IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 16, 2020, accepted January 15, 2021, date of publication January 22, 2021, date of current version January 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3053607

LLISP: Low-Light Image Signal Processing Net via Two-Stage Network

HONGJIN ZHU[1,2], YANG ZHAO[2,3], (Member, IEEE), RONGJIE WANG[4], RONGGANG WANG[1,2], (Member, IEEE), WEIQIANG CHEN[5], AND XUESONG GAO[5]

[1]School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen 518055, China
[2]Peng Cheng Laboratory, Shenzhen 518055, China
[3]School of Computer and Information, Hefei University of Technology, Hefei 230000, China
[4]Department of Computer Science, City University of Hong Kong, Hong Kong
[5]State Key Laboratory of Digital Multimedia Technology, Hisense Group Company Ltd., Qingdao 266071, China

Corresponding author: Rongjie Wang (rjwang.hit@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61672063 and Grant 61972129, and in part by the Shenzhen Research Projects under Grant JCYJ20180503182128089 and Grant 201806080921419290.

ABSTRACT Images taken in extremely low light suffer from various problems such as heavy noise, blur, and color distortion. Assuming the low-light images contain a good representation of the scene content, current enhancement methods focus on finding a suitable illumination adjustment but often fail to deal with heavy noise and color distortion. Recently, some works try to suppress noise and reconstruct low-light images from raw data. But these works apply a network instead of an image signal processing pipeline (ISP) to map the raw data to enhanced results which leads to heavy learning burden for the network and get unsatisfactory results. In order to remove heavy noise, correct color bias and enhance details more effectively, we propose a two-stage Low Light Image Signal Processing Network named LLISP. The design of our network is inspired by the traditional ISP: processing the images in multiple stages according to the attributes of different tasks. In the first stage, a simple denoising module is introduced to reduce heavy noise. In the second stage, we propose a two-branch network to reconstruct the low-light images and enhance texture details. One branch aims at correcting color distortion and restoring image content, while another branch focuses on recovering realistic texture. Experimental results demonstrate that the proposed method can reconstruct high-quality images from low-light raw data and replace the traditional ISP.

INDEX TERMS Low-light enhancement, image enhancement, artifacts removal, image signal processing, deep learning.

I. INTRODUCTION

Typically, the raw sensor data we captured will be processed by an in-camera image signal processing pipeline (ISP) to generate JPEG-format images. And the key steps in the ISP include: ISO gain, denoising, demosaicing, detail enhancing, white balance, color manipulation and color mapping. The quality of these JPEG-format images is very important both for our daily life and for many computer vision tasks, e.g., video surveillance, segmentation, and object detection [1], [2]. However, images captured in low-light environments suffer from various problems such as heavy noise, color distortion and blur. And these problems will be aggravated by quantization, clipping, and other processing in the traditional ISP. High ISO, large aperture, or long exposure time can be used to brighten the images, but they also lead to various drawbacks, for example, the amplified noise or inevitable blur.

Researchers have proposed lots of techniques to restore low-light images. Retinex [3], [4] and histogram equalization [5] are traditional methods to brighten images. Due to the lack of content understanding, these methods may produce unnatural results. Recently, deep learning-based approaches have revealed their superior performance in image enhancement. Some methods [6], [7] directly kindle low-light images without special consideration about noise or blur. Other methods focus on some challenges which are related to low-light image enhancement such as denoising [8], [9], demosaicing [10], deblurring [11], multiexposure image fusion [12], [13]. However, these methods still cannot produce high-quality

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

enhanced images for the following reasons: First, most low-light enhancement methods cannot handle images taken in extremely dark conditions that contain severe noise and color degradation. Under these conditions, JPEG-format images cannot provide enough information due to the information loss during the traditional ISP. What's more, heavy noise often leads to inaccurate white balance and blurred results. Second, sequentially denoising, deblurring, and correcting color bias may accumulate errors. Hence, we need an effective method that can operate directly on raw sensor data and produce pleasant enhanced images.

In this paper, we propose a Low Light Image Signal Processing Network (LLISP) to address the extremely low-light enhancement problem. As the traditional ISP cannot work well in such conditions, we reconstruct the images directly from raw sensor data to avoid further information loss. Inspired by the traditional ISP, we firstly use a U-net-based module [14] to remove noise as heavy noise is one of the most challenging problems in dark conditions, which also influences detail enhancement and white balance. Then, a two-branch network is proposed to reconstruct images and refine textural details simultaneously. Specifically, different network architectures are used in different branches. The reconstruction branch aims at correcting color distortion and restoring image content. Hence, we use a U-net [14] to learn high-level features. The enhancing branch aims at recovering texture and focuses on detailed information. In this branch, the resolution of features is not reduced to persevere structural integrity and the dilated convolution [15] is applied to enlarge the receptive field.

In summary, we make the following contributions:

- We propose a novel two-stage low-light enhancement net which can directly brighten extremely low-light images from raw data and replace the traditional ISP. The proposed method inherits the benefits of both end-to-end network and traditional multistage ISP.
- A two-stream structure is presented in the second stage, which consists of a reconstruction branch and a texture enhancing branch. The reconstruction branch restores images from both original input and pre-denoised features. The texture enhancing branch utilizes gradient information to reduce artifacts and enhance details.
- Experimental results demonstrate that, to enhance extremely dark images, a pre-denoising module is indispensable and can improve the robustness of the proposed method.

The rest of the paper is organized as follows. Section II briefly introduces the related works. Section III describes the proposed method in detail. Experimental results are shown in Section IV. Finally, Section V concludes this paper.

II. RELATED WORK

Low-light image enhancement has a long history and it covers lots of aspects such as denoising and demosaicing. We provide a short review of previous arts closely related to our task.

A. LOW-LIGHT IMAGE ENHANCEMENT

Classic approaches can be roughly divided into two main categories: histogram equalization (HE) [16]–[18] and gamma correction (GC) [19]. These methods ignore the relationship between individual pixels and their neighbors. As a result, they often produce artifacts and compromised aesthetic quality. Another technical line is based on the Retinex theory [4], [20]–[22], which decomposes the image into two components, i.e., reflectance and illumination, and enhances the illumination component. But a global adjustment tends to over-/under- enhance local regions. To further improve the adaptability of enhancement and avoid local over/under enhancement due to uneven illumination, Wang *et al.* [23] enhances the image via multi-scale image fusion. Unfortunately, these approaches still cannot handle heavy noise and color bias. Besides, the lack of understanding of the image content causes unnatural enhancement.

Deep learning-based methods perform more global analysis and try to understand image content. Some works use paired data to learn the mapping function from low-light images to high-quality outputs [6], [24], [25]. Other works use unpaired data to train the models which release the necessity for collecting paired data [7]. However, these approaches generally assume that the images do not suffer from heavy noise and color distortion. As a consequence, under extremely low-light conditions, they may either enhance both the noise and scene details, or fail to recover the low visibility of low-light images. Compared with these methods, our LLISP brightens up the image while preserving the inherent color and details via a proper image processing pipeline and efficient utilization of the raw data.

More recently, some approaches [26]–[28] use neural networks to replace the traditional ISP and directly reconstruct high-quality images from raw data. By using raw data, they avoid information loss caused by the traditional ISP. However, these works tend to learn the ISP pipeline as a black-box, which increases the learning burden of networks and causes the inefficient utilization of data. Different from those approaches, our LLISP pays more attention to model a proper image processing pipeline and make full use of the raw data.

B. IMAGE DENOISING METHODS

Image denoising is a hot topic in low-level visual tasks and is very essential for further image processing. Classic approaches [8], [9] use specific priors of natural clean images such as pixel-wise smoothness and non-local similarity. Recently, deep convolutional neural networks have led to significant improvement in denoising. Some works focus on applying effective network structure to learn the mapping between noisy images and clean images, e.g., auto-encoders [29], residual block [30] and non-local attention block [31]. Other works focus on simulating realistic noise models for better performance on real-world denoising tasks [30].

In our work, we adopt a simple but effective pre-denoising module so that we can avoid the disruption of severe noise on the subsequent enhancement.

C. IMAGE SIGNAL PROCESSING PIPELINE

In order to reconstruct the images from raw data more accurately, it's necessary to be clear of the in-camera ISP. Typical ISP in our daily used cameras includes: ISO gain, denoising, demosaicing, detail enhancing, white balance, color manipulation, then mapping the data to sRGB color space and finally saving to file. There are many classical approaches for the above steps [32]. Recently, lots of deep learning-based methods have been proposed and outperform those classical approaches. Some works focus on applying convolutional neural networks (CNN) for specific steps in the ISP, such as demosaicing [10] or white balance [33]. Other works [26], [34] use deep learning models to replace the entire ISP pipeline. In this paper, we propose a deep network to replace the entire ISP for low-light image reconstruction. Inspired by the typical ISP, the proposed net also adopts a multi-stage enhancement strategy.

III. METHOD

The proposed LLISP aims at removing noise, correcting color bias and reconstructing high-quality images from raw data. As illustrated in Fig. 1, the proposed LLISP network consists of two components: a Denoising Module (DNM), an Enhancement Net (EN).

A. DATA PREPARING

In the training stage, four types of data are used, i.e., low-light raw data (I_{raw}), amplification ratio k, ground truth raw data (GT_{raw}), and ground truth sRGB data (GT_{sRGB}). The data can be collected from commonly used digital cameras or smartphones. In our experiment, we use the SID dataset [26], which consists of raw short-exposure images and the corresponding long-exposure images both in raw and RGB format. The corresponding exposure time for these images is also provided in the dataset. Following SID [26], the amplification ratio k is set to be the exposure difference between the input and reference images (e.g., x100, x250, or x300) for both training and testing. We scale the low-light raw data (I_{raw}) by the desired amplification ratio k to get the inputs (I^*_{raw}) for our LLISP. Specially, in the testing phase, k can be specified by users.

B. STAGE I: DENOISING MODULE

Denoising is very essential and important in the image processing pipeline, especially for low-light images that suffer from heavy noise. Because heavy noise significantly influences subsequent processes, e.g., deblurring, white balance, and color mapping, we put the DNM in the first stage to obtain relatively clean data and reduce the difficulty for the following stages. Formally, given the scaled low-light raw inputs (I^*_{raw}), we can generate clean raw data (C_{raw}) as,

$$C_{raw} = DNM(I^*_{raw}) \tag{1}$$

The architecture of this module can be seen in Table 1. Commonly used U-net [14] is selected as the backbone of the DNM for its effectiveness in denoising tasks. The input and output channels are set to 4 to suit for raw data. As a trade-off between efficiency and restoration performance, the kernel size is set to (3,3) following SID [26]. Considering the fact that, in extremely low-light conditions, even the long-exposure ground truth data still has noise, besides the pixel-wise $Loss_{L1}$, we also add the $Loss_{TV}$ to further smooth the denoised output. $Loss_{L1}$ is defined as the l_1 distance between the output of the denoising module and ground truth raw data (2). $Loss_{TV}$ is defined as a total variation regularizer to constrain the smoothness of outputs (3)

$$Loss_{L1} = ||C_{raw} - GT_{raw}||_1 \tag{2}$$

$$Loss_{TV} = ||\nabla_h C_{raw}||_2^2 + ||\nabla_v C_{raw}||_2^2 \tag{3}$$

where∇_h and ∇_v denote the gradients along the horizontal and the vertical directions.

The total loss function for DNM is defined as $Loss_{DNM}$ (4). We empirically set $\alpha_1 = 1, \alpha_2 = 0.05$. Note that the DNM is firstly pre-trained via GT_{raw} and then fixed during the training stage of the following module.

$$Loss_{DNM} = \alpha_1 Loss_{L1} + \alpha_2 Loss_{TV} \tag{4}$$

C. STAGE II: ENHANCEMENT NET

After obtaining pre-denoised raw data from DNM, the EN aims at mapping the raw data to final sRGB outputs, which corresponds to the processes that need global information in traditional ISP as shown in Fig. 2. To produce high-quality outputs, the EN consists of two branches, i.e., the Reconstruction Branch (RB) and the Texture Enhancing Branch (TEB).

1) RECONSTRUCTION BRANCH

The RB is responsible for global color mapping which is similar to white balance and color space mapping steps in the traditional ISP. The architecture of the RB_{net} can be seen in Fig. 1(b). For accurate color mapping, a global understanding of the whole images is required. U-net architecture, which has a large receptive field, is used to extract high-level features. Specifically, to avoid checkerboard artifacts, we use bilinear interpolation for upsampling. Considering the loss of details caused by the denoising module, we input the original images and the denoised images together to this branch to get reconstructing features ($RB_{feature}$). The input channel is set to 8 and the output channel is set to 12. Formally:

$$RB^{\in \mathbb{R}^{H,W,12}}_{feature} = RB_{net}([C_{raw}, I_{raw}]) \tag{5}$$

where [,] denotes the channel-wise concatenation operation.

2) TEXTURE ENHANCING BRANCH

The TEB aims at reducing artifacts and preserving high-frequency details which may be ignored in the RB net. The architecture of this branch can be seen in Fig. 1(c). In this branch, we use dense connection [8] and dilated convolutions [15] to make full use of multi-scale features and keep a large receptive field. Instead of using denoised images as input, we simply calculate the gradients of denoised images

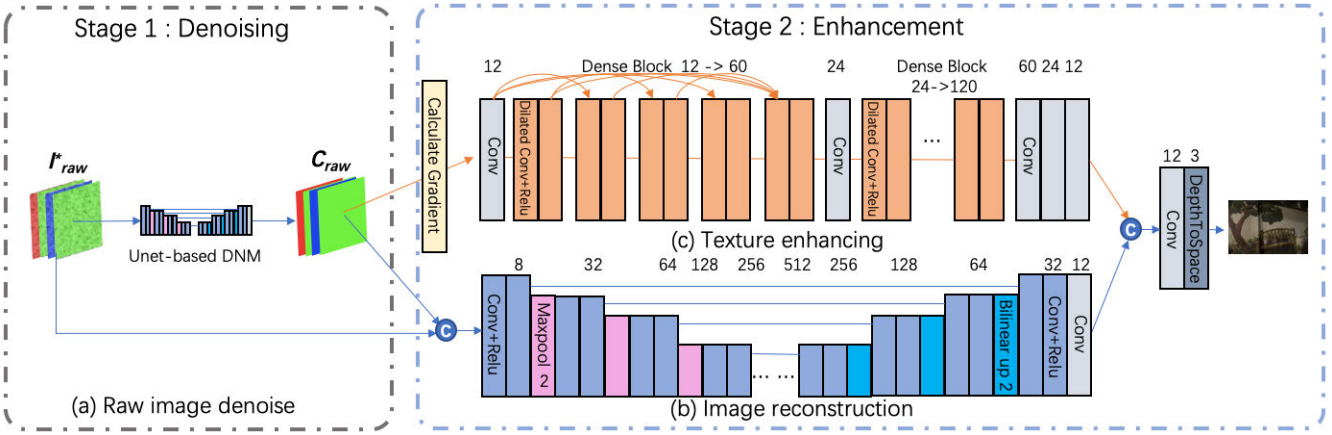


FIGURE 1. The architecture of our proposed LLISP. Our proposed LLISP consists of two stages: The first stage is responsible for denoising. In the second stage, the divide and conquer network is responsible for producing high-quality images in sRGB color space. The image reconstruction branch takes denoised raw data and original raw data as input to reduce color bias and recover image content. Using gradient information as input, the texture enhancing branch pays more attention to texture details and cooperates with the reconstruction branch to generate images with fewer artifacts.

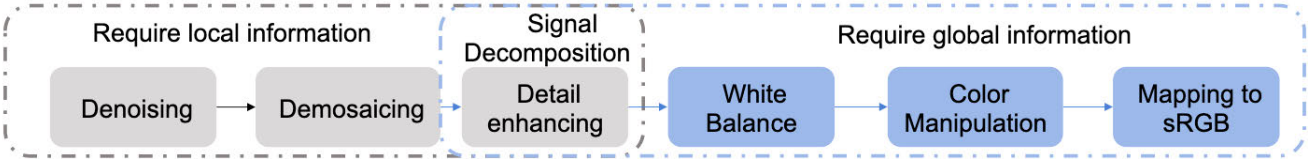


FIGURE 2. The key steps in the traditional image processing pipeline. Although different cameras may apply different algorithms in the detail enhancing step, most of them use frequency filters to decompose the signal into different layers.

as inputs(I_{TEB}). Formally:

$$I_{TEB}^{\in\mathbb{R}^{H,W,4}} = ||\nabla_h C_{raw}||_2^2 + ||\nabla_v C_{raw}||_2^2 \tag{6}$$

where∇_h and ∇_v denote the gradients along the horizontal and the vertical directions respectively. The input channel for TEB_{net} is set to 4 and the output channel is set to 12. Formally, the output of TEB can be written as (7)

$$TEB_{feature}^{\in\mathbb{R}^{H,W,12}} = TEB_{net}(I_{TEB}) \tag{7}$$

3) FUSION AND DEMOSAICING

After concatenating the features generated from the above two branches, we use convolution layers and a sub-pixel layer [35] to fuse them and up-sample data to the original resolution. The final output O_{RGB} is written as (8)

$$O_{RGB} = FD([TEB_{feature}, RB_{feature}]) \tag{8}$$

where [,] denotes the channel-wise concatenation operation. We train the Enhancing Net using l_1 distance defined as $Loss_{EN}$

$$Loss_{EN} = ||O_{RGB} - GT_{sRGB}||_1 \tag{9}$$

IV. EXPERIMENTS

A. DATASET

We adopt the Sony set in [26]. This set is captured by Sony $\alpha 7IIS$. It includes 2697 raw short-exposure images and 231 long-exposure images. The resolution of images is 4280 × 2832. The exposure time for low-light images is set between 1/30 and 1/10 second and the corresponding long-exposure ground truth images are captured with 100 to 300 times longer. We use the same training and testing set following [26]. In their public dataset, approximately 20% of the images with different exposure time are selected to form the test set.

B. IMPLEMENTATION DETAILS

Our proposed framework is implemented with Pytorch and an Nvidia TITAN-V GPU is used in experiments. The architecture of the denoising module is listed in Table 1, and the architecture of the enhancement net can be seen in Fig. 1. We train the denoising module with a learning rate 10^{-4} for 2k epochs. Then, we fix the weights of the denoising module and train the Enhancing Net for 3k epochs using ADAM [36] optimizer. The learning rate is set to 10^{-4} and is reduced to 10^{-5} after 1500 epochs. We randomly crop 512 × 512 patches for training and apply random flipping and rotation for data augmentation. Following Chen *et al.* [26], we subtract the black level and divide the maximal pixel value to map the data between 0 and 1. It takes 30 hours to train the whole net in which about 10 hours are used for pretraining. It takes about 0.5s to process one full-resolution image (4280 × 2832). Our code is available at https://github.com/Aacrobat/LLISP.

C. AMPLIFICATION RATIO k

The amplification ratio determines the brightness of the outputs. In our network, we firstly scale the low-light raw data

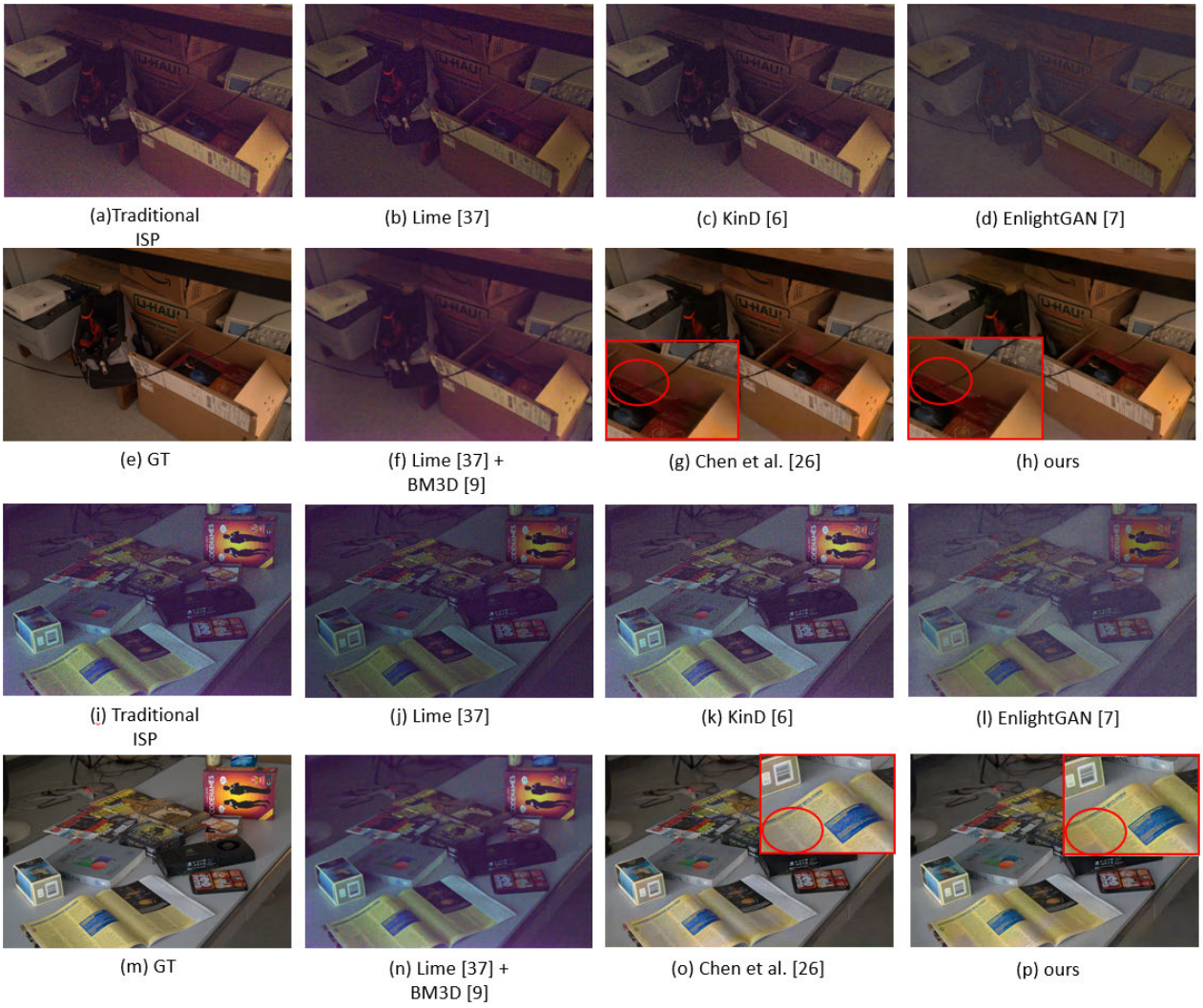


FIGURE 3. Qualitative results of state-of-the-art methods and our proposed LLISP evaluated on the SID test set. As we can see, the traditional ISP breaks down in extremely dark conditions, and most existing enhancing methods cannot reconstruct images successfully. Focusing on severe noise and the extremely dark conditions, both Chen *et al.* [26] and our method get much better results. Compared with Chen *et al.*, our method can recover color distortion accurately and suppress artifacts.

by the desired amplification ratios. This is similar to the ISO gain in cameras. During the training stage, the amplification ratios are set to be the difference between the exposure time for inputs and their ground truth images. During the test stage, users can adjust the brightness of the output images by setting different amplification factors. In Fig. 4, we show the effect of the amplification factors on images captured by smartphones.

By choosing different amplification ratios, we can test the amplification range in which our method can produce high-quality results. Images with different exposure time and different amplification ratios are fed into the network. As shown in Fig. 5, longer exposure time and smaller amplification ratios will produce better results. Our method can reconstruct high-quality results with an amplification ratio up to 100. However, the enhanced results with an amplification ratio of 300 still suffer from color bias and blur.

D. QUALITATIVE EVALUATION

We firstly compare our model with the traditional ISP. We use the in-camera auto-bright to kindle the dark inputs. As we can see in Fig. 3(a,i), in extremely dark conditions, the traditional ISP breaks down. Most existing low-light enhancement methods [6], [7], [37] only focus on adjusting illumination without considering noise and other degradations. It can be seen in Fig. 3(b-d,j-l), heavy noise and color bias seriously spoil the enhanced results. Applying an existing denoising algorithm [9] after the enhanced images cannot produce promising results, which can be seen in Fig. 3(f,n). Taking heavy noise into consideration, Chen *et al.* [26] and our method start from raw data and get much better results. Compared with Chen *et al.*, our method can recover color distortion accurately and suppress artifacts.

Since previous methods designed for JPEG-format images cannot handle extremely dark images, we mainly compare

TABLE 1. The architecture of the denoising module.

Input name	Input channels	Operator	Kernel	Output name	Output channels
in1	4	Conv&Relu	(3,3)	out1	32
out1	32	Conv&Relu	(3,3)	out2	32
out2	32	Maxpool	(2,2)	out3	32
out3	32	Conv&Relu	(3,3)	out4	64
out4	64	Conv&Relu	(3,3)	out5	64
out6	64	Maxpool	(2,2)	out7	64
out7	64	Conv&Relu	(3,3)	out8	128
out8	128	Conv&Relu	(3,3)	out9	128
out9	128	Maxpool	(2,2)	out10	128
out10	128	Conv&Relu	(3,3)	out11	256
out11	256	Conv, Relu	(3,3)	out12	256
out12	256	Maxpool	(2,2)	out13	256
out13	256	Conv&Relu	(3,3)	out14	512
out14	512	Conv&Relu	(3,3)	out15	512
out15	512	Conv	(2,2)	out16	256
out16&out12	512	Conv&Relu	(3,3)	out17	256
out17	256	Conv&Relu	(3,3)	out 18	256
out18	256	Conv	(2,2)	out 19	128
out19&out8	256	Conv&Relu	(3,3)	out20	128
out20	128	Conv&Relu	(3,3)	out 21	128
out21	128	Conv	(2,2)	out22	64
out21&out4	128	Conv&Relu	(3,3)	out22	64
out22	64	Conv&Relu	(3,3)	out23	64
out23	64	Conv	(2,2)	out24	32
out24&out1	64	Conv&Relu	(3,3)	out25	32
out25	32	Conv&Relu	(3,3)	out26	32
out26	32	Conv&Relu	(3,3)	out27	4

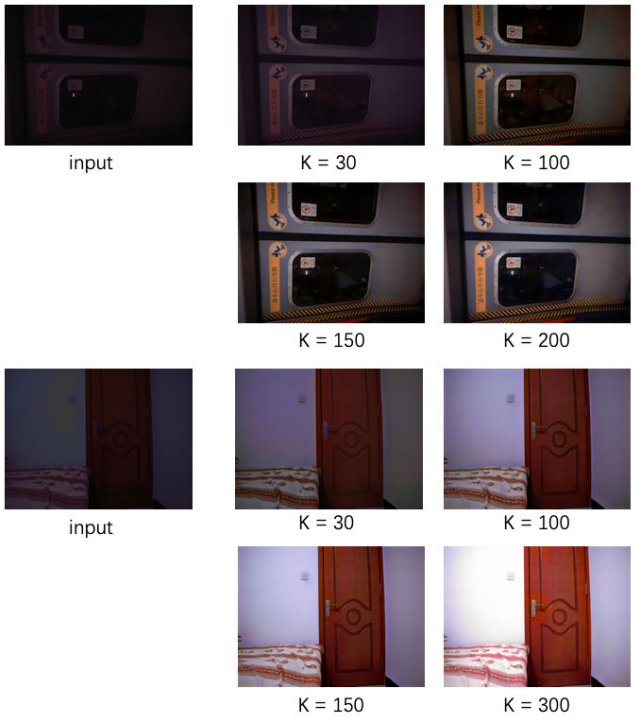


FIGURE 4. The effect of different amplification ratios on the same images captured by smartphones.

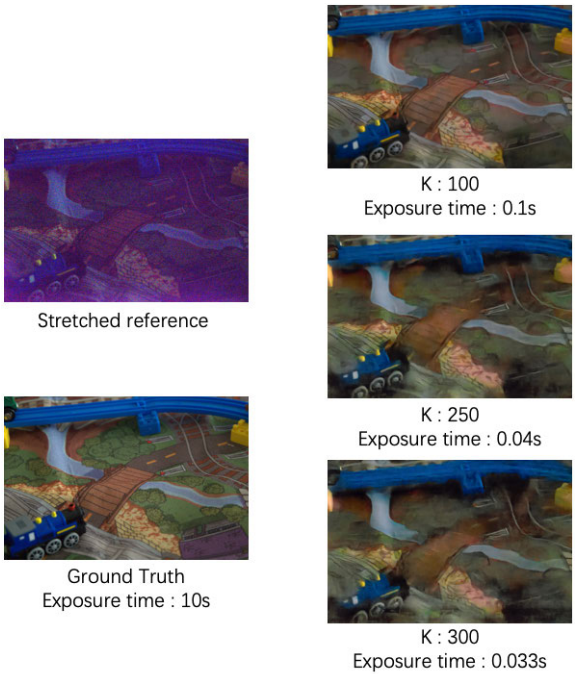


FIGURE 5. The effect of different amplification ratios on images with different exposure time. The images were chosen from the SID test set.

with Chen *et al.* [26] to show our improvements in detail. It can be seen in Fig. 6a, because of the heavy noise, it is easy to produce artifacts during the enhancement. Owing to the denoising module and the texture enhancing branch, we can reduce artifacts during enhancing and produce more realistic images. Fig. 6b and Fig. 6c show that our method can correct color bias and preserve details.

As shown in Fig. 7, we test our model on three common cameras. We can see that, when there is a domain gap between training and testing data, our two-stage model has a stronger generalization ability. By using the denoising module, we can get clearer results (the third row of Fig. 7), and eliminate the influence of noise on white balance (the first row of Fig. 7). Thanks to our effective two-branch enhancing module, our results can preserve more details (the second row of Fig. 7).

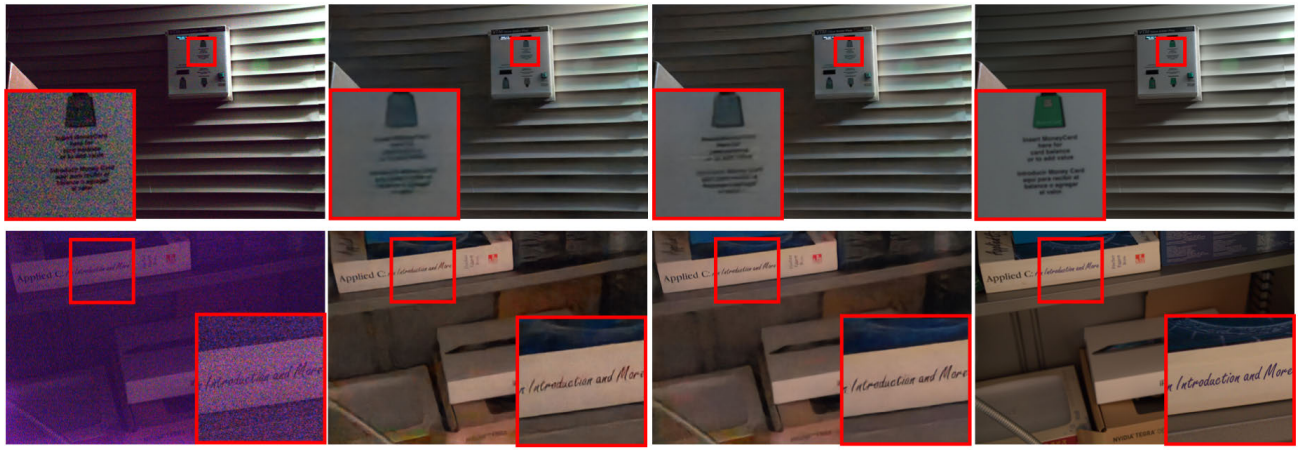
E. QUANTITATIVE EVALUATION

In this section, we compare our approach with the state-of-the-art methods [6], [7], [26], [28], [37]–[39]. We also use the existing denoising method BM3D [9] post-hoc to the results produced by Lime [37]. Besides, a baseline that simply duplicates the U-net is introduced. The first U-net learns to denoise the low-light raw data, and the second U-net learns to map raw data to sRGB outputs.

Table 2 reports quantitative results for different low-light enhancing methods. It can be seen from the first five rows, the traditional ISP cannot handle extremely dark scenes. Using the spoiled sRGB images produced by traditional ISP as inputs, most existing enhancing methods cannot remove heavy noise and color bias. It is necessary to begin with raw data and suppress the heavy noise. Our baseline outperforms CAN and Chen *et al.*, which means that simply denoising the data before enhancing it is very helpful for extremely low-light image enhancement tasks. Thanks to our

(a) Removing artifacts. From left to right: Stretched reference, Chen et al. [26], Ours, Ground truth

(b) Correcting color bias. From left to right: Stretched reference, Chen et al. [26], Ours, Ground truth



(c) Enhancing details. From left to right: Stretched reference, Chen et al. [26], Ours, Ground truth

FIGURE 6. **Qualitative results for our proposed LLISP. As we can see, our method can accurately reconstruct low-light images.**

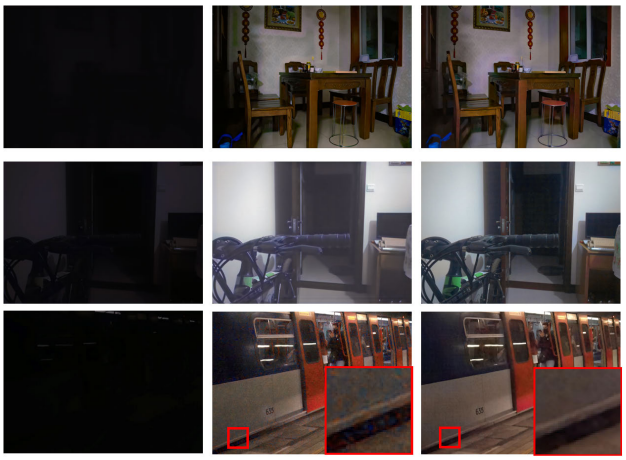
effective two-branch Enhancement Net, we further improve the accuracy from 29.18/0.815 to 29.68/0.832 with respect to PSNR and SSIM. We also employ the LPIPS metric [40] to measure perceptual distance. Higher distance means further different and lower means more similar. As we can see from Table 2, in terms of SSIM and LPIPS, our proposed method outperforms the state-of-the-art methods by a large margin. The experimental results demonstrate we can achieve state-of-the-art results both in pixelwise distance and perceptual similarity.

F. ABLATION STUDY

Ablation experiments are performed in order to have a better understanding of our model and prove the indispensability of each module.

1) DENOISING MODULE

In this part, we show the importance of the DNM and compare the impact of different architectures and loss functions for this module. A single network can theoretically complete denoising and color space conversion at the same time. But heavy noise affects accurate color reconstruction and it is difficult for networks to optimize both tasks at the same time. Learning denoising and color reconstruction in separate stages improves the final accuracy. As we can see from the second row of Table 3, we use the state-of-the-art denoising model RNAN [31] and retrain it using our dataset for denoising. However, due to the large memory consumption of the non-local module, we have to chop the input images into blocks which will result in uneven brightness and poor results. Note that although the addition of TV regularization term leads to



(a) inputs (b)Chen et al. [26] (c) ours

FIGURE 7. Qualitative results of state-of-the-art methods and our proposed LLISP evaluated on daily used cameras (Canon Eos 80 D: 1st row, iPhone7: 2nd row, Huawei meta20: 3rd row.

TABLE 2. Quantitative evaluation of low-light image enhancement algorithms in terms of PSNR/SSIM/MAE/NIQE/LPIPS. The best results are highlighted in bold. Note that a ∗ indicates that we use the PSNR, SSIM and LPIPS values reported in their original papers.

Method	PSNR	SSIM	LPIPS	MAE	NIQE
Traditional ISP	18.23	0.674	1.083	0.135	8.077
Lime [37]	17.76	0.351	1.132	0.142	8.071
KinD [6]	18.070	0.205	1.327	0.122	8.063
EnlightingGAN [7]	18.148	0.364	1.107	0.120	8.052
Lime [37]+BM3D [9]	17.90	0.361	1.045	0.079	7.998
CAN [38]	27.40	0.792	–	–	–
Chen et al. [26]	28.88	0.787	0.476	0.030	8.044
baseline	29.18	0.815	0.437	0.028	8.002
Ke Xu et al.* [39]	29.56	0.7991	–	–	–
EEMEFN* [28]	29.60	0.791	0.458	–	–
LLISP	**29.68**	**0.832**	**0.409**	**0.027**	**7.880**

TABLE 3. Ablation study on the denoising module. The results are in terms of PSNR/SSIM. We also compare the L1 distance between denoised images and corresponding ground truths in denoising stage. The best results are highlighted in bold.

Method	L1 distance for stage I	PSNR	SSIM
w/o denoising module	-	29.01	0.799
Using RNAN to denoise	**0.005**	29.17	0.815
w/o TV loss	0.006	29.45	0.823
LLISP	0.01	**29.68**	**0.832**

higher l_1 error between denoised images and corresponding ground truths in the denoising stage, the smoothened images with TV loss can help subsequent enhancements and thus obtain better results.

2) TEXTURE ENHANCING BRANCH

In this part, we show the indispensability of the TEB and compare different types of inputs for this branch. An interesting result is shown in the third row of Table 4. If we input the original images into the TEB, the final results are even worse than removing this branch, which indicates that the improvement of this branch is not because of increased parameters but because of more reasonable utilization of gradient features. We have also tried to use a simple edge detection algorithm such as Canny to extract the edges of denoised images and input them to the network. However, the edge detection algorithm will ignore the texture details and only retain the edge information, which is not conducive to texture enhancement and artifact removal.

TABLE 4. Ablation study on the texture enhancing branch. The results are in terms of PSNR/SSIM. The best results are highlighted in bold.

Method	PSNR	SSIM
w/o texture enhancing branch	29.33	0.816
Using denoised image as input	29.43	0.818
Using orignal image as input	28.92	0.804
Using edge as input	29.40	0.823
LLISP	**29.68**	**0.832**

TABLE 5. Ablation study on the reconstruction branch. The best results are highlighted in bold.

Method	PSNR	SSIM
Using denoised image as input	29.49	0.826
LLISP	**29.68**	**0.832**

3) RECONSTRUCTION BRANCH

As shown in Table 5, due to the loss of details caused by the denoising process, putting the original images and the denoised images into the network together can obtain better results.

V. CONCLUSION

In this paper, we present a novel low-light enhancement method LLISP. Inspired by the traditional ISP, our network firstly focuses on image denoising, and then finishes other image processing steps by a two-branch enhancement net. Extensive experiments depict the effectiveness and indispensability of different modules of the network. The proposed method is not only applicable to the training dataset but also applicable to raw data captured by different devices.

REFERENCES

[1] X. Zhu, Y. Wang, J. Dai, L. Yuan, and Y. Wei, "Flow-guided feature aggregation for video object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 408–417.

[2] W. Wang, X. Wu, X. Yuan, and Z. Gao, "An experiment-based review of low-light image enhancement methods," *IEEE Access*, vol. 8, pp. 87884–87917, 2020.

[3] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–128, Dec. 1977.

[4] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.

[5] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5372–5384, Dec. 2013.

[6] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 1632–1640.

[7] 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," 2019, *arXiv:1906.06972*. [Online]. Available: https://arxiv.org/abs/1906.06972

[8] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.
[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[10] J. Zhang, C. An, and T. Nguyen, "Deep joint demosaicing and super resolution on high resolution bayer sensor data," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 619–623.
[11] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 769–777.
[12] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, and N. Mazur, "A novel fast single image dehazing algorithm based on artificial multiexposure image fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–23, 2021.
[13] M. Zheng, G. Qi, Z. Zhu, Y. Li, H. Wei, and Y. Liu, "Image dehazing by an artificial image fusion method based on adaptive structure decomposition," *IEEE Sensors J.*, vol. 20, no. 14, pp. 8062–8072, Jul. 2020.
[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.
[15] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: https://arxiv.org/abs/1511.07122
[16] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *J. Digit. Imag.*, vol. 11, no. 4, pp. 193–200, Nov. 1998.
[17] H. D. Cheng and X. J. Shi, "A simple and effective histogram equalization approach to image enhancement," *Digit. Signal Process.*, vol. 14, no. 2, pp. 158–170, Mar. 2004.
[18] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 593–600, May 2007.
[19] L. M. I. Leo Joseph and S. Rajarajan, "Reconfigurable hybrid vision enhancement system using tone mapping and adaptive gamma correction algorithm for night surveillance robot," *Multimedia Tools Appl.*, vol. 78, no. 5, pp. 6013–6032, Mar. 2019.
[20] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.
[21] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.
[22] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.
[23] W. Wang, Z. Chen, X. Yuan, and X. Wu, "Adaptive image enhancement method for correcting low-illumination images," *Inf. Sci.*, vol. 496, pp. 25–41, Sep. 2019.
[24] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6849–6857.
[25] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf.*, Newcastle, U.K.: Northumbria Univ., Sep. 2018, p. 155.
[26] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.
[27] C. Chen, Q. Chen, M. Do, and V. Koltun, "Seeing motion in the dark," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3184–3193.
[28] M. Zhu, P. Pan, W. Chen, and Y. Yang, "Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network," in *Proc. AAAI*, 2020, pp. 13106–13113.
[29] F. Agostinelli, M. R. Anderson, and H. Lee, "Adaptive multi-column deep neural networks with application to robust image denoising," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 1493–1501.
[30] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1712–1722.
[31] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," in *Proc. ICLR Poster)*. [Online]. Available: OpenReview.net, 2019.
[32] E. Dubois, "Filter design for adaptive frequency-domain bayer demosaicking," in *Proc. Int. Conf. Image Process.*, Oct. 2006, pp. 2705–2708.
[33] Y. Hu, B. Wang, and S. Lin, "FC^4: Fully convolutional color constancy with confidence-weighted pooling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 330–339.
[34] E. Schwartz, R. Giryes, and A. M. Bronstein, "DeepISP: Toward learning an end-to-end image processing pipeline," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 912–923, Feb. 2019.
[35] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent., ICLR*, San Diego, CA, USA, May 2015.
[37] X. Guo, Y. Li, and H. Ling, "LIME: low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.
[38] Q. Chen, J. Xu, and V. Koltun, "Fast image processing with fully-convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2516–2525.
[39] K. Xu, X. Yang, B. Yin, and R. W. H. Lau, "Learning to restore low-light images via decomposition-and-enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2281–2290.
[40] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

HONGJIN ZHU received the B.S. degree in software engineering from the University of Sichuan of China, in 2019. She is currently pursuing the M.S. degree in computer engineering with Peking University. Her research interests include image processing, computer vision, and deep learning.

YANG ZHAO (Member, IEEE) received the B.E. and Ph.D. degrees from the Department of Automation, University of Science and Technology of China, in 2008 and 2013, respectively. From September 2013 to October 2015, he was a Postdoctoral Fellow with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, China. He is currently a Research Associate Professor with the School of Computer and Information, Hefei University of Technology. His research interests include image processing and pattern recognition.

RONGJIE WANG received the B.S. degree in mathematics and applied mathematics from Heilongjiang University, China, in 2006, and the M.S. and Ph.D. degrees in applied mathematics and computer applied technology from the Harbin Institute of Technology, China, in 2009 and 2019, respectively. His research interests include genome compression, deep learning in bioinformatics, and image/video processing.

RONGGANG WANG (Member, IEEE) received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, in 2006.

From 2006 to 2010, he was a Research Staff with Orange (France telecom) Labs. He is currently a Professor with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School. He has authored more than 60 papers in international journals and conferences, and held more than 40 patents. His research interest includes video coding and processing. He has done many technical contributions to ISO/IEC MPEG and China AVS. He leaded the MPEG Internet Video Coding (IVC) Standard. Since 2012, he has been serving as the Co-Chair for MPEG IVC AHG. Since 2015, he has been serving as the AVS Implementation Sub-Group Co-Chair.

WEIQIANG CHEN received the Ph.D. degree from the Faculty of Computing, Harbin Institute of Technology, in 1998. He is currently the Senior Vice President with Hisense Group Company Ltd. His research interests include artificial intelligence and multimedia computing.

XUESONG GAO is currently pursuing the Ph.D. degree with the College of Intelligence and Computing, Tianjin University. He is an Adjunct Professor with Shandong University. He is also an Executive Deputy Director with the State Key Laboratory of Digital Multimedia Technology, Hisense Group Company Ltd. His research interests include artificial intelligence and privacy computing.