2022 26th International Conference on Pattern Recognition (ICPR)
August 21-25, 2022, Montréal, Québec, Canada

A Deep Retinex Framework for Light Field Restoration under Low-light Conditions

Shansi Zhang and Edmund Y. Lam
Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam, Hong Kong, China
Email: sszhang@eee.hku.hk, elam@eee.hku.hk

2022 26th International Conference on Pattern Recognition (ICPR) | 978-1-6654-9062-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICPR56361.2022.9956107

***Abstract*—Light field (LF) images can record the scene from multiple directions and have many applications, such as refocusing and depth estimation. However, these applications can be heavily influenced by poor light condition and noise. This work aims to recover the high-quality LF images from their low-light detection. First, a decomposition network is employed to decompose each LF image into its reflectance and illumination with the Retinex theory. Then, two enhancement networks are designed to denoise the reflectance and enhance the illumination, respectively. They adopt alternate spatial-angular feature extractions and process all the views synchronously with high efficiency. A parallel dual attention mechanism is integrated to both the spatial and angular feature extractions, to encode more important information. Moreover, a discriminator is introduced during the training to generate more realistic LF images by making judgment according to both the spatial and angular characteristics. Experimental results have demonstrated the superior performance of our method, which can restore the content, luminance, color and geometric structures of LF images effectively.**

I. Introduction

Light field (LF) cameras have the advantage of recording not only intensities but also directions of the light rays to generate LF images with the spatial-angular information [1]–[3]. Each LF image comprises of multiple sub-aperture images (SAIs), each of which captures the scene from a different viewpoint. LF images have many applications, such as depth estimation [4]–[6], post-capture refocusing [7], [8] and novel view synthesis [9]–[11]. Usually, there is a trade-off between the spatial resolution and angular resolution, and dense angular sampling may lead to low resolution of each SAI. Thus, many researches about LF focus on the super-resolution [12]–[18]. In common with the conventional images, LF images may also be subjected to various degradations, such as noise and poor illumination, which are either due to the physical limitations of the LF camera or inadequate light during image acquisition. These degradations can seriously affect the applications of LF. Therefore, recovering the high-quality contents from the degraded LF images is significant for their better applications.

The techniques of denoising [19], [20], low-light enhancement [21]–[23] or other low-level computer vision tasks [24]–[26] for the conventional images can also be applied to the LF images. Moreover, the geometric structures of LF images should be fully exploited to assist restoration. There are mainly two approaches to address the multi-view information. The first approach is to handle one view at each forward process by utilizing multiple auxiliary views, and the second approach is to handle all the views synchronously at each forward process with joint spatial-angular feature encoding.

So far, several methods have been proposed to restore low-light LF images. Ge et al. [27] developed a deep neural network by using 4D convolutions with a color compensation sub-module to brighten the LF images. However, 4D convolutions involve large computational cost with high GPU memory consumption. Lamba et al. [28] proposed a framework, which restores the central view each time by incorporating its surrounding views, and incorporates a global representation block for encoding the LF geometries. Zhang and Lam [29] developed a two-stage framework to restore the raw LF images, where a multi-to-one network first restores each view individually by exploiting the features of its neighboring views, and then an all-to-all network further refines the contents of all the views synchronously while improving the angular geometries. They extended their work to restore RGB LF images in low-light environment [30] by a decomposition-enhancement approach, which can effectively restore the LF images but involves relatively complex network architecture.

In this paper, we aim to restore LF images with poor light conditions and severe noise. It may be difficult to learn a direct mapping from the low-light noisy LF images to the normal-light clean LF images. Thus, we divide this task into several sub-tasks by utilizing the Retinex theory [31]. We design different networks, each of which solves a particular sub-task. More specifically, there is a Decomposition network (DecomNet) [30] to first decompose the input LF image into its corresponding reflectance and illumination. Then, the reflectance is denoised by an RENet and the illumination is adjusted by an IENet. The final restored LF image is obtained by multiplying the enhanced reflectance and illumination. In order to achieve fast inference, all these networks adopt a synchronous approach to handle all the SAIs through one forward process. We design an efficient spatial-angular residual module to extract the spatial features and angular features alternately. Considering the severity of noise handled by the RENet, we introduce an attention mechanism to the original modules, which integrates parallel dual attention applied to both the spatial and angular features to encode more important information. Furthermore, a discriminator is employed during

the joint training of RENet and IENet to produce more visually compelling LF images by making judgment in terms of both the spatial and angular aspects. Comparison experiments and ablation studies are performed, and the quantitative and qualitative results demonstrate the superior performance of our method for low-light LF restoration. In addition, our network architecture is simpler and more lightweight than that in [30], but can achieve comparable performance with it.

II. Proposed Method

Fig. 1. SAI-array image and macro-pixel image.

An LF image is represented as $L \in \mathbb{R}^{U\times V\times H\times W}$, with angular dimensions U and V, and spatial dimensions H and W. It can be displayed by two patterns, as shown in Fig. 1, a $U\times V$ SAI array, where each SAI with a shape $H\times W$ provides one view of the scene, and a macro-pixel image, consisting of $H \times W$ macro-pixels, each of which has a shape $U \times V$.

A. Network architecture

The degraded LF images we deal with are under extreme low-light conditions and also corrupted by severe noise. Therefore, it is a challenging task that involves luminance recovery, noise suppression and color correction. Considering the complexity of our task, a single end-to-end network may not be effective to achieve complete restoration. Thus, we use the Retinex theory to resolve this complex task into several smaller tasks, each of which are handled by a separate network. The Retinex theory supposes that any RGB image can be decomposed into its reflectance that describes the intrinsic properties, and illumination that reflects the light distribution. It is also applicable to the LF images by decomposing each SAI. The overview of our method is depicted in Fig. 2. A low-light LF images L_l is first decomposed into its reflectance R_l and illumination I_l by the DecomNet [30]. Then, the reflectance is fed to the RENet for denoising, and the illumination passes through the IENet for luminance adjustment. Finally, the enhanced reflectance R_e and illumination I_e are multiplied together to obtain the restored LF image L_e. We also introduce a discriminator during the training aiming to produce more realistic LF images. These networks process all the SAIs synchronously to preserve better angular geometries and also realize more efficient inference.

The DecomNet, RENet and IENet all adopt an encoder-decoder architecture. The main components of the DecomNet and IENet are the spatial-angular modules. For the RENet, we introduce an attention mechanism to the original modules since it needs to address severe noise, a more difficult subtask compared with others. Here, we mainly introduce the architecture of RENet, as shown in Fig. 3. Its main components are the spatial-angular attention modules (Fig. 4), where the feature maps with the SAI-array pattern are first input to a spatial attention residual block for spatial feature extraction, and then fed to an angular attention residual block after converting to the macro-pixel pattern for angular feature extraction. The features with macro-pixel pattern need to be transformed to the SAI-array pattern again before passing through the next module. Each residual block contains two 3×3 convolution layers and uses leaky ReLU (LReLU) activation function. For extracting more useful features, we design an attention mechanism applied to both the spatial and angular features. It integrates dual attentions in parallel to encode important features along both the spatial and channel dimensions. The spatial attention map is obtained by two convolution layers and a sigmoid activation function, and the channel attention vector is obtained by the squeeze and excitation operations [32] with global average pooling (GAP), two fully-connected (FC) layers and the sigmoid activation. The features after multiplying with the dual attentions are added together to obtain the calibrated features.

Throughout the network, the spatial and angular features are extracted alternately to fully explore the spatial-angular information. The resolution of the feature maps for each SAI is gradually reduced during the encoding process, and is recovered to the original size during the decoding stage. Skip connections are employed to fuse the more precise features in the encoder and more abstract features in the decoder. Our network design for processing LF is more efficient than 4D CNN [16] and decoupled spatial-angular branches [17] in terms of GPU memory consumption and inference speed.

In order to generate more realistic LF images, we introduce a discriminator D during training. The discriminator is built by several spatial-angular modules, which enable it to make judgment according to both the spatial and angular characteristics. Within the network, the resolution of the spatial features is gradually reduced, and the final layer uses the mean operation to calculate the output that measures the quality of the restored LF image.

B. Loss function

The DecomNet takes as input the paired low-light LF L_l and normal-light LF L_n during training. Their decomposed reflectances are expected to be consistent, leading to the consistency loss

$$\ell_c = \|R_l - R_n\|_1 , \tag{1}$$

where $\|\cdot\|$ is the ℓ_1 norm.

The illuminations are expected to be smooth while retaining the boundary structures of objects. Thus, smoothness loss is required to constrain I_l and I_n, with

$$\ell_s = |\nabla I_l| \times \exp(-\eta|\nabla R_l|) + |\nabla I_n| \times \exp(-\eta|\nabla R_n|), \tag{2}$$

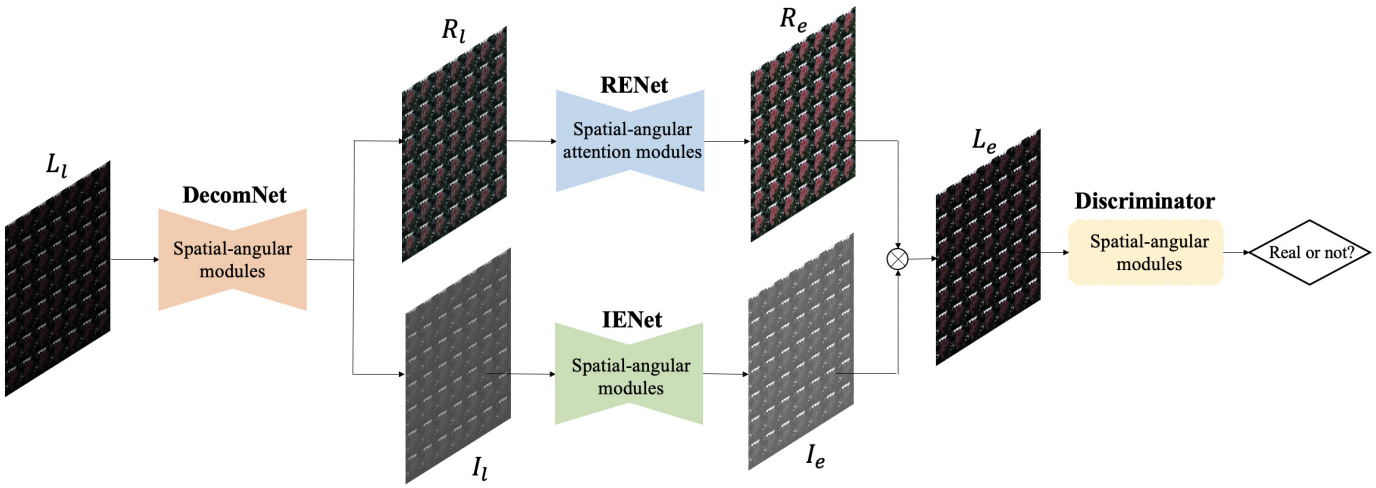

Fig. 2. The overall pipeline of our method, with the DecomNet for decomposition, the RENet and IENet for reflectance and illumination enhancement, and a discriminator during training to make judgment.

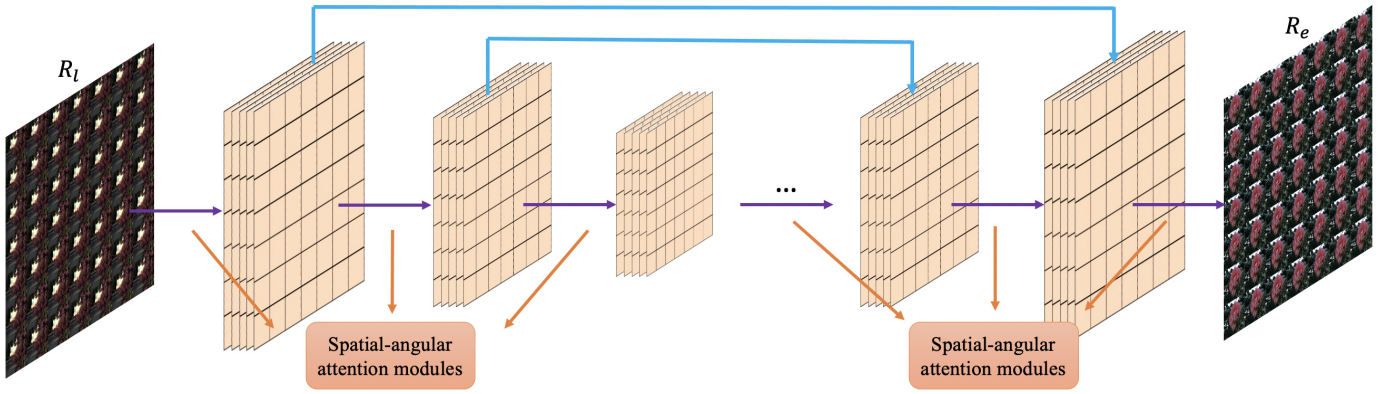

Fig. 3. The encoder-decoder architecture of the RENet, with spatial-angular attention modules for fully extracting the spatial and angular features.

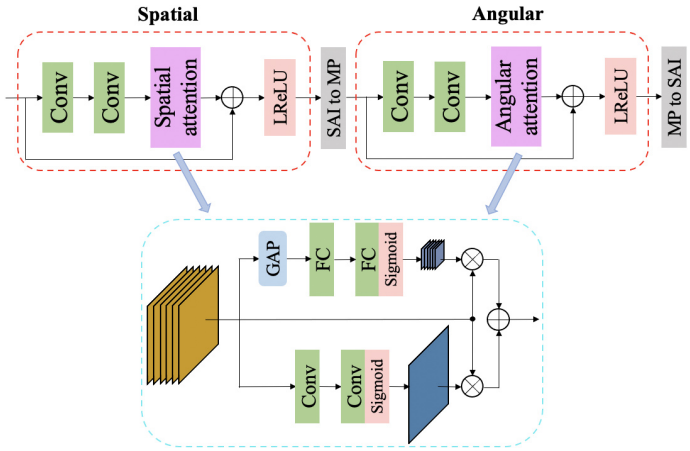

Fig. 4. The structure of spatial-angular attention module.

where ∇ means calculating the gradients along both the horizontal and vertical directions, and η is a hyper-parameter to control the degree of retaining structures by referring to the reflectances.

Furthermore, the decomposed reflectance and illumination need to reconstruct the original LF image, introducing the reconstruction loss

$$\ell_r = \|R_l \times I_l - L_l\|_1 + \|R_n \times I_n - L_n\|_1. \quad (3)$$

The full loss to train the DecomNet is expressed as

$$\ell_{\text{Decom}} = \ell_r + \alpha_1 \ell_s + \alpha_2 \ell_c, \quad (4)$$

where α_1 and α_2 are the coefficients of smoothness loss and consistency loss, respectively.

The reflectance and illumination of the normal-light LF can be treated as the ground truths of the RENet and IENet, and their respective losses are written as

$$\ell_R = \|R_e - R_n\|_1, \quad (5)$$
$$\ell_I = \|I_e - I_n\|_1. \quad (6)$$

These two networks are also jointly trained by the ℓ_1 loss, SSIM [33] loss and adversarial loss.

$$\ell_{\text{joint}} = \|L_e - L_n\|_1 + (1 - \text{SSIM}(L_e, L_n)) - \beta D(L_e), \quad (7)$$

where $L_e = R_e \times I_e$, and $-D(L_e)$ is the adversarial loss with coefficient β, which aims to maximize the output of L_e passing through the discriminator so that the restored LF could be indistinguishable with the real LF.

The discriminator D is trained by the hinge loss [34], expressed as

$$\ell_D = \max(0, 1 - D(L_n)) + \max(0, 1 + D(L_e)). \quad (8)$$

It aims to maximize the margin between the generated LF images and real LF images, and can make the training more stable.

III. Experiments

A. Datasets and implementation details

We synthesize low-light LF images by simulating the imaging process with the Kalantari [9] and Stanford [35] LF datasets. The synthetic method is formulated as

$$L_l = K\Big[\text{Poisson}\Big(\frac{L_{\text{Bayer}}}{\text{Mean}(L_{\text{Bayer}})} \times p\Big) + n_r\Big], \quad (9)$$

where L_{Bayer} is the Bayer pattern of the original LF image L_n, 'Poisson' means the Poisson process of photon arrival, leading to the shot noise, n_r is the readout noise following Gaussian distribution, p can be tuned to control the light levels, and K means scaling operation. Moreover, a demosaicing operation is leveraged to yield the RGB LF images, which are corrupted by Gaussian-Poisson mixed noises. The range of p was set to $0.5 \sim 1.5$, and the central 7×7 SAIs were preserved in the experiments.

We first trained the DecomNet for decomposition. During the training, each input LF was synthesized from the ground truth with a random p between 0.5 and 1.5, and all the SAIs were cropped into 128×128 patches randomly. The weights α_1 and α_2 in (4) were set to 0.1 and 0.01, respectively. The learning rate was set to 5×10^{-4} initially and gradually decreased by multiplying 0.9 after every 20 epochs. The DecomNet was trained for 100 epochs with a batch size of 1.

Next, the RENet was trained for reflectance denoising by taking the low-light reflectance as input and the normal-light reflectance as ground truth, with the loss in (5). The initial learning rate was 1×10^{-3} and decayed by multiplying 0.8 every 30 epochs. The training stopped after about 150 epochs when the validation performance was not improved anymore. The SAIs were also cropped to 128×128 and batch size was still 1 during the training.

Then, we trained the IENet with the loss in (6) and fine-tuned the RENet with a smaller learning rate 1×10^{-4} meanwhile. The two networks are jointly trained by (7), where β was set to 0.1. This training process lasted for about 200 epochs. The initial learning rate of the IENet was 1×10^{-3} and decreased by a decay rate 0.8 every 40 epochs. Adam optimizer were used in all the training stages.

The DecomNet has 0.909M parameters, and the RENet and IENet have 1.030M and 0.954M parameters, respectively. Thus, our framework only has 2.893M parameters in total, indicating its lightweight characteristic.

TABLE I
THE QUANTITATIVE RESULTS OF DIFFERENT METHODS ON THE KALANTARI TEST SET

Method	Light level	PSNR	SSIM
KinD [22]	0.5	19.90	0.6272
	1.0	20.19	0.6314
	1.5	20.22	0.6677
LFSAS [15]	0.5	22.71	0.7291
	1.0	23.48	0.7279
	1.5	23.94	0.7446
LFInterNet [17]	0.5	22.01	0.7541
	1.0	22.98	0.7931
	1.5	23.38	0.7992
LFLowLight [30]	0.5	26.57	**0.8230**
	1.0	**27.12**	**0.8581**
	1.5	**27.72**	0.8540
Ours	0.5	**26.64**	0.8214
	1.0	26.85	0.8534
	1.5	27.54	**0.8572**

B. Experimental results

First, we present some intermediate and final outputs by our method, as depicted in Fig. 5, including the decomposed low-light reflectances R_l and illuminations I_l by the DecomNet, the enhanced reflectances R_e and illuminations I_e by the RENet and IENet, and the final restored LF images. The central SAIs and epipolar-plane images (EPIs) that imply the angular geometries are shown to represent the LFs. We can see that the input LF images with $p = 1.0$ and 1.2 are under very poor light conditions and corrupted by severe noise. After decomposition, the noises are displayed on the reflectances, and the low light levels are reflected on the illuminations. The RENet can effectively suppress the noises to obtain clean reflectances, and the IENet can brighten the illuminations to recover the light levels. The final restored LF images have good visual qualities while preserving the angular geometries well.

Then, we make comparison with several other methods for low-light or LF enhancement, including KinD [22], which is based on the Retinex theory but operates on the single images, LFSAS [15] and LFInterNet [17], which are end-to-end networks and achieve good performance on the LF super-resolution task, and LFLowLight [30], which also applies the Retinex theory to LF but with more complex network architecture. We modify the network structures and hyper-parameters of LFSAS and LFInterNet slightly in order to adapt to our task. Then, we test different methods on the Kalantari and Stanford test sets, and the quantitative results are recorded in Table I and II. The metrics used for evaluation are the PSNR and SSIM, and the light levels for testing are 0.5, 1.0 and 1.5. It can be seen that our method can achieve comparable performance with LFLowLight [30], and obviously outperform the other methods at various light levels on these two test sets.

Some visual results of different methods are presented in Fig. 6 for qualitative comparison. The input LF images are under different light levels with $p = 0.7, 1, 0$ and 1.5,

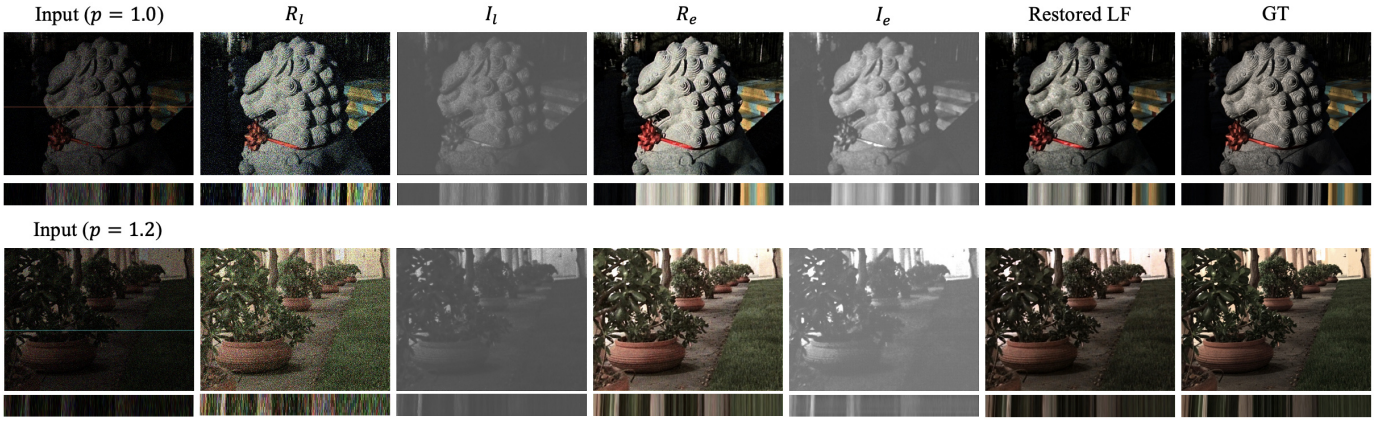

Fig. 5. The intermediate and final outputs by our method at $p = 1.0$ and 1.2, including the decomposed low-light reflectances and illuminations by the DecomNet, the enhanced reflectances and illuminations by the RENet and IENet, and the final restored LF images.

TABLE II
THE QUANTITATIVE RESULTS OF DIFFERENT METHODS ON THE STANFORD TEST SET

Method	Light level	PSNR	SSIM
KinD [22]	0.5	20.44	0.6461
	1.0	21.38	0.6572
	1.5	21.52	0.6862
LFSAS [15]	0.5	23.16	0.7310
	1.0	23.63	0.7449
	1.5	24.01	0.7513
LFInterNet [17]	0.5	23.03	0.7832
	1.0	23.97	0.8147
	1.5	24.89	0.8275
LFLowLight [30]	0.5	26.98	**0.8353**
	1.0	**27.66**	0.8577
	1.5	**27.76**	0.8705
Ours	0.5	**27.14**	0.8345
	1.0	27.46	**0.8623**
	1.5	27.67	**0.8719**

respectively. Two cropped patches are enlarged to display the details more clearly. As the KinD restores each SAI separately without handling any angular information, it can not preserve the angular geometries effectively, as depicted in its EPIs. Moreover, its noise suppression effect is not good, leading to poor visual qualities. The LFSAS and LFInterNet can recover relatively better geometric structures since they both incorporate angular feature extractions. However, their restored LF images suffer from obvious color distortion, and the zoomed patches also show the residual noises. The reason may be that it is difficult for a single network to learn multiple tasks, including denoising, luminance recovery and color correction. In contrast, our method and LFLowLight [30], both of which are based on the Retinex theory, can achieve better noise suppression, color and luminance recovery with accurate angular geometries, and therefore obtain higher-quality LF images.

In addition, we compare the network parameters and average run time of different methods, as shown in Table. III. All the methods were implemented on a NVIDIA Tesla P100 GPU to obtain the average complete inference time of each $7\times7\times384\times512$ LF image. We can find that our framework is more or equivalently lightweight and efficient compared with others. In particular, we can achieve comparable performance with LFLowLight [30] but with fewer network parameters and faster inference speed.

TABLE III
COMPARISON OF NETWORK PARAMETERS AND AVERAGE RUN TIME OF DIFFERENT METHODS.

Method	Parameters	Run time (s)
KinD [22]	2.006M	5.473
LFSAS [15]	2.226M	3.443
LFInterNet [17]	4.706M	8.022
LFLowLight [30]	4.012M	5.926
Ours	2.893M	3.704

We also validate our method design by performing several ablation studies. We train another two models by removing the adversarial loss and attention mechanism, respectively. From Table IV, we can observe that their corresponding quantitative results on the test LF images are worse than our default configurations. Therefore, introducing adversarial loss contributes to obtaining higher-quality LF images, and the attentions for spatial and angular features also help to improve the restoration performance.

IV. CONCLUSION

In this paper, we propose a deep Retinex framework to restore LF images under low-light conditions. The LF images that we deal with are degraded severely with noises, poor luminance and color distortion. To reduce learning difficulty, we resolve this complicated task into several sub-tasks by designing different networks, including a DecomNet for decomposition, a RENet for reflectance denoising, and an IENet for illumination enhancement. All these networks adopt alternate spatial-angular feature extractions to handle all the

Fig. 6. The qualitative results of different methods at $p = 0.7, 1.0$ and 1.5, with central SAIs, EPIs and zoomed patches.

TABLE IV
ABLATION STUDY

Methods	Light level	PSNR	SSIM
Ours	0.5	**26.79**	**0.8253**
	1.0	**27.03**	**0.8561**
	1.5	**27.58**	**0.8616**
w/o adversarial loss	0.5	26.05	0.8188
	1.0	26.29	0.8430
	1.5	26.52	0.8557
w/o attention	0.5	26.08	0.8201
	1.0	26.18	0.8409
	1.5	26.37	0.8520

views synchronously. A parallel dual attention mechanism is designed to encode more useful spatial and angular features. In addition, we introduce a discriminator during the joint training of the RENet and IENet to produce more realistic LF images. The experimental results have demonstrated that our method can achieve superior performance on the LF restoration.

ACKNOWLEDGMENT

The work is supported in part by the Research Grants Council of Hong Kong (GRF 17201818, 17200019, 17201620) and the University of Hong Kong (104005864).

REFERENCES

[1] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report*, vol. 2, no. 11, pp. 1–11, 2005.

[2] E. Y. Lam, "Computational photography with plenoptic camera and light field capture: tutorial," *Journal of the Optical Society of America A*, vol. 32, no. 11, pp. 2021–2032, Nov. 2015.

[3] N. Chen, Z. Ren, D. Li, and E. Y. Lam, "Analysis of the noise in backprojection light field acquisition and its optimization," *Applied Optics*, vol. 56, no. 13, pp. F20–F26, May 2017.

[4] C. Shin, H.-G. Jeon, Y. Yoon, I. S. Kweon, and S. J. Kim, "EPINET: A fully-convolutional neural network using epipolar geometry for depth from light field images," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[5] J. Shi, X. Jiang, and C. Guillemot, "A framework for learning depth from a flexible subset of dense and sparse light field views," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5867–5880, 2019.

[6] Y.-J. Tsai, Y.-L. Liu, M. Ouhyoung, and Y.-Y. Chuang, "Attention-based view selection networks for light-field disparity estimation," in *AAAI Conference on Artificial Intelligence*, 2020, pp. 12 095–12 103.

[7] J. Fiss, B. Curless, and R. Szeliski, "Refocusing plenoptic images using depth-adaptive splatting," in *IEEE International Conference on Computational Photography*, 2014, pp. 1–9.

[8] Y. Wang, J. Yang, Y. Guo, C. Xiao, and W. An, "Selective light field refocusing for camera arrays using bokeh rendering and super-resolution," *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 204–208, Jan. 2019.

[9] N. K. Kalantari, T. C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Transactions on Graphics*, vol. 35, no. 6, 2016.

[10] J. Jin, J. Hou, H. Yuan, and S. Kwong, "Learning light field angular super-resolution via a geometry-aware network," in *AAAI Conference on Artificial Intelligence*, 2020, pp. 11 141–11 148.

[11] N. Meng, K. Li, J. Liu, and E. Y. Lam, "Light field view synthesis via aperture disparity and warping confidence map," *IEEE Transactions on Image Processing*, vol. 30, pp. 3908–3921, 2021.
[12] S. Zhang, Y. Lin, and H. Sheng, "Residual networks for light field image super-resolution," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11 046–11 055.
[13] Y. Wang, F. Liu, K. Zhang, G. Hou, Z. Sun, and T. Tan, "LFNet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4274–4286, Sept. 2018.
[14] J. Jin, J. Hou, J. Chen, and S. Kwong, "Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2260–2269.
[15] H. W. F. Yeung, J. Hou, X. Chen, J. Chen, Z. Chen, and Y. Y. Chung, "Light field spatial super-resolution using deep efficient spatial-angular separable convolution," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2319–2330, May 2019.
[16] N. Meng, H. K.-H. So, X. Sun, and E. Y. Lam, "High-dimensional dense residual convolutional neural network for light field reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 3, pp. 873–886, March 2021.
[17] Y. Wang, L. Wang, J. Yang, W. An, J. Yu, and Y. Guo, "Spatial-angular interaction for light field image super-resolution," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 290–308.
[18] N. Meng, Z. Ge, T. Zeng, and E. Y. Lam, "LightGAN: A deep generative model for light field reconstruction," *IEEE Access*, vol. 8, pp. 116 052–116 063, 2020.
[19] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 492–511.
[20] S. Zhang and E. Y. Lam, "An effective image restorer: Denoising and luminance adjustment for low-photon-count imaging," *arXiv preprint*, 2021.
[21] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *British Machine Vision Conference (BMVC)*, 2018.
[22] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *ACM International Conference on Multimedia*, 2019, pp. 1632–1640.
[23] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.
[24] T. Zeng, H. K.-H. So, and E. Y. Lam, "Redcap: residual encoder-decoder capsule network for holographic image reconstruction," *Optics Express*, vol. 28, no. 4, pp. 4876–4887, 2020.
[25] P. Zhang and E. Y. Lam, "From local to global: Efficient dual attention mechanism for single image super-resolution," *IEEE Access*, vol. 9, pp. 114 957–114 964, 2021.
[26] L. Song and E. Y. Lam, "Mbd-gan: Model-based image deblurring with a generative adversarial network," in *International Conference on Pattern Recognition (ICPR)*, 2021, pp. 7306–7313.
[27] Z. Ge, L. Song, and E. Y. Lam, "Light field image restoration in low-light environment," in *Future Sensing Technologies*, ser. Proceedings of the SPIE, vol. 11525, 2020, p. 115251H.
[28] M. Lamba, K. Kumar, and K. Mitra, "Harnessing multi-view perspective of light fields for low-light imaging," *IEEE Transactions on Image Processing*, vol. 30, pp. 1501–1513, 2021.
[29] S. Zhang and E. Y. Lam, "Learning to restore light fields under low-light imaging," *Neurocomputing*, vol. 456, pp. 76–87, 2021.
[30] S. Zhang and E. Y. Lam, "An effective decomposition-enhancement method to restore light field images captured in the dark," *Signal Processing*, vol. 189, p. 108279, 2021.
[31] E. H. Land, "The retinex theory of color vision," *Sci. Amer*, vol. 237, no. 6, pp. 108–129, 1977.
[32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7132–7141.
[33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.
[34] J. H. Lim and J. C. Ye, "Geometric gan," *arXiv preprint*, 2017.
[35] R. Shah, G. Wetzstein, A. S. Raj, and M. Lowney, "Stanford lytro light field archive," 2016.