IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 May 2023, accepted 17 May 2023, date of publication 22 May 2023, date of current version 1 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3278734

RESEARCH ARTICLE

A Variational Retinex Model With Structure-Awareness Regularization for Single-Image Low-Light Enhancement

DAWEI ZHANG[1], YANTING HUANG[1], XIAOYANG XIE[1], AND XIAOYONG GUO[1,2]
[1]College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin 300457, China
[2]Xingtai Key Laboratory for Research and Application of Robot Intelligent Detection and Sorting Technology, Xingtai 054001, China

Corresponding author: Xiaoyong Guo (gxyauthor@tust.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 72104177.

ABSTRACT Low-light image enhancement (LLIE) is a method of improving the visual quality of images captured in weak illumination conditions. In such conditions, the images tend to be noisy, hazy, and have low contrast, making them difficult to distinguish details. LLIE techniques have many practical applications in various fields, including surveillance, astronomy, medical imaging, and consumer photography. The total variational method is a sound solution in this field. However, requirement of an overall spatial smoothness of the illumination map leads to the failure of recovering intricate details. This paper proposes that the interaction between the global spatial smoothness and the detail recovery in the total variational Retinex model can be optimized by adopting a structure-awareness regularization term. The resultant non-linear model is more effective than the original one for LLIE. As a model-based method, its performance does not rely on architecture engineering, super-parameter tuning, or specific training dataset. Experiments of the proposed formulation on various challenging low-light images yield promising results. It is shown that this method not only produces visually pleasing pictures, but it is also quantitatively superior in that the calculated full-reference, no-reference, and semantic metrics are beyond most of state-of-the-art methods. It has a better generalization capability and stability than learning-based methods. Due to its flexibility and effectiveness, the proposed method can be deployed as a pre-processing subroutine for high-level computer vision applications.

INDEX TERMS Low-light image enhancement, total variational retinex model, structure-awareness.

I. INTRODUCTION

High-contrast and high-visibility images not only reflect abundant details for target scenes, but are also crucial for many high-level computer vision tasks, such as object detection [1], and segmentation [2]. Unfortunately, most photos are captured from sub-optimal lighting conditions suffering from back-lit, non-uniform lighting, weak or extremely low-illumination, color cast, and intensive noise. The low visual quality of these images would have detrimental effect to high-level algorithms. Low-light image enhancement (LLIE) refers to the research activities that devote to solve this problem [3]. Both model-based and learning-based methods are included and prosperous, such as histogram equalization (HE) [4], [5], Retinex based methods [6], [7], [8], deep neural networks (DNNs) [9], as well as network architecture searching [10]. However, the HE-based techniques can result in over-saturation or under-saturation of certain pixels. This may lead to an unrealistic looking without natural color balance. The HE also cannot handle high contrast scenes well and can make the image look unnatural. On the other hand, the Retinex-based conventional algorithms may lead to color shifts and artifacts. Moreover, such a technique is time consuming to process large images or videos, making it impractical for real-time applications.

The recent progress of deep-learning (DL) enables researchers to take great advantage of this kind of learning-based methods to solve LLIE problems. However, the

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

performance of these DNNs heavily relies on elaborately designed network architectures and carefully selected paired/unpaired large-scale training data. When presented with real-world images whose statistical distribution is not included in the training data, most of these DNNs would give rise to unsatisfactory visibility. Moreover, learning-based methods should be regarded as a black box that fits to certain training data. It would be inevitably suffered from a lack of interpretability. This could bring difficulties in analyzing the nature of LLIE itself so as to find potential cue for improvement.

Due to these existing limitations in learning-based methods, this work turns to the model-based method for the seeking of improvements. Compared with the learning-based method, the model-based has good interpretability and it can better explore the structure of low-light images. Since it does not rely on specific training data, the model-based method may be more robust than the learning-based, leading to a satisfactory performance in the wild. In this way, we propose TV_SA which is a model-based LLIE by integrating a **s**tructure-**a**wareness (SA) regularization term into a **t**otal **v**ariational (TV) Retinex model. Previous TV methods [6], [11] require an overall spatial smoothness of the illumination field. Although the assumption is effective for image denoising, this smoothing process can also remove important image features, making the frameworks noneffective. Another issue is that the frameworks can lead to over-smoothing of the image in regions of high contrast. This over-smoothing can result in loss of details and information, especially in regions such as edges and corners [12]. As a matter of fact, how to restore the details of the enhanced image is a primary concern of many LLIE investigations. In order to solve this problem, various SA regularization terms have been proposed in the previous works [7], [13], which can help to recover the details of the enhanced images. Inspired by these works, this paper designs a new SA regularization term which can be integrated into the TV Retinex model. The modified TV model not only improves the brightness and reduces the noise, but also retains more details. To make an unbiased quantitative and qualitative evaluation of TV_SA, we collect 207 images that suffer inevitable noise and poor visibility from real environments. These images compose a challenging low-light image dataset. In addition, TV_SA is also evaluated on the well-known benchmark datasets. In summary, the contributions of this work are as follows:

- A new SA regularization term which can be integrated into a TV Retinex model is proposed. Different from the previous works, this SA regularization term uses the square of l_2 norm of the gradient of reflection to weight the gradient of the illumination. This not only enables the derivation of a differential equation that can be solved via the multi-resolution projected normalized steepest descent (PNSD) method, but it also associates with the lighting property of the image. Therefore, the present assumption is more physically motivated than previous ones.
- In contrast to existing learning-based LLIE, the present method dose not contain any trainable parameters. As a result, TV_SA will not be biased to specific training data. The quantitative and qualitative evaluations guarantee the universality of the present method with superior performance than DL ones.
- We propose a low-light image dataset that contains photos captured by different mobile devices under diverse illumination conditions to evaluate the generalization capability of LLIE methods. It is shown that the dataset is challenging and the performance of most LLIE methods is barely satisfactory.

The remainder of this paper is organized as follows. Section II summarizes the recent works regarding the progress of Retinex-based LLIE methods and recently proposed SA regularization terms. In section III, the proposed new variational framework with its mathematical formulas is presented and the associate numerical method is also provided. In Sec. IV, the experiments are preformed on three benchmark datasets and comparison is made between existing methods. The conclusions are drawn in the final section.

II. RELATED WORKS

A. RETINEX THEORY IN LLIE METHODS

A model-based LLIE method often relies on a certain mathematical model with assumption about human visual system (HVS) or imaging process of specific device. The most commonly used assumption is the well-known Retinex theory which addresses the problem of separating the illumination from the reflectance in a given image [14]. Since it is a mathematically ill-posed problem, several decomposition strategies have been proposed under variational frameworks [6], [11], [15]. In these works, the decomposition is solved by a variational equation with certain priors such as the global smoothness of the illumination map. The illumination is enhanced by a Gamma correction and then it synthesizes with the reflectance. Although the brightness is increased and the noise is reduced, the details are inevitably blurred. In order to preserve naturalness and in the mean time enhance details, the lightness-order error measure, the bright-pass filter, and the bi-log transformation are adopted into the Retinex-based algorithm to encourage naturalness preservation [16]. To optimize the image restoration, model-based methods whose energy functions are under maximum a posteriori framework are proposed [7], [17], [18]. With the recent progresses in DL-based methods, sophisticated loss functions are designed to train the DNNs for the Retinex decomposition. RetinexNet is an end-to-end trainable network with a decomposition module and an illumination adjustment module [13]. To address the lack of paired training data, a self-supervised learning framework combining the Retinex theory with maximum entropy to impose the constraint on reflectance is proposed [19]. With the recent development of auto-machine learning, authors make great efforts to explore the searching method for most efficient DNNs that may be suitable for LLIE tasks. Since the searching space for DNNs is huge, the

Retinex theory can play the role as the heuristics to guide the searching [10]. Furthermore, effort is also devoted to design new attention mechanism, such as signal-to-noise-ratio awareness network [20]. In this way, the algorithm could adaptively focus on which patches need to be enhanced most and which need not.

Although these attempts have shown promising performance on LLIE, two main shortcomings are still remained. On the one hand, they cannot cope with heavily noisy image captured in extremely low-light conditions. The outputs often look unnatural with artifact and blocks. On the other hand, the details in the enhanced images are lost. Therefore, a method that can simultaneously handle above issues is highly desirable.

B. STRUCTURE-AWARENESS REGULARIZATION

As a matter of fact, the concept of SA is not new. It is a regularization term designed for restoring image details and it can be found in certain publications. LIME proposed in [7] includes a SA regularization to refine the illumination map based on the initial one. For Retinex-Net in [13], a SA term is added to the loss function which would guide the training process to preserve more details. For LIME [7], the SA constraint is weighted by a constant illumination map, which is the maximum intensity of each pixel in R, G, and B channels. Such a static initial estimation may not depict the image structure. Instead, Retinex-Net uses reflectance to weight the smoothness loss. This makes the model more sensitive to image domain with a large reflectance gradient. However, as a learning-based method, Retinex-Net will be inevitably fitted to certain training data. Its performance would be degenerated when feeds on real world images characterized by various data distributions. Additionally, the noise is inevitably enhanced even if it uses a post-processing layer that is based on BM3D [21]. To integrate the strengths from the above methods, we formulate a modified SA constraint into a TV Retinex model which would be effective for the noise reduction and structure preservation.

III. METHODOLOGY

To provide a comprehensive understanding of our methodology, the pipeline of the proposed method will be elaborated in this section. The overall method is attribute as TV_SA for short.

A. VARIATIONAL RETINEX MODEL WITH STRUCTURE-AWARENESS

In Retinex theory [11], a source image S can be decomposed into two factors, say a reflectance image R and an illumination image L. At a given point (x, y) in the image domain, it can be formulated as

$$S(x, y) = R(x, y) \cdot L(x, y), \tag{1}$$

where $\cdot$ denotes the element-wise matrix multiplication. Recovering the illumination form a given image is a highly mathematically ill-posed problem. Solving it needs additional priors or constraints. The first step taken by most methods is the conversion to the logarithmic domain by $s = \log S$, $l = \log L$, and $r = \log R$, and thereby $s = l + r$. In the following, we list the known priors about the illumination image [11].

- The overall spatial smoothness is the first and most important assumption about the illumination image.
- Since R is restricted to the unit interval, the constraint $L \geq S$ should be added. It is also had that $l \geq s$ because the log function is monotone.
- A trivial solution $l = Const$, where $Const$ is any constant above the maximal value of s, satisfies the two previous assumptions. Therefore, the assumption that the illumination image is close to the intensity image s should be added. It minimizes a penalty term of the form $dist(l, s)$, e.g., the L_2 norm $(l - s)^2$.
- Boundary conditions are necessary for solving the partial differential equation (PDE) which is derived from minimizing the functional. It is assumed that the illumination is continuous as a constant beyond the image boundaries. This artificial assumption would have minor effect on the final results [11].

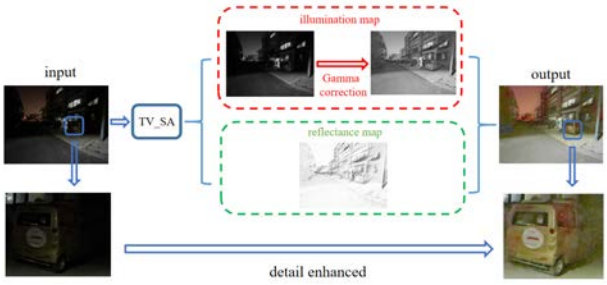

FIGURE 1. (Color online) Overall framework of TV_SA-based LLIE method. The source image is decomposed into a reflectance image and an illumination image via TV_SA. The illumination image is enhanced by a Gamma correction and then recombines with the reflectance.

By these assumptions, the solution of Eq. (1) can be obtained via the following variational problem:

Minimize:

$$F[l] = \int_{\Omega} \left[|\nabla l|^2 + \alpha (l - s)^2 + \beta |\nabla (l - s)|^2 \right] dxdy, \tag{2}$$

Subject to:

$$l \geq s \quad and \quad \langle \nabla l, \vec{n} \rangle = 0 \quad on \quad \partial\Omega,$$

where ∇ is the gradient operator, Ω is the support of the image, $\partial\Omega$ is its boundary, and $\vec{n}$ is the normal to the boundary. In Eq. (2), α and β are free non-negative real parameters. The numerical inner product $\langle G, F \rangle$ is defined as

$$\langle G, F \rangle = \sum_{n=1}^{N} \sum_{m=1}^{M} G[n, m] F[n, m]. \tag{3}$$

Terms in functional $F[l]$ are explained as follows:

- The first penalty term $(|\nabla l|^2)$ forces spatial smoothness on the illumination image.

- The second penalty term $(l-s)^2$ forces a proximity between l and s. The difference between these images is exactly r, which means that the norm of r should be small (i.e., R tends to white).
- The third term represents a Bayesian penalty expression. It forces r to be a "visually pleasing" image.
- In addition, the solution l forces to be $l \geq s$. In practice, this penalty term should be weak in order not to pull l down too much towards s.

By inserting the functional $F[l]$ into the Euler-Lagrange (EL) equation, a PDE can be obtained whose solution is the l. When l is solved, it would be tune up by a Gamma correction

$$L' = W \cdot \left(\frac{L}{W}\right)^{\frac{1}{\gamma}}. \tag{4}$$

Here, L' is the enhanced illumination, $L = \exp(l)$, W is the white value, and γ is a free parameter. The final result S' yields

$$S' = L' \cdot R, \tag{5}$$

where $R = \exp(r)$.

In the TV problem defined in Eq. (2), both l and r are required to be smooth, it fails at regions where the image has strong structures or where lightness changes drastically. To deal with it, we modify hypothesis 1 by making a replacement:

$$\int_\Omega |\nabla l|^2 \to \int_\Omega |\nabla l e^{-\frac{\lambda}{2}|\nabla(s-l)|^2}|^2, \tag{6}$$

where λ is the coefficient balancing between the strength of SA regularization and other terms. With the weight $e^{-\frac{\lambda}{2}|\nabla(s-l)|^2}$, functional $F[l]$ loosens the constraint for l smoothness where the gradient of reflectance is steep. In other words, the illumination should be discontinuous or not so smooth in the region where image structures locate. With Eq. (6), the TV problem in Eq. (2) is reformulated as

Minimize:

$$F[l] = \int_\Omega [|\nabla l e^{-\frac{\lambda}{2}|\nabla(s-l)|^2}|^2 + \alpha(l-s)^2 + \beta|\nabla(l-s)|^2]dxdy, \tag{7}$$

Subject to:

$$l \geq s \quad and \quad \langle \nabla l, \vec{n} \rangle = 0 \quad on \quad \partial\Omega,$$

Since Eq. (7) is featured by SA, it is attributed as TV_SA for short. An overall pipeline of TV_SA-based LLIE is provided in Fig. 1. In following sections, it is shown that the numerical solution can still be obtained based on a multi-resolution PNSD method as the original TV. The illumination and reflectance are updated simultaneously at each iteration.

B. EULER-LAGRANGE EQUATION

The solution of TV_SA in Eq. (7) can be found by inserting functional $F[l]$ into the EL equation,

$$\frac{\partial F}{\partial l} = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial l_x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial l_y}\right). \tag{8}$$

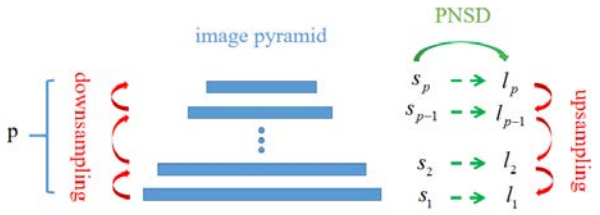

FIGURE 2. (Color online) A sketch of multi-resolution PNSD method. Here, s_i ($i \in [1, p]$) denotes the source image coarse-grained at ith resolution layer with l_i being the corresponding illumination image.

Here, $l_{x/y} = \frac{\partial l}{\partial x/y}$ is the partial derivatives of l with respect to x or y, and $\frac{\partial F}{\partial l_{x/y}}$ are the functional derivatives of F with respect to $l_{x/y}$. Using these notions, Eq. (7) can be formulated as

$$F[l] = \int_\Omega (l_x^2 + l_y^2)e^{-\lambda[(s_x-l_x)^2+(s_y-l_y)^2]} + \alpha(l-s)^2 + \beta[(l_x-s_x)^2 + (l_y-s_y)^2]dxdy, \tag{9}$$

where $s_{x/y} = \frac{\partial l}{\partial x/y}$ is the partial derivatives of s with respect to x or y. With Eq. (9), derivatives in Eq. (8) are calculated as follows:

$$\frac{\partial F}{\partial l_{x/y}} = 2l_{x/y}(1+\lambda|\nabla l|^2)e^{-\lambda|\nabla(s-l)|^2}, \tag{10}$$

and

$$\begin{aligned}\frac{\partial}{\partial x/y}\left(\frac{\partial F}{\partial l_{x/y}}\right) &= 2l_{xx/yy}(1+\lambda|\nabla l|^2)e^{-\lambda|\nabla(s-l)|^2} \\ &+ 2l_{x/y}\lambda\frac{\partial}{\partial x/y}|\nabla l|^2 e^{-\lambda|\nabla(s-l)|^2} \\ &+ 2l_{x/y}(1+\lambda|\nabla l|^2)\frac{\partial}{\partial x/y}e^{-\lambda|\nabla(s-l)|^2} \\ &\approx 2l_{xx/yy}(1+\lambda|\nabla l|^2)e^{-\lambda|\nabla(s-l)|^2}.\end{aligned} \tag{11}$$

Here, $l_{xx/yy} = \frac{\partial^2 l}{(\partial x/y)^2}$ is the second order partial derivatives of l with respect to x or y. To obtain a closed-form solutions, the spatial derivatives on gradients $|\nabla l|^2$ and $|\nabla(s-l)|^2$ are neglected in Eq. (11). By inserting Eqs. (10) and (11) into Eq. (8), one finds that

$$\nabla^2 l(1+\lambda|\nabla l|^2)e^{-\lambda|\nabla(s-l)|^2} - \alpha(l-s) + \beta\nabla^2(l-s) = 0. \tag{12}$$

Here, ∇^2 is the Laplacian. In the next section, the numerical solution of the above equation is discussed.

C. NUMERICAL SOLUTION

A multi-resolution PNSD method [11] is used to solve Eq. (12). This method constructs a Gaussian pyramid from the source image s by the kernel κ_{PYR} at a downsampling rate of 2:1. Here, the kernel κ_{PYR} reads

$$\kappa_{PYR} = \begin{pmatrix} \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \end{pmatrix}. \tag{13}$$

This kernel would be applied p times and gives rise to a pyramid of images $\{s_k\}_{k=1}^{p}$, where $s_1 = s$ and s_p is the most coarse-grained image. The iteration should be start at s_p layer and the initial condition is $l_0 = \max\{s_p\}$. For the kth resolution layer, the solution l_k can be obtained by first calculating the Laplacian $\mathcal{L}_s = \nabla_k^2 s_k$ and gradient $\mathcal{G}_s = \nabla_k s_k$. The numerical Laplacian is defined as

$$\nabla_k^2 s_k = s_k * \kappa_{Lap} 2^{-2(k-1)}, \tag{14}$$

where the kernel κ_{Lap} is given as

$$\kappa_{Lap} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{15}$$

The numerical gradient is defined as

$$\nabla_k s_k = s_k * \kappa_{Grad}^{x} e_x + s_k * \kappa_{Grad}^{y} e_y, \tag{16}$$

where the kernels κ_{Grad}^{x} and κ_{Grad}^{x} are given as

$$\kappa_{Grad}^{x} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \tag{17}$$

and

$$\kappa_{Grad}^{y} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}, \tag{18}$$

and $e_{x/y}$ are the unit vectors along x/y directions.

For $j \in \{1, \ldots, T_k\}$, the following steps should be repeated T_k times:

- Calculate gradient and update:

$$\begin{aligned} \mathcal{L}_l &= \nabla_k^2 l_{j-1} \\ \mathcal{G}_l &= \nabla_k l_{j-1} \\ G &\leftarrow -\mathcal{L}_l(1 + \lambda|\mathcal{G}_l|^2)e^{-\lambda|\mathcal{G}_s - \mathcal{G}_l|^2} \\ &\quad + \alpha(l_{j-1} - s_k) - \beta(\mathcal{L}_l - \mathcal{L}_s). \end{aligned} \tag{19}$$

- Calculate update step size μ_{NSD} according to

$$\mu_{NSD} \leftarrow \frac{\mu_A}{\alpha\mu_A + (1+\beta)\mu_B}, \tag{20}$$

  where $\mu_A = \langle G, G\rangle$ and $\mu_B = \langle G, \nabla_k^2 G\rangle$.
- Complete the NSD iteration

$$l_j \leftarrow l_{j-1} - \mu_{NSD} \cdot G. \tag{21}$$

- Project onto the constraint

$$l_j = \max\{l_j, s_k\}. \tag{22}$$

The above procedure solves Eq. (12) on the kth resolution layer. The solution is l_{T_k}. For the layer $k \neq 1$, the l_{T_k} is up scaled with 1:2 ratio by pixel replication into the new l_0, that is the initialization for the $k-1$ resolution layer. The algorithm proceeds by going again to the above steps (19)-(22). When it reaches $k = 1$, the result l_{T_1} is regarded as the final output of the algorithm.

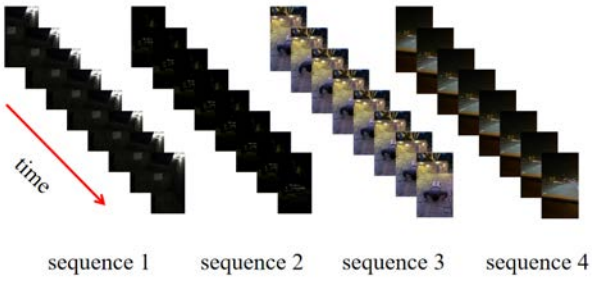

FIGURE 3. (Color online) Demonstration of image sequences in LoLi-Phone dataset. These sequences are highly correlated. Thus, only one representative image is selected from each sequence.

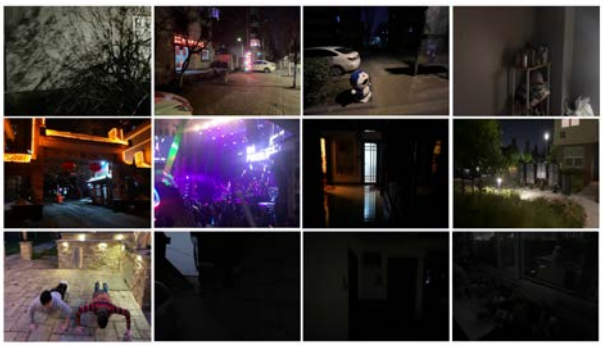

FIGURE 4. (Color online) Several images sampled from proposed Real-Scene-v2 dataset. The images are taken by different mobile devices under diverse lighting conditions and scenes.

IV. EXPERIMENTS

In this section, we are in a position to verify the universality and effectiveness of TV_SA on real low-light images via intensive experiments. For datasets, the well-known benchmark datasets LOL and LIME, as well as a custom real scene dataset collected and arranged by authors are included. We qualitatively and quantitatively compare TV_SA with various sound LLIE methods, including a traditional model-based algorithm (LIME) [7], a deep Retinex model (KinD++) [22], a zero-shot-learning DL algorithm (Zero-DCE++) [23], and a no-reference DL algorithm (RetinexDIP) [24]. Furthermore, comparison is also make between different variational frameworks, including original TV and STAR [25].

A. DATASET AND METRICS

The datasets involved in this paper are elaborated as follows:

- **LOL and LIME.**[1] LOL [13] is the first paired low-/normal-light image dataset taken in real scenes. The low-light images are collected by changing the exposure time and ISO. LOL contains 500 pairs of low-/normal-light images of size 600 × 400 saved in RGB format. In our experiments, we use the LOL-test set which contains 15 paired low-/normal-light images. Moreover,

[1]These datasets can be downloaded from URL https://daooshee.github.io/BMVC2018website/ and https://github.com/estija/LIME

LIME dataset is proposed in Ref. [7] and it contains 8 weak illuminate images without ground truth.

- **Real-Scene-v2.**[2] The flexibility and effectiveness of TV_SA means that the algorithm itself and the associate parameters are applicable to any real world scenes without structure fine-tuning or parameter adjustment. In order to verify it, we introduce another real world low-light image dataset. This dataset consists of two parts. The first part is the one that we have introduced in our previous work [26]. It contains 94 real low-light images collected over six different cities of China. These images cover multiple scenes, including indoor, country field, inside building corridor, residential square, street scene, etc. The images are taken using the mobile phone cameras with various resolutions, ranging from 4000×3000, 1920×873, to 1368×1824, and they are saved in RGB format. This dataset is named as Real-Scene. The second part of the dataset is collected from the LoLi-Phone dataset proposed in Ref. [3]. LoLi-Phone is also constituted by images and video captured by various mobile phones. It is the largest and most challenging real-world testing dataset, including 120 short videos and 55,148 images. However, these videos are usually continuous shot of a same scene, and the images are screenshots from these videos. A demonstration of such an image sequence is provided in Fig. 3. These sequences have significant time relevance which may be useful for burst imaging algorithms. Since the proposed method is based on a single RGB image, we filter LoLi-Phone and selected 113 images from different scenarios and add them to the previous Real-Scene dataset. Therefore, the expanded dataset contains 207 images and it is named as Real-Scene-v2. Several samples of Real-Scene-v2 are provided in Fig. 4.

Comprehensive evaluation of a LLIE algorithm is more intricate then evaluating a high-level task. One could not get a conclusion from a single indicator such as mean average precision for object detection algorithms [1]. In quantitative comparisons, various metrics are used including full-reference metrics (**s**tructural **s**imilarity **i**ndex **m**easurement, SSIM, **p**eak **s**ignal-to-**n**oise **r**atio, PSNR, **m**ean **s**quare **e**rror, MSE, and **m**ean **a**bsolute **e**rror, MAE), no-reference metrics (**n**aturalness **i**mage **q**uality **e**valuator, NIQE [27], BNBT [28], S-CCR [29], [30], and **p**erceptual **i**ndex, PI [31]), and semantic metrics(**s**martphone **p**hotography **a**ttribute and **q**uality, SPAQ [32]). For SPAQ, there are three sub-indexes: SPAQ-BL, SPAQ-IA, and SPAQ-SS. Here, SPAQ-BL estimates image quality by a baseline model (**r**esidual **n**etwork 50, ResNet-50), SPAQ-IA works for by input image attributes, and SPAQ-SS accounts for input image semantic information. The details of these indicator are elaborated in a recent survey [3], and we do not intend to reproduce them here.

[2]This dataset can be downloaded from URL https://pan.baidu.com/s/1Y8gYqyVsplPv7ycX_NG3vg?pwd=d5lc/ (code: d5lc)

B. IMPLEMENTATION DETAILS

The numerical solution of TV_SA is realized by python-3.6 with numpy-1.18.4 for matrix manipulation and opencv-4.2.0 for image reading and writing. The experiment is performed on a computer with ubuntu-18.04, Intel i5-9400F CPU, and 8GB memory. A same set of empirically parameters in which $\alpha = 0.1$, $\beta = 100$, $\gamma = 4$, and $T_k = 10$ is used over all of our experiments. Since the parameter λ controls the intensity of SA term, its influence would be discussed in details. The white value in 8-bit images is $W = 255$. For color images, one can apply either the RGB Retinex method or the HSV Retinex method. The former treats the R, G, and B channels separately and usually yields a color shifting effect. The latter only applies Eq. (7) to the intensity channel V, and then it maps the enhanced V together with H and S channels back to the RGB domain. The HSV Retinex method not only suppresses the color shifts, but also has the advantage to process a single channel [11]. Thus, we prefer to the HSV Retinex method in this work.

C. RESULTS AND DISCUSSIONS

1) EVALUATION ON LOL

The qualitatively and quantitatively results of TV_SA on LOL-test set and LIME dataset are provided in Figs. 5 and 6, and in Tab. 1. Here, λ is fixed as 0.1. Since the visual results of TV and STAR are very similar with TV_SA for most of images without strong structure, it is not provided here. A detailed visual inspection between TV and TV_SA will be provided afterward. The quantitative results of full-reference metrics and no-reference metrics mentioned in Sec. IV-A are given in Tab. 1. As can be seen from full-reference metrics, TV_SA is the best one in MSE, MAE, and SSIM in that these indicators outperform other methods by a wide margin. However, the PSNR for Zero-DCE++ is slightly better than other methods. For the no-reference metrics (NIQE, PI, and SPAQ), a similar trend is shown. The NIQE of TV_SA is better than other methods by an average of 25.7% (i.e., 32.2%, 31.0%, 26.7%, and 12.9% enhanced with respect to LIME, Zero-DCE++, RetinexDIP, and STAR). For PI, TV_SA achieves a significant improved performance by an average of 49.15% better than other methods (i.e., 65.4%, 48.4%, 63.4%, and 19.4% enhanced with respect to LIME, KinD++, Zero-DCE++, and RetinexDIP). For SPAQ, it is shown that TV_SA is better than other methods in SPAQ-BL and SPAQ-SS, and it gets a nearly equal performance in SPAQ-IA compared with TV.

Visual comparisons on the LOL-test set and LIME dataset are shown in the Figs. 5 and 6. In Fig. 5, it is shown that the brightness of LIME algorithm enhanced images is less than ground truth (GT) images, and the RetinexDIP results are even lower brightened. Compared with these two methods, TV_SA outputs the closest results to GT images. This is also consistent with the quantitative results in Tab. 1. For the LIME dataset, the enhanced images given by LIME algorithm are still slightly less bright than

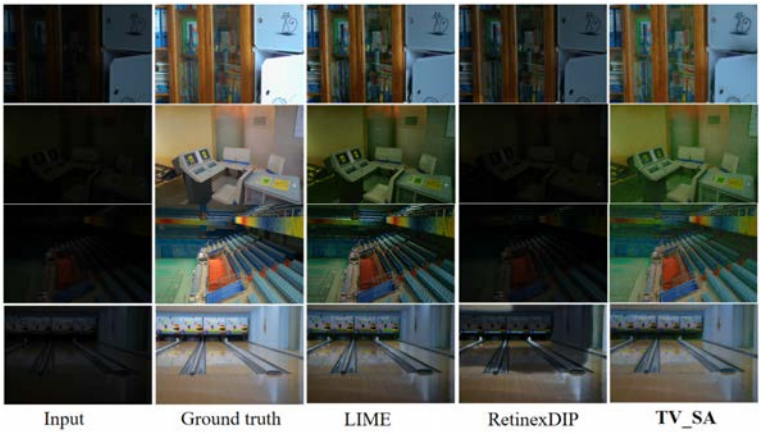

FIGURE 5. (Color online) Visual results of different methods on LOL-test. Here, λ is fixed as 0.1.

TABLE 1. Quantitative comparisons on LOL-test in terms of SSIM, PSNR, MSE, MAE, NIQE, PI, and SPAQ. Here, the best results are boldfaced and $\lambda = 0.1$.

	LIME	KinD++	Zero-DCE++	RetinexDIP	TV	STAR	TV_SA
MSE↓	100.81	0.06	2.41	105.64	0.034	0.195	**0.032**
MAE↓	61.86	0.19	12.26	87.95	0.15	0.787	**0.14**
PSNR↑	28.19	21.3	**65.78**	27.91	64.44	62.17	64.81
SSIM↑	0.62	0.82	**0.99**	0.36	**0.99**	0.906	**0.99**
NIQE↓	8.09	**5.12**	8.02	7.69	7.07	6.91	6.12
PI↓	8.95	8.03	8.84	6.46	6.32	**5.40**	5.41
SPAQ-BL↑	55.28	54.43	55.75	54.84	56.41	57.41	**58.92**
SPAQ-IA↑	54.41	54.02	54.31	54.17	**54.67**	51.77	53.18
SPAQ-SS↑	54.32	52.86	54.48	53.98	55.79	55.69	**56.61**

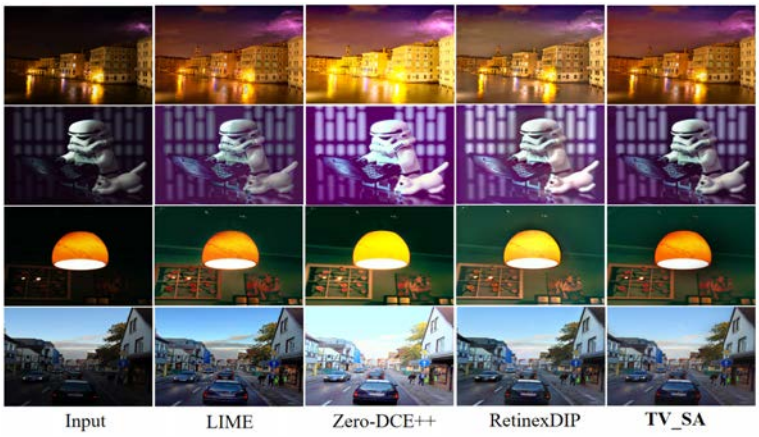

FIGURE 6. (Color online) Visual results of different methods on LIME dataset. Here, λ is fixed as 0.1.

TV_SA. However, Zero-DCE++ outputs are over enhanced. The RetinexDIP outputs are similar to those of TV_SA. It is worth noting that for the images in the first row, LIME algorithm shows that the reflection of the building in the river appears overexposure and artifact, while TV_SA obviously fixes this problem. As a result, it is obvious to learn that TV_SA gives rise to the more visually pleasing images and they are even more closer to GT.

2) EVALUATION ON REAL-SCENE-v2

In Tab. 2 and Fig. 7, the qualitatively and quantitatively comparisons between various methods on the Real-Scene-v2 are reported. Here, λ is fixed as 0.1. Different from the previous LOL-test set and LIME dataset, it is shown in Tab. 2 that TV_SA achieves promoting performance in NIQE, PI, SPAQ-IA, SPAQ-SS, BNBT, and S-CCR. TV_SA also outperforms the DL methods. For NIQE, TV_SA is reduced by

FIGURE 7. (Color online) Visual results of different methods on Real-Scene-v2. Here, λ is fixed as 0.1.

TABLE 2. Quantitative comparisons on Real-Scene-v2 in terms of NIQE, PI, SPAQ, BNBT, and S-CCR. Here, the best results are boldfaced and $\lambda = 0.1$.

	Input	KinD++	Zero-DCE++	RetinexDIP	TV	STAR	TV_SA
NIQE↓	5.5910	4.2772	4.2566	4.3777	6.2133	5.187	**4.2564**
PI↓	8.0851	**7.192**	7.955	8.3984	8.1087	7.282	7.52
SPAQ-BL↑	49.637	49.071	50.785	51.986	**52.334**	51.071	52.087
SPAQ-IA↑	49.867	50.625	50.783	50.146	51.674	50.806	**51.762**
SPAQ-SS↑	49.931	50.776	48.847	51.652	49.573	51.267	**53.059**
BNBT↑	0.341	0.651	0.512	0.261	0.355	0.71	**0.721**
S-CCR↑	0.282	0.543	0.514	0.346	0.561	0.715	**0.722**

23.9%, 0.5%, 21.9% and 2.8% relative to the input, KinD++, RetinexDIP, and STAR. For SPAQ-BL, SPAQ-IA, and SPAQ-SS, TV_SA are increased by 2.6%, 1.93%, and 8.6% with respect to Zero-DCE++, and 0.2%, 3.2%, and 2.7% with respect to RetinexDIP. This means that TV_SA can obtain an evidently improvement with respect to Zero-DCE++, and

FIGURE 8. (Color online) Several visual results for demonstration of detail restoration. Here, comparison between TV and TV_SA is made, and λ is fixed as 0.1.

a nearly equal SPAQ with respect to RetinexDIP. Compared with KinD++ and STAR, TV_SA achieves a nearly 3% performance improvement in SPAQ. For indictors BNBT and S-CCR, TV_SA also achieves a significant performance improvement with respect to other methods. Compared with TV, indicator SPAQ-BL has decreased a little, but SPAQ-IA and SPAQ-SS of TV_SA are still the best of all. Therefore, TV_SA performs better results than all other methods in terms of restoring semantic information of images. This is also consistent with the results on LOL-test set and LIME dataset. Since TV_SA is modified base on TV, a comprehensive comparison among these two methods for preserving small details will the subject of the next section.

For the Real-Scene-v2, a similar tendency can be found from the visualization results in Fig. 7. Zero-DCE++ as a learning-based LLIE method has a limited capacity to achieve noise suppression and reaches a satisfactory result. Although RetinexDIP removes noise, it may bring other problems such as losing details, blurring or even worse image quality. TV may retain details and brightness. It also leads to a color distortion which gives rise to an uncomfortable visualization. In comparison to all of these approaches, our model is capable of noise suppression and detail preservation in the meantime sufficiently revealing low-light domains. However, for certain images suffered by severe noise or extremely low-lighting, all of these methods produce output images which are featured by noise, color blocks, and artifacts. In this case, the information is even not captured by the sensor or lost via camera's ISP. As a result, the methods based on RGB images would fail and one may have to turn to use raw data for enhancement [33], [34].

TABLE 3. NIQE, PI, SPAQ, BNBT, and S-CCR of TV_SA outputs evaluated on Real-Scene-v2 for various values of λ. Here, the best results are boldfaced.

	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 10$
NIQE↓	6.34	**4.2564**	5.175	7.698
PI↓	8.019	**7.52**	8.917	10.569
$\langle SPAQ \rangle$↑	51.62	**52.303**	49.607	48.712
BNBT↑	0.345	**0.721**	0.6317	0.431
S-CCR↑	0.281	**0.722**	0.6317	0.441

The influence of parameters α, β, and λ will be discussed here. As a matter of fact, the chosen value of parameters α and β is adopted from Ref. [11], where the value of β is one thousand times of α. Therefore, we directly borrow the ratio into our experiments for a fair comparison. However, parameter λ describes the intensity of the SA term, and it is a newly introduced parameter featuring TV_SA. Thus, the influence of λ on the model's performance should be discussed in details. To this end, we evaluate TV_SA on Real-Scene-v2 for various values of λ. The experimental result is reported in Tab. 3. Here, three sub-indicators for SPAQ are averaged as $\langle SPAQ \rangle$ for simplicity. It is shown that when $\lambda = 0.01$, the SA term has relatively minor effect.

In this case, the performance of TV_SA is very similar to that of original TV. When $\lambda = 0.1$, TV_SA achieves a great performance improvement. With the further increasing of λ, the improvement becomes less obvious once again. When λ is taken to be a larger value, the performance of TV_SA will be degraded greatly. These results demonstrate that the contribution of SA term is best when the value of λ has a same magnitude as that of α. As λ approaches the magnitude of β, the SA term will give a negative contribution to the model's performance.

3) COMPARISON WITH TV

The proposed TV_SA is modified by the inclusion of a SA term into the Retinex-based TV method. An extensive visual comparison on Real-Scene-v2 is given among TV and TV_SA. From Tab. 2, it is found that TV_SA outperforms TV by 31.5%, 7.3%, 7.0%, 50.8%, and 22.3% for indictors NIQE, PI, SPAQ-SS, BNBT, and S-CCR. However, they achieve a nearly equal performance for SPAQ-IA. For SPAQ-BL, TV_SA is slightly lower than TV. The visual comparison results are plotted in the Fig. 8. It can be seen that the detailed structures are better enhanced along with the brightness and contrast enhancement. For example, in the image provided in the upper panel, details of the Doraemon's eyes and beard almost disappear in the TV's output. However, these details are better recovered and sharpened by TV_SA. The same can be found in the other panels, such as the grass, the lion, the door of the zoo, and the words on the yellow jar. These indicate that the SA term can balance the global smoothness enforced by TV. It enables a TV method making fully usage of the information in the original image to better recover the details of the enhanced image.

V. CONCLUSION

In this study, a new LLIE algorithm (TV_SA) is proposed by integrating a SA term into the Retinex-based TV model. By EL equation, we derive a non-linear PDE whose solution is the optimized illumination image. Only V channel in the HSV color space is enhanced by a Gamma correction and then reunions with other channels. By a multi-resolution PNSD method, TV_SA is closed mathematically and can be numerically solved. The experiments are performed on existing benchmark datasets (i.e., LOL and LIME) as well as on a custom dataset. Various metrics are included and comparisons are made among existing model-based and learning-based methods. The proposed TV_SA outperforms these methods on both publicly available and custom datasets in that it has promoting performance on most of those metrics. TV_SA yields not only state-of-the-art quantitative results but also applicable visualizations. Since TV_SA is not data-driven, its parameters are universal and the generalization capability could be better than that of the learning-based methods. Detailed comparison and analysis demonstrate the effectiveness of TV_SA in restoring image details.

This work only realizes a simple implementation by Python to verify the idea of TV_SA. In this realization, TV_SA is programmed serially. Since a multi-resolution pyramid is involved, it is time-consuming in the iterative optimization. Therefore, TV_SA cannot achieve real time running speed. How to optimize its running speed would be our future research topic. It is also interesting to explore the deployment in low-end devices such as mobile phones. In this way, one may programme in a paralleled manner taking full usage of the computing capability of the device.

REFERENCES

[1] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, Feb. 2020.

[2] C. Yu, J. Wang, C. Gao, G. Yu, C. Shen, and N. Sang, "Context prior for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12416–12425.

[3] C. Li, C. Guo, L. Han, J. Jiang, M. Cheng, J. Gu, and C. C. Loy, "Low-light image and video enhancement using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9396–9416, Dec. 2022.

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

[5] 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, Jul. 2007.

[6] M. K. Ng and W. Wang, "A total variation model for Retinex," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 345–365, Jan. 2011.

[7] 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.

[8] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5891–5900.

[9] 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.

[10] 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/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10556–10565.

[11] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for Retinex," *Int. J. Comput. Vis.*, vol. 52, pp. 7–23, Apr. 2003.

[12] F. Ma, J. Chai, and H. Wang, "Two-dimensional compact variational mode decomposition-based low-light image enhancement," *IEEE Access*, vol. 7, pp. 136299–136309, 2019.

[13] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf.*, 2018, pp. 1–12.

[14] E. H. Land, "The Retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, 1977.

[15] X. Fu, D. Zeng, Y. Huang, X. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2782–2790.

[16] 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.

[17] S. Hao, X. Han, Y. Guo, X. Xu, and M. Wang, "Low-light image enhancement with semi-decoupled decomposition," *IEEE Trans. Multimedia*, vol. 22, no. 12, pp. 3025–3038, Dec. 2020.

[18] X. Ren, W. Yang, W. Cheng, and J. Liu, "LR3M: Robust low-light enhancement via low-rank regularized Retinex model," *IEEE Trans. Image Process.*, vol. 29, pp. 5862–5876, 2020.

[19] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5637–5646.

[20] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "SNR-aware low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17714–17724.

[21] K. Dabov, A. Foi, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," *Proc. SPIE*, vol. 6064, pp. 354–365, Mar. 2006.
[22] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, "Beyond brightening low-light images," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1013–1037, Apr. 2021.
[23] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4225–4238, Aug. 2022.
[24] Z. Zhao, B. Xiong, L. Wang, Q. Ou, L. Yu, and F. Kuang, "RetinexDIP: A unified deep framework for low-light image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1076–1088, Mar. 2022.
[25] J. Xu, Y. Hou, D. Ren, L. Liu, F. Zhu, M. Yu, H. Wang, and L. Shao, "STAR: A structure and texture aware Retinex model," *IEEE Trans. Image Process.*, vol. 29, pp. 5022–5037, 2020.
[26] X. Guo, K. Zhang, J. Peng, X. Chen, and G. Guo, "Opt2Ada: A universal method for single-image low-light enhancement," *J. Intell. Fuzzy Syst.*, vol. 10, pp. 1–14, Apr. 2023.
[27] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[28] T. Xiang, Y. Yang, and S. Guo, "Blind night-time image quality assessment: Subjective and objective approaches," *IEEE Trans. Multimedia*, vol. 22, no. 5, pp. 1259–1272, May 2020.
[29] M. Wang, Z. Xu, Y. Gong, and W. Xie, "S-CCR: Super-complete comparative representation for low-light image quality inference in-the-wild," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 5219–5227.
[30] M. Wang, Y. Huang, J. Xiong, and W. Xie, "Low-light images in-the-wild: A novel visibility perception-guided blind quality indicator," *IEEE Trans. Ind. Informat.*, vol. 19, no. 4, pp. 6026–6036, Apr. 2023.
[31] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Comput. Vis. Image Understand.*, vol. 158, pp. 1–16, May 2017.
[32] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang, "Perceptual quality assessment of smartphone photography," in *Proc. Int. Conf. Comput. Vis.*, 2020, pp. 3677–3686.
[33] 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.
[34] H. Jiang and Y. Zheng, "Learning to see moving objects in the dark," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7324–7333.

DAWEI ZHANG received the Ph.D. degree from Harbin Engineering University, in 2019. He is currently an Assistant Professor with the College of Electronic Information and Automation, Tianjin University of Science and Technology. His research interests include computational imaging and complex system simulation.

YANTING HUANG, photograph and biography not available at the time of publication.

XIAOYANG XIE, photograph and biography not available at the time of publication.

XIAOYONG GUO received the Ph.D. degree from Nanjing University, in 2012. He is currently an Associate Professor with the College of Electronic Information and Automation, Tianjin University of Science and Technology. His research interests include computational imaging and deep learning.