design.

Impact of the CFTL number. We further investigate the influence of the CFTL number. The corresponding quantitative number K comparison from 1 to 4 is reported in Fig. 7. As can be seen, only incorporating one CFTL would lead to a significant performance improvement. By increasing the number of CFTL, the results are further improved apparently than other comparison operations. Overall, our CFTL exhibits well extensibility.

5 Conclusion

In this paper, we first propose a new channel-based Fourier transform learning mechanism for image enhancement. We propose the CFTL that enhances the discriminative ability of global information, meanwhile serving as a representative space for global information adjustment. The CFTL comprises several formats for implementation, which can be easily integrated into existing backbones by introducing few computation costs, Extensive experiments demonstrate the effectiveness and scalability of applying the CFTL and its variants in various image enhancement tasks.

REFERENCES

Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9157–9167, 2021.

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 97–104, 2011.

Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018.

Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. HDRUnet: Single image hdr reconstruction with denoising and dequantization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 354–363, 2021a.

Xiangyu Chen, Zhengwen Zhang, Jimmy S Ren, Lynhoo Tian, Yu Qiao, and Chao Dong. A new journey from SDRTV to HDRTV. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4500–4509, 2021b.

Xiangyu Chen, Zhengwen Zhang, Jimmy S Ren, Lynhoo Tian, Yu Qiao, and Chao Dong. A new journey from sdrtv to hdrtv. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4500–4509, 2021c.

Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, 2020.

Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2154–2164, 2020. doi: 10.1109/CVPR42600.2020.00223.

Xuan Dong, Xiaoyan Hu, Weixin Li, Xiaojie Wang, and Yunhong Wang. Miehdr cnn: Main image enhancement based ghost-free high dynamic range imaging using dual-lens systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 1264–1272, 2021.

M. Frigo and S. G. Johnson. Fftw: An adaptive software architecture for the fft. *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, 3, 1998.

Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2782–2790, 2016.

Dario Fuoli, Luc Van Gool, and Radu Timofte. Fourier space losses for efficient perceptual image super-resolution, 2021.

Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics (TOG)*, 36 (4):1–12, 2017.

Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1780–1789, 2020a.

Chunle Guo Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1780–1789, 2020b.

Xiaojie Guo and Qiming Hu. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, 131(1):48–66, 2023.

Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. *arXiv preprint arXiv:2106.14501*, 2021.

Jingwen He, Yihao Liu, Yu Qiao, and Chao Dong. Conditional sequential modulation for efficient global image retouching. In *Proceedings of the European Conference on Computer Vision*, pp. 679–695. Springer, 2020a.

Jingwen He, Yihao Liu, Yu Qiao, and Chao Dong. Conditional sequential modulation for efficient global image retouching. In *European Conference on Computer Vision*, pp. 679–695. Springer, 2020b.

Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. Exposure: A white-box photo post-processing framework. *ACM Transactions on Graphics (TOG)*, 37(2):1–17, 2018.

Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (eds.), *Computer Vision – ECCV 2022*, 2022a.

Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial frequency interaction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022b.

Yongqing Huo, Fan Yang, Le Dong, and Vincent Brost. Physiological inverse tone mapping based on retina response. *The Visual Computer*, 30:507–517, 2014.

Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. DSLR-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3277–3285, 2017.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134, 2017.

Jingxia Jiang, Tian Ye, Jinbin Bai, Sixiang Chen, Wenhao Chai, Shi Jun, Yun Liu, and Erkang Chen. Five a $\hat{}\{+\}$ network: You only need 9k parameters for underwater image enhancement. *arXiv preprint arXiv:2305.08824*, 2023.

Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

Soo Ye Kim, Jihyong Oh, and Munchurl Kim. Deep sr-itm: Joint learning of super-resolution and inverse tone-mapping for 4k uhd hdr applications. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3116–3125, 2019.

Soo Ye Kim, Jihyong Oh, and Munchurl Kim. JSI-GAN: GAN-based joint super-resolution and inverse tone-mapping with pixel-wise task-specific filters for UHD HDR video. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 11287–11295, 2020.

Sangrok Lee, Jongseong Bae, and Ha Young Kim. Decompose, adjust, compose: Effective normalization by playing with frequency for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11776–11785, June 2023.

Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28 (1):492–505, 2019a. doi: 10.1109/TIP.2018.2867951.

Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2019b.

Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *arXiv preprint arXiv:2103.00860*, 2021.

Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. *International Conference on Learning Representations*, 2023.

Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust Retinex model. *IEEE Transactions on Image Processing*, 27 (6):2828–2841, 2018.

Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. *CoRR*, abs/1705.08086, 2017. URL `http://arxiv.org/abs/1705.08086`.

Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10561–10570, 2021.

Kun Lu and Lihong Zhang. TBEFN: A two-branch exposure-fusion network for low-light image enhancement. *IEEE Transactions on Multimedia*, 23:4093–4105, 2020.

Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. MBLLEN: Low-light image/video enhancement using CNNs. In *Proceedings of the British Machine Vision Conference*, pp. 1–13, 2018.

Sean Moran, Pierre Marza, Steven McDonagh, Sarah Parisot, and Gregory Slabaugh. DeepLPF: Deep local parametric filters for image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12823–12832, 2020.

Ntumba Elie Nsamp, Zhongyun Hu, and Qing Wang. Learning exposure correction via consistency modeling. In *Proceedings of the British Machine Vision Conference*, pp. 1–12, 2021.

Jongchan Park, Joon-Young Lee, Donggeun Yoo, and In So Kweon. Distort-and-recover: Color enhancement using deep reinforcement learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5928–5936, 2018.

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. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.

Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 11908–11915, 2020.

Yongming Rao, Wenliang Zhao, Zheng Zhu, Jiwen Lu, and Jie Zhou. Global filter networks for image classification. *Advances in Neural Information Processing Systems*, 34, 2021.

Ali M Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *IEEE Transactions on Multimedia*, pp. 35–44, 2004.

Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 2149–2159, 2022.

Haoyuan Wang, Ke Xu, and Rynson W.H. Lau. Local color distributions prior for image enhancement. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6849–6857, 2019a.

Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6849–6857, 2019b.

Wenjing Wang, Chen Wei, Wenhan Yang, and Jiaying Liu. GLADNet: Low-light enhancement network with global awareness. In *Proceedings of 13th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 751–755, 2018.

Yudong Wang, Jichang Guo, Huan Gao, and Huihui Yue. UIECˆ 2-Net: CNN-based underwater image enhancement using two color space. *Signal Processing: Image Communication*, 96:116250, 2021.

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep Retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.

Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5901–5910, June 2022.

Yuxin Wu and Kaiming He. Group normalization. *CoRR*, abs/1803.08494, 2018. URL `http://arxiv.org/abs/1803.08494`.

Fengze Liu Qingli Li Wei Shen Xintian Mao, Yiming Liu and Yan Wang. Intriguing findings of frequency selection for image deblurring. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, 2023.

Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. A fourier-based framework for domain generalization. In *CVPR*, pp. 14383–14392, 2021a.

Shuang Xu, Jiangshe Zhang, Zixiang Zhao, Kai Sun, Junmin Liu, and Chunxia Zhang. Deep gradient projection networks for pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1366–1375, June 2021b.

Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. SNR-aware low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17714–17724, June 2022a.

Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17714–17724, 2022b.

Junfeng Yang, Xueyang Fu, Yuwen Hu, Yue Huang, Xinghao Ding, and John Paisley. Pannet: A deep network architecture for pan-sharpening. In *IEEE International Conference on Computer Vision*, pp. 5449–5457, 2017.

Kai-Fu Yang, Cheng Cheng, Shi-Xuan Zhao, Hong-Mei Yan, Xian-Shi Zhang, and Yong-Jie Li. Learning to adapt to light. *International Journal of Computer Vision*, pp. 1–20, 2023.

Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3063–3072, 2020a.

Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision (CVPR)*, pp. 3063–3072, 2020b.

Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. Frequency and spatial dual guidance for image dehazing. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (eds.), *Computer Vision – ECCV 2022*, pp. 181–198, Cham, 2022a. Springer Nature Switzerland.

Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. Frequency and spatial dual guidance for image dehazing. In *European Conference on Computer Vision*, pp. 181–198. Springer, 2022b.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5728–5739, 2022.

Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):2058–2073, 2020.

Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the ACM International Conference on Multimedia*, pp. 1632–1640, 2019.

Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021a.

Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021b.

Lin Zhao, Shao-Ping Lu, Tao Chen, Zhenglu Yang, and Ariel Shamir. Deep symmetric network for underexposed image enhancement with recurrent attentional learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12075–12084, 2021.

Naishan Zheng, Man Zhou, Yanmeng Dong, Xiangyu Rui, Jie Huang, Chongyi Li, and Feng Zhao. Empowering low-light image enhancer through customized learnable priors, 2023.

Man Zhou, Jie Huang, Chongyi Li, Hu Yu, Keyu Yan, Naishan Zheng, and Feng Zhao. Adaptively learning low-high frequency information integration for pan-sharpening. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, pp. 3375–3384, New York, NY, USA, 2022a. Association for Computing Machinery. ISBN 9781450392037. doi: 10.1145/3503161.3547924. URL https://doi.org/10.1145/3503161.3547924.

Man Zhou, Jie Huang, Keyu Yan, Hu Yu, Xueyang Fu, Aiping Liu, Xian Wei, and Feng Zhao. Spatial-frequency domain information integration for pan-sharpening. In *Computer Vision – ECCV 2022*, pp. 274–291, Cham, 2022b. Springer Nature Switzerland.

Man Zhou, Keyu Yan, Jie Huang, Zihe Yang, Xueyang Fu, and Feng Zhao. Mutual information-driven pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1798–1808, June 2022c.

man zhou, Hu Yu, Jie Huang, Feng Zhao, Jinwei Gu, Chen Change Loy, Deyu Meng, and Chongyi Li. Deep fourier up-sampling. In *Advances in Neural Information Processing Systems*, volume 35, pp. 22995–23008. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/91a23b3e6a2ebaad62e17d0269f88c6b-Paper-Conference.pdf.

Man Zhou, Jie Huang, Chun-Le Guo, and Chongyi Li. Fourmer: An efficient global modeling paradigm for image restoration. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 42589–42601. PMLR, 23–29 Jul 2023a.

Man Zhou, Jie Huang, Naishan Zheng, and Chongyi Li. Learned image reasoning prior penetrates deep unfolding network for panchromatic and multi-spectral image fusion, 2023b.

Anqi Zhu, Lin Zhang, Ying Shen, Yong Ma, Shengjie Zhao, and Yicong Zhou. Zero-shot restoration of underexposed images via robust Retinex decomposition. In *IEEE International Conference on Multimedia and Expo*, pp. 1–6, 2020.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2223–2232, 2017.

Appendix

In this appendix, we provide additional details and results.

In Sec. A, we present more implementation details.

In Sec. B, we extend the CFTL to other tasks, including image dehazing and pan-sharpening.

In Sec. C, we present more illustrations about the global information for image enhancement.

In Sec. D, we present more discussions about the toy experiment.

In Sec. E, we present more illustrations of the mechanism of our proposed method.

In Sec. F, we present the discussion about our work, including the reason for its design formats, the reason for its effectiveness, its limitations, and potential extension formats.

In Sec. G, we present detailed information about experimental settings.

In Sec. H, we present more ablation studies for investigating the CFTL.

In Sec. I, we present more results of applying the CFTL in other backbones.

In Sec. J, we present more results by comparing with more methods.

In Sec. K, we show more visualization results on multiple image restoration enhancement tasks.

A More Implementation Details

Pseudo code of the CFTL. For the implementation of the CFTL, we first provide the pseudo-code of the original CFTL and High-order CFTL in Fig. 8 as follows:

```
def CFTL(F):

 # F: input with shape [N, C,
    H, W]
   FG = GlobalPooling(F)
   # Calculation in Eq.5
   #FG: [N, C, 1, 1]

   FA, FP = CFFT(FG)
   # Applying Channel-based
      FFT in Eq.3 and Eq.4
   # FA and FP are the derived
      amplitude and phase
   FA = Seq1(FG)*FA
   FP = Seq2(FG)*FP
   # Process amplitude and
      phase in Eq.6
   FI = iCFFT(FA,FP)
   # iCFFT is the
      Channel-based iFFT
   # FI: [N, C, 1, 1]
   Y = Repeat(FI)
   # Repeat FI to the original
      resolution [N, C, H, W]


   Return Y #[N, C, H, W]
```

```
def High-order CFTL(F):

 # F: input with shape [N, C,
    H, W]
   FG = GlobalPooling(F)
   # Calculation in Eq.5
   #FG: [N, C, 1, 1]
   F^k = Korder(F)
   # Calculation in Eq.8
   #F^k: [N, C, 1, 1]
   Fsum = FG+F^k
   # Integrate FG and F^k in
      Eq.9

   FA, FP = CFFT(Fsum)
   # Applying Channel-based
      FFT in Eq.3 and Eq.4
   FA = Seq1(FG)*FA
   FP = Seq2(FG)*FP
   # Process amplitude and
      phase in Eq.6
   FI = iCFFT(FA,FP)
   # iCFFT is the
      Channel-based iFFT
   # FI: [N, C, 1, 1]
   Y = Repeat(FI)
   # Repeat FI to the original
      resolution [N, C, H, W]

   Return Y #[N, C, H, W]
```

Figure 8: **Pseudo-code of the two variants of the proposed CFTL.** The left is the *Original CFTL* while the right is the *High-oreder CFTL*.

```
def Group CFTL(F):

 # F: input with shape [N, C,
    H, W]
   FG = GlobalPooling(F)
   # Calculation in Eq.5
   #FG: [N, C, 1, 1]
   [FG_1,..FG_K] = Split(FG)
   # Split FG into K groups in
       Eq.10
   # FG_i: [N,C/K,1,1]
   for i in [1, K]:
      FA_i, FP_i = CFFT(FG_i)
      # Applying Channel-based
          FFT in Eq.3 and Eq.4
      # FAi and FPi are the
          derived amplitude and
          phase
      FA_i = Seq1i(FG_i)*FA_i
      FP_i = Seq2i(FG_i)*FP_i
      # Process amplitude and
          phase in Eq.6
      FI_i = iCFFT(FA_i,FP_i)
      # iCFFT is the
          Channel-based iFFT
      # FI_i: [N, C, 1, 1]
   FI = cat([FI_1,..FI_K])
   # Concatention in Eq.11
   Y = Repeat(FI)
   # Repeat FI to the original
       resolution [N, C, H, W]

   Return Y #[N, C, H, W]
```

```
def Spatial-Fourier CFTL(F):

 # F: input with shape [N, C,
    H, W]
   FAS, FPS = FFT(F)
   # Applying Spatial-based
       FFT in Eq.1 and Eq.2
   #FAS, FPS: [N, C, H, W]

   FASA, FASP = CFFT(FAS)
   # Applying Channel-based
       FFT in Eq.3 and Eq.4
   FASA = Seq1(FASA)
   FASP = Seq2(FASP)
   # Process amplitude and
       phase
   FAS =
       Conv_1x1(cat[FASA,FASP])
   # Fuse FASA and FASP as the
       processed amplitude

   FI = iFFT(FSA,FPS)
   # iFFT is the Spatial-based
       iFFT
   # FI: [N, C, H, W]
   Y = FI

   Return Y #[N, C, H, W]
```

Figure 9: **Pseudo-code of another two variants of the proposed CFTL.** The left is the *Group CFTL* while the right is the *Spatial-Fourier CFTL*.

We also provide the pseudo-code of the Group CFTL and Spatial-Fourier CFTL in Fig. 9.

Illustration of Spatial-Fourier CFTL. We present the detailed illustration of CFTL in Fig. 10. As also described in the main body and Fig. 9, we do not apply the channel-based iFFT on the processed spatial-based Fourier amplitude FAS in Fig. 9. Instead, we directly convert it back to the original space with the spatial-based Fourier phase. **The reason** is that we find applying the channel-based iFFT on FAS would lead to unstable training and result in "NAN" phenomenon. This issue could be attributed to the unstable gradient in two sequential inverse Fourier transforms.

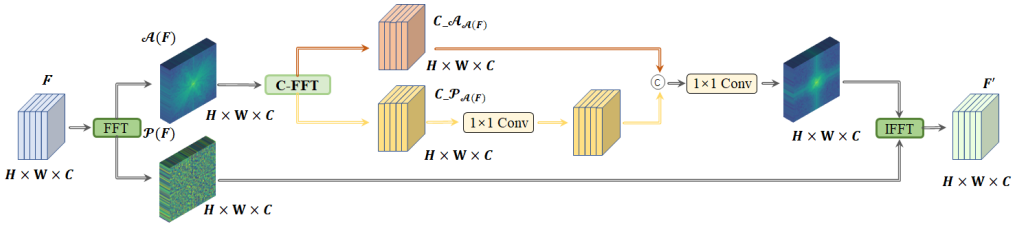


Figure 10: The illustration of Spatial-Fourier CFTL.

Since the spatial-based Fourier component is an effective global information representation, delving into its channel representation with a channel-based Fourier transform would enhance this representation. Meanwhile, the spatial-based Fourier component is a generalized representation of global pooling as described in Sec. 3.3. Therefore, it is reasonable to apply the channel-based Fourier

transform to the spatial-based Fourier component. Finally, our Spatial-based CFFT achieves better performance than the previous Spatial-based Fourier transform, depicting its effectiveness.

Implementation details of CFTL-Net. We implement the CFTL-Net in an encoder-decoder architecture, which consists of four scales with a feature channel number of 8. Its implementation structure is shown in Fig. 11, the CFTL-Net employs Spatial-Fourier CFTL and High-order CFTL to formulate the basic unit for processing features in one scale.

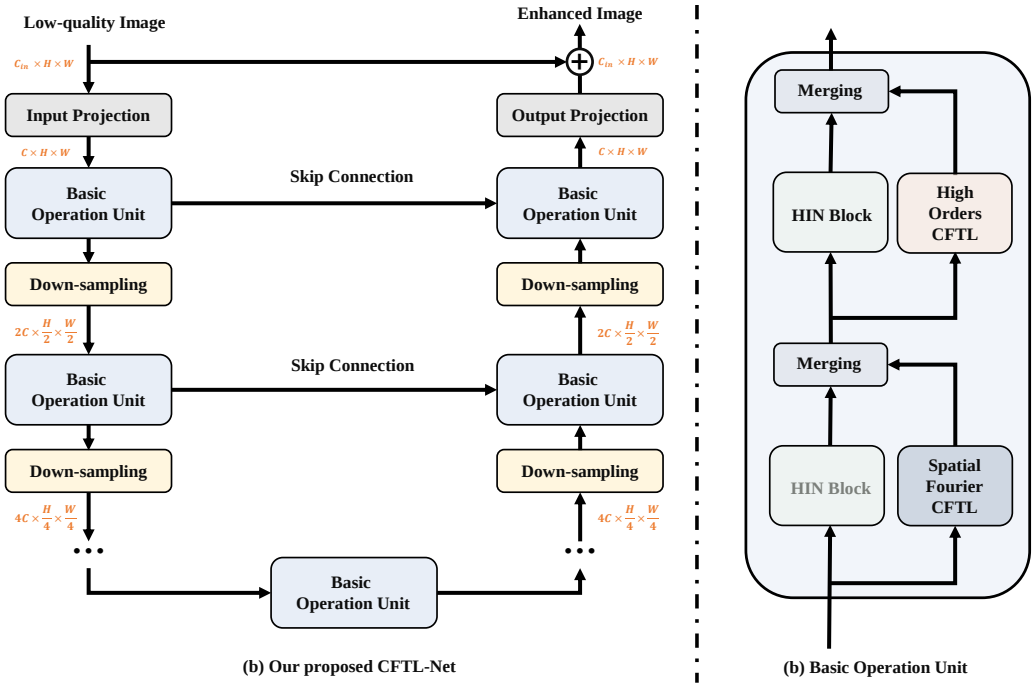


Figure 11: The illustration of the CFTL-Net.

Specifically, the basic unit consists of two parts borrowed from the transformer, the former part focuses on processing features in the spatial dimension, while the latter aims to process channel-dimension information. We apply Spatial-Fourier CFTL in the former part and the High-order CFTL in the latter part. In each part, besides the two operations proposed in this paper, we introduce the half-instance normalization block as the backbone block. Note that the performance of the CFTL-Net could be further improved if other effective blocks can replace the backbone block.

We train the CFTL-Net on a single GTX3090 GPU with a batch size of 4 and total epochs of 1000, the learning rate is set as 8e-4 and decays to half every 200 epochs. The loss function is set as the L1 loss and the training process is end-to-end.

Details of the inverse Fourier transform. In the main body of Sec. 3.4, we illustrate the operation of the Fourier transform, here, we also depict the inverse Fourier transform. Given the amplitude component $\mathcal{A}(y)(f)$ and phase component $\mathcal{P}(y)(f)$, the real and image parts of the Fourier representation are obtained:

$$\begin{aligned} R(y) &= \mathcal{A}(y)(f) \odot cos(\mathcal{P}(y)(f)), \\ I(y) &= \mathcal{A}(y)(f) \odot sin(\mathcal{P}(y)(f)). \end{aligned} \tag{13}$$

Then, $\mathcal{F}(y)(f)$ is formulated by $R(y)$ and $I(y)$, and perform inverse Fourier transform as:

$$y = \mathcal{F}^{-1}(y)(f) = \frac{1}{C}\sum_{f=0}^{C-1} \mathcal{F}(y)(f) e^{j2\pi \frac{c}{C} f}, \tag{14}$$

B EXTEND CFTL TO OTHER TASKS.

Extension on Image Dehazing. Following (Dong et al., 2020), we employ the RESIDE dataset (Li et al., 2019a) consisting of Indoor and Outdoor parts for evaluations. We adopt the network of MSBDN (Dong et al., 2020) and FFA-Net (Qin et al., 2020) as the baseline for validation. The results are presented in Table 6.

Extension on Guided Image Super-resolution. We apply the original CFTL to the GPPNN (Xu et al., 2021b) and PANNet (Yang et al., 2017) baselines in the pan-sharpening task, which is a common task in guided image super-resolution. We integrate it when fusing pan and multi-spectral features, and experimental results on WorldView II dataset (Zhou et al., 2022c) in Fig. B suggest the effectiveness of the CFTL.

Settings	#Param	Flops	RESIDE(ITS)	RESIDE(OTS)
MSBDN (Baseline)	31.35M	166.02G	29.77/0.9591	28.88/0.9581
+Original CFTL	31.36M	166.63G	**30.20**/0.9632	**29.26**/0.9588
+Group CFTL	31.36M	166.63G	29.96/**0.9665**	29.02/0.9600
+High-order CFTL	31.36M	166.63G	30.13/0.9611	29.05/0.9598
+Spatial-Fourier CFTL	31.36M	166.76G	29.89/0.9596	29.12/**0.9602**
FFA (Baseline)	4.46M	1.15T	36.39/0.9886	33.57/0.9840
+Original CFTL	4.50M	1.16T	36.46/0.9902	33.89/**0.9912**
+Group CFTL	4.49M	1.16T	37.03/0.9936	**34.28**/0.9901
+High-order CFTL	4.50M	1.16T	**37.25/0.9917**	33.70/0.9842
+Spatial-Fourier CFTL	4.51M	1.16T	36.61/0.9913	33.79/0.9904
CFTL-Net	0.028M	3.64G	33.91/0.9829	31.62/0.9772

Table 6: Results of image dehazing.

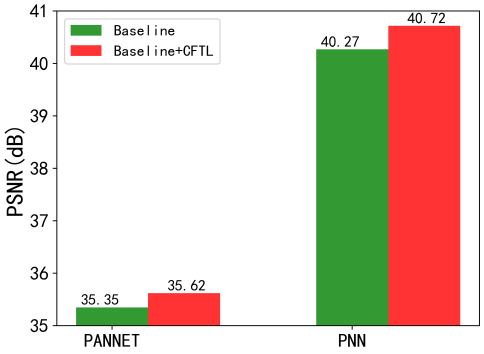


Figure 12: Quantative results of Pan-sharpening.

C MORE ILLUSTRATION ABOUT GLOBAL INFORMATION FOR IMAGE ENHANCEMENT

In the main body, we describe that the global average pooling equals $\mathcal{A}(0,0)$ in the amplitude. Here, we further verify this both from two sides. Typically, the Spatail Fourier transform is expressed as:

$$F(x)(u,v) = \frac{1}{\sqrt{HW}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} x(h,w) e^{-j2\pi(\frac{h}{H}u + \frac{w}{W}v)} \tag{15}$$

The center point of the amplitude spectrum means that u and v are 0. The formula is as follows:

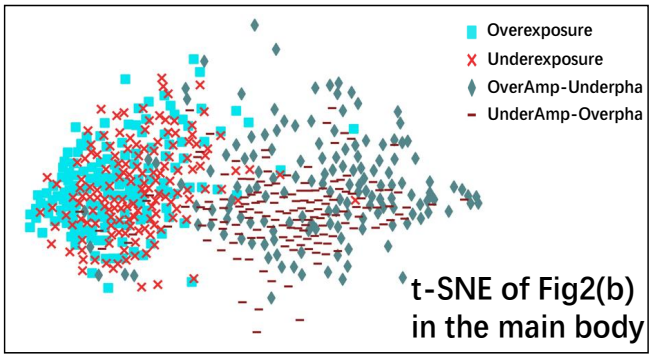
$$F(x)(0,0) = \frac{1}{\sqrt{HW}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} x(h,w) \tag{16}$$

It can be seen that the above formula is essentially to find the average value of the entire feature map. Therefore, taking the center point of the amplitude spectrum is equivalent to global average pooling (GAP).

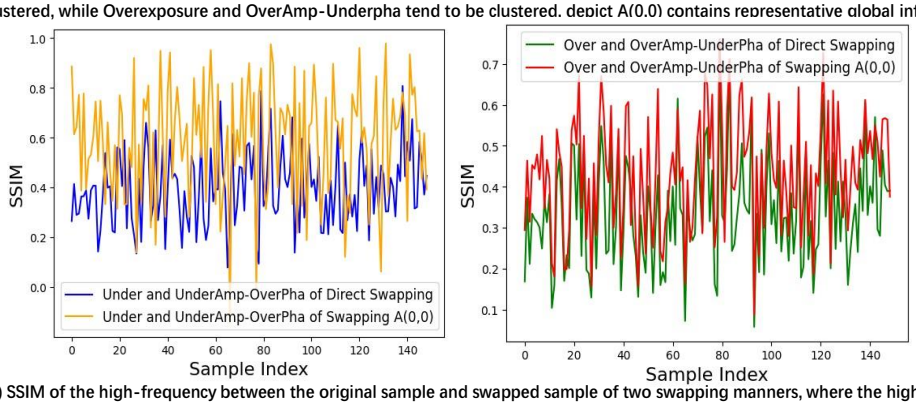
In Fig. 2 of the main body, we illustrate that $\mathcal{A}(0,0)$ comprises most global information about image enhancement by swapping $\mathcal{A}(0,0)$ of the underexposure and overexposure images, and their lightness becomes much similar to the swapped image. The statistic result is presented in Fig. 13 (a), also indicating that global pooling is an effective global information format related to image enhancement.

Moreover, this phenomenon also depicts the advantages of our proposed method. As shown in Fig. 2 (a), previous swapping amplitude would cause artifacts in the swapped results, while in Fig. 2 (b), swapping $\mathcal{A}(0,0)$ can alleviate this issue to much extent. We further verify this by comparing SSIM between the original sample and the swapped sample in terms of their high-frequency component. As shown in Fig. 13 (b), swapping $\mathcal{A}(0,0)$ leads to higher SSIM in terms of most samples, depict there exist less artifacts of swapping $\mathcal{A}(0,0)$.

To discuss the reason, we describe it as follows. **This is because there exists a mismatch between the swapped amplitude and the original phase components, which is also referred to as disconjugacy in signal processing.** Therefore, we argue that previous methods directly processing



(a) t-SNE of samples in Fig 2 of the main bofy, by swapping A(0,0) in the amplitude, the Underexposure and UnderAmp-Overpha tend to be clustered, while Overexposure and OverAmp-Underpha tend to be clustered. depict A(0,0) contains representative global information. .



(b) SSIM of the high-frequency between the original sample and swapped sample of two swapping manners, where the high frequency is extracted by substracting the blurred part. Swapping A(0,0) tend to achieve higher SSIM due to its less resulted artifacts.

Figure 13: Different from swapping the amplitude component in (a), we find swapping $\mathcal{A}(0,0)$ in (b) can also swap lightness, while the swapped results contain fewer artifacts.

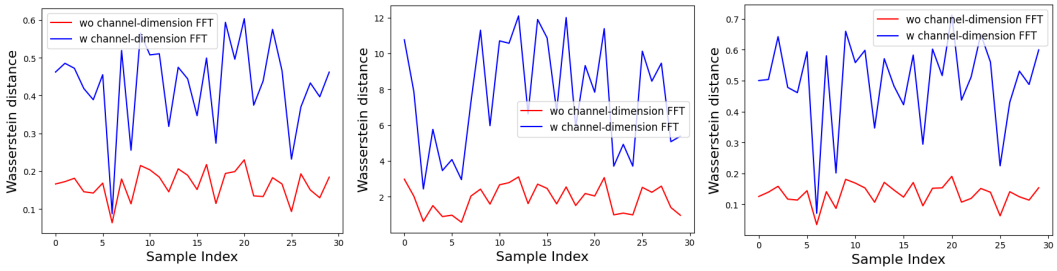
information in the Fourier domain ignore this issue, and thus may limit the further improvement of performance. **Instead, our proposed information processing format avoids this issue, while also processing global information that is highly related to image enhancement.**

D MORE DISCUSSION ABOUT TOY EXPERIMENT

We provide more discussion about toy experiments here.

More illustration of enhancing the discriminability of features using channel-dimension FFT.

In the supplement of Fig. 4 in the main body, which measures the distribution distance of all samples, we provide the distribution distance of each test sample on the SICE dataset for the toy experiment in Fig. 14. As can be seen, the channel-dimension Fourier transform increases the differences between underexposure and overexposure samples.



(a) Original space (b) Spatial-FFT space (c) Global pooling space

Figure 14: Toy experiment results on sample levels in different spaces, which measures the distribution distances between underexposure and overexposure w/wo using channel-dimension FFT.

Feature visualization of applying channel-dimension Fourier transform.

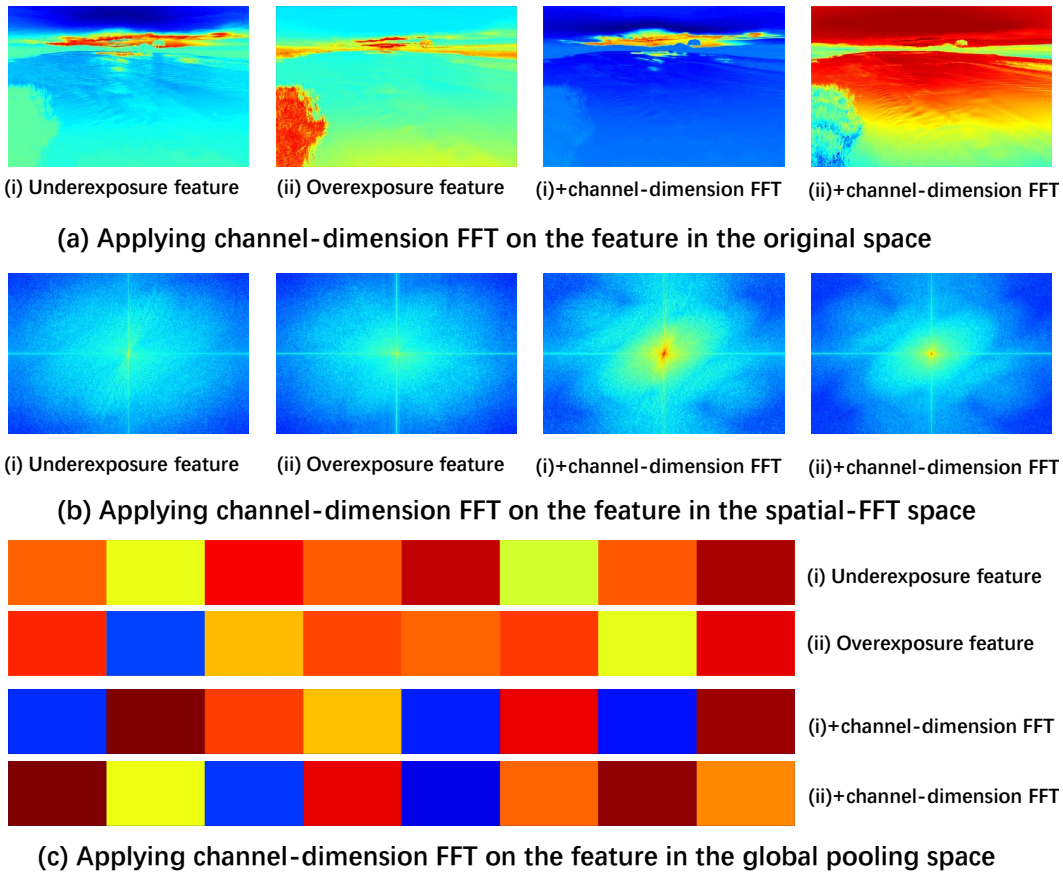


Figure 15: Feature visualization of applying channel-dimension Fourier transform on features in different spaces, since the global pooling space is hard to show in one channel, we present all channels of their features.

In supplement to the above statistic results of the toy experiment, we provide feature visualization of applying channel-dimension Fourier transform in different spaces. As shown in Fig. 15, applying channel-dimension Fourier transform leads to a more discriminative appearance between underexposure and overexposure features, which correspond to the previous results.

E More illustrations of the CFTL's mechanism

Making the other part focus on learning local representation. Since the CFTL mainly focuses on learning the global representation, the other part can better capture local representation. We present the feature visualization in Fig. 16, where the features come from the backbone of DRBN trained on the exposure correction task. It can be seen that the introduce of the CFTL brings more texture learning than other comparison operations, depicting its effectiveness.

Leading to adjust lightness information effectively. We present the visualization results of underexposure and overexposure features before and after processing by the CFTL. The features come from the backbone of DRBN trained on the exposure correction task. The results are presented in Fig. 17. As can be seen, **the CFTL can effectively reduce the gap between underexposure and overexposure, depicting its effectiveness in adjusting the lightness representation.** Moreover, we compare the distribution distance between underexposure and overexposure in different operations, the CFTL gets the lowest Wasserstein distance of 0.1682, while pooling attention and Sptail Fourier get 0.2622 and 0.1991, respectively. This result also depict the effectiveness of the proposed CFTL.

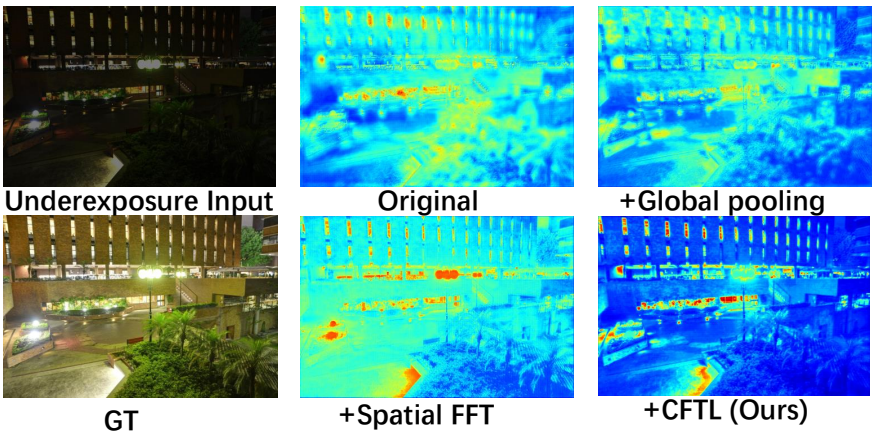


Figure 16: Feature visualization of the local branch on the DRBN-based backbone. the sample comes from the SICE dataset. Our proposed CFTL enables the local part captures more textures.

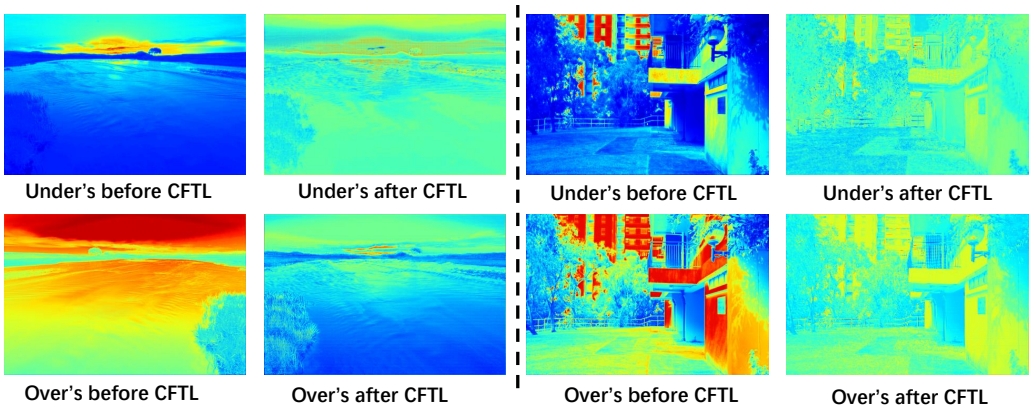


Figure 17: Feature visualization of underexposure and overexposure samples before and after processing by the CFTL, the gap between underexposure and overexposure features is reduced significantly after CFTL.

F More discussion about this work

The reason why CFTL is designed in this format. We perform CFTL in the global information representation, where most of them are based on the global pooling-based representation. The reasons for its design formats are summarized as follows: **(1)** global pooling-based information can represent the most information about global representation, we have detailed its advantages in Sec. C; **(2) one question is why not apply channel-based Fourier transform on the original feature?** We have attempted to do it, but the performance is not good as shown in Table 5 in the ablation study. Moreover, we have also presented the feature in Fig. 18, and the resulting features appear to have similar properties to the original feature, which would not help improve the learning process. Instead, global pooling information is more suitable for existing convenient operations such as 1x1 convolution to conduct learning. **(3)** We perform operations on the amplitude and phase components because they have explicit meanings in signal processing, and here they correspond to the information energy distribution and information position distribution along the channel dimension.

Discussion why CFTL is effective. Besides the experiments and other sections about the mechanism of why CFTL works, here, we present other explanations: **(1)** As depicted in the main body, the channel-based relationship reflects the global style information, such as the Gram matrix. Based on this, the CFTL also learns the relationship of channel information, thus affecting the global style information effectively. **(2)** Applying channel-based Fourier transform derives the global representation of channel information, therefore, only changing a point in the transformed feature would lead to significant information change in the original channel information, thus leading to the ef-

Figure 18: Feature visualization of applying channel-dimension Fourier transform on the spatial feature (bottom), which also exhibits spatial diversity like original features (top).

fectiveness of CFTL. **(3)** The proposed mechanism also decouples the learning of global and local information, facilitating the learning of the whole framework. **(4)** Many image enhancement tasks are based on channel prior, such as dark channel prior or rank-one prior, we believe our designed format also attempt to exact the channel prior in the feature level for image enhancement.

Some limitations of CFTL. However, there are also some limitations of CFTL: **(1)** it still occupies some computation costs, which need to be further improved; **(2)** its effects on some popular techniques such as diffusion model, have not explored; **(3)** the more comprehensive experiments on broader Low-level vision tasks (*e.g.*, image de-noising and image de-blurring) have not been explored; **(3)** we mainly apply the CFTL on the lightweight and classic backbones, while some recent backbones with huge parameters have not been explored with the CFTL. **Note that the focus of this work is not to bring a new operation to improve the state-of-the-art, and we aim to introduce a new global information representation format to depict its effectiveness and facilitate the application of classic backbones.**

Other potential application formats. According to the above discussions, there are some potential formats for the CFTL: **(1)** conducting attention operations on the channel-based Fourier transformed features without pooling, which could extend the proposed mechanism to the spatial dimension features. **(2)** applying CFTL on the wavelet-based features, which could learn the coefficients of wavelet conveniently. **(3)** applying CFTL to fuse features from different architectures or positions, facilitating connecting their channel relationship. **(4)** applying CFTL as a loss function, which could facilitate the optimization of learning global information.

G More details about experimental settings

Illustration of the comparison operator. In Fig. 19, We present the illustration of the two comparison operators in the experiments.

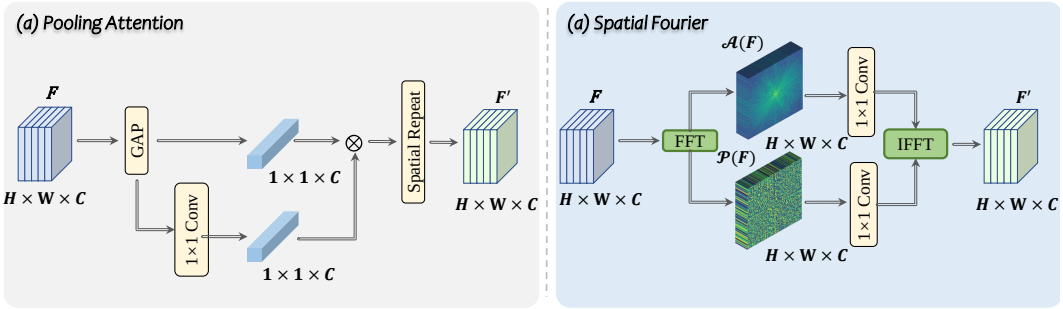


Figure 19: The illustration of the two comparison operators in the experiments.

Illustration of how to integrate the CFTL in the backbone. In Fig. 20, we present how to integrate the CFTL in the backbone. For the backbone with sequential blocks, we usually place the CFTL on the first part of the block. While for the backbone with an encoder-decoder backbone, we place the CFTL on its shallow layers.

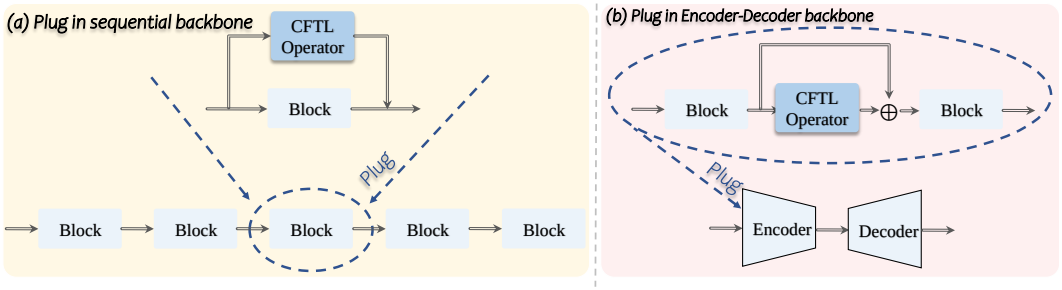


Figure 20: The illustration of how to integrate the CFTL in existing backbones.

H MORE ABLATION STUDIES FOR INVESTIGATING THE CFTL.

Refer to the main body, we also perform ablation studies in the DRBN backbone on the exposure correction task (SICE dataset).

Ablation study of different orders in High-order CFTL. Defautly, we set k in High-order CFTL as 2. Here, we also present other combinations of different orders' results in Table 7. As can be seen, most of them achieve comparable performance with the default setting, but outperform the original CFTL. These results demonstrate the reasonableness of the default setting and also the effectiveness of the High-order CFTL.

Settings (order)	Baseline (DRBN)	+CFTL	+CFTL (1+2)	+CFTL (1+4)	+CFTL (1+2+4)
PSNR/MS-SSIM	17.65/0.6798	21.32/0.7250	**21.64**/ 0.7243	21.53/ 0.7247	21.56/ **0.7262**

Table 7: Investigating different orders (denoted as 1+k) in High-order CFTL on the SICE dataset for exposure correction, where "+CFTL (1+2)" is the default High-order CFTL in the manuscript.

Ablation study of different groups in Group CFTL. Defautly, we set K in Group CFTL as 4. Here, we also present setting other group numbers' results in Table. 8. It can be seen that setting the group numbers as 1, 2, and 4 has similar results, while increasing the number to 8 would lead to a performance drop. This result suggests the reasonableness of the default setting, where setting group as 4 achieves comparable performance with introducing fewer parameters.

Settings (order)	Baseline (DRBN)	+CFTL (K=1)	+CFTL (K=2)	+CFTL (K=4)	+CFTL (K=8)
PSNR/MS-SSIM	17.65/0.6798	21.32/0.7250	21.26/0.7209	21.30/0.7177	20.86/0.7093
Parameters	0.532M	0.534M	0.533M	0.532M	0.532M

Table 8: Investigating the number of groups K in Group CFTL on the SICE dataset for exposure correction, where "+CFTL(K=1)" is the default Group CFTL in the manuscript.

Ablation study of different settings in Spatial-Fourier CFTL. In Sec. A, we present the details of the Spatial-Fourier CFTL. In Table 9, we present other configurations of the Spatial-Fourier CFTL. As can be seen, applying CiFFT (channel-based inverse Fourier transform) would lead to unstable training and results in NAN, while processing both amplitude and phase components derived from the spatial-based Fourier transform brings performance drop. Overall, the final setting of the Spatial-Fourier CFTL achieves the best performance.

Ablation study of investigating the channel number in the CFTL-Net. We set the channel number of CFTL-Net as 8 in the manuscript. We also investigate setting other channel numbers, and results conducted on the exposure correction and low-light image enhancement are presented in Fig. 21. It can be seen that increasing the channel number would lead to more performance improvement in exposure correction than low-light image enhancement, while reducing the channel number results in a significant performance drop. This could be attributed to that exposure correction requires a stronger ability to adjust different lightness. The results suggest the potential extensive ability of the CFTL-Net with more channel numbers.

Settings	Baseline (DRBN)	+Spatial-Fourier CFTL	+Spatial-Fourier CFTL (+CiFFT)	(a)
PSNR/MS-SSIM	17.65/0.6798	21.33/0.7201	NAN/NAN	21.20/0.7159
Parameters	0.532M	0.536M	0.536M	0.537M

Table 9: Investigating the configuration of Spatial-Fourier CFTL on the SICE dataset for exposure correction, where "(a)" denotes processing both amplitude and phase components derived from the spatial-based Fourier transform.

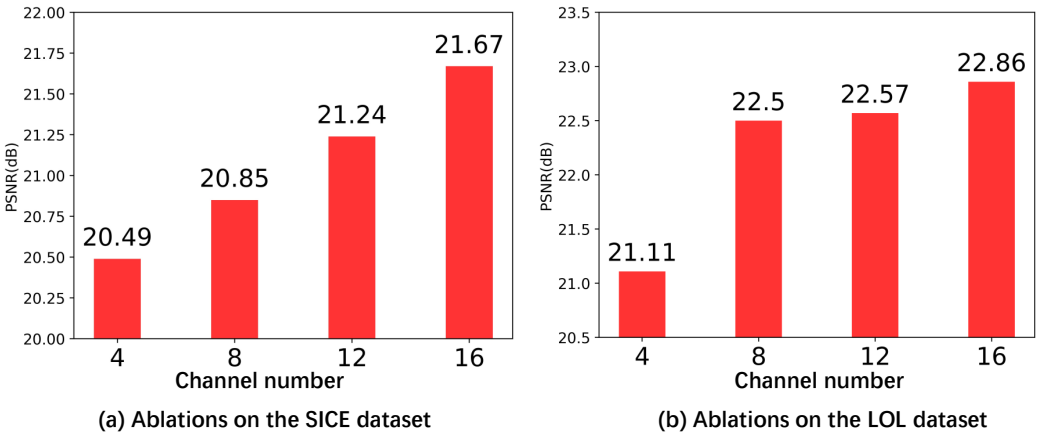


Figure 21: Abaltion studies for investigating different channel numbers of the CFTL-Net in exposure correction (left) and low-light image enhancement (right).

I More results of applying the CFTL in other backbones

In the main body, we adopt a few networks as the backbone to integrate the CFTL. Here, we employ more networks as the backbone to demonstrate the scalability and effectiveness of the CFTL, which are presented in Table 10, Table 11, Table 12 and Table 13.

For Low-light image enhancement.

We further employ the Bread (Guo & Hu, 2023) as the backbone and perform the experiments on the LOL dataset (Wei et al., 2018). The extensive results presented in Table 10 validate the effectiveness of the proposed CFTL.

Settings	Baseline (Bread)	+Spatial Fourier	+Attention Pooling	+Original CFTL	+Group CFTL	+Spatial-Fourier CFTL
PSNR/SSIM	22.96/0.8383	23.22/0.8399	23.16/0.8407	23.31/0.8403	23.37/**0.8414**	**23.42**/0.8409

Table 10: More Comparison over low-light image enhancement on the LOL dataset.

For exposure correction. Moreover, we adopt FECNet (Huang et al., 2022b) as the backbone for exposure correction. We perform the experiments in the SICE dataset, and the experimental results in Table 11 demonstrate the effectiveness of our method. Note that FECNet is based on the design of spatial Fourier transform as the basic unit, this result depicts our method is compatible with existing spatial Fourier transform design.

For SDR2HDR translation. Moreover, we adopt the the first stage of the AGCM (Chen et al., 2021c) as the backbone for SDR2HDR translation. We perform the experiments in the HDRTV dataset (Chen et al., 2021c), and the experimental results in Table 12 demonstrate the effectiveness of our method.

For underwater image enhancement. Moreover, we adopt Five A plus Jiang et al. (2023) as the backbone for underwater image enhancement. We perform the experiments in the UIEB dataset (Li et al., 2019b), and the experimental results in Table 13 demonstrate the effectiveness of our method.

Settings	Baseline (FECNet)	+Spatial Fourier	+Attention Pooling	+Original CFTL
PSNR/SSIM	20.96/0.6849	21.06/0.6913	21.14/0.6955	21.60/0.7302

Table 11: More Comparison over exposure correction on the SICE dataset.

Settings	Baseline (AGCM)	+Spatial Fourier	+Attention Pooling	+Original CFTL
PSNR/SSIM	36.53/0.9624	36.58/0.9635	36.79/0.9655	36.83/0.9657

Table 12: More Comparison over SDR2HDR translation on the HDRTV dataset.

J MORE RESULTS BY COMPARING WITH MORE METHODS

We provide more quantitative results of the CFTL-Net and other comparison methods on different image enhancement tasks, including low-light image enhancement, exposure correction, and SDR2HDR translation. The results are presented in Table 14 (low-light image enhancement), Table 15 (low-light image enhancement), Table 16 (exposure correction), and Table 17 (SDR2HDR translation). which demonstrates the CFTL-Net achieves an elegant balance between performance and efficiency. Note that the improved version of the method in the main body also has higher performance than most comparison methods in these tables.

K MORE QUALITATIVE RESULTS

Due to the page limit of the main body, we provide more visualization results here. In the main body, we have presented the visual results of exposure correction in Fig. 6. Here, we further respectively present the results of low-light image enhancement (Fig. 22) underwater image enhancement (Fig. 23) as follows. As can be seen, our CFTL can help enhance more correct lightness and color, or reduce the structure artifacts.

Settings	Baseline (Five A plus)	+Spatial Fourier	+Attention Pooling	+Original CFTL	+Group CFTL	+Spatial-Fourier CFTL
PSNR/SSIM	23.63/0.9138	23.96/0.9157	23.79/0.9134	**24.24/0.9172**	24.07/0.9157	23.98/0.9145

Table 13: More Comparison over underwater image enhancement on the UIEB dataset.

Method	LOL		Huawei		# Param	GFlops
	PSNR	SSIM	PSNR	SSIM		
SRIE (Fu et al., 2016)	12.28	0.5962	13.04	0.4770	-	-
RobustRetinex (Li et al., 2018)	13.88	0.6643	14.60	0.5593	-	-
RetinexNet (Wei et al., 2018)	16.77	0.4257	16.65	0.4857	0.84M	148.54
MBLLEN (Lv et al., 2018)	17.56	0.7293	16.63	0.5264	0.45M	21.37
EnGAN (Jiang et al., 2021)	17.48	0.6746	17.03	0.5140	8.37M	72.61
GLADNet (Wang et al., 2018)	19.72	0.6802	17.76	0.5214	1.13M	275.32
Xu et al. (Xu et al., 2022a)	16.78	0.7665	16.12	0.5862	8.62M	68.45
TBEFN (Lu & Zhang, 2020)	17.35	0.7817	16.88	0.5759	0.49M	24.11
KIND (Zhang et al., 2019)	20.86	0.8023	16.48	0.5406	8.54M	36.57
ZeroDCE (Guo et al., 2020a)	15.29	0.5182	12.46	0.4074	0.08M	20.24
DRBN (Yang et al., 2020a)	20.13	0.8011	18.46	**0.6355**	0.53M	42.41
RUAS (Liu et al., 2021)	16.41	0.5004	13.76	0.5167	0.003M	0.86
KIND++ (Zhang et al., 2021b)	21.30	<u>0.8221</u>	15.78	0.4523	8.28M	2970.50
URetinex (Wu et al., 2022)	21.32	**0.8358**	<u>18.79</u>	0.6078	1.23M	68.37
LA-Net (Yang et al., 2023)	<u>21.71</u>	0.8149	18.15	0.5941	0.55M	185.79
CFTL-Net (Ours)	**22.50**	0.8139	20.91	0.6941	0.028M	3.64

Table 14: Quantitative results of different methods on the LOL and Huawei datasets for low-light image enhancement. #Param denotes the parameter number. The best and second results are highlighted in **bold** and <u>underline</u>, respectively.

Method	PSNR	SSIM	#Param
Whitebox (Hu et al., 2018)	18.59	0.7973	8.17M
Distort (Park et al., 2018)	19.54	0.7998	247.25M
HDRNet (Gharbi et al., 2017)	22.65	0.8802	0.46M
SID (Chen et al., 2018)	21.49	0.8425	7.40M
DUPE (Wang et al., 2019a)	20.22	0.8287	0.95M
DeepLPF (Moran et al., 2020)	23.21	0.8863	0.80M
DRBN (Yang et al., 2020a)	22.11	0.8684	0.53M
CSRNet (He et al., 2020a)	23.69	<u>0.8981</u>	0.034M
LA-Net (Yang et al., 2023)	19.94	0.8057	0.55M
DSN (Zhao et al., 2021)	<u>23.84</u>	**0.9002**	4.42M
CFTL-Net (Ours)	**24.03**	0.8904	0.028M

Table 15: Quantitative results of different methods on the MIT-FiveK dataset for low-light image enhancement. #Param denotes the parameter number. The best and second results are highlighted in **bold** and <u>underline</u>, respectively.

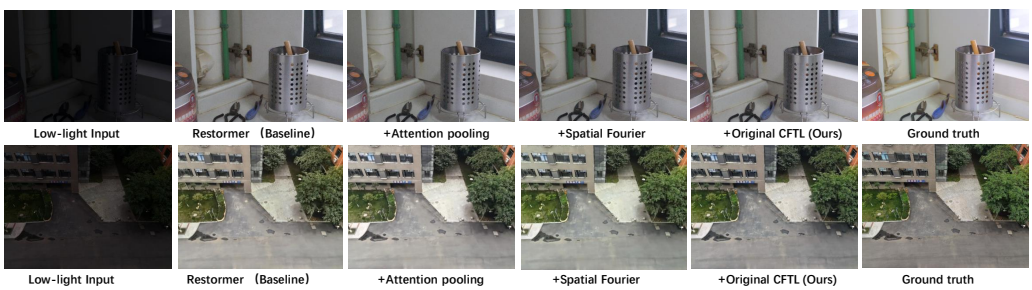


Figure 22: The visualization results on the LOL dataset (top) and Huawei dataset (bottom) for low-light image enhancement.

Method	MSEC						SICE						#Param	GFlops
	Underexposure		Overexposure		Average		Underexposure		Overexposure		Average			
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		
HE (Pizer et al., 1987)	16.52	0.6918	16.53	0.6991	16.53	0.6959	14.69	0.5651	12.87	0.4991	13.78	0.5376	-	-
CLAHE (Reza, 2004)	16.77	0.6211	14.45	0.5842	15.38	0.5990	12.69	0.5037	10.21	0.4847	11.45	0.4942	-	-
RetinexNet (Wei et al., 2018)	12.13	0.6209	10.47	0.5953	11.14	0.6048	12.94	0.5171	12.87	0.5252	12.90	0.5212	0.84M	148.54
DPED (Ignatov et al., 2017)	20.06	0.6826	13.14	0.5812	15.91	0.6219	16.83	0.6133	7.99	0.4300	12.41	0.5217	0.39M	94.64
SID (Chen et al., 2018)	19.37	0.8103	18.83	0.8055	19.04	0.8074	19.51	0.6635	16.79	0.6444	18.15	0.6540	7.40M	53.12
URetinexNet (Wu et al., 2022)	13.85	0.7371	9.81	0.6733	11.42	0.6988	17.39	0.6448	7.40	0.4543	12.40	0.5496	1.32M	68.37
Zero-DCE (Guo et al., 2020a)	14.55	0.5887	10.40	0.5142	12.06	0.5441	16.92	0.6330	7.11	0.4292	12.02	0.5311	0.079M	20.24
Zero-DCE++ (Li et al., 2021)	13.82	0.5887	9.74	0.5142	11.37	0.5583	11.93	0.4755	6.88	0.4088	9.41	0.4422	0.010M	0.17
RUAS (Liu et al., 2021)	13.43	0.6807	6.39	0.4655	9.20	0.5515	16.63	0.5589	4.54	0.3196	10.59	0.4393	0.0014M	0.86
DRBN (Yang et al., 2020b)	19.74	0.8290	19.37	0.8321	19.52	0.8309	17.96	0.6767	17.33	0.6828	17.65	0.6798	0.53M	42.41
MSEC (Afifi et al., 2021)	20.52	0.8129	19.79	0.8156	20.08	0.8210	19.62	0.6512	17.59	0.6560	18.58	0.6536	7.04M	35.87
CMEC (Nsamp et al., 2021)	22.23	0.8140	22.75	0.8336	22.54	0.8257	17.68	0.6592	18.17	0.6811	17.93	0.6702	5.40M	35.71
LA-Net (Yang et al., 2023)	21.84	0.8264	21.02	0.8164	21.35	0.8207	19.22	0.6514	17.65	0.6025	18.44	0.6270	0.55M	185.79
LCDPNet (Wang et al., 2022)	22.35	0.8650	22.17	0.8476	22.30	0.8552	20.71	0.6822	20.21	0.6863	20.46	0.6843	0.96M	9.40
CFTL-Net (Ours)	**23.04**	**0.8612**	**22.63**	**0.8567**	**22.88**	**0.8594**	**22.26**	**0.6927**	**20.71**	**0.7071**	**21.24**	**0.6999**	0.028M	3.61

Table 16: Quantitative results of different methods on the MSEC and SICE datasets for exposure correction. #Param denotes the parameter number. The best and second results are highlighted in **bold**.

Method	PSNR	SSIM	#Param
HuoPhyEO (Huo et al., 2014)	25.90	0.9296	-
Pixel2Pixel (Isola et al., 2017)	25.80	0.8777	11.38M
CycleGAN (Zhu et al., 2017)	21.33	0.8496	11.38M
HDRNet (Gharbi et al., 2017)	35.73	0.9664	0.46M
JSI-GAN (Kim et al., 2020)	<u>37.01</u>	**0.9694**	1.06M
Ada-3DLUT (Zeng et al., 2020)	36.22	0.9658	0.59M
DRBN (Yang et al., 2020a)	36.44	0.9671	0.53M
CSRNet (He et al., 2020a)	35.34	0.9625	0.034M
LA-Net (Yang et al., 2023)	31.52	0.9427	0.55M
AGCM (Plus) (Chen et al., 2021b)	36.88	0.9655	0.034M
CFTL-Net (Ours)	**37.37**	<u>0.9683</u>	0.028M

Table 17: Quantitative results of different methods on the HDRTV dataset for SDR2HDR translation. #Param denotes the parameter number. The best and second results are highlighted in **bold** and <u>underline</u>, respectively.

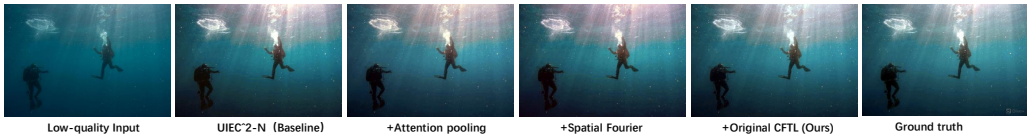


Figure 23: The visualization results on the UIEB dataset for underwater image enhancement.