Low-Light Hyperspectral Image Enhancement

Xuelong Li, Fellow, IEEE, Guanlin Li, and Bin Zhao, Member, IEEE

Abstract—Due to inadequate energy captured by the hyperspectral camera sensor in poor illumination conditions, low-light hyperspectral images (LHSIs) usually suffer from low visibility, spectral distortion, and various noises. A range of hyperspectral image (HSI) restoration methods have been developed, yet their effectiveness in enhancing low-light HSIs is constrained. This work focuses on the low-light HSI enhancement task, which aims to reveal the spatial-spectral information hidden in darkened areas. To facilitate the development of low-light HSI processing, we collect an LHSI dataset of both indoor and outdoor scenes. Based on Laplacian pyramid decomposition and reconstruction, we developed an end-to-end data-driven low-light HSI enhancement (HSIE) approach trained on the LHSI dataset. With the observation that illumination is related to the low-frequency component of HSI, while textural details are closely correlated with the high-frequency component, the proposed HSIE is designed to have two branches. The illumination enhancement branch is adopted to enlighten the low-frequency component with reduced resolution. The high-frequency refinement branch is utilized for refining the high-frequency component via a predicted mask. In addition, to improve information flow and boost performance, we introduce an effective channel attention block (CAB) with residual dense connection, which served as the basic block of the illumination enhancement branch. The effectiveness and efficiency of HSIE both in quantitative assessment measures and visual effects are demonstrated by experimental results on the LHSI dataset. According to the classification performance on the remote-sensing Indian Pines dataset, downstream tasks benefit from the enhanced HSI. Datasets and codes are available at https://github.com/guanguanboy/HSIE.

Index Terms—Denoising, hyperspectral images (HSIs), Laplacian pyramid, low-light enhancement.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) is composed of substantial discrete bands for each spatial pixel. Therefore, it contains more copious information than a natural image, which benefits massive applications in HSI fusion [1], classification [2], [3], [4], [5], [6], remote sensing [7], change detection [8], [9], visual question answering [10], and so on. In particular, hyperspectral imaging technology is increasingly employed for outdoor surveillance and environmental monitoring [11], such as real-time water quality and atmospheric pollution monitoring, which are highly significant for environmental safety. This type of full-day surveillance requires hyperspectral cameras to capture high-quality HSIs even at night. However, owing to insufficient light reaching hyperspectral camera sensors at night, the captured HSIs are often affected by poor visibility, spectral distortion, and varied noises (emphasis on Gaussian, impulse, and stripe). These degradations bury the useful spatial and spectral signals of the captured HSIs, which consequently affect the performance of the aforementioned HSI applications. To alleviate such degradations, it will be beneficial to take advantage of more advanced HSI imaging hardware devices equipped with specialized photographic techniques which are not easily affordable. However, even with advanced HSI imaging devices, it is still hard to prevent the presence of noises and spectral distortion. Therefore, it is essential to design an effective algorithm to solve the low-light HSI (LHSI) enhancement problem. This work focuses on low-light HSI enhancement, which aims to reduce spectral distortion, suppress noises, and reveal hidden information in low light simultaneously.

The most straightforward strategy for enhancing low-light HSIs is to process the HSI through band scanning with a natural image enhancement method. In line with this strategy, traditional model-driven methods such as methods using statistical characteristic [12], [13] and methods with Retinex theory [14], [15] can be directly employed for low-light HSI enhancement. Apart from traditional model-driven methods, deep-learning-based methods, such as SID [16], Retinex-Net [17], EnlightenGAN [18], and DRBN [19], can also be utilized to solve this problem. However, without considering the correlation between different HSI bands and noise suppression, this strategy usually results in spectral distortion and amplified noises. This statement can be verified by the comparison results in Section IV-D, where the low-light HSI enhancement results with many methods of this strategy are reported. Another strategy is adopting well-studied HSI denoising algorithms such as BM4D [20], total variation-based (TV) methods (LRTV) [21], and low rank methods [22], [23]. However, these model-driven HSI denoising methods concentrate on denoising only and ignore promoting HSI contrast and visibility. Most recently, the success of deep learning also stimulates the development of HSI denoising [24], [25], [26], and [27]. Existing deep-learning-based methods usually perform well on HSI denoising tasks. However, they are not perfectly suitable to enhance low-light HSIs, since they are not considering the intrinsic properties of low-light HSIs. Therefore, how to leverage the intrinsic properties of low-light HSIs is a fundamental problem for enhancement tasks.

The intrinsic properties of a low-light HSI are required to be exploited for designing an appropriate enhancement approach. Here, we discover two important intrinsic properties of a low-light HSI according to its statistical characteristics. On the one hand, in a low-light HSI, domain-specific attributes,
such as illumination, are mainly related to low-frequency components, while the details of textures are relevant to high-frequency components [28]. As depicted in Fig. 1, we capture paired HSIs in the same scene with short and long exposure times, respectively. In Fig. 1, the left column displays the HSIs captured in low-light conditions, while the right column shows their counterpart in the normal-light condition. The mean squared errors between the low-frequency components [see Fig. 1(c)] under the two conditions are about five times greater than those between the high-frequency [see Fig. 1(b)] components. Paying attention to the histograms and pseudo-color visual appearance of the paired HSIs, a similar conclusion can be drawn. On the other hand, since information among different bands of the HSI is redundant and complementary, we conclude that the averaged high-frequency component of adjacent bands in low-light HSIs contains more textural information than that of a single band. Hence, we can restore the lost textures with low luminance values by averaging high-frequency components, leaving only textures with high luminance values to recover. This inspires us that the averaged high-frequency components of adjacent bands in low-light HSIs are stronger prior than the high-frequency component of a single band.

Based on the findings, we propose the low-light HSI enhancement (HSIE) model. This model can first enlighten the dark areas of a low-light HSI. Then, it is capable of suppressing various noises and keeping spectral fidelity at the same time. Specifically, we build a two-branch network, of which the overall structure is depicted in Fig. 2. The illumination enhancement branch aims to enlighten the low-frequency component of a low-light HSI. The high-frequency refinement branch intends to restore the textural details. The illumination enhancement branch consists of three submodules. The first submodule is responsible for extracting multiscale spatial and spectrum features of HSIs. Then, the second submodule is used to enlighten the dark areas and remove various noises in low-light HSIs. Finally, the purpose of the third submodule is to reconstruct the low-frequency component of a low-light HSI. We restore the high-frequency components of the low-light HSI through the high-frequency refinement branch. For the sake of efficiency, the high-frequency refinement branch is a lightweight network. It is composed of three cascaded residual blocks to predict a mask with which the textual details of HSIs can be adaptively adjusted. In addition, the input of the high-frequency refinement branch is set as the average of the high-frequency components of adjacent bands instead of a single band. This design aims to sufficiently leverage the complementary properties among adjacent bands of HSIs and ease the learning of mapping between high-frequency components.

Below is an overview of our major contributions.

1) We present a two-branch low-light HSIE network based on Laplacian pyramid decomposition and reconstruction, with two intrinsic properties of low-light HSIs taken into account. HSIE can effectively boost the brightness of low-light HSIs, suppress noises, and efficiently keep spectral fidelity.

2) We have gathered a new LHSI dataset containing both indoor and outdoor scenes. To our knowledge, the LHSI dataset is the first to allow for the training and testing of low-light HSI enhancement approaches.

3) Promising results are achieved on the new LHSI dataset, which confirms the effectiveness of the proposed HSIE. Furthermore, some HSI classification experiments on the Indian Pine dataset are conducted to show that low-light HSIs preprocessed by the proposed approach benefit the downstream tasks.

The remainder of this work is arranged in the following manner. In Section II, we overview three related research areas, such as low-light natural image enhancement, HSI denoising, and the application of the Laplacian pyramid in deep-learning-based algorithms. The proposed HSIE is discussed in depth in Section III. In Section IV, the experimental results and accompanying analysis of LHSI datasets are provided. In addition, the experimental results of a downstream classification task on the Indian Pine dataset and a denoising task on the Washington DC Mall (WDC) dataset are also illustrated in Section IV. Lastly, in Section V, we conclude this work.

II. RELATED WORK

A. Low-Light Natural Image Enhancement

In recent years, low-light natural image enhancement has achieved significant progress. There are essentially two types
of solutions to this problem: traditional model-driven methods, and deep-learning-based data-driven methods. Traditional model-driven methods are mainly based on statistical characteristics like [12][13] and Retinex theory [14][15],[29][30]. In Retinex-based methods, they first obtain an illumination component as well as a reflectance component through decomposing the observed low-light natural image. Then, according to the Retinex theory [31], the reflectance component holds steady under low-light scenes. Thus, the estimation of the illumination component dominates the enhancement results. Furthermore, in [32], the authors propose a classical variational Bayesian Retinex (VBR) method, which provides a new framework for Retinex from the perspective of Bayesian. However, most of the traditional methods only tackle the low-visibility problem without considering noise suppression, which leads to the amplification of noise in the enhanced results. Due to impressive performance gains and robustness over traditional methods, deep-learning-based low-light natural image enhancement approaches have attracted continuous attention [33]. Lore et al. [34] proposed LLNet made up of a deep autoencoder to simultaneously enlighten and denoise low-light natural images. Thereafter, a U-Shaped convolutional network is introduced by Chen et al. [16] to map the short-exposure image with the raw format to its full-exposed counterpart with sRGB format. In addition, Chen et al. established a low-light dataset named SID. An unsupervised pioneering work EnlightenGAN [18] was developed to eliminate the dependency on paired data. The generator of EnlightenGAN was based on an attention-guided U-Net, while the discriminator takes both global and local information into account [35] to guarantee the reality of the enhanced image. Recently, a lightweight yet effective low-light enhancement method RUAS was proposed by Liu et al. [36]. By considering the Retinex rule, RUAS introduced a label-free learning strategy to exploit low-light prior characteristics for the illumination map in a carefully designed compact search space. Above all, the low-light natural image enhancement methods only consider three channels (R, G, and B) and cannot guarantee the spectral consistency between bands in HSIs. This article focuses on designing a low-light HSI enhancement model, which intends to restore all bands of an HSI.

B. HSI Denoising

HSI denoising aims to distill clean data from its noisy counterpart [37]. HSI denoising methods can be mainly categorized into two types: traditional model-driven, and deep-learning-based data-driven. BM4D [20], an extended version of the well-known image denoising method BM3D [38], is a classical traditional model-driven method for HSI denoising. However, BM4D suffers from an unacceptable long processing time for HSIs with large spatial resolution. Yuan et al. [24] proposed a data-driven CNN model with residual learning (HSID-CNN). The HSID-CNN can simultaneously extract multiscale spatial and spectral joint features and achieve excellent performance. However, when the noise distribution is irregular among HSI bands, HSID-CNN fails to suppress noise efficiently due to a lack of feature representation. To achieve better denoising performance, Ma et al. [25] proposed an attention-based [39] enhanced nonlocal cascading network. Another shortcoming of HSID-CNN is that it needs to train different models for varying noise levels. To overcome this limitation, Liu et al. [36] introduced Partial-DNet with strong generalization ability, which estimates the noise map of every band of HSIs as guidance for adaptively fitting different datasets. These methods are usually trained with small patches (20 × 20) for the sake of training cost, thus these methods always have a limited receptive field, leading to performing well only on fixed distribution noises instead of inconsistent noises. To exploit the underlying global and local spatial–spectral characteristics of HSIs, Wei et al. [40] proposed QRNN3D, where RNN-based attention [41] was employed. The architecture of QRNN3D is a typical encoder–decoder model with a residual connection. Recently, Shi et al. [42] proposed 3-D-ADNet, which is an advanced version of HSID-CNN. To fully exploit the global correlation between spectral and spatial, 3-D-ADNet employed the self-attention mechanism and a multiscale structure. Although all of these methods can be immediately expanded to enhance low-light HSIs, neither one of them considers the intrinsic properties of low-light HSIs. Besides, they are designed for different purposes, and the HSI denoising model is mainly for denoising but not for enhancing the brightness of the HSI. Thus, the performance of these methods is limited when applied to enhance the low-light HSI. We intend to specially design an algorithm for enhancing low-light HSIs by taking the intrinsic properties of low-light HSIs into account.

C. Laplacian Pyramid

The classical hierarchical structure of the Laplacian pyramid benefits several deep-learning-based tasks such as high-resolution image translation [28], unsupervised image generation [43], and image super-resolution [44]. Denton et al. [43] proposed Laplacian generative adversarial networks, which aim to generate sharp images by training generative adversarial networks at each level of the Laplacian pyramid. Lai et al. [44] first generate multiple high-frequency residuals serving as different levels of the Laplacian pyramid and then reconstruct the final image gradually using the Laplacian pyramid. To alleviate the heavy computation burden when translating high-resolution images, Liang et al. [28] designed Laplacian pyramid translation network (LPTN). LPTN is composed of two main branches, of which one is used to translate the low-frequency component of an image with reduced resolution, while the other aims to refine the high-frequency component through a progressive masking strategy. The proposed HSIE differs from LPTN in that LPTN intends to translate natural RGB images, while HSIE concentrates on enhancing low-light HSIs.

III. PROPOSED APPROACH

A. Overview

To overcome the low-light HSI enhancement challenge, we present an end-to-end model called the HSIE network. Fig. 2 depicts the overall structure of the proposed HSIE. The output of HSIE can be formulated as

\[ I_E = H_{HSIE}(I_0, C_0) \]  

(1)
Inspired by the above properties of the Laplacian pyramid and low-light HSIs, we propose to enhance the low-resolution low-frequency component of the data cube, which is denoted as \( C_L \), where most pixels have an intensity value close to 0. At the same time, we carry out this decomposition operation on the low-rank properties of HSIs \( I \) to stabilize and speed up the convergence process. For the second branch, we learn a mask of the high-frequency component \( I_{Mean} \), which is applied in this branch by iterating over each band of the low-light HSI.

As depicted in Fig. 2, we first decompose \( I_0 \) into a Laplacian pyramid through the Laplacian decomposition module, obtaining a high-frequency component represented by \( I_H \) and a low-frequency component \( I_L \). The resolution of \( I_H \) is \( h \times w \), while the width and height of \( I_L \) are \( h/2 \) and \( w/2 \), respectively. The Laplacian decomposition module stands for a standard Laplacian decomposition process. Due to the invertible characteristic of the Laplacian pyramid, we can reconstruct the original image by sequenced mirror operations. According to Burt and Adelson [45], \( I_L \) is blurred by a Gaussian filter and reflects the global attributes of an image. Concurrently, \( I_H \) demonstrates detailed textures of the image where most pixels have an intensity value close to 0. At the same time, we carry out this decomposition operation on data cube \( C_0 \). As a result, we obtain \( C_H \) and \( C_L \), which denote the resolution-kept high-frequency component and the resolution-reduced low-frequency component of the data cube, respectively. To leverage the low-rank properties of HSIs [23], we start from the averaged \( I_{Mean} \) of \( I_H \) and \( C_H \) instead of \( I_H \) to refine the high-frequency component, since \( I_{Mean} \) involves parts of texture information lost in \( I_H \).

Inspired by the above properties of the Laplacian pyramid and low-light HSIs, we propose to enhance \( I_L \) to recover the illumination; meanwhile, refine \( I_{Mean} \) adaptively to reduce artifacts in reconstruction. The proposed HSIE model is therefore composed of two branches. For the first branch, we convert the low-resolution \( I_L \) to \( \hat{I}_L \) using a convolution network comprised of three modules: a shallow feature extraction (SFE) module, an enlightening module (EM), and a reconstruction module. Residual learning strategy [46], which was tested to be effective for restoration tasks, is applied in this branch to stabilize and speed up the convergence process. For the second branch, we learn a mask of the high-frequency component through a lightweight convolution network with the concatenation of \([I_{Mean}, up(I_L), up(\hat{I}_L)]\) as input, where \( up(\cdot) \) represents the bilinear upsampling operation. To refine the high-frequency component \( I_{Mean} \), the mask is applied to \( I_{Mean} \) by pixel-wise multiplication. We introduce the two branches in detail in Sections III-B and III-C.

### B. Illumination Enhancement Branch

1) **Shallow Feature Extraction**

The SFE module is composed of two submodules as shown in Fig. 3. The first submodule takes a low-frequency component \( I_L \) of the low-light band \( I_0 \) as its input and is composed of three different kernel sizes for 2-D convolution to extract multiscale spatial information. The three convolutions are executed concurrently and the extracted features are concatenated to form a feature map \( F_{SL} \). This process can be represented as

\[
F_{SL} = [H_{L3}(I_L), H_{L5}(I_L), H_{L7}(I_L)]
\]

where \( H_{L3}, H_{L5}, \) and \( H_{L7} \) denote 2-D convolution with kernel size \( 3 \times 3, 5 \times 5, \) and \( 7 \times 7 \), respectively. The second submodule takes a low-frequency component \( C_L \) of the low-light data cube \( C_0 \) as input. This is motivated by the fact that the redundant spectral information in an HSI is beneficial for
restoring HSI [24]. The second submodule extracts multiscale joint spatial–spectral representation and shares the same configuration as the first submodule. The extracted representation is then concatenated to form a representation \( F_{SC} \), the same size as the output of the first submodule. This process can be represented as

\[
F_{SC} = [H_{C3}(C_L), H_{C5}(C_L), H_{C7}(C_L)]
\]  

(3)

where \( H_{C3}, H_{C5}, \) and \( H_{C7} \) denote 2-D convolution with different kernel sizes executed on the low-frequency component \( C_L \) of the low-light data cube \( C_0 \).

Finally, \( F_S \) is obtained by concatenating the output of the two submodules. This process can be represented as

\[
F_S = \text{ReLU}([F_{SL}, F_{SC}])
\]  

(4)

where \( \text{ReLU} \) denotes the rectified linear units (ReLU) function.

The output feature map \( F_S \) is introduced directly to a convolution layer to half the channels and further extracts features. This operation can be denoted as

\[
F_0 = \text{ReLU}(H_{\text{Conv}}(F_S))
\]  

(5)

where \( H_{\text{Conv}} \) denotes a \( 3 \times 3 \) convolution layer with padding 1 to maintain the feature map’s spatial resolution. The feature map \( F_0 \) extracted by this operation is set as the input to the EM.

2) Enlightening Module: The EM is mainly composed of several channel attention blocks (CABs), as shown in Fig. 4 (top). Supposing we have \( N \) CABs, the output \( F_i \) of the \( i \)th CAB can be obtained by

\[
F_i = H_{\text{CAB},i}(F_{i-1}) = H_{\text{CAB},i}([H_{\text{CAB},i-1}(...(H_{\text{CAB},1}(F_0))...)])
\]  

(6)

where \( H_{\text{CAB},i} \) represents the operation of the \( i \)th CAB. \( H_{\text{CAB},i} \) is composed of several standard operations, namely 2-D convolution, global average pooling, and activation function. The details of the CAB will be discussed in the following sections. Following the extraction of deep features with a list of CABs, a feature fusion function is applied to integrate the output features from all of the previous CABs. This process can be denoted as

\[
F_D = H_{DF}([F_0, F_1, \ldots, F_N])
\]  

(7)

where \([F_0, F_1, \ldots, F_N] \) denotes the concatenation of feature maps. \( H_{DF} \) is a standard convolution with kernel size 1 \( \times \) 1, which is leveraged to integrate features from different levels.

3) Residual Dense Effective CABs: The detailed structure of a CAB is depicted in Fig. 4 (bottom). The CAB consists of several dense connected layers [47], an effective channel attention [48], module and a residual connection. In Fig. 4 (bottom), \( F_{n-1} \) and \( F_n \) symbolize the input and output feature maps of the \( n \)th CAB, respectively. Feature maps generated from the \( c \)th convolution layer of the \( n \)th CAB can be obtained by

\[
F_{n,c} = \text{ReLU}(W_{n,c} [F_{n-1}, F_{n,1}, \ldots, F_{n,c-1}])
\]  

(8)

where \( W_{n,c} \) denotes the shared parameters of the \( c \)th convolution layer of which the bias term is neglected to simplify the expression. The feature maps generated by the \((n-1)\)th CAB were concatenated, which is denoted as \([F_{n-1}, F_{n,1}, \ldots, F_{n,c-1}] \). There exist direct connections between the precedent CAB’s output feature map and every layer of the current CAB. In addition, each layer in the same CAB has direct links to all succeeding layers. This type of dense connection can strengthen feature propagation and encourage feature extraction.

An in-block feature fusion operation is followed to merge all the feature maps produced by each convolution layer and the output feature map of the proceeding CAB, which can be formulated as

\[
F_{n,F} = H_T([F_{n-1}, F_{n,1}, \ldots, F_{n,c}, \ldots, F_{n,C}])
\]  

(9)

where \( H_T \) represents a transition layer composed of a single convolution layer with kernel size 1 \( \times \) 1, which is used to decrease the dimension of the concatenated feature maps. This transition layer alleviates the training and computational burden of the whole network and makes the network easy to train.

After obtaining the fused feature map, an effective channel attention module is adopted to capture interchannel interaction efficiently. Assume we obtained a fused feature map \( F_{n,F} \in \mathbb{R}^{W \times H \times C} \), where \( W \) and \( H \) denote the feature map’s width and height, respectively, and the channel dimension of \( F_{n,F} \) is represented by \( C \). The weights of the channels in the effective channel attention module can be obtained by

\[
W_A = \sigma(H_{1D}(g(F_{n,F})))
\]  

(10)

where \( F_{n,F} \) is the output of the in-block feature fusion operation and \( g(F_{n,F}) = (1/WH) \sum_{i=1,j=1}^{W,H} F_{n,F} \) indicates global average pooling. The \( H_{1D} \) denotes 1-D convolution with kernel size 3, which avoids dimensionality reduction and captures channel attention in an efficient way. The \( \sigma \) is a Sigmoid function. We can obtain attention weighted feature map by

\[
F_{n,W} = F_{n,F} \odot W_A
\]  

(11)

where \( \odot \) denotes element-wise multiplication.

To further strengthen the representation ability of our model, the inner-block residual learning strategy is applied. The \( n \)th CAB’s final output may be written as

\[
F_n = F_{n,W} + F_{n-1}.
\]  

(12)
4) Reconstruction Module: We reconstruct the enlightened low-frequency component of the HSI band by a single $3 \times 3$ convolution layer denoted as $H_R$, which is used to restore the residual $I_R$ directly (see Fig. 2). This procedure can be represented as

$$I_R = H_R(F_D)$$

where $F_D$ denotes the EM’s output and $I_R$ represents the restored residual.

Finally, the restored low-frequency component is obtained by simply adding the restored residual $I_R$ and the input low-light band $I_L$, which is formulated as

$$\hat{I}_L = I_R + I_L \quad (14)$$

where $\hat{I}_L$ is the enlightened low-frequency component of an HSI band.

C. High-Frequency Refinement Branch

To reach a reliable reconstruction, we refine the averaged high-frequency component $I_{\text{Mean}}$ on the basis of $I_L$ and $\hat{I}_L$. We intend to generate a single-channel mask for the averaged high-frequency component $I_{\text{Mean}}$ and adjust $I_{\text{Mean}}$ using the generated mask in this branch.

To match the resolution of $I_{\text{Mean}} \in \mathbb{R}^{h \times w \times 1}$, we first upsample $I_L \in \mathbb{R}^{(h/2) \times (w/2) \times 1}$ and $\hat{I}_L \in \mathbb{R}^{(h/2) \times (w/2) \times 1}$. Then, we feed the concatenated $[I_{\text{Mean}}, I_L, \hat{I}_L]$ into a lightweight network composed of three residual blocks, where the detailed composition of the lightweight network is depicted in Fig. 2. A mask of $I_{\text{Mean}}$ is generated by the lightweight network, which is denoted as $I_{\text{Mask}} \in \mathbb{R}^{h \times w \times 1}$. We regard the mask as a global adjustment to the high-frequency component which is easier to be optimized than the images without decomposition. Hence, we adjust the $I_{\text{Mean}}$ by

$$\hat{I}_H = I_{\text{Mean}} \otimes I_{\text{Mask}}$$

where $\otimes$ indicates the pixel-wise multiplication. We can then reconstruct the resulting image $I_E$ using the refined $\hat{I}_H$ and the enhanced $\hat{I}_L$ according to the Laplacian pyramid reconstruction strategy. This can be formulated as

$$I_E = H_R(\hat{I}_H + \text{Upscale}(\hat{I}_L)) \quad (16)$$

where Upscale is an upsampling process, in which we first double the size of $\hat{I}_L$ by padding zero, then we convolve the resized image with a Gaussian kernel, which is the same kernel used in the Laplacian pyramid decomposition. $H_R$ is a simple same convolution layer to further refine the reconstructed band.

IV. EXPERIMENTAL ANALYSIS

A. Datasets

A new LHSI dataset is gathered for the development of low-light HSI enhancement methods. Each sample in the LHSI dataset contains a short-exposure low-light HSI captured by setting the exposure time to 1 ms and a long-exposure reference HSI captured with an exposure time of 15 ms. The long-exposure HSI is sufficient to serve as ground truth. The LHSI dataset contains both indoor and outdoor sceneries. A small fraction of samples in the LHSI dataset is shown in Fig. 5.

The indoor dataset is captured using the SPECIM FX10\footnote{Detailed information about the SPECIM FX10 can be found at https://www.specim.fi/ftx/}, which works in a line-scan mode and is equipped with a fixed platform. During image acquisition in indoor scenes, we use a halogen light source to stabilize the lighting conditions. To make the HSI images acquired by the camera have accurate reflectance values, we first use a whiteboard for reflectance calibration before starting the acquisition. After calibration, the camera automatically converts the digital number (DN) values to reflectance values and outputs the reflectance spectrum image. The indoor dataset contains six pairs of HSIs, of which five pairs are used for training and one pair is used for testing. The spatial resolution of each indoor HSI is 390 × 512, and each HSI has 224 bands with wavelengths ranging from 400 to 1000 nm. Since the first several bands and the last several bands are subjected to greater interference, we remove the first 20 bands and the last 12 bands. Among the remaining 192 bands, every three bands are equally spaced to choose one band, resulting in an HSI with the size of $390 \times 512 \times 64$, where 64 represents the band number.

The outdoor dataset contains 20 pairs of HSIs and is acquired using the SPECIM IQ\footnote{Detailed information about the SPECIM IQ can be found at https://www.specim.fi/iq/}, which is a compact scanning-based hyperspectral camera. To obtain an accurate image of the reflectance spectrum, we recalibrate the reflectance using a whiteboard whenever the lighting conditions change. The outdoor dataset is captured on campus and contains images of various objects, such as cars, trees, stones, and buildings. Each of the outdoor HSI has 204 bands with wavelengths ranging from 400 to 1000 nm and a spatial resolution of 512 × 512. For unification with indoor datasets, the first six bands and the last six bands are removed. Then, among the remaining 192 bands, we uniformly sample one band in every three bands, reserving about 64 bands. The final shape of the outdoor HSI is $512 \times 512 \times 64$. In the case of the
outdoor dataset, we randomly chose 80 percent of the samples for training and the rest 20 percent for testing.

A training sample is created according to the following procedure. We first normalize all the HSIs in the range of [0,1]. Then, for each low-light HSI, a $64 \times 64$ patch on each band without overlapping is cropped, and a $64 \times 64 \times 24$ cubic is cropped at the same position of its adjacent 24 bands. Following this preprocessing procedure, 48 training samples can be cropped from an HSI with spatial resolution $390 \times 512$. The label is cropped from the same position as the corresponding long-exposure HSI using the same preprocessing method. In the end, 152,000 training samples are obtained on the indoor dataset and 20,480 training samples on the outdoor dataset.

B. Implementation Details

Our model is implemented using the prevalent Pytorch framework and is trained with a single NVIDIA GeForce RTX 3090 graphics processing unit (GPU). The training epoch is set as 600. The Adam optimizer is adopted to optimize the proposed model, and the momentum parameters of the optimizer are set as $0.9, 0.999, and 10^{-8}$, respectively. We initialize the weights of the proposed model by the Kaiming initialization method [49]. The initial learning rate for training our model is set as $2 \times 10^{-4}$. We use a step learning rate scheduler that decreases in half for every 200 epochs. We take the $L_1$ loss function as the optimizing objective, which is proved to provide better convergence than the $L_2$ [50]. In our case, $L_1$ loss also performs better than $L_2$ loss in terms of noise suppression. A detailed discussion of $L_1$ and $L_2$ loss is presented in Section IV-F.

C. Evaluation Metrics

We use three objective assessment measures to quantitatively assess the proposed HSIE’s performance, including the mean peak-signal-to-noise ratio (MPSNR) [51], mean structural similarity (MSSIM) [51], and spectral angle mapper (SAM) [52]. MPSNR and MSSIM are metrics for spatial features of HSIs, SAM is used for evaluating spectral consistency. In general, better low-light HSI enhancement results are indicated by lower SAM and higher MPSNR and MSSIM values.

D. Experiments on Indoor Dataset

We compare the proposed HSIE with current mainstream HSI denoising methods such as ENCAM [25] and 3-D-ADNet [42]. Also, the proposed method is compared with classic low-light natural image enhancement methods such as HE [53], CLAHE [54], MSR [55], the McCann’s Retinex (MR) method [56], and Retinex-Net [17]. In the MSR method, the scale parameters are 15, 80, and 360, respectively. In the MR method, we set the number of iterations to 3. By regarding each HSI band as an independent grayscale image, it is straightforward to utilize the aforementioned classic image enhancement methods to enhance the low-light HSI band. For a fair comparison, all deep-learning-based algorithms are trained on our collected LHSI dataset.

Table I displays the quantitative evaluations of several comparison approaches. In Table I, the top performance is emphasized in bold, while the second best is underlined. In Table I, the MPSNR value of the 3-D-ADNet method is 25.268, which is 13.360 dB smaller than the MPSNR value of the proposed approach. Furthermore, as can be observed from Table I, our approach HSIE performs favorably against the second-best ENCAM in all three metrics. Compared with ENCAM, the accuracy of our method in MPSNR and MSSIM are increased by 5.919 dB and 0.0131, respectively, while the SAM is decreased by 0.9423. The results from 3-D-ADNet and ENCAM reveal that the current state-of-the-art HSI denoising approach is ineffective for low-light HSI enhancement. We conjecture that this is caused by ignoring the intrinsic properties of low-light HSIs. In addition, we can observe that data-driven neural network methods outperform model-driven traditional algorithms by a significant margin.

Fig. 6 shows visual comparisons of the pseudo-color enhancement results of all the comparison methods. From Fig. 6, we can see clearly that the model-driven image
To further demonstrate the superiority of our approach, Fig. 8 exhibits the PSNR curves evaluated on a validation low-light HSI in the indoor LHSI dataset with different comparison methods. The proposed HSIE outperforms other comparison methods in all bands.

The proposed model uses the current low-light band and its $k$ adjoining bands as inputs. We set the value of $k$ as 24 during all the training procedures on the indoor and outdoor LHSI datasets. A detailed discussion for the value of $k$ is provided in Section IV-F.

In addition, to show the variation of the spectral reflectance, the spectral distortion between the spectral curves in three positions obtained by the comparison models and ground truth is illustrated in Fig. 9. After whiteboard calibration, the cameras used can directly output both the DN value and the reflectance. We select the reflectance for processing, so the curves illustrated in Fig. 9 are computed based on spectral reflectance values. From Fig. 9, we can see that in most circumstances, the spectral curve generated by the proposed HSIE (red curve) overlaps the curve generated by the label (cyan curve) in all three positions. It implies that the proposed HSIE is the best to keep spectral fidelity.

E. Experiments on Outdoor Dataset

The collected outdoor LHSI dataset is employed to support the training and testing of HSIE to further verify its efficacy. The enhancement results of the specific outdoor low-light HSI using different algorithms are demonstrated in Fig. 10 for visual comparison. Fig. 11 demonstrates the magnified version of the highlighted red rectangle region of interest in Fig. 10(a). The result in Fig. 7(i) shows that the proposed approach gains the best result. It also demonstrates that the HSIE is capable of properly enlightening the dark regions of the low-light HSI while keeping textural details.

To further demonstrate the superiority of our approach, Fig. 8 exhibits the PSNR curves evaluated on the validation of low-light HSIs with different comparison methods. One can observe that the proposed HSIE gains the top performance among all bands. The excellent performance attributes to all components of HSIE. For one thing, the multiscale feature fusion mechanism in the SFE module adequately extracted the joint spatial and spectral information. For another thing, the concatenated CABs show the benefits of enlightening the low-light band. Finally, the high-frequency refinement branch successfully adjusts textural details of low-light HSIs.

The enhancement MR method is unable to enlighten the extreme dark areas of the HSI, which is consistent with the conclusion drawn from Table I. The results of the HE and MSR show that the two methods can successfully enlighten the original image. However, they are unable to suppress the noise and keep spectral fidelity. CLAHE is also capable of enlightening dark areas. However, it introduces obvious textural distortion. The result of Retinex-Net demonstrates that it can enhance the low-light HSI to some extent, but the generated image looks unrealistic. From Fig. 6(g), we can see clearly that the restoration result of the 3-D-ADNet introduces heavy artifacts and looks blurry. The enhanced image of ENCAM loses some textural details and looks blurry too. This result can be highlighted by the observation in Fig. 7(h). Fig. 7 demonstrates the magnified version of the highlighted red rectangle region of interest in Fig. 6(a). The result in Fig. 7(i) shows that the proposed approach gains the best result. It also demonstrates that the HSIE is capable of properly enlightening the dark regions of the low-light HSI while keeping textural details.

The collected outdoor LHSI dataset is employed to support the training and testing of HSIE to further verify its efficacy. The enhancement results of the specific outdoor low-light HSI using different algorithms are demonstrated in Fig. 10 for visual comparison. Fig. 11 demonstrates the magnified version of the highlighted red rectangle region of interest in Fig. 10(a). For better viewing, the darker images in Figs. 10 and 11 are linearly stretched using the Matlab `imadjust` function with consistent parameters. Unless specifically mentioned, images marked with an asterisk in this article denote that they are linearly stretched. From Fig. 10, we can draw a similar conclusion that the traditional HSI denoising algorithm LRMR is unable to enhance low-light HSIs, while deep-learning-based HSI denoising methods ENCAM and 3-D-ADNet can successfully enlighten the darkened HSIs. However, the enhanced result of ENCAM looks darker. This result demonstrates that ENCAM cannot improve the illumination sufficiently. The enhanced result of 3-D-ADNet introduces some strip artifacts on the whole image [see Fig. 11(h)], indicating that

---

1. https://ww2.mathworks.cn/help/images/ref/imadjust.html?lang=en
Fig. 9. Visual comparison of spectral distortion at pixel positions (100, 100), (200, 200), and (340, 340) on the validation HSI. The spectral curve generated by the proposed HSIE (in red) and by the label (in cyan) almost overlapped. Best viewed in color and zoomed in.

Fig. 10. Vision comparison of the pseudo-color enhanced results on the low-light HSI. Bands 57, 27, and 17 are selected to simulate red, green, and blue, respectively. (a) Low light. (b) LRMR. (c) HE. (d) CLAHE. (e) MSR. (f) Label. (g) ENCAM. (h) 3-D-ADNet. (i) HSIE (Ours). * denotes that the image is linearly stretched for better viewing. Best viewed in color and zoomed in.

Fig. 11. Vision comparison of the highlighted region of interest. (a) Low light. (b) LRMR. (c) HE. (d) CLAHE. (e) MSR. (f) Label. (g) ENCAM. (h) 3-D-ADNet. (i) HSIE (Ours). * denotes that the image is linearly stretched for better viewing. Best viewed in color and zoomed in.

3-D-ADNet cannot restore the textural details. These results once again proved that HSI denoising methods are not the perfect solutions for the low-light HSI enhancement problems. The most straightforward reason is that HSI denoising methods are not designed for low-light HSI enhancement. We can generate another conclusion from Figs. 10 and 11, where model-driven natural image enhancement methods (such as HE, CLAHE, and MSR) are competent to boost the illumination of low-light HSIs. However, these methods only focus on improving the illumination, resulting in high brightness while ignoring the preservation of spectral fidelity and noise suppression. Finally, when compared to other competing approaches, the proposed HSIE produced the best visual results.

We also list the quantitative results evaluated with the aforementioned metrics on four scenes in the outdoor LHSI dataset in Table II, where for each metric, the top performance is highlighted in bold and the suboptimal is underlined. The four scenes are with different light conditions and objects of various materials, which comprise stone, tree, building, and car, as demonstrated in Fig. 5(c)–(f), respectively. As shown in Table II, the top performance in all three metrics is achieved by the proposed HSIE in almost all four scenes, demonstrating that the proposed HSIE can generalize well in different light conditions.

F. Ablation Study

In this section, we first examine each of the HSIE module’s effectiveness, then we investigate the impact of some key model design hyperparameters, such as the adjoining spectral band number, and the CAB number in the EM. Finally, the results of the analysis of model complexity and the impact of different loss functions are demonstrated. Unless other-
TABLE II

| Methods | Car | Tree | Stone | Building |
|---------|-----|------|-------|----------|
| MR [56] | 30.360 | 0.9042 | 4.909 | 20.667 |
| LMR [22] | 29.059 | 0.9302 | 4.965 | 20.052 |
| HE [53] | 7.977 | 0.3677 | 17.708 | 6.880 |
| CLAIRE [54] | 32.609 | 0.9293 | 7.377 | 25.734 |
| MSER [55] | 4.428 | 0.2554 | 6.844 | 11.819 |
| Retinex-Net [17] | 7.384 | 0.4681 | 5.408 | 7.417 |
| ENCAME [25] | 33.665 | 0.9658 | 4.737 | 23.737 |
| 3D-ADNet [42] | 29.401 | 0.9382 | 2.021 | 22.491 |

Fig. 12. Ablation study on different settings of HSIE. Results are generated from our LHSI dataset. (a) Dimensions of feature maps in CAB. (b) CAB number of the EM. (c) Convolution layer number in CAB.

TABLE III

| Evaluation metric | MPSNR↑ | MSSIM↑ | SAM↓ |
|-------------------|--------|--------|------|
| RDN | 37.278 | 0.9727 | 2.061 |
| RDN+SFE | 37.494 | 0.9737 | 1.958 |
| RDN+SFE+CAB | 37.699 | 0.9746 | 1.749 |
| RDN+SFE+CAB+HIGH | 38.628 | 0.9794 | 1.390 |

TABLE IV

| Adjacent Band Number (k) | MPSNR↑ | MSSIM↑ | SAM↓ |
|--------------------------|--------|--------|------|
| k = 8 | 37.466 | 0.9715 | 1.660 |
| k = 12 | 37.885 | 0.9756 | 1.662 |
| k = 18 | 37.847 | 0.9756 | 1.666 |
| k = 24 (Ours) | 38.628 | 0.9794 | 1.390 |
| k = 36 | 37.175 | 0.9722 | 1.868 |

wise specified, the ablation experiments in this section are performed on the indoor LHSI datasets.

The results of the ablation investigation into the effects of each module are reported in Table III. A simplified residual dense network (RDN) [57] is adopted as our initial backbone, which is considered to be a very efficient module for image restoration. We observe that “RDN+SFE” outperforms “RDN”, which shows the benefit of using multiscale SFE. Another observation is that “RDN+SFE+CAB” performs more favorably than “RDN+SFE”, which indicates the advantage of employing CAB. Finally, our full approach “RDN+SFE+CAB+HIGH” achieves the best results on the LHSI dataset, where “HIGH” stands for the high-frequency refinement branch. The results in Table III show that each module can enhance restoration performance significantly since each module has a positive effect on information and gradient flow. Furthermore, the CAB can make use of cross-channel interaction to adaptively inhibit less useful features and stress significant features.

For the rest of this section, we concentrate on investigations into some key model design hyperparameters. The adjoining spectral band number $k$ plays a vital role in the whole enhancement process. Hence, we investigate $k$ to evaluate its influence. In Table IV, we demonstrate the investigation result of adjacent spectral band number $k$. The best performance is highlighted in bold. When the value of $k$ is set as 24, our model gains the best performance evaluated with MPSNR. In the meanwhile, the performance of the proposed HSIE decreases when the value of $k$ is larger or smaller than 24. We conjecture that the intrinsic association between distinct bands of an HSI is primarily responsible. The adjoining 24 bands of a single band are already sufficient to encode all the correlations along with spectral information. Less adjoining bands cause loss of information, while more adjoining bands introduce noisy information.

Fig. 12 investigates the impact of feature map dimensions in CAB, the CAB number of the EM, and the convolution layer number of CAB in Fig. 12(a)–(c), respectively. It can be seen that the MPSNR is positively correlated with feature map dimensions in CAB. While the feature map dimensions are set as 60, 90, and 180, the MPSNR reaches a comparable good performance. For the sake of model complexity, the feature map dimensions are set as 60 in the final model. For the CAB number, the MPSNR becomes saturated when it equals 4. Finally, Fig. 12(c) shows that the proposed HSIE gains the best performance when the convolution layer count equals 4 in CAB.

To evaluate the impact of different loss functions, the proposed HSIE is trained on an indoor LHSI dataset using $L_1$ loss and $L_2$ loss, respectively. Fig. 13 shows the MPSNR computed on the validation HSI at different epochs in the training process. The $L_1$ loss function leads to a quicker convergence speed than the $L_2$ loss function, which is another reason we chose the $L_1$ loss as the HSIE loss function. Furthermore, because the proposed HSIE is trained with the
Fig. 13. Investigation into different loss functions. The usage of the $L_1$ loss function results in a faster convergence speed and a more stable training process than the $L_2$ loss function.

TABLE V

| Methods          | Complexity (GFLOPs) |
|------------------|----------------------|
|                  | 200 × | 384 × | 512 × | 1024 × |
| ENCAM [25]       | 105.8  | N.A.  | N.A.  | N.A.   |
| 3D-ADNet [42]    | 16.1   | N.A.  | N.A.  | N.A.   |
| HSIE (w/o high)  | 16.6   | 61.2  | 108.8 | 435.2  |
| HSIE (w/ high)   | 13.1   | 48.4  | 85.9  | 343.9  |

$L_1$ loss function rather than the $L_2$ loss function, the training process is more stable.

To ensure that the proposed HSIE is efficient, a computation complexity analysis experiment is conducted on the HSIE and corresponding deep-learning-based models. The model complexity is assessed by giga floating-point operations per second (GFLOPs). As shown in Table V, four input spatial resolutions are exploited for the analysis, where the N.A. implies the algorithm cannot deal with the HSI of a certain spatial resolution on a single GPU with 24 G RAM. Results show that the proposed HSIE with a high-frequency refinement branch is more efficient than other methods on all four input spatial resolutions, and this superiority in computation efficiency becomes more and more obvious as the input spatial resolution increases.

G. HSI Classification Results

As mentioned before, one purpose of enhancing low-light HSIs is to boost the performance of downstream tasks. We designed a classification task on the enhanced HSIs using the classic support vector machine (SVM) algorithm [58] to certify the efficacy of the proposed HSIE. To carry out this experiment, we employ the classical remote-sensing Indian Pine dataset, which is first darkened by a reversed HSIE network. This network is trained by exchanging the input and the label used for training the original HSIE. The network is trained by exchanging the input and the label used for training the original HSIE. The enhanced results of varying methods on the darkened India Pine dataset are classified by SVM. We conduct this classification task without dimension reduction [59]. The classification results on different enhanced HSIs and the original darkened India Pine data are presented in Fig. 14. The visualization of classification results on the original darkened HSI appears to have more corrupted and discontinuous areas, as illustrated in Fig. 14(a). After the original darkened HSI is enhanced by the proposed HSIE, the number of corrupted areas is reduced and more continuous areas are turned up [depicted in Fig. 14(i)].

In this article, we adopt the Kappa coefficient [60] and overall accuracy (OA) to quantitatively evaluate the classification results, which are presented in Table VI. We can observe from Table VI that the Kappa and OA values of HSIE reach 0.8594% and 87.70%, respectively, which outperforms all the listed comparison methods.

H. HSI Denoising

To reveal the applicability of our HSIE approach, we conduct an HSI denoising experiment on the remote-sensing WDC dataset. We crop a cubic slice with shape $1000 \times 303 \times 191$ from the WDC dataset for training different methods and the rest for evaluation.

The results of the denoising performance on the WDC dataset are depicted in Table VII. The top performance is
We offer an LHSI dataset in this work, which is created to support the evolution of low-light HSI enhancement approaches. We developed an end-to-end two-branch deep-learning-based network, dubbed HSIE, to raise the quality of low-light HSIs. According to the decomposition strategy of the Laplacian pyramid, we first decompose the input HSI into a low-resolution domain-specific component and a textural-related component. Then, the illumination enhancement branch is applied to boost the illumination with reduced resolution, while the lightweight high-frequency refinement branch aims to improve the textural details via a predicted mask. Substantial experiments on the LHSI dataset indicate that the proposed approach is capable of reaching promising performance in both evaluation metrics and visual effect, while at the same time suppressing noise and maintaining spectral fidelity. In addition, the accuracy of a downstream HSI classification task using SVM on enhanced low-light HSI demonstrated the benefits of HSIE as a low-light HSI preprocessing tool. This work offers many opportunities for future investigation into low-light HSI processing.

V. Conclusion

We offer an LHSI dataset in this work, which is created to support the evolution of low-light HSI enhancement approaches. We developed an end-to-end two-branch deep-learning-based network, dubbed HSIE, to raise the quality of low-light HSIs. According to the decomposition strategy of the Laplacian pyramid, we first decompose the input HSI into a low-resolution domain-specific component and a textural-related component. Then, the illumination enhancement branch is applied to boost the illumination with reduced resolution, while the lightweight high-frequency refinement branch aims to improve the textural details via a predicted mask. Substantial experiments on the LHSI dataset indicate that the proposed approach is capable of reaching promising performance in both evaluation metrics and visual effect, while at the same time suppressing noise and maintaining spectral fidelity. In addition, the accuracy of a downstream HSI classification task using SVM on enhanced low-light HSI demonstrated the benefits of HSIE as a low-light HSI preprocessing tool. This work offers many opportunities for future investigation into low-light HSI processing.

References

[1] X. Li, Y. Yuan, and Q. Wang, “Hyperspectral and multispectral image fusion via nonlocal low-rank tensor approximation and sparse representation,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 1, pp. 550–562, Jan. 2021.

[2] L. Mou, X. Lu, X. Li, and X. X. Zhu, “Nonlocal graph convolutional networks for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 12, pp. 8246–8257, Dec. 2020.

[3] Q. Wang, X. He, and X. Li, “Locality and structure regularized low rank representation for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 2, pp. 911–923, Feb. 2019.

[4] B. Yang, F. Cao, and H. Ye, “A novel method for hyperspectral image classification: Deep network with adaptive graph structure integration,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–12, 2022, doi: 10.1109/TGRS.2021.3150349.

[5] Z. Lv, X.-M. Dong, J. Peng, and W. Sun, “ESSINet: Efficient spatial–spectral interaction network for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–15, 2022, doi: 10.1109/TGRS.2021.3162721.

[6] X. Zheng, T. Gong, X. Li, and X. Lu, “Generalized scene classification from small-scale datasets with multitask learning,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–11, 2022, doi: 10.1109/TGRS.2021.3161474.

[7] X. Wang et al., “A double-dictionary-based nonlinear representation model for hyperspectral subpixel target detection,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–16, 2022, doi: 10.1109/TGRS.2021.3153308.

[8] X. Li, Z. Yuan, and Q. Wang, “Unsupervised deep noise modeling for hyperspectral image change detection,” Remote Sens., vol. 11, no. 3, p. 258, 2019.

[9] X. Zheng, X. Chen, X. Lu, and B. Sun, “Unsupervised change detection by cross-resolution difference learning,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–16, 2022, doi: 10.1109/TGRS.2021.3179907.

[10] X. Zheng, B. Wang, X. Du, and X. Lu, “Mutual attention inception network for remote sensing visual question answering,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–14, 2022, doi: 10.1109/TGRS.2021.3179918.

[11] M. Stuart, A. McDonigle, and J. Willmott, “Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems,” Sensors, vol. 19, no. 14, p. 3071, Jul. 2019.

[12] H. Ibrahim and N. S. P. Kong, “Brightness preserving dynamic histogram equalization for image contrast enhancement,” IEEE Trans. Consum. Electron., vol. 53, no. 4, pp. 1752–1758, Nov. 2007.

[13] 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, July 2007.

[14] 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.

[15] X. Fu, Y. Liao, D. Zeng, Y. Huang, X. Zhang, and X. Ding, “A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation,” IEEE Trans. Image Process., vol. 24, no. 12, pp. 4965–4977, Dec. 2015.

[16] C. Chen, Q. Chen, J. Xu, and V. Koltun, “Learning to see in the dark,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 3291–3300.

[17] C. Wei, W. Wang, W. Yang, and J. Liu, “Deep Retinex decomposition for low-light enhancement,” in Proc. Brit. Mach. Vis. Conf., 2018, p. 155.

[18] Y. Jiang, X. Gong, D. Liu, Y. Cheng, and C. Fang, “EnlightenGAN: Deep light enhancement without paired supervision,” IEEE Trans. Image Process., vol. 30, pp. 2340–2349, 2021.

[19] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, “From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2020, pp. 3063–3072.

[20] M. Maggioni, V. Katiukovnik, K. Egiazarian, and A. Foi, “Nonlocal transform-domain filter for volumetric data denoising and reconstruction,” IEEE Trans. Image Process., vol. 22, no. 1, pp. 119–133, Apr. 2013.

[21] W. He, H. Zhang, L. Zhang, and H. Shen, “Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 1, pp. 178–188, Jan. 2016.

[22] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, “Hyperspectral image restoration using low-rank matrix recovery,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[23] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, “Non-local meets global: An integrated paradigm for hyperspectral denoising,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 6868–6877.

[24] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, “Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 2, pp. 1205–1218, Sep. 2019.

[25] H. Ma, G. Liu, and Y. Yuan, “Enhanced non-local cascading network with attention mechanism for hyperspectral image denoising,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), May 2020, pp. 2423–2427.

[26] Y. Yuan, H. Ma, and G. Liu, “Partial-DNet: A novel blind denoising model with noise intensity estimation for HSI,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–13, 2022.
1020 University, Xi'an, China.
1019 (Fellow, IEEE) is a Full Professor with the School of Artificial Intelligence, Optics and Electrotechnics (OPEN), Northwestern Polytechnical University, Xi'an, China.
1021 His research interests include introducing physics models and cognitive science to artificial intelligence.

Xuelong Li (Fellow, IEEE) is a Full Professor with the School of Artificial Intelligence, Optics and Electrotechnics (OPEN), Northwestern Polytechnical University, Xi'an, China.
1025 His research interests include generative image modeling, computer vision, and machine learning.

Guanglin Li is currently pursuing the Ph.D. degree with the School of Computer Science and School of Artificial Intelligence, Optics and Electrotechnics (OPEN), Northwestern Polytechnical University, Xi'an, China.
1028 His research interests include introducing physics models and cognitive science to artificial intelligence.

Bin Zhao (Member, IEEE) is an Associate Professor with the School of Artificial Intelligence, Optics and Electrotechnics (OPEN), Northwestern Polytechnical University, Xi'an, China.
1026 His research interests include introducing physics models and cognitive science to artificial intelligence.

[27] E. Pan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, “SQAD: Spatial–spectral quasi-attention recurrent network for hyperspectral image denoising,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–14, 2022, doi: 10.1109/TGRS.2021.3156646.

[28] J. Liang, H. Zeng, and L. Zhang, “High-resolution photorealistic image translation in real-time: A Laplacian pyramid translation network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2021, pp. 9392–9400.

[29] 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. 2016.

[30] S. Park, S. Yu, B. Moon, S. Ko, and J. Paik, “Low-light image enhancement using variational optimization-based Retinex model,” IEEE Trans. Consum. Electron., vol. 63, no. 2, pp. 178–184, May 2017.

[31] E. H. Land, “The retinex theory of color vision,” Sci. Amer., vol. 237, no. 6, pp. 108–129, Dec. 1977.

[32] L. Wang, L. Xiao, H. Liu, and Z. Wei, “Variational Bayesian method for Retinex,” IEEE Trans. Image Process., vol. 23, no. 8, pp. 3381–3396, Aug. 2014.

[33] C. Li et al., “Low-light image and video enhancement using deep learning: A survey,” 2021, arXiv:2104.10729.

[34] K. G. Lore, A. Akintayo, and S. Sarkar, “LLNet: A deep autoencoder approach to natural low-light image enhancement,” Pattern Recognit., vol. 61, pp. 650–662, Jan. 2017.

[35] B. Zhao, H. Li, X. Lu, and X. Li, “Reconstructive sequence-graph network for video summarization,” IEEE Trans. Pattern Anal. Mach. Intell., early access, Apr. 9, 2021, doi: 10.1109/TPAMI.2021.3072117.

[36] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, “Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2021, pp. 10561–10570.

[37] X. Li and B. Zhao, “Video distillation,” Sci. China Inf. Sci., vol. 51, no. 5, pp. 695–734, 2021.

[38] 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.

[39] B. Zhao, X. Li, and X. Lu, “HSA-RNN: Hierarchical structure-adaptive RNN for video summarization,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7405–7414.

[40] K. Wei, Y. Fu, and H. Huang, “3-D quasi-recurrent neural network for hyperspectral image denoising,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 1, pp. 363–375, Jan. 2021.

[41] B. Zhao, X. Li, and X. Lu, “CAM-RNN: Co-attention model based RNN for video captioning,” IEEE Trans. Image Process., vol. 28, no. 11, pp. 5552–5565, Nov. 2019.

[42] Q. Shi, X. Tang, T. Yang, R. Liu, and L. Zhang, “Hyperspectral image denoising using a 3-D attention denoising network,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 12, pp. 10348–10363, Dec. 2021.

[43] E. Denton, S. Chintala, A. Szlam, and R. Fergus, “Deep generative image models using a Laplacian pyramid of adversarial networks,” in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 1486–1494.

[44] W. Lai, J. Huang, N. Ahuja, and M. Yang, “Deep Laplacian pyramid networks for fast and accurate super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 624–632.

[45] P. Burt and E. Adelson, “The Laplacian pyramid as a compact image code,” IEEE Trans. Commun., vol. 31, no. 4, pp. 532–540, Apr. 2003.

[46] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising,” IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[47] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4700–4708.

[48] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, “ECA-Net: Efficient channel attention for deep convolutional neural networks,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 11534–11542.

[49] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification,” in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2015, pp. 1026–1034.