Frequency-Consistent Optimization for Image Enhancement Networks

Bing Li*
bing0123@mail.ustc.edu.cn
Naishan Zheng*
nszheng@mail.ustc.edu.cn
Qi Zhu
zqcrafts@mail.ustc.edu.cn
Jie Huang
hj0117@mail.ustc.edu.cn
Feng Zhao ✉
fzhao956@ustc.edu.cn

Department of Automation
University of Science and Technology of China
Hefei 230027, China

Abstract

Image enhancement aims at enhancing the overall contrast (low frequency) while reconstructing details (high frequency). Existing studies typically achieve these two objectives with a heuristically constructed complex architecture (*i.e.*, two-stage or two-branch). In contrast, we attempt to perform the image enhancement task within a single-stage and single-branch network. However, directly employing a single plain network to optimize the two objectives simultaneously will lead to an optimization conflict between contrast enhancement and texture restoration, resulting in suboptimal performances. To alleviate this problem, we construct a frequency-independent feature space for maintaining optimization consistency. Specifically, we propose a Frequency Decorrelation and Integration (FDI) module with two core insights: 1) formulating a frequency-independent space via decorrelation normalization to bridge the frequency discrimination; 2) integrating the initial frequency-dependent features with a channel shuffle operation for information complement and reducing the sensitivity to frequency during optimization. Therefore, these two designs encourage networks to learn along the optimization direction of frequency consistency. In addition, the proposed FDI is a plug-and-play module that can be incorporated into the existing methods with negligible parameters. Extensive experiments on various image enhancement benchmarks demonstrate consistent performance gains by utilizing our proposed module.

1 Introduction

The rapid advancement of smartphones has facilitated access to photography. However, the captured photographs will inevitably suffer from low contrast and serious noise pollution due to insufficient illumination, imaging equipment, and other factors. These degradations

* Equal contribution. ✉ Corresponding author.

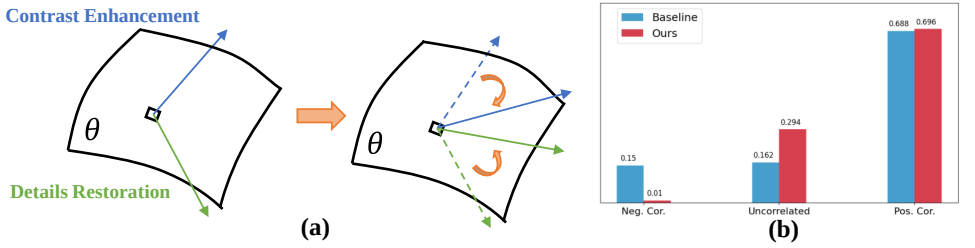


Figure 1: The motivation of our method. (a) We point out that there is an optimization inconsistency between contrast enhancement and details restoration in the image enhancement task. To alleviate it, we formulate a frequency-independent feature space with the proposed Frequency Decorrelation and Integration (FDI) module; and (b) compared to the baseline, incorporating our proposed FDI mitigates the ratio of negative correlation while maintaining that of positive correlation.

not only reduce the visual quality but also burden the performance in the downstream recognition tasks (*e.g.*, classification, image detection, and segmentation). Therefore, image enhancement is urgently required to improve the visual quality of photos by enhancing contrast, correcting color, and reconstructing details. The current image enhancement methods can be coarsely divided into two categories: conventional methods and learning-based methods.

Conventional enhancement methods rely on hand-crafted priors to constrain the solution space of latent high-quality images. For example, Dong *et al*. [6] developed a bright channel prior, and Guo *et al*. [8] designed a structure prior to refine the illumination map. However, these manually designed priors have limited generalization ability, hindering the practical application of the corresponding algorithms.

Recently, deep learning-based methods have achieved remarkable progress in the image enhancement community. The representative approach is attributed to the Retinex theory [16], which assumes that the observed image can be decomposed into illumination and reflectance components. Following this, various works propose to enhance contrast (low frequency) and reconstruct details (high frequency) in these two terms, respectively. RetinexNet [23] develops a two-stage training framework to decompose the low-light image and enhance the illumination component sequentially. KinD [31] further introduces a sub-network to recover the reflectance component for detail reconstruction. RUAS [19] first designs a bi-level optimization model adhering to the Retinex theory and then establishes a holistic propagation structure by unrolling the optimization process. Undeniably, with the advanced design of the training strategies and architectures, recent methodologies have achieved significant success in the image enhancement task. However, the complex and specific designs result in unsatisfactory flexibility and transferability of these algorithms. Furthermore, employing two-stage methodologies contributes to the accumulation of errors, subsequently leading to suboptimal outcomes.

In this paper, we attempt to explore the potential of a simple single-stage and single-branch network on contrast enhancement and detail reconstruction simultaneously. Unfortunately, since the optimization of contrast enhancement (low frequency) and details reconstruction (high frequency) is inconsistent (see Fig. 1) during the training process [25], directly employing a single plain network to achieve the above two objectives simultaneously will lead to suboptimal performances. We therefore wonder, "Is it possible to alleviate the adverse effects of such optimization inconsistency within a straightforward network?"

To achieve this goal, as described in Fig. 1, we formulate a frequency-independent feature space where the optimization inconsistency between frequencies will be eliminated. To this end, we design a Frequency Decorrelation and Integration (FDI) module, as shown in Fig. 2. It includes a frequency decorrelation part and an integration part. Specifically, the frequency decorrelation component initially produces a high-frequency counterpart from the original features, and then removes the correlation between two features with distinct frequencies via a decorrelation normalization for generating the target feature space. Since the normalization operation will result in the loss of the image discriminative features for image enhancement, we introduce an integration component to aggregate the normalized features with the original frequency-dependent features for information complement. To further reduce the network sensitivity to frequency during optimization, a channel shuffle operation is employed to randomly disturb the relationship between the frequency-independent and -dependent features. Such designs guarantee that networks are optimized in a frequency-consistent manner.

To highlight, the proposed FDI can be used as a plug-and-play module for existing enhancement networks with negligible parameters. Extensive experiments on diverse image enhancement datasets have demonstrated the flexibility and effectiveness of our proposed approach. Our contributions are summarized as follows:

- We point out the optimization inconsistency between contrast enhancement and texture restoration in image enhancement. To this end, we construct a frequency-independent feature space with a Frequency Decorrelation and Integration (FDI) module.
- Within the FDI module, we design a frequency decorrelation part for mapping features to a frequency-independent space, and an integration part for reducing the sensitivity to the frequency during the optimization process.
- Our FDI module is general and can be integrated into the existing image enhancement methods with negligible parameters. Extensive experiments demonstrate consistent performance gains by introducing our proposed module.

2 Related Work

2.1 Conventional Methods

Image enhancement has been studied for a long time. The earliest methods mainly rely on hand-crafted operations or filters to enhance images. Histogram Equalization-based methods [13, 17] enhance the light by expanding the dynamic range of an image. Gamma Correction uses the non-linear function to enhance contrast [2]. However, these methods ignore the connection between pixels and tend to yield unnatural results. In addition, several methods leverage prior knowledge to assist enhancement. Dong *et al*. [6] proposed an enhancement algorithm using a bright channel prior. Guo *et al*. [8] refined the initial coarse illumination map by imposing a structure prior. Nevertheless, these methods have limited the representation capacity in complex real-world scenes, hindering their practical applications.

2.2 Learning-based Methods

Recently, with the emergence and development of deep learning, image enhancement task has benefited from the deep learning models. Inspired by the Retinex theory [16] that assumes that images can be decomposed into illumination and reflectance, RetinexNet [23]

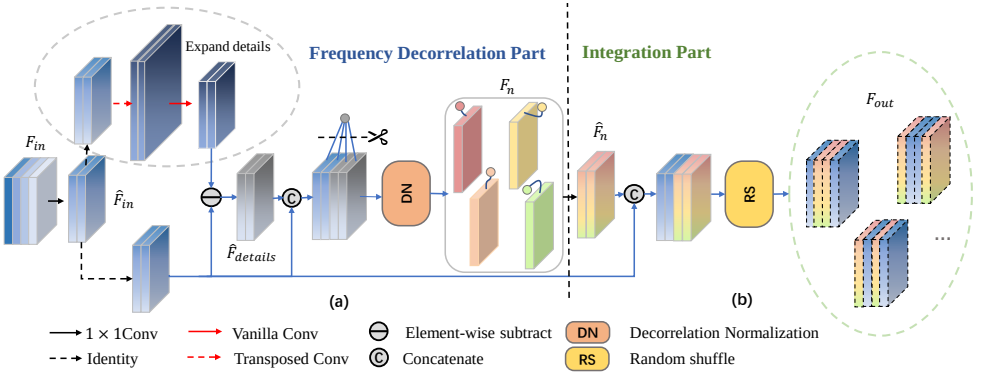


Figure 2: The overview of our proposed FDI module, which formulates a frequency-independent feature space with (a) a frequency decorrelation part and (b) an integration part. The frequency decorrelation part maps frequency-dependent features F_{in} to the frequency-independent features F_n. The integration part compensates for the information loss caused by decorrelation normalization, which introduces a random channel shuffle operator to reduce the sensitivity to frequency during the training process.

develops a two-stage training framework to sequentially decompose the inputs and restore the illumination (low frequency). KinD [31] further introduces an additional sub-network for reflectance recovery (high frequency). RUAS [19] first designs optimization models based on the Retinex rule and then establishes the holistic propagation structure by unrolling the optimization process. As another form of component decomposition, DRBN [26] decomposes features into different bands in the first stage and then recomposes them towards fitting visual properties in the second stage. More recently, Zhao *et al.* [33] employed the invertible neural network to perform bidirectional feature learning between the low-light and normal-light images. Xu *et al.* [24] proposed a signal-to-noise-aware framework that simultaneously adopts a transformer structure and a convolutional model for achieving spatial-varying enhancement with an SNR prior. Undeniably, these algorithms have achieved significant success in image enhancement. However, they rely on the complicated designs of the training strategy or architecture and do not explore the potential of simpler architecture. As opposed to them, we hope to achieve image enhancement within a more straightforward architecture.

3 Method

3.1 Overview

Images captured in underexposure scenes inevitably suffer from degradation, including low contrast and loss of details. Since the optimization of contrast enhancement (*i.e.*, low frequency) and details reconstruction (*i.e.*, high frequency) is inconsistent during the training process, directly employing a plain network to simultaneously achieve these two objectives will result in suboptimal performance. To eliminate the inconsistency between different frequencies, we design a Frequency Decorrelation and Integration (FDI) module to formulate a frequency-independent feature space for consistent optimization across frequencies. As

described in Fig. 2, we implement our FDI module with two parts: a frequency decorrelation part maps distinct frequency features to frequency-independent space, and an integration part is introduced to assimilate the original unprocessed features for compensating the loss of image discriminative information caused by decorrelation. To highlight, the FDI module can be used as a plug-and-play module with negligible parameters to assist existing approaches training within a frequency-consistent optimization.

3.2 Frequency Decorrelation Part

Given an input feature $F_{in} \in \mathcal{R}^{C\times H\times W}$ (C denotes the number of channels, H, W denotes the feature height, width.), the frequency decorrelation part first employ a 1×1 convolution to reduce the channel number of F_{in} to $\frac{C}{2}$, which can be formulated as:

$$\hat{F}_{in} = Conv_{1\times 1}(F_{in}). \tag{1}$$

Then we extract the high-frequency related feature from the $\hat{F}_{in}$. Specifically, we expand the initial resolution feature to a higher resolution with $2H\times 2W$ spatial dimension by a transposed convolution with 2 strides and 3×3 kernel. Then, we employ a vanilla convolution (stride is 2 and kernel size is 3) to project the high-resolution feature to the same spatial dimension as $\hat{F}_{in}$, and the difference between the projected feature $\hat{F}_{down}$ and initial $\hat{F}_{in}$ is the high-frequency related feature for details reconstruction. Every convolution layer is followed by a nonlinear ReLU. The above operations can be expressed as follows:

$$\begin{aligned} \hat{F}_{up} &= f_{up}(\hat{F}_{in}) = ReLU(TransConv(\hat{F}_{in})), \\ \hat{F}_{down} &= f_{down}(\hat{F}_{up}) = ReLU(Conv(\hat{F}_{up})), \\ \hat{F}_{details} &= \hat{F}_{down} - \hat{F}_{in}, \end{aligned} \tag{2}$$

where $\hat{F}_{up}$, $\hat{F}_{down}$, and $\hat{F}_{details}$ denote the expanded feature, projected feature, and high-frequency related feature for details reconstruction, respectively. $TansConv(\cdot)$ denotes the transposed convolution and $Conv(\cdot)$ denotes the vanilla convolution. Note that we can expand the resolution to $N\times$ times by simply repeating $f_{up}(\cdot)$, *i.e.*, $\hat{F}^N_{up} = f^N_{up}(\hat{F}_{in}) = f_{up}(f_{up}(\cdots(\hat{F}_{in})))$. Experiments in Sec. 4.4 indicate that it works well with just $2\times$ resolution.

Inspired by the decorrelation property of the whitening operation (discussed in detail in Sec. 3.5), we introduce the Decorrelation Normalization (DN) based on ZCA-whitening [11]. Formally, given the low-frequency feature $\hat{F}_{in}$ which is relevant to contrast enhancement and the high-frequency feature $\hat{F}_{details}$ that is relevant to details reconstruction, we concatenate them and then employ the DN in the channel dimension to eliminate the correlation of features across different frequencies, thus formulating a frequency-independent feature space. The frequency-independent feature $F_n \in \mathcal{R}^{C\times H\times W}$ can be formulated by:

$$F_n = DN(Concat[\hat{F}_{in}, \hat{F}_{details}]), \tag{3}$$

where $DN(\cdot)$ denotes the Decorrelation Normalization which will be discussed in detail in Sec. 3.5, and *Concat* indicates the concatenation operation.

3.3 Integration Part

As normalization inevitably erases the discriminative information, which is crucial to image restoration, we propose an integration part to compensate the original features unprocessed by the DN to ensure the information completeness [2]. Specifically, as illustrated

in Fig. 2, we first employ a 1×1 convolution to reduce the channel number of F_n to $\frac{C}{2}$: $\hat{F}_n = Conv_{1\times1}(F_n)$. Considering that directly concatenating $\hat{F}_{in}$ and $\hat{F}_n$ would trigger the network to tend to derive knowledge from simpler features, we introduce a random shuffle operation across the channel dimension after concatenation. It prevents the network from excessively relying on a subset of the frequency-dependent or -independent features and reduces the sensitivity to the frequency during optimization. The final output is written by:

$$F_{out} = RandomShuffle(Concat[\hat{F}_{in}, \hat{F}_n]). \tag{4}$$

Since features are randomly shuffled on the channel dimension throughout the training process, the network will be not biased to features of a particular frequency. However, we find that randomly shuffling all channels will result in unstable training. To this end, we investigate the influence of the shuffle ratio on performance in Sec. 4.4, and we empirically set the shuffle ratio as 20% in all experiments. Note that the shuffle operation is not applied during the testing phase to ensure reproducible results without affecting performance.

3.4 Plugging into Existing Networks

As a plug-and-play module, the proposed FDI can be flexibly integrated into existing networks. We choose two representative baseline networks, SID [3] and MPRNet [27], as the backbones. They are U-Net-like architectures with an encoder and a decoder. As described in Fig. 3, we present three distinct strategies for the incorporation of the FDI module: (i) at the head of the encoder, (ii) between the encoder and decoder, and (iii) at the rear of the decoder. All of them achieve performance progress compared to the plain network, and strategy (ii) has the best performance (refer to Sec. 4.4 for details).

3.5 Decorrelation Normalization

Whitening operation is a data processing technique commonly used in machine learning for decorrelating the data [17]. In this paper, we employ the whitening operation as the decorrelation normalization to remove the correlation of features between different channels. Specifically, given the input features $\mathbf{X} \in \mathcal{R}^{C\times HW}$, we perform Decorrelation Normalization by:

$$\hat{\mathbf{X}} = DN(\mathbf{X}) = \Sigma^{-\frac{1}{2}}(\mathbf{X} - \mu \cdot \mathbf{1}^T), \tag{5}$$

where $\mu = \frac{1}{HW}\mathbf{X}\cdot\mathbf{1}$ is the mean on each row of X, $\Sigma = \frac{1}{HW}(\mathbf{X} - \mu\cdot\mathbf{1}^T)(\mathbf{X} - \mu\cdot\mathbf{1}^T)^T + \varepsilon\mathbf{I}$ is the covariance matrix of the zero-centered $\mathbf{X}$, $\mathbf{1}$ is a column vector of all ones, and $\varepsilon > 0$ is a small positive number for numerical stability. It is easy to verify that $\hat{\mathbf{X}}$ is the whitened features, i.e., $\hat{\mathbf{X}}\hat{\mathbf{X}}^T = \mathbf{I}$.

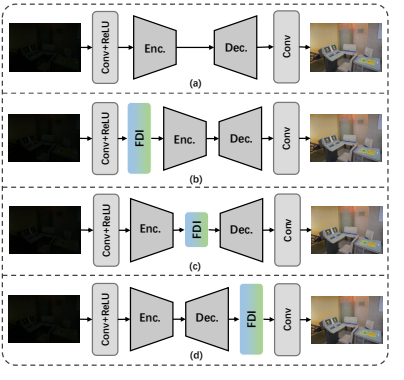


Figure 3: The proposed plug-in strategies. (a) The initial plain network; and (b)-(d) the networks obtained by plugging the FDI module into (a), respectively.

However, the above whitening transformation is not unique due to $\Sigma^{-\frac{1}{2}}$ is defined up to rotation. In this paper, we adopt the widely used ZCA whitening, which preserves the original distribution of each feature. The formulation of ZCA whitening is:

$$\hat{\mathbf{X}} = \Sigma^{-\frac{1}{2}}(\mathbf{X} - \mu\cdot\mathbf{1}^T) = \mathbf{U}\Lambda^{-\frac{1}{2}}\mathbf{U}^T(\mathbf{X} - \mu\cdot\mathbf{1}^T), \tag{6}$$

Figure 4: Visual comparison on LOL dataset. M-SID and M-MPRNet denote SID and MPRNet incorporated with the proposed FDI module by strategy (ii) mentioned in Sec. 3.4.

where $\Lambda = \text{diag}(\sigma_1, \cdots, \sigma_C)$ and $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_C]$ are the eigenvalues and eigenvectors of Σ, i.e. $\Sigma = \mathbf{U}\Lambda\mathbf{U}^T$. As illustrated in Eq. 6, classical ZCA whitening heavily relies on eigen-decomposition, resulting in poor efficiency. Therefore, we employ the Iterative Normalization proposed in [11] for efficient ZCA whitening, which calculates $\Sigma^{-\frac{1}{2}}$ by Newton's Iteration.

3.6 Loss Function

In this paper, we focus on providing an efficient plug-and-play module to assist existing methods to be optimized in a frequency consistency feature space for better performance, thus, we simply utilize L_1 loss and perceptual loss [15], which are widely-used loss functions in the image enhancement community. We employ the pre-trained VGG19 [21] as the feature extractor of the perceptual loss. The total loss function is expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_1 + \lambda \cdot \mathcal{L}_p, \tag{7}$$

where λ denotes the coefficient to balance the $\mathcal{L}_1$ loss and the perceptual loss $\mathcal{L}_p$.

4 Experiments and Results

4.1 Settings

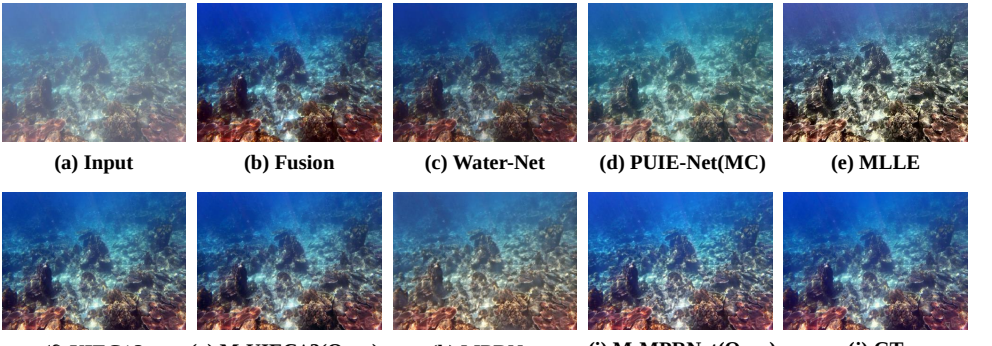
We evaluate the proposed module on two typical image enhancement tasks: Low-light Image Enhancement (LLIE) and Underwater Image Enhancement (UIE).

Datasets and baseline. *Low-light Image Enhancement*: We construct LLIE experiments on the LOL [23] dataset. The LOL dataset is a real captured dataset containing 485 low/normal light image pairs for training and 15 for testing. For performance comparison, we compare our method with the baseline networks and the most representative methods, including Retinex-Net [23], KinD [31], KinD++ [32], RUAS [19], DRBN [26], EnlightenGAN [14], SCI [20]. Due to the introduction of more parameters in FDI, we expand our baseline networks by increasing the number of channels for a fair comparison, which is denoted as SID-L and MPRNet-L. We further compare our method with SOTA methods for

Table 1: Quantitative results on LOL dataset. The best results are highlighted in bold.

Method	RetinexNet	KinD	KinD++	RUAS	DRBN	EnlightenGAN	SCI
PSNR(↑)	16.77	20.87	18.97	18.23	19.86	17.48	14.78
SSIM(↑)	0.5671	0.7988	0.8441	0.7170	0.8342	0.7330	0.6350
LPIPS(↓)	0.474	0.207	0.175	0.257	0.155	0.306	0.3334
#Param	0.62M	8.03M	9.63M	45K	2.21M	8.64M	43k
Method	SID	SID-L	M-SID	MPRNet	MPRNet-L	M-MPRNet	LEDNet
PSNR(↑)	19.16	18.99	21.07(+1.91)	20.13	20.22	21.59(+1.46)	20.94
SSIM(↑)	0.7862	0.8242	0.8360(+0.0498)	0.8170	0.8095	0.8512(+0.0342)	0.8506
LPIPS(↓)	0.440	0.258	0.227(+0.213)	0.266	0.261	0.154(+0.112)	0.2609
#Param	29.6M	118M	35.1M	17.5M	34.3M	18.2M	28.4M
Method	LEDNet-FDI	Restormer	Restormer-FDI	NAFNet	NAFNet-FDI	IAT	IAT-FDI
PSNR(↑)	21.59(+0.65)	20.67	20.79(+0.12)	22.44	22.79(+0.35)	23.38	**23.59**(+0.21)
SSIM(↑)	0.8618(+0.0112)	0.8193	0.8212(+0.0019)	0.8608	0.8620(+0.0012)	0.8675	**0.8704**(+0.0029)
LPIPS(↓)	0.2484(+0.0125)	0.2145	0.2107(+0.0038)	0.1482	**0.1467**(+0.0015)	0.2158	0.2049(+0.0109)
#Param	28.7M	99.8M	102.9M	65.5M	71.0M	0.41M	0.43M

image enhancement, including LEDNet [34], Restormer [28], NAFNet [4], IATNet [5], to demonstrate the superiority of our method. *Underwater Image Enhancement*: We select UIEB [18] to conduct UIE experiments. The UIEB dataset includes 950 real-world underwater images, 890 of which have the corresponding reference images. 800 paired images are randomly selected for training, while the remaining images are allocated for testing. We choose MPRNet and UIEC^2 [22] as the baseline networks. We compare our method against baseline networks and four representative UIE methods: Fusion [1], Water-Net [18], PUIE-Net [7], MLLE [30].



(a) Input **(b) Fusion** **(c) Water-Net** **(d) PUIE-Net(MC)** **(e) MLLE**

(f) UIEC^2 **(g) M-UIEC^2(Ours)** **(h) MPRNet** **(i) M-MPRNet(Ours)** **(j) GT**

Figure 5: Visual comparison on UIEB dataset.

Table 2: Quantitative comparison on UIEB dataset. The best results are highlighted in bold.

Method	Fusion	Water-Net	PUIE-Net(MC)	MLLE	UIEC^2	M-UIEC^2	MPRNet	M-MPRNet
PSNR(↑)	21.47	18.92	22.09	18.58	21.41	22.27(+0.86)	21.33	**23.39**(+2.06)
SSIM(↑)	0.8739	0.8533	0.8441	0.7706	0.9357	**0.9433**(+0.0076)	0.9154	0.9387(+0.0233)
LPIPS(↓)	0.158	0.158	0.156	0.312	0.128	**0.125**(+0.003)	0.178	0.147(+0.031)
#Param	N/A	4.16M	61.5M	N/A	2.05M	2.32M	17.5M	18.2M

Implementation details. Our experiments are implemented in PyTorch, and the training is carried out on a single NVIDIA 3070 GPU. Specifically, our SID-based architectures are trained with a batch size of 8, while the MPRNet-based and UIEC^2 architectures are trained with a batch size of 4 and patch size of 256×256. During training, we adopt the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$ for a total of 200 epochs. The coefficient λ is set to 0.01 empirically. The initial learning rate is set to 1×10^{-4} and the cosine-annealing strategy gradually reduces the learning rate to 1×10^{-7} in the total training epochs.

4.2 Quantitative Evaluation

We employ three metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) [29] as numerical evaluation metrics. The evaluation results on the LOL dataset are reported in Table 1. As can be seen, with the assistance of our FDI module, the M-SID, M-MPRNet, LDENet-FDI, Restormer-FDI, NAFNet-FDI, IAT-FDI networks all achieve better performance. What's more, the performance of the M-MPRNet surpasses the baseline MPRNet significantly, exhibiting an improvement of 1.47 dB in PSNR, 0.0342 in SSIM, and 0.112 in LPIPS, while incurring a mere 4% increase in size. Table 2 describes the results on the UIEB dataset, which indicates our methods achieve the best performance with 23.39dB PSNR, 0.9433 SSIM, and 0.154 LPIPS.

Table 3: Ablation study for investigating the components of the FDI module. DN, DE, and RS stand for Decorrelation Normalization, Details Expanding, and Random Shuffle, respectively.

DN	DE	RS	PSNR (↑)	SSIM(↑)	LPIPS(↓)
			20.13	0.8170	0.266
	✓	✓	20.37	0.8199	0.250
✓		✓	20.82	0.8202	0.243
✓	✓		20.61	0.8243	0.244
✓	✓	✓	**21.59**	**0.8512**	**0.154**

4.3 Qualitative Evaluation

We present the visual comparison of the LOL and UIEB datasets in Fig. 4 and 5, respectively. With the employment of our FDI, the baseline networks achieve superior ability in both contrast enhancement and details recovery compared to other methods. Other competitive baselines suffer from either color distortion or details loss. In contrast, our module can assist the networks in generating more consistent color and natural details. Specifically, our FDI module corrects the unnatural color introduced by SID, reduces the artifact caused by MPRNet, and mitigates the color shift created by IAT. More visualization results are provided in the supplementary material.

4.4 Ablation Studies

To demonstrate the rationality of the core components of our design, we conduct ablation studies on the LOL dataset with MPRNet. As illustrated in Table 3, the performance drops significantly without decorrelation normalization, demonstrating the effectiveness of mapping different features to the frequency-independent space. The random shuffle operation in the integration part can also improve performance since it reduces the networks' sensitivity to different frequencies. Additionally, Table 4 presents the impact of the incorporation position on performance. Incorporating into the head of the network causes performance degradation in the SSIM metric, while performance gains on all metrics are obtained in both the middle and the rear, with the best performance in the middle. We believe that this is because the strongest correlation between high and low frequency occurred in the middle of the network.

Table 4: Quantitative results on LOL dataset. H-, R-, and M- denote incorporating the FDI module at the head, rear, and middle of networks, respectively.

Method	PSNR(↑)	SSIM(↑)	LPIPS(↓)	#Param(M)
SID	19.16	0.7862	0.439	29.6
H-SID	19.24(+0.08)	0.7846(-0.0016)	0.288(+0.218)	29.6
R-SID	20.18(+1.02)	0.8330(+0.0468)	**0.221**(+0.218)	29.6
M-SID	**21.07**(+1.91)	**0.8360**(+0.0498)	0.227(+0.212)	35.1
MPRNet	20.13	0.8170	0.266	17.5
H-MPRNet	20.72(+0.59)	0.8040(-0.0130)	0.259(+0.007)	17.6
R-MPRNet	20.63(+0.50)	0.8192(+0.0022)	0.231(+0.035)	17.6
M-MPRNet	**21.59**(+1.46)	**0.8512**(+0.0342)	**0.154**(+0.112)	18.2

We further investigate the influence of the expanding resolution in the frequency decorrelation part and shuffle ratio in the integration part. As illustrated in Fig. 6, since higher resolution expansion may be detrimental to convergence, expanding to a higher resolution will not improve performance. Moreover, a small shuffle ratio will alleviate the frequency bias of networks, while a large shuffle ratio will result in unstable training and performance degradation. To demonstrate the superiority of the ZCA-whitening normalization, we present layer normalization (LN), instance normalization (IN), and PCA-whitening normalization substitute in Tab. 5. The ZCA-whitening normalization outperforms both LN and IN, emphasizing the importance of the correlation capacity introduced by ZCA-whitening in the optimization process. For PCA-whitening, it causes severe stochastic axis swapping [10], resulting in performance degradation. More results are provided in supplementary materials.

Table 5: Ablation study with decorrelation normalization on LOL dataset.

Method	PSNR(↑)	SSIM(↑)	LPIPS(↓)
baseline	20.13	0.8170	0.266
LN	21.14(+1.01)	0.8281(+0.0111)	0.204(+0.062)
IN	21.10(+0.97)	0.8330(+0.0208)	0.208(+0.058)
PCA	19.51(-0.62)	0.8241(+0.0071)	0.191(+0.075)
ZCA	**21.59**(+1.91)	**0.8360**(+0.0498)	**0.154**(+0.112)

5 Conclusions

In this paper, we point out that the optimization of different frequencies in the image enhancement task is inconsistent within a single neural network. Thus, we develop a Frequency Decorrelation and Integration (FDI) module to facilitate the network to be optimized along a frequency-consistent direction. In particular, our method constructs a frequency-independent feature space to reduce the correlation between different frequencies and introduces a channel shuffle operation to encourage the network to leverage diverse features for avoiding particular frequency bias. Our frequency decorrelation design can be integrated into existing image enhancement approaches with high flexibility. Extensive experiments demonstrate that our approach achieves consistent performance improvement on various image enhancement tasks. However, our method lacks a specific design for handling severe noise that typically arises in extremely dark conditions, which can be explored in the future.

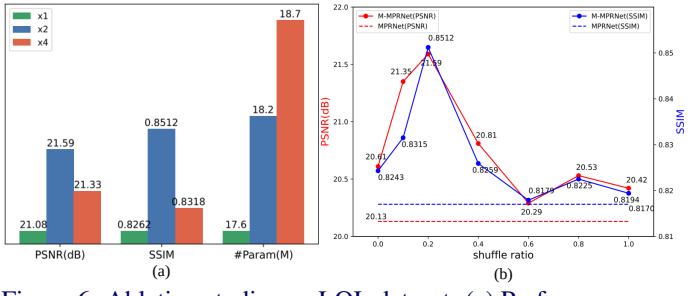


Figure 6: Ablation studies on LOL dataset. (a) Performance versus expanding resolution; (b) performance versus shuffle ratio.

Acknowledgments

This work was supported by the JKW Research Funds under Grant 20-163-14-LZ-001-004-01, and 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] Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. Enhancing underwater images and videos by fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 81–88, 2012.

[2] Kenneth R Castleman. *Digital image processing*. Prentice Hall Press, 1996.

[3] 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*, pages 3291–3300, 2018.

[4] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Proceedings of the European Conference on Computer Vision*, pages 17–33. Springer, 2022.

[5] Ziteng Cui, Kunchang Li, Lin Gu, Shenghan Su, Peng Gao, ZhengKai Jiang, Yu Qiao, and Tatsuya Harada. You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. In *Proceedings of the 33rd British Machine Vision Conference*, pages 1–16, 2022.

[6] Xuan Dong, Yi Pang, and Jiangtao Wen. Fast efficient algorithm for enhancement of low lighting video. In *ACM SIGGRAPH Posters*, pages 1–1. 2010.

[7] Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Uncertainty inspired underwater image enhancement. In *Proceedings of the European Conference on Computer Vision*, pages 465–482. Springer, 2022.

[8] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2016.

[9] Jie Huang, Yajing Liu, Xueyang Fu, Man Zhou, Yang Wang, Feng Zhao, and Zhiwei Xiong. Exposure normalization and compensation for multiple-exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6043–6052, 2022.

[10] Lei Huang, Dawei Yang, Bo Lang, and Jia Deng. Decorrelated batch normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 791–800, June 2018.

[11] Lei Huang, Yi Zhou, Fan Zhu, Li Liu, and Ling Shao. Iterative normalization: Beyond standardization towards efficient whitening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4874–4883, 2019.

[12] Lei Huang, Jie Qin, Yi Zhou, Fan Zhu, Li Liu, and Ling Shao. Normalization techniques in training DNNs: Methodology, analysis and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10173–10196, 2023.

[13] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758, 2007.

[14] 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.

[15] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the European Conference on Computer Vision*, pages 694–711. Springer, 2016.

[16] Edwin H Land. The Retinex theory of color vision. *Scientific American*, 237(6):108–129, 1977.

[17] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE Transactions on Image Processing*, 22(12):5372–5384, 2013.

[18] 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, 2019.

[19] 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/CVF Conference on Computer Vision and Pattern Recognition*, pages 10561–10570, 2021.

[20] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5637–5646, 2022.

[21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

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

[23] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep Retinex decomposition for low-light enhancement. In *Proceedings of the British Machine Vision Conference*, pages 1–12, 2018.

[24] 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*, pages 17714–17724, 2022.

[25] Zhi-Qin John Xu, Yaoyu Zhang, and Yanyang Xiao. Training behavior of deep neural network in frequency domain. In *Proceedings of the 26th International Conference on Neural Information Processing*, pages 264–274. Springer, 2019.

[26] 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*, pages 3063–3072, 2020.

[27] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021.

[28] 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*, pages 5718–5729, 2022.

[29] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.

[30] Weidong Zhang, Peixian Zhuang, Hai-Han Sun, Guohou Li, Sam Kwong, and Chongyi Li. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing*, 31:3997–4010, 2022.

[31] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1632–1640, 2019.

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

[33] 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*, pages 12075–12084, 2021.

[34] Shangchen Zhou, Chongyi Li, and Chen Change Loy. LEDNet: Joint low-light enhancement and deblurring in the dark. In *Proceedings of the European Conference on Computer Vision*, pages 573–589. Springer, 2022.