Deep Fourier-based Exposure Correction Network with Spatial-Frequency Interaction

Jie Huang[1*], Yajing Liu[2*], Feng Zhao[1**], Keyu Yan[1], Jinghao Zhang[1], Yukun Huang[1], Man Zhou[1**], and Zhiwei Xiong[1]

[1] University of Science and Technology of China
[2] JD Logistics, JD.com
{hj0117,lyj123}@mail.ustc.edu.cn, fzhao956@ustc.edu.cn,
{keyu,jhaozhang,kevinh,manman}@mail.ustc.edu.cn, zwxiong@ustc.edu.cn

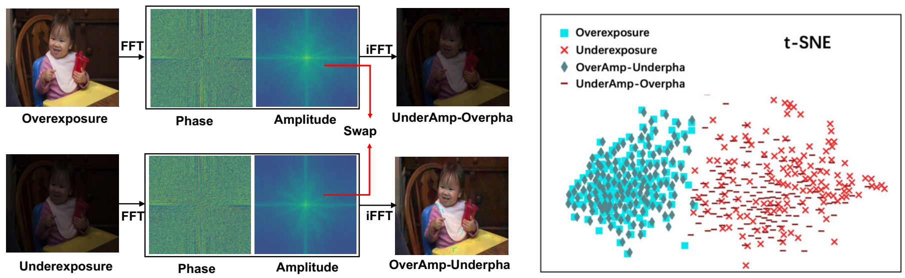
Abstract. Images captured under incorrect exposures unavoidably suffer from mixed degradations of lightness and structures. Most existing deep learning-based exposure correction methods separately restore such degradations in the spatial domain. In this paper, we present a new perspective for exposure correction with spatial-frequency interaction. Specifically, we first revisit the frequency properties of different exposure images via Fourier transform where the amplitude component contains most lightness information and the phase component is relevant to structure information. To this end, we propose a deep Fourier-based Exposure Correction Network (FECNet) consisting of an amplitude sub-network and a phase sub-network to progressively reconstruct the representation of lightness and structure components. To facilitate learning these two representations, we introduce a Spatial-Frequency Interaction (SFI) block in two formats tailored to these two sub-networks, which interactively process the local spatial features and the global frequency information to encourage the complementary learning. Extensive experiments demonstrate that our method achieves superior results than other approaches with fewer parameters and can be extended to other image enhancement tasks, validating its potential in wide-range applications. Code will be available at https://github.com/KevinJ-Huang/FECNet.

Keywords: Exposure Correction, Fourier Transform, Spatial-Frequency Interaction

1 Introduction

With the wide-range applications of camera devices, images can be captured under scenes with varying exposures, which could result in unsatisfactory visual results including lightness and structure distortions. Thus, it is necessary to correct such exposures of these images, which not only improves their visual

* Equal contribution.
** Corresponding author.



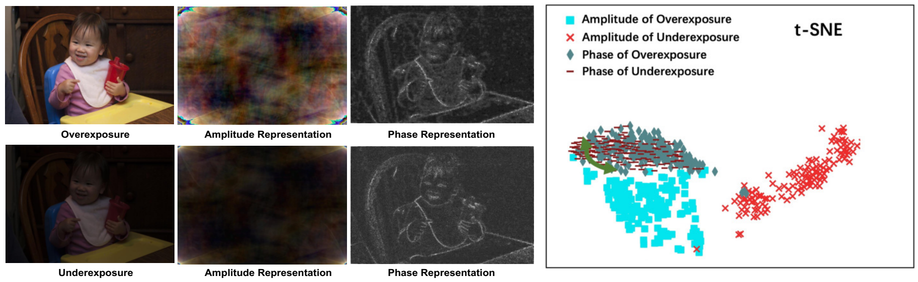
(a) (b)

Fig. 1. (a) We swap the amplitude and phase components of different exposures of the same context. The recombined result of the amplitude of underexposure and the phase of overexposure (UnderAmp-Overpha) has similar lightness appearance with underexposure, while the recombined result of the amplitude of overexposure and the phase of underexposure (OverAmp-Underpha) has similar lightness appearance with overexposure. **(b) The t-SNE [28] for images of overexposure, underexposure, UnderAmp-Overpha, and OverAmp-Underpha.** The distributions of images in UnderAmp-Overpha and Underexposure are matched, while the distributions of images in OverAmp-Underpha and Overexposure are matched, which indicate that the swapped amplitude components include the most lightness information.

qualities but also benefits other sub-sequential high-level vision tasks such as image detection and segmentation [41, 46].

The mixed degradations of both lightness and structure components may lead to difficulties in conducting exposure correction [26], which may cause structure distortions and ineffective lightness adjustments [2, 25]. To solve this problem, since different exposures share similar structure representations but different lightness depictions [22], it is natural to decompose and restore the lightness and structure components of the input image, respectively. Retinex theory-based methods [42, 51, 16] decompose images into illumination and reflectance components, and then separately recover the lightness and structure information. Multi-scale decomposition-based approaches [2, 23, 25] intend to decompose and recover the coarse-scale lightness and fine-scale structures in a progressive manner. With the advanced design of deep neural networks, recent techniques have significantly improved the visual quality. However, most of them rarely explore the potential solutions in the frequency domain, which is quite crucial for improving the image quality [13, 20].

In this work, we introduce a novel Fourier-based perspective to conduct exposure correction, which facilitates utilizing and restoring the frequency-domain information. From [43], the amplitude and phase components of Fourier space correspond to the style and semantic information of an image. This property can be extended in exposure correction, i.e., the amplitude component of an image reflects the lightness representation, while the phase component corresponds to structures and is less related to lightness. As shown in Fig. 1, we first swap the amplitude and phase components of different exposures of the same context. The recombined result of the amplitude of underexposure and the phase of overexposure has similar lightness appearance with underexposure, while the other



(a) (b)

Fig. 2. (a) The visualization for the amplitude and phase components of the same context. We apply the iFFT to the phase and amplitude to compare them in the spatial domain. The amplitude representations differ significantly between different exposures, while the phase representations are very similar across exposures and present structure representation. **(b) The t-SNE of amplitude and phase of different exposures.** The distributions of phase representations across different exposures are matched, while distributions of amplitude representations across different exposures vary greatly. It means that the phase component includes the most structure information and is less affected by lightness.

behaves conversely. This implies that the swapped amplitude components include the most lightness information, and the phase component may correspond to the structure representation and is less affected by lightness.

To validate this, as shown in Fig. 2, we apply the inverse Fast Fourier Transform (iFFT) [31] to the phase and amplitude components to visualize them in the spatial domain. The appearance of the phase representation is more similar to the structure representation, and the distribution of the phase components is less affected by lightness. To this end, the phase component is more related to structures that are less affected by lightness in the spatial domain. Therefore, following existing works that separately restore lightness and structure degradations, we intend to restore the amplitude and phase components progressively.

Based on the above analysis, we propose a Fourier-based Exposure Correction Network (FECNet), as shown in Fig. 3. It consists of an amplitude sub-network and a phase sub-network that are arranged sequentially. Specifically, the amplitude sub-network learns to restore the amplitude representation to improve the lightness appearance, while the phase sub-network learns to reconstruct the phase representation that refines the structures. To guide the learning of these two sub-networks, in addition to the constraint of the ground truth, we supervise them with corresponding amplitude and phase components of the ground truth.

To further facilitate the representation learning of the amplitude and phase, we introduce a Spatial-Frequency Interaction (SFI) block (see Fig. 5). It is tailored in two formats (amplitude and phase) with two sub-networks as the basic units to learn the corresponding representation, and the SFI block of each format is composed of a frequency branch and a spatial branch to complement the global and local information. On the one hand, for the amplitude or phase sub-network, the amplitude/phase format of SFI processes the corresponding amplitude/phase component in the Fourier space and bypasses the other component. On the other hand, since the Fourier transform allows the image-wide receptive field to cover the whole image [11, 21], the frequency-domain representation focuses on global attributions. Meanwhile, the local attribution can be learned in the spatial branch by normal convolutions. To this end, we interact with these two branches to obtain the complementary information, which benefits the learning of corresponding representations.

Moreover, our proposed FECNet is lightweight and can be extended to other enhancement tasks like low-light image enhancement and retouching, showing its potential in wide-range applications. In summary, our contributions include:

1. We introduce a new perspective for exposure correction by restoring the representation of different components in the frequency domain. Particularly, we propose a Fourier-based Exposure Correction Network (FECNet) consisting of an amplitude sub-network and a phase sub-network, which restores the amplitude and phase representations that correspond to improving lightness and refining structures progressively.
2. Tailored with the learning of the amplitude and phase sub-networks, we design a Spatial-Frequency Interaction (SFI) block in two formats that correspond to the two sub-networks as their basic units. The interaction of spatial and frequency information helps integrate the global and local representations that provide complementary information.
3. Our FECNet is lightweight, and we validate its effectiveness on several datasets. Furthermore, we extend our method to other enhancement tasks, including low-light enhancement and retouching, which demonstrate its superiority ability in wide-range applications.

2 Related Work

2.1 Exposure Correction

Exposure correction has been studied for a long time. Several conventional methods apply histogram adjustment for correcting the lightness and contrast [36, 1, 33, 48]. Another line of works is based on the Retinex theory [22], which improves the lightness through enhancing the illumination component, and regularizes the reflectance component to recover the texture [16, 24, 35, 6, 49].

In recent years, deep learning-based methods have been developed for exposure correction [9, 27, 39, 26, 45]. Most exposure correction works are dedicated to enhancing underexposure images. Based on the Retinex theory, RetinexNet [42] and KinD [51] decompose the image into illumination and reflectance components and then restore them in a data-driven manner. As another form of component decomposition, DRBN [44] decomposes features into different band representations and then recursively recovers them. More recently, targeting at correcting both underexposure and overexposure images, MSEC [2] proposes to correct varieties of exposures with a pyramid structure to restore different-scale components in a coarse-to-fine manner. CMEC [30] employs an encoder to map different exposures to an exposure-invariant space with the assistance of a transformer for exposure correction. However, existing methods rarely consider correcting exposures by frequency-domain representations. Compared with these methods, our algorithm focuses on processing information in the Fourier space to recover frequency representations, which is a new perspective in this area.

2.2 Fourier Transform in Neural Networks

Recently, information processing in the Fourier space of frequency domain has attracted increasing attentions [11, 34, 43, 10, 37], which is capable of capturing global frequency representation effectively [21]. A line of works leverages the Fourier transform to improve the generalization of neural networks. For example, FDA [47] develops a data augmentation strategy by swapping the amplitude and phase components in the Fourier space across images, enabling the network to learn robust representations for image segmentation. Similarly, Xu *et al.* [43] proposed a Fourier-based augmentation strategy with the combing of a mix-up for generalized image classification. Another line of works employs the Fourier transform to improve the representation ability of neural networks. For instance, GFNet [34] attempts to transform features to the Fourier space before fully-connected layers to improve the network stability. FFC [10] introduces paired spatial-frequency transforms and devises several new layers in the Fourier space. Besides, a few works adopt Fourier-based loss functions for image restoration [13] and image translation [20], achieving pleasant visual results. Motivated by the success of these works, we propose a deep Fourier-based exposure correction network, which learns to recover different components of frequency representations.

3 Method

3.1 Motivation and Background

Images captured under improper exposures often suffer from unsatisfactory visual problems, including lightness and structure distortions. Previous works rarely restore these distortions in the frequency domain, which has been proved crucial for improving the visual qualities [13]. To this end, we design a deep Fourier-based exposure correction network to capture and restore the frequency representations effectively.

Firstly, we revisit the operation and property of the Fourier transform. Given a single channel image x with the shape of $H \times W$, the Fourier transform $\mathcal{F}$ converts to the Fourier space as a complex component X, which is expressed as:

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

and $\mathcal{F}^{-1}$ denotes the inverse Fourier transform. Since an image or feature may contain multiple channels, we separately apply Fourier transform to each channel in our work with the FFT [31].

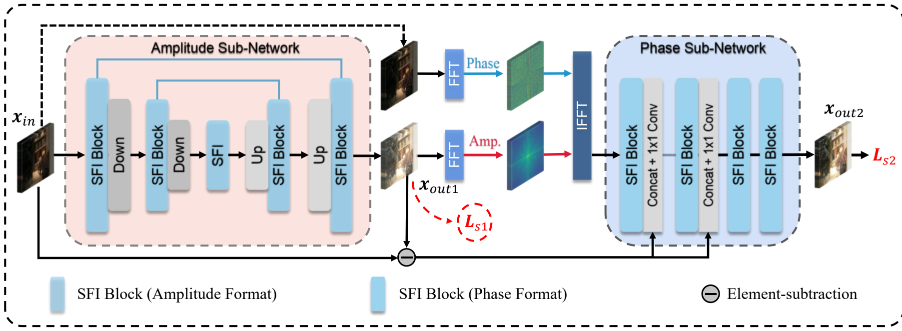


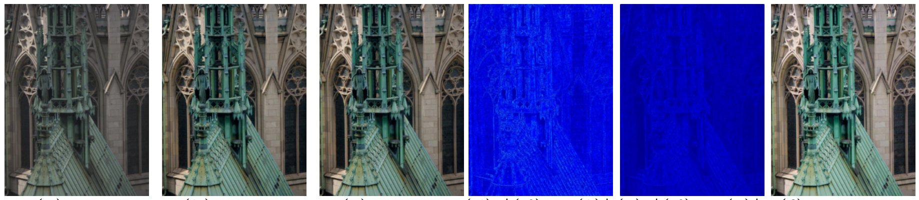
Fig. 3. The overview of our proposed FECNet consisting of an amplitude sub-network that restores the amplitude representation and a phase sub-network that restores the phase representation. The phase sub-network takes the recombined results of $\mathcal{F}^{-1}(\mathcal{A}(X_{out1}), \mathcal{P}(X_{in}))$ as the input, with the lightness changing residual of the amplitude sub-network to guide its learning. Both sub-networks employ the corresponding amplitude and phase components of the ground truth as the supervision signal, and the two formats of the SFI block are set as their basic units correspondingly.

In the Fourier space, each complex component $X(u, v)$ can be represented by the amplitude component $\mathcal{A}(X(u, v))$ and the phase component $\mathcal{P}(X(u, v))$, which provides an intuitive analysis of the frequency components [13]. These two components 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 parts of $X(u, v)$.

According to the Fourier theory, the amplitude component $\mathcal{A}$ reflects the style information of an image in the frequency domain, while the phase component $\mathcal{P}$ represents the semantic information [43, 47]. For exposure correction, we explore whether $\mathcal{A}$ and $\mathcal{P}$ could correspond to the frequency-domain representations of lightness and structure components. To visualize the amplitude and phase components, we swap the amplitude and phase components of different exposures of the same context, then we observe the phenomenon as shown in Fig. 1. Denoting the underexposure and overexposure image as x_{under} and x_{over}, and their Fourier representations as X_{under} and X_{over}, respectively. The recombined result $\mathcal{F}^{-1}(\mathcal{A}(X_{under}), \mathcal{P}(X_{over}))$ has similar lightness appearance with x_{under}, while $\mathcal{F}^{-1}(\mathcal{A}(X_{over}), \mathcal{P}(X_{under}))$ behaves conversely. Furthermore, we convert the amplitude (phase) components of x_{under} and x_{over} to spatial domain by replacing the amplitude (phase) components with a constant c, and we observe the phenomenon in Fig. 2. The appearance of the converted phase representation looks like more similar to structures than the converted amplitude one. Besides, the difference between the converted result $\mathcal{F}^{-1}(c, \mathcal{P}(X_{under}))$ and $\mathcal{F}^{-1}(c, \mathcal{P}(X_{over}))$ is small, while the difference between $\mathcal{F}^{-1}(\mathcal{A}(X_{under}), c)$ and $\mathcal{F}^{-1}(\mathcal{A}(X_{over}), c)$ is larger. It proves the phase component responds more to the structure information and is less affected by the lightness.



(a) x_{in} (b) x_{out1} (c) x_{out2} (d) $|(f)-(b)|$ (e) $|(f)-(c)|$ (f) x_{normal}

Fig. 4. Visualization of different components in FECNet. As can be seen, with the amplitude sub-network, the overall lightness representations are improved. After the processing of the phase sub-network, the structures are refined with lower residual error. $|\cdot|$ denotes the "absolute" operation. Darker areas in the residual map denote lower errors.

Based on the above observations, we can draw the conclusion that the amplitude component of an image reflects the lightness representation, while the phase component corresponds to the structures and is less affected by lightness. Following existing works that respectively restore degradations of lightness and structures, we restore the amplitude and phase components progressively, which facilitate the recovering the frequency representations of lightness and structures that benefit improving the image quality.

3.2 Deep Fourier-based Exposure Correction Network

Based on the above analysis, we design a simple but effective FECNet net as shown in Fig. 3. The entire network consists of two sub-networks: an amplitude sub-network and a phase sub-network, progressively restoring the amplitude and phase representations. Specifically, both sub-networks employ the SFI block as the basic unit, which will be described in Sec. 3.3.

We design an encoder-decoder format for the amplitude sub-network, consisting of five SFI blocks of its amplitude format. Let us denote x_{in} and x_{out1} as the input and output of the amplitude sub-network, x_{normal} represents the ground truth normal exposure image, and their representations in Fourier space are denoted as X_{in}, X_{out1} and X_{normal}, respectively. To guarantee this sub-network learns the amplitude representation, it is supervised by the recombined component $\mathcal{F}^{-1}(\mathcal{A}(X_{normal}), \mathcal{P}(X_{in}))$, as well as the amplitude component of the ground truth $\mathcal{A}(X_{normal})$. The loss function for this sub-network $\mathcal{L}_{s1}$ is expressed as:

$$L_{s1} = ||x_{out1} - \mathcal{F}^{-1}(\mathcal{A}(X_{normal}), \mathcal{P}(X_{in}))||_1 + \alpha||\mathcal{A}(X_{out1}) - \mathcal{A}(X_{normal})||_1, \quad (3)$$

where $||\cdot||_1$ denotes the mean absolutely error, α is the weight factor and we set it as 0.2.

While for the phase sub-network, we formulate it sequentially with four SFI blocks of its phase format. Specifically, we use the recombined component $\mathcal{F}^{-1}(\mathcal{A}(X_{out1}), \mathcal{P}(X_{in}))$ as the input of this sub-network instead of x_{out1}, avoiding introduce the altered phase component [47]. In addition, since the distortion of structures are relevant to the lightness changing [22, 24], and the residual

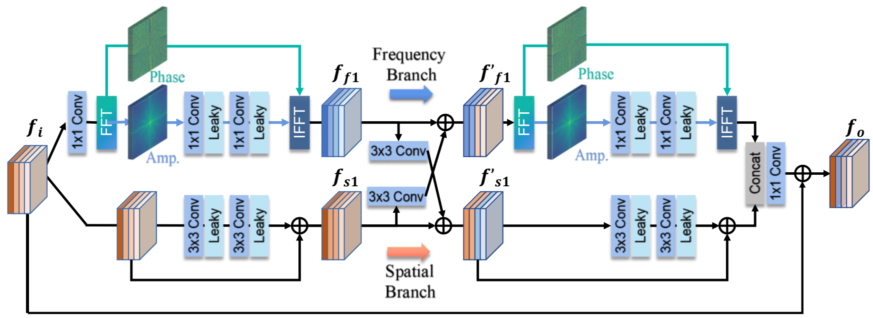


Fig. 5. The illustration of the amplitude format of the SFI block, which consists of a frequency branch and a spatial branch. The frequency branch processes the amplitude component and bypasses the phase component, while the spatial one utilizes a residual block. There exist interactions across representations of these two branches for complementary information. The phase format of the SFI block is similar except for the frequency branch, and we illustrate it in the supplementary material.

of the amplitude sub-network can represent the lightness changing, we utilize the residual between x_{out1} and x_{in} to guide the learning of this sub-network. It is implemented by concatenating this residual with the features in the phase sub-network, following by a 1×1 convolution to integrate them. Denoting the output of the phase sub-network as x_{out2}, we set the loss function L_{s2} for this sub-network as:

$$L_{s2} = ||x_{out2} - x_{normal}||_1 + \beta||\mathcal{P}(X_{out2}) - \mathcal{P}(X_{normal})||_1, \quad (4)$$

where β is the weight factor and we set it as 0.1. The phase sub-network can learn the recovery of the phase representation, and x_{out2} is the final output of FECNet. In this way, the FECNet is able to conduct exposure correction in a coarse to fine manner as shown in Fig. 4.

The overall network comprised of these two sub-networks is training in an end-to-end manner, and the overall loss L_{total} is the combination of L_{s1} and L_{s2}, which is formulated as:

$$L_{total} = L_{s2} + \lambda L_{s1}, \quad (5)$$

where λ is the weight factor and is empirically set as 0.5.

3.3 Spatial-Frequency Interaction Block

To further facilitate learning the amplitude and phase representations, we propose the SFC block in two formats as the basic unit of the two sub-networks correspondingly. According to Fourier theory [21], processing information in Fourier space is capable of capturing the global frequency representation in the frequency domain. In contrast, the normal convolution focuses on learning local representations in the spatial domain. In this way, we propose the interactive block to combine these two representations, which can learn more representative features.

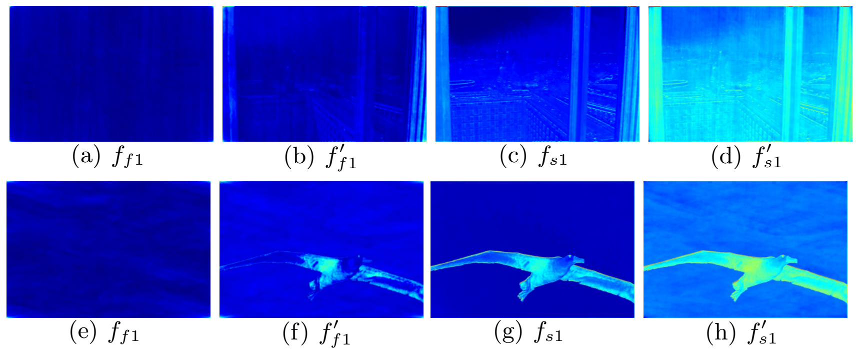


Fig. 6. Feature visualization of different representations in the SFI block. As can be seen, since features after interaction can obtain complementary representations from each other, features with and without interaction across the frequency branch and spatial branch are quite different. f_{f1} is more spatial invariant and f_{s1} keeps more spatial information, while f'_{f1} obtains the spatial information and the details in f'_{s1} are enhanced.

We illustrate the amplitude format of the SFI block as shown in Fig. 5. Specifically, it comprises a spatial branch and a frequency branch for processing spatial and frequency representations. Denoting f_i as the input features of SFI block, the spatial branch first adopts a residual block with 3×3 convolution layers to process information in the spatial domain and obtain f_{s1}. While the frequency branch uses a 1×1 convolution to process f_i first that obtains f_{f0}, and then adopts Fourier transform to convert it to the Fourier space as F_{f0} by Eq. 1. To process frequency-domain representation F_{f0}, we adopt the operation $Op(\cdot)$ that consists of 1×1 convolution layers on its amplitude component, and then recompose the operated result with the phase component that obtain f_{f1}, which is expressed as:

$$f_{f1} = \mathcal{F}^{-1}(Op(\mathcal{A}(F_{f0})), \mathcal{P}(F_{f0})). \tag{6}$$

Thus, f_{f1} is the processed result of the frequency-domain representation. Next, we interact the features from spatial branch f_{s1} and frequency branch f_{f1} as:

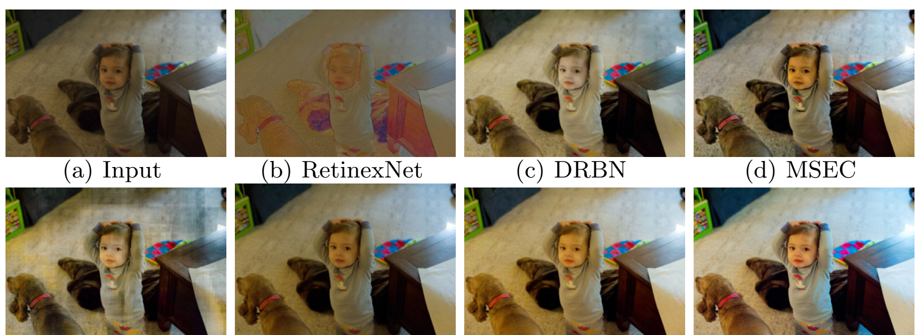
$$\begin{aligned} f'_{s1} &= f_{s1} + W_1(f_{f1}), \\ f'_{f1} &= f_{f1} + W_2(f_{s1}), \end{aligned} \tag{7}$$

where both $W_1(\cdot)$ and $W_2(\cdot)$ denote the 3×3 convolution operation, f'_{s1} and f'_{f1} are the output of the interacted spatial branch and frequency branch. As illustrated in Fig. 6, both f'_{s1} and f'_{f1} get the complementary representation, which benefits for these two branches to obtain more representational features. The following spatial and frequency branches are formulated in the same way as above and output the results f_{s2} and f_{f2}, respectively.

Finally, we concatenate f_{s2} and f_{f2} and then apply a 1×1 convolution operation to integrate them as f_o, which is the output of SFI block. Similarly, in the phase format of the SFI block, we replace the operation on the amplitude component in Eq. 6 with the phase component, while other parts keep unchanged.

Table 1. Quantitative results of different methods on the ME and SICE datasets in terms of PSNR and SSIM. #Param denotes the number of parameters.

Method	ME						SICE						#Param
	Under		Over		Average		Under		Over		Average		
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
CLAHE [36]	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 [42]	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
Zero-DCE [15]	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
DPED [19]	13.14	0.5812	20.06	0.6826	15.91	0.6219	16.83	0.6133	7.99	0.4300	12.41	0.5217	0.39M
DRBN [44]	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
SID [8]	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
RUAS [38]	13.43	0.6807	6.39	0.4655	9.20	0.5515	16.63	0.5589	4.54	0.3196	10.59	0.4393	**0.003M**
MSEC [2]	20.52	0.8129	19.79	0.8156	20.35	0.8210	19.62	0.6512	17.59	0.6560	18.58	0.6536	7.04M
CMEC [30]	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
FECNet (Ours)	**22.96**	**0.8598**	**23.22**	**0.8748**	**23.12**	**0.8688**	**22.01**	0.6737	**19.91**	**0.6961**	**20.96**	**0.6849**	0.15M



(a) Input (b) RetinexNet (c) DRBN (d) MSEC
(e) SID (f) CMEC (g) Ours (h) GT

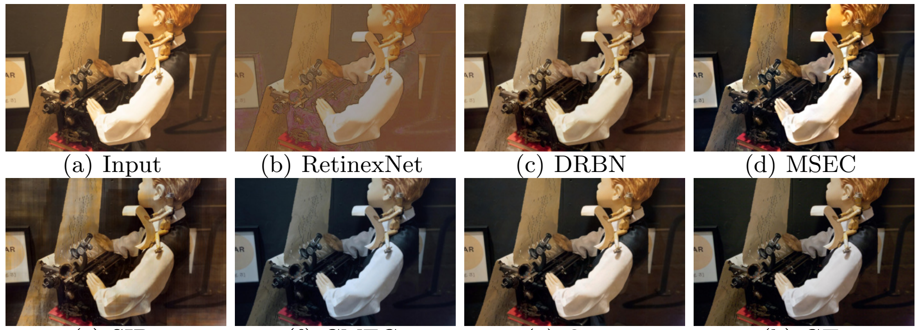
Fig. 7. Visualization results on the ME dataset of underexposure correction. There exist color and lightness shift as well as artifact generation problems in other methods, while our method can simultaneously achieve good context and lightness recovery.

4 Experiment

4.1 Settings

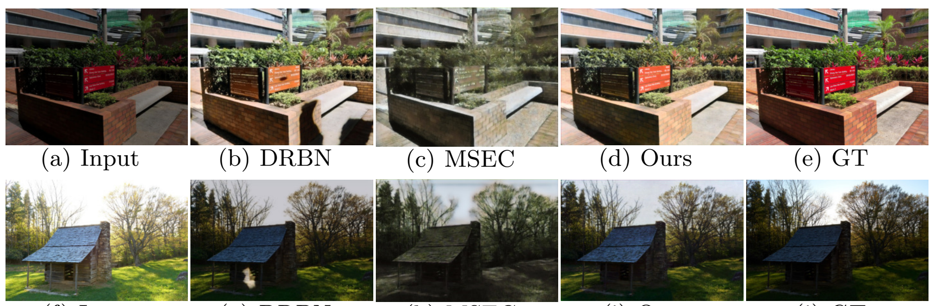
Datasets. We train our network on two representative multiple exposure datasets, including the multiple exposure (ME) dataset proposed in MSEC [2] and SICE dataset [7]. The ME dataset contains exposure images of 5 exposure levels, including 17675 images for training, 750 images for validation, and 5905 images for testing. For the SICE dataset, we derive the middle-level exposure subset as the ground truth and the corresponding second and last-second exposure subsets are set as underexposed and overexposed images, respectively. We adopt 1000 images for training, 24 images for validation and 60 images for testing respectively.

Implementation Details. The implement of our proposed method is based on PyTorch framework with one NVIDIA 3090 GPU. During the training, we adopt the Adam optimizer with the patch size of 384×384 and batch size of 4. For the ME and SICE datasets, the total number of epochs is set as 120 and 240, respectively. The initial learning rate of our FECNet is $1e^{-4}$, which decays



(a) Input (b) RetinexNet (c) DRBN (d) MSEC
(e) SID (f) CMEC (g) Ours (h) GT

Fig. 8. Visualization results on the ME dataset of overexposure correction. As can be seen, the context and lightness can be well recovered in our method.



(a) Input (b) DRBN (c) MSEC (d) Ours (e) GT
(f) Input (g) DRBN (h) MSEC (i) Ours (j) GT

Fig. 9. Visualization results on the SICE dataset of (top) underexposure correction and (bottom) overexposure correction.

by a factor value of 0.5 every 40 epochs and 80 epochs for the ME and SICE datasets. We adopt the commonly used metrics PSNR and SSIM for evaluation.

Table 2. Ablation study of investigating different settings of FECNet on the SICE dataset.

Option	(a)	(b)	(c)	(d)	(e)	(f)	FECNet
PSNR	19.98	20.03	19.35	20.78	20.67	20.77	20.96
SSIM	0.6698	0.6712	0.6643	0.6795	0.6773	0.6809	0.6849

4.2 Performance Evaluation

In this paper, we compare our algorithm with several state-of-the-art exposure correction methods, including MSEC [2], DRBN [44], SID [8], RetinexNet [42], Zero-DCE [15], CMEC [30] and RUAS [38]. We provide more comparison results with other methods in the supplementary material.

Quantitative Evaluation. The quantitative results are shown in Table 1. For the ME dataset, following MSEC, we average the results of the exposures of the first two levels and the remaining levels of exposures as the underexposure and overexposure results, respectively. As can be observed, our method achieves the

Table 3. Ablation study of investigating the loss functions on the SICE dataset. (a) denotes removing the second term of L_{s1} on the base of L_{total}.

Options	Baseline	L_{s2}	Baseline+L_{s1}	(a)	L_{total}
PSNR/SSIM	20.01/0.6682	20.03/0.6713	20.83/0.6827	20.43/0.6785	20.96/0.6849

Table 4. Ablation study of investigating the SFI block on the SICE dataset. 2-SPB represents both branches are set to the spatial branches, and 2-FRB represents both branches are set to the frequency branches.

Option	2-SPB	2-FRB	w/o Interaction	SFI block
PSNR/SSIM	18.57/0.6593	18.64/0.6602	19.82/0.6676	20.96/0.6849

best performance among these methods. Specifically, MSEC significantly outperforms other methods except ours due to its well-designed architecture, while our FECNet has superior results than MSEC using its 2.1% network parameters, demonstrating the effectiveness and efficiency of our methods.

Qualitative Evaluation. In addition, we provide the visualization results of the ME dataset in Fig. 7 and Fig. 8, and the results of the SICE dataset in Fig. 9, respectively. It can be seen that our FECNet produces the more pleasing results with corrected lightness and color appearance while maintaining the detailed structures. We provide more visualization results in the supplementary material.

4.3 Ablation Studies

In this section, we conduct the experiments to demonstrate the effectiveness of our method. More ablation studies are provided in supplementary materials.

Investigation of FECNet. To demonstrate the effectiveness of the overall setting of FECNet, we set several settings as ablations and present the results in Table 2. Particularly, (a) denotes removing the amplitude sub-network in FECNet; (b) represents removing the phase sub-network in FECNet; (c) denotes recovering the phase representation first and then restoring the amplitude representation; (d) represents swapping the two formats of SFI block in the two sub-networks; (e) denotes replacing the input of the phase sub-network with the output of the amplitude sub-network; (f) represents removing the lightness residual guidance for the phase sub-network.

As can be seen, both amplitude and phase sub-networks are effective for exposure correction, and the sequential order of arranging them are important, demonstrating the reasonableness of recovering the amplitude component first and then refine the phase component. In addition, the two formats of SFI block are proved to be coupled with these two sub-networks. For the input of the phase sub-network, the recombination with the phase component of the original input is more effective than the amplitude sub-network output. The residual map of the amplitude sub-net also helps improve performance.

Investigation of Losses. To validate the effectiveness of loss functions, we conduct experiments with different losses. The baseline set the L_1 loss on the final output, and we present results in Table 3. As can be seen, without the constraint of L_{s1}, the performance drops significantly, while the amplitude constraint and

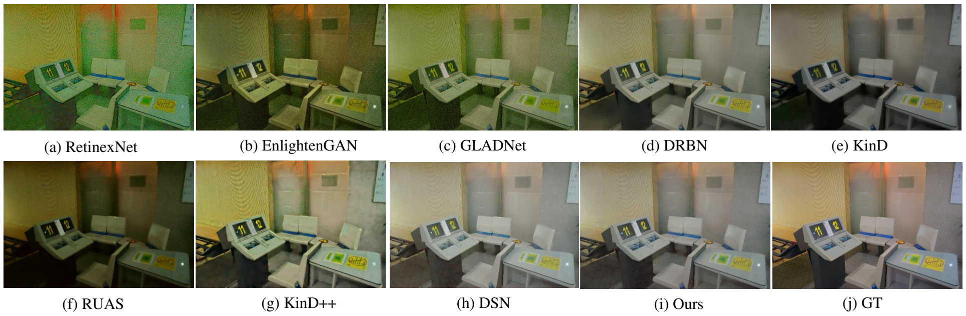


Fig. 10. Visualization results on the LOL dataset.

Table 5. Quantitative results of different methods on the LOL dataset in terms of PSNR and SSIM. #Param denotes the number of parameters.

Method	LIME	RetinexNet	MBLLEN	EnGAN	GLADNet	KinD	DRBN	RUAS	KinD++	DSN	FECNet (Ours)
PSNR	17.18	16.77	17.56	17.48	19.72	20.38	18.65	16.41	21.80	22.04	**23.44**
SSIM	0.5621	0.4249	0.7293	0.6737	0.6803	0.8248	0.8008	0.5001	0.8285	0.8334	**0.8383**
#Param	-	0.84M	0.45M	8.37M	1.13M	8.54M	0.58M	**0.003M**	8.23M	4.42M	0.15M

phase constraint in L_{s1} and L_{s2} are also proved to be effective, demonstrating the reasonableness of the supervision manner.

Investigation of SFI block. We validate the effectiveness of the design of the SFI block in Table 4. As can be seen, both replacing the spatial branch with frequency branch or replacing the frequency branch with spatial branch results in a significant performance drop. While interacting these two branches can further improve performance remarkably, demonstrating the effectiveness of integrating these two complementary representations.

4.4 Extensions on Other Image Enhancement Tasks

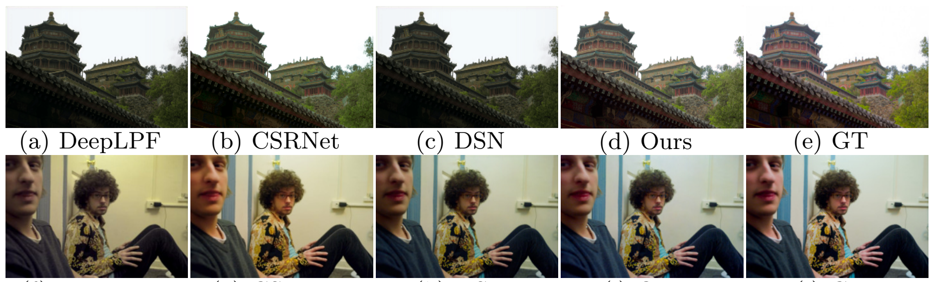
To demonstrate the potential of our FECNet, we extend it to other image enhancement tasks, including low-light image enhancement and image retouching.

Extension on low-light image enhancement. Low-light image enhancement mainly focuses on lighting the darkness of a scene and removing the amplified noise. We adopt LOL dataset [42] to train and evaluate different methods, consisting of 485 images for training and 15 images for testing. Several low-light image enhancement methods are selected for comparison: LIME [16], RetinexNet [42], MBLLEN [12], DRBN [44], KinD [51], GLADNet [3], EnGAN [4], RUAS [38], KinD++ [50] and DSN [52]. The quantitative and qualitative results are shown in Table 5 and Fig 10, respectively. As can be seen, our FECNet achieves the best performance both quantitatively and qualitatively.

Extension on image retouching. Image retouching aims to improve the color and lightness of an image to the expert manipulated effect. In this task, we apply the MIT-FiveK dataset [5] that is adopted by CSRNet [17], which contains 4500 images for training and 500 images for testing. Specifically, we compare our FECNet with several methods, including CSRNet [17], HDRNet [14] DUPE [40], Distort-Recover [32], White-box [18], DeepLPF [29] and DSN [52].

Table 6. Quantitative results of different methods on the MIT-FiveK dataset in terms of PSNR and SSIM. #Param denotes the number of parameters.

Method	White-Box	Distort-Recover	HDRNet	DUPE	DeepLPF	CSRNet	DSN	FECNet (Ours)
PSNR	18.59	19.54	22.65	20.22	23.21	23.69	23.84	**24.18**
SSIM	0.7973	0.7998	0.8802	0.8287	0.8863	0.8951	0.9002	**0.9030**
#Param	8.17M	247.25M	0.46M	0.95M	0.80M	**0.034M**	4.42M	0.15M



(a) DeepLPF (b) CSRNet (c) DSN (d) Ours (e) GT
(f) DeepLPF (g) CSRNet (h) DSN (i) Ours (j) GT

Fig. 11. Visualization results on the MIT-FiveK dataset. Images processed by other methods exist color and lightness shift and the details cannot be well recovered, while our method can obtain better visual qualities.

We give the quantitative evaluation in Table 6, and present the visualization results in Fig. 11. It can be seen that the generated results of our FECNet achieves the best performance with high quantitative performance and visual effects.

5 Conclusion

In this paper, we develop a new perspective for exposure correction with spatial-frequency information interaction in the spatial and frequency domain. We propose a deep Fourier-based Exposure Correction network (FECNet), which consists of two sub-networks: amplitude sub-network and phase sub-network. Specifically, the former aims to restore the amplitude, thus improving the lightness, while the latter is responsible for phase reconstruction, corresponding to refining structures. We further design a Spatial-Frequency Interaction (SFI) block as the basic unit of the FECNet to facilitate the learning of these two components with complementary representations. Extensive experimental results show that our method achieves superior performance for exposure correction. Moreover, the proposed approach can be extended to other image enhancement tasks, demonstrating its potential usage in wide-range applications. Although there exists color shift problem in some cases, we believe that the dynamic mechanism could be leveraged to relieve this issue. Considering that the mainstream of related works is still based on the spatial domain, we hope that the validity of our work will provide some insights into this community.

Acknowledgments. This work was supported by the Anhui Provincial Natural Science Foundation under Grant 2108085UD12. We acknowledge the support of GPU cluster built by MCC Lab of Information Science and Technology Institution, USTC.

References

1. Abdullah-Al-Wadud, M., Kabir, M.H., Akber Dewan, M.A., Chae, O.: A dynamic histogram equalization for image contrast enhancement. IEEE Transactions on Consumer Electronics **53**(2), 593–600 (2007)
2. Afifi, M., Derpanis, K.G., Ommer, B., Brown, M.S.: Learning multi-scale photo exposure correction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
3. Wang et al., W.: Gladnet: Low-light enhancement network with global awareness. In: Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition (FG) (2018)
4. Jiang et al., Y.: Enlightengan: Deep light enhancement without paired supervision. IEEE Transactions on Image Processing (TIP) (2021)
5. Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input output image pairs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2011)
6. Cai, B., Xu, X., Guo, K., Jia, K., Hu, B., Tao, D.: A joint intrinsic-extrinsic prior model for retinex. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4000–4009 (2017)
7. Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. IEEE Transactions on Image Processing (TIP) **27**(4), 2049–2062 (2018)
8. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. arXiv preprint arXiv:1805.01934 (2018)
9. Chen, Y.S., Wang, Y.C., Kao, M.H., Chuang, Y.Y.: Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6306–6314 (2018)
10. Chi, L., Jiang, B., Mu, Y.: Fast fourier convolution. Advances in Neural Information Processing Systems (NIPS) **33**, 4479–4488 (2020)
11. Chi, L., Tian, G., Mu, Y., Xie, L., Tian, Q.: Fast non-local neural networks with spectral residual learning. In: Proceedings of the 27th ACM International Conference on Multimedia (MM). pp. 2142–2151 (2019)
12. Feifan Lv, Feng Lu, J.W.C.L.: Mbllen: Low-light image/video enhancement using CNNs. In: Proceedings of the The British Machine Vision Conference (BMVC) (2018)
13. Fuoli, D., Van Gool, L., Timofte, R.: Fourier space losses for efficient perceptual image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 2360–2369 (2021)
14. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. ACM Transactions on Graphics (TOG) **36**(4), 118 (2017)
15. Guo, C.G., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1780–1789 (2020)
16. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing (TIP) **26**(2), 982–993 (2016)

17. He, J., Liu, Y., Qiao, Y., Dong, C.: Conditional sequential modulation for efficient global image retouching. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 679–695. Springer (2020)
18. Hu, Y., He, H., Xu, C., Wang, B., Lin, S.: Exposure: A white-box photo post-processing framework. ACM Transactions on Graphics (TOG) **37**(2), 1–17 (2018)
19. Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L.: Dslr-quality photos on mobile devices with deep convolutional networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3277–3285 (2017)
20. Jiang, L., Dai, B., Wu, W., Loy, C.C.: Focal frequency loss for image reconstruction and synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13919–13929 (2021)
21. Katznelson, Y.: An introduction to harmonic analysis. Cambridge University Press (2004)
22. Land, E.H.: The retinex theory of color vision. Scientific American **237**(6), 108–129 (1977)
23. Li, J., Li, J., Fang, F., Li, F., Zhang, G.: Luminance-aware pyramid network for low-light image enhancement. IEEE Transactions on Multimedia (TMM) pp. 1–1 (2020)
24. Li, M., Liu, J., Yang, W., Sun, X., Guo, Z.: Structure-revealing low-light image enhancement via robust retinex model. IEEE Transactions on Image Processing (TIP) **27**(6), 2828–2841 (2018)
25. Lim, S., Kim, W.: Dslr: Deep stacked laplacian restorer for low-light image enhancement. IEEE Transactions on Multimedia (TMM) pp. 1–1 (2020)
26. Liu, J., Xu, D., Yang, W., Fan, M., Huang, H.: Benchmarking low-light image enhancement and beyond. International Journal of Computer Vision (IJCV) **129** (2021)
27. Lv, F., Li, Y., Lu, F.: Attention guided low-light image enhancement with a large scale low-light simulation dataset. International Journal of Computer Vision (IJCV) **129** (2021)
28. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research (JMLR) **9**(11) (2008)
29. Moran, S., Marza, P., McDonagh, S., Parisot, S., Slabaugh, G.: Deeplpf: Deep local parametric filters for image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
30. Nsamp, N.E., Hu, Z., Wang, Q.: Learning exposure correction via consistency modeling. In: Proceedings of the British Machine Vision Conference (BMVC). pp. 1–12 (2018)
31. Nussbaumer, H.J.: The fast fourier transform. In: Fast Fourier Transform and Convolution Algorithms, pp. 80–111. Springer (1981)
32. Park, J., Lee, J.Y., Yoo, D., Kweon, I.S.: Distort-and-recover: Color enhancement using deep reinforcement learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5928–5936 (2018)
33. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer Vision, Graphics, and Image Processing **39**(3), 355–368 (1987)
34. Rao, Y., Zhao, W., Zhu, Z., Lu, J., Zhou, J.: Global filter networks for image classification. Advances in Neural Information Processing Systems (NIPS) **34** (2021)

35. Ren, X., Yang, W., Cheng, W.H., Liu, J.: Lr3m: Robust low-light enhancement via low-rank regularized retinex model. IEEE Transactions on Image Processing (TIP) **29**, 5862–5876 (2020)
36. Reza, A.M.: Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. IEEE Transactions on Multimedia (TMM) pp. 35–44 (2004)
37. Rippel, O., Snoek, J., Adams, R.P.: Spectral representations for convolutional neural networks. Advances in neural information processing systems (NIPS) **28** (2015)
38. Risheng, L., Long, M., Jiaao, Z., Xin, F., Zhongxuan, L.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
39. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6849–6857 (2019)
40. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6849–6857 (2019)
41. Wang, W., Yang, W., Liu, J.: Hla-face: Joint high-low adaptation for low light face detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
42. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: Proceedings of the British Machine Vision Conference (BMVC). pp. 155–165 (2018)
43. Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.: A fourier-based framework for domain generalization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14383–14392 (2021)
44. Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3063–3072 (2020)
45. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Transactions on Image Processing (TIP) **30**, 2072–2086 (2021)
46. Yang, W., Yuan, Y., Ren, W., Liu, J., Scheirer, W.J., Wang, Z., Zhang, et al.: Advancing image understanding in poor visibility environments: A collective benchmark study. IEEE Transactions on Image Processing (TIP) **29**, 5737–5752 (2020)
47. Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4085–4095 (2020)
48. Ying, Z., Li, G., Ren, Y., Wang, R., Wang, W.: A new image contrast enhancement algorithm using exposure fusion framework. In: ICCP. pp. 36–46. Springer (2017)
49. Zhang, Q., Yuan, G., Xiao, C., Zhu, L., Zheng, W.S.: High-quality exposure correction of underexposed photos. In: ACM International Conference on Multimedia (ACM MM). p. 582–590 (2018)
50. Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J.: Beyond brightening low-light images. International Journal of Computer Vision (IJCV) **129** (2021)

51. Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: A practical low-light image enhancer. In: ACM International Conference on Multimedia (ACM MM). pp. 1632–1640 (2019)
52. Zhao, L., Lu, S.P., Chen, T., Yang, Z., Shamir, A.: Deep symmetric network for underexposed image enhancement with recurrent attentional learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 12075–12084 (2021)