IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 May 2024, accepted 28 May 2024, date of publication 3 June 2024, date of current version 10 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3408248

RESEARCH ARTICLE

PESiT: Progressive Joint Enhancement and Blind Super-Resolution for Low-Light and Low-Resolution Images Under Total Variation Constraints

HE DENG, KAI CHENG, AND YUQING LI

School of Computer Science and Technology, Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan 430081, China

Corresponding author: He Deng (denghe@wust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62071456 and in part by the 14th Five Year Plan Hubei Provincial Advantaged Characteristic Disciplines (Groups) Project of Wuhan University of Science and Technology under Grant 2023D0502.

ABSTRACT Traditional enhancement techniques can improve the contrast of low-light and low-resolution images, but they fail to recover their resolution. Conversely, traditional super-resolution (SR) algorithms can enhance resolution but not restore contrast. To address this issue, a novel progressive joint enhancement and SR tactic, named PESiT, is proposed to synchronously improve contrast and resolution in low-light and low-resolution images. PESiT comprises an enhanced multi-scale Retinex module followed by a blind SR module with regularization optimization. In the first module, the common logarithm is replaced with an S-function to expand the intensity distribution of images and prevent color inversion. In the latter module, those merits of reconstruction- and learning-based tactics are combined to tackle various unknown degradations by imposing consistency constraints on high- and low-resolution image pairs. Extensive experiments on public datasets demonstrate the robustness and superiority of PESiT in processing low-light and low-resolution images under various scenarios. Compared with state-of-the-art techniques, PESiT achieves superior performance, e.g., the highest peak signal-to-noise ratio, structural similarity index, feature similarity index, and the lowest learned perceptual image patch similarity, highlighting its validity in achieving optimal image quality improvements.

INDEX TERMS Image enhancement, low-light enhancement, low-resolution image, super-resolution.

I. INTRODUCTION

High-quality image that has excellent contrast, precise details, and high spatial or temporal resolution is critical in numerous fields, such as remote sensing, computer vision, scene analysis, and image understanding [1]. Despite the importances, image quality is usually compromised by various constraints, such as limited acquisition time and/or inadequate illumination [1]. As a result, the image suffers from diminished contrast, resolution and clarity of details [2], which significantly affects the visual assessment and advanced processing tasks, for instance, object recognition or target detection. Therefore, it is crucial to better the quality of degraded images, encompassing but not limited to improvement in contrast, resolution and details.

Image enhancement involves various tactics, each focusing on different aspects of images e.g., global/local characteristic, filtering out irrelevant information, and highlighting regions of interest (ROIs). Common enhancement techniques include histogram equalization (HE) and its variants like adaptive HE and bi-HE [3], inverse implementation [4], DL (deep learning) [5], and Retinex decomposition [6], [7]. HE and variants work by adjusting the distribution of histogram using regularization terms [8], predominantly enhancing image contrast rather than correcting for illumination. This possibly brings over- / under-enhanced outcomes [9]. Another tactic treats

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

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

low-light images similarly to hazy images [4]. Although this strategy results in acceptable results, it lacks robust physical justification for its underlying model [9]. DL techniques, such as low-light net [5] and deep convolutional network, have been extensively used in image enhancement. Howbeit, such strategy may not fully capture intrinsic properties of natural low-light images, due to challenges in collecting a substantial dataset of images paired with low-light and standard lighting conditions [10]. This may lead to unnatural results.

Drawing inspiration from biological mechanisms, Retinex-based enhancement techniques assume that a visual image can be decomposed into reflectance and illumination components. This hypothesis results in the development of various Retinex strategies, e.g., single-scale Retinex (SSR) [6] and multi-scale Retinex (MSR) [7], which principally focus on extracting the reflectance component. These tactics typically use logarithmic transformation owing to its beneficial modeling characteristics [9]. However, the logarithmical form is not the most suitable for the regularization term, as it has tendency to emphasize the influence of pixels with low intensities at the expense of those with high intensity, probably leading to suboptimal results [6], [9]. To address this problem, a weighted variational model that incorporates prior knowledge into regularization term, such as structural priors. Although this approach produces impressive enhancements, it may also yield over-exposure and the loss of detail in brighter areas of the image [6].

Despite their strengths, the strategies previously mentioned usually fall short in enhancing the spatial resolution of images, which is essential for maintaining the richness of detail. Super-resolution (SR) techniques are designed to reconstruct a high-resolution image from a low-resolution one, and DL-based SR strategies have consistently topped performance charts across various benchmarks. Nevertheless, most methods are tailored to specific, fixed forms of degradation, such as bicubic down-sampling or single Gaussian blurring, which limits the general applicability [11]. When the distribution of test images differs from that of training set, a phenomenon known as distribution shift, those DL-based strategies can suffer a marked decline in performance [18]. In response, some non-blind and blind SR methods are advanced to tackle multiple forms of degradations [12], [13]. Non-blind strategies often require ground-truth (GT) degradation maps as an additional input [12], but they may not perform well when the actual degradation does not match the assumed kernels [13]. Blind methods, which do not rely on GT maps, become the focus of increased attention in recent years [13], [14], [15], though their focus tends to be on the blur and down-sampling degradations [13]. While SR methods yield visually pleasing results, they often struggle to ensure the consistence between the reconstructed images and input images during the testing period [16].

Furthermore, current SR techniques are generally not well-suited for low-light images due to their limited ability to adjust contrast effectively. For instance, if SR is applied directly to a low-light image, the contrast improvement is insufficient, as shown in Fig. 1(c) that adopts Cria-CL (criteria comparative learning in realSR) [17] to perform SR of Fig. 1(a). Howbeit, if SR is implemented on the enhanced image (e.g., Fig. 1(b)), both contrast and structural details are significantly improved, yielding a higher quality outcome, as displayed in Fig. 1(d). From quantitative comparisons, the measures in Fig. 1(d) is higher to that in Fig. 1(c), with significant improvements in lift ratios. The respective lift ratio (%) is 74.51, 52.52, 10.25, and −45.16 with regard to PSNR (peak signal-to-noise ratio), SSIM (structural similarity index), FSIM (feature similarity index), and LPIPS (learned perceptual image patch similarity).

To markedly improve the contrast and spatial resolution of images taken in low-light environments, an effective approach involves incorporating image enhancement with SR within a single cohesive process. Lyu et al. propose an end-to-end DL-based strategy, termed as JSENet, which unifies enhancement and SR through a bilateral grid processing mechanism for joint optimization [18]. Recently, a lightening SR (LSR) network is presented to tackle the concurrent challenges of enhancement and SR, aiming to gain seamless zooming in low-light settings [19]. Despite these advancements, LSR requires considerable training time because of its dense architecture and extensive parameters. Moreover, both JSENet and LSR have yet to fully conquer the complexity associated with the enhancement and SR tasks. Simultaneously executing enhancement and SR for those images with poor color, contrast, or resolution remains a significant and unresolved challenge in the field.

In this study, we propose a new progressive fusion tactic of enhancement and blind SR (named PESiT), which is tailored for the enhancement of low-light and low-resolution images under total variation (TV) constraints. PESiT incorporates a dedicated enhancement module (namely, improved MSR via S-function) into a SR module (blind SR via TV constraints). This integrated approach accepts low-light and low-resolution images as input and generates outputs that feature enhanced contrast and spatial resolution. The contributions of this work are summarized as follows.

(1) To gradually improve the contrast and spatial resolution of low-light images, PESiT is proposed, which consists of an improved MSR module followed by a blind SR strategy that incorporates with regularization optimization. This integrated pipeline leverages the strengths of both enhancement and SR techniques to produce the high quality, superior outcomes.

(2) In the improved MSR module, we replace the traditional logarithm typically used in Retinex-based methods with an S-shaped function. This innovative modification, coupled with V-channel correction, successfully combats issues e.g., color distortion and detail loss. Furthermore, it markedly enhances the overall image contrast.

(3) The SR module leverages TV consistency constraints between high- and low-resolution images, blending strengths of both reconstruction-based and learning-based SR methods.

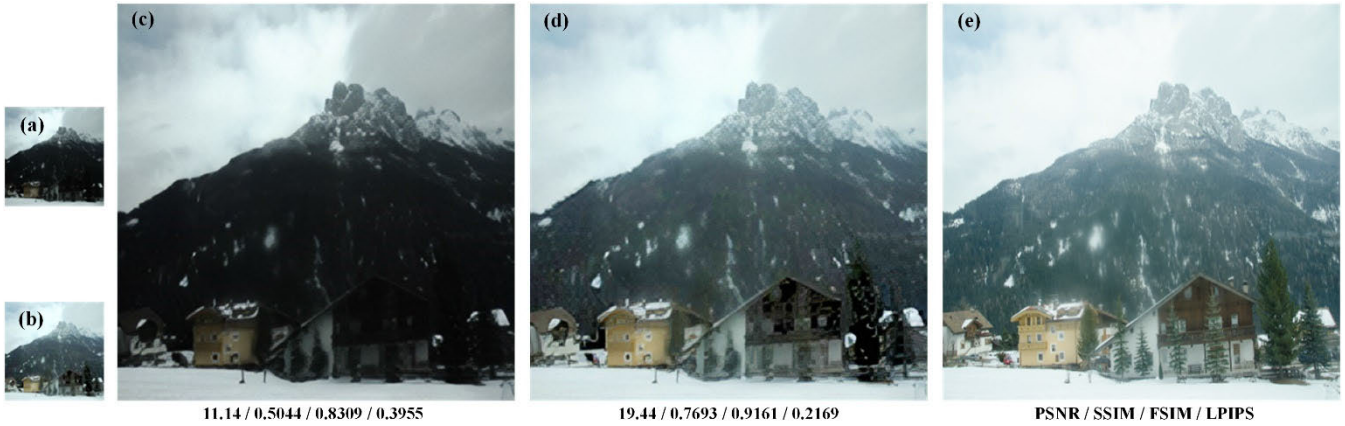

FIGURE 1. For a representative low-light image with low-resolution (a), enhanced result using proposed strategy (b), SR result of (a) using Cria-CL (c), SR result of (b) using Cria-CL (d), and ground truth with high contrast and resolution (e).

This tactic effectively tackles difficulties which presented by a range of unknown degradation processes. As a result, it not only elevates image quality but also can bolster the robustness of SR tasks.

The remainder of this paper is as follows. In Section II, we review some related works about image enhancement and SR. We explicate details about PESiT in Section III. In Section IV, extensive experimental results are provided, and conclusion and perspectives are presented in Section V.

II. RELATED WORK

This section reviews some related works, including low-light enhancement, SR, and joint enhancement and SR tactics.

A. LOW-LIGHT ENHANCEMENT

Low-light enhancement is challenging owing to severe noise, loss of information and poor visibility [19]. Many traditional and DL-based strategies strive to recover submerged and lost details as well as to brighten underexposed areas. Techniques like HE and gamma correction are usually applied to expand dynamic range and contrast of images, but those methods may yield unwanted artifacts that degrade visual quality. Retinex-based models, which possess definite physical meanings, can improve visual quality of images [6], [7], [9], and a weighted variational model (WVmSE) is applied to concurrently gauge reflectance and illumination components [20]. Among them, a logarithmic function is usually used to decrease computational complexity [8]. Yet, those models predominantly concentrate on contrast enhancement, which may not adequately consider to preserve natural color attributes in real-world scenarios [4].

In line with advancements in other computer vision tasks, DL-based techniques have made dominant progresses in low-light enhancement [21], [22], [23], [24], [25], [26], [27], [28]. Generative adversarial network (GAN)-based methods, such as Deep Photo Enhancer [24] and EnlightenGAN [25], become increasingly popular for their ability to improve image quality. Despite their effectiveness, GAN-based methods often suffer from training instability and require extensive datasets to produce satisfactory results [22]. Drawing inspiration from Retinex principle, some DL models aim to approximate the illumination components of an image, such as LightenNet [26] and LVENet [27]. Nevertheless, their performance is not always optimal. Recently, light channel enhancement network (LiCENt) has been employed to recover lost details using a neural model [22], and zero-reference deep curve estimation (Zero-DCE) has been used to estimate pixel-wise and high-order curves for dynamic range adjustment [21]. However, these tactics often result in oversaturated colors [22]. Furthermore, a signal-to-noise ratio (SNR)-aware tactic has been proposed to enhance the low-light images with spatially varying adjustments.

B. IMAGE SUPER-RESOLUTION

DL-based SR strategies are roughly divided into non-blind and blind SR algorithms [17], [29], [30], [31]. The majority of non-blind methods [29] operate under the assumption that the blur kernel is a pre-specified degradation kernel, for example, the bicubic interpolation. Recently, some methods are proposed to address multiple degradation issues. Nonetheless, if actual degradation kernels, whether single or multiple, deviate from the assumed ones, non-blind tactics may experience a significant decline in performance [13]. Hence, there is a growing need to focus on unknown blur kernels, which is the domain of blind SR tactics.

Most blind SR approaches include two stages (viz., kernel estimation and kernel-based reconstruction) [13]. Bell-Kliger et al. presented KernelGAN [30], which estimates degradation kernels from individual images, and then uses these estimates to facilitate image reconstruction. Luo et al. introduced a deep alternating network (DAN) [14] that iteratively estimates blur kernel and restores high-resolution images, though the process is time-intensive and computationally expensive [13]. KXNet is a more recent method that jointly estimates blur kernels and high-resolution images [15]. However, these tactics face two major

limitations [13], [31]: 1) The estimated blur kernels lack a strong physical interpretation. 2) The degradation model is adopted heuristically, without fully leveraging the information from both the estimated kernels and the images themselves.

C. JOINT ENHANCEMENT AND SR

None of the previously mentioned methods effectively address the challenges of jointly performing SR and enhancement for low-light and low-resolution images [19]. JSENet is the first to attempt incorporate SR and enhancement via a two-stream deep fusion scheme [18]. Subsequently, LSR proposed a deep lightening network to concurrently address the enhancement and SR issues, though it requires considerable time for training. However, both JSENet and LSR are initial efforts and do not comprehensively solve the issue associated with enhancement and SR, such as color distortion, under- or over-exposure, and the inconsistencies between high- and low-resolution image pairs. In response, this work introduces a new progressive joint of enhancement and blind SR with regularization optimization, specifically designed for low-light and low-resolution images. This aims to eliminate noise and blurriness, enhance quality, and ensure consistency in the mapping between low- and high-resolution images.

III. PROPOSED METHOD

In this section, we supply a comprehensive overview of PESiT, a tactic designed to concurrently enhance the contrast, details, and spatial resolution of low-light and low-resolution images.

A. FRAMEWORK OF PESIT

Although a single enhancement model effectively boosts the contrast of low-light images, it falls short in improving their spatial resolution. High-resolution images, howbeit, are rich in texture details, supplement higher pixel density, and are more reliable. Conversely, a standalone SR model increases image resolution but cannot address the issue of low contrast in low-light images. To overcome the limitation, we propose PESiT, a progressive joint enhancement and SR tactic, which aims to simultaneously enhance contrast and resolution of low-light and low-resolution images. The framework of PESiT is shown in Fig. 2, comprising an improved MSR module via an S-function, followed by a blind SR module via TV constraints. In the former module, the classical logarithm is replaced with an S-shaped function, which stretches the intensity distribution and prevents the color inversion. The enhanced results are then processed by the blind SR, which imposes the consistencies between the reconstructed and input images by incorporating regularization term during testing, effectively tackling various unknown degradations.

It has been proven that traditional Retinex enhancement in RGB color space is equivalent to performing the enhancement operation solely on the V channel in HSV space (H means hue, S means saturation, and V denotes value), while leaving H and S channels unaltered [32]. Hence, the enhancement process is specifically used to the V channel of the image (see Fig. 2). This involves initially converting low-light color images (such as RGB images) into HSV images. The V channel images are then processed via large-, medium-, and small-scale Gaussian filters and subsequently undergo an S-function transformation. Following the V channel correction, the corrected images are converted back to RGB space, gaining final enhanced images.

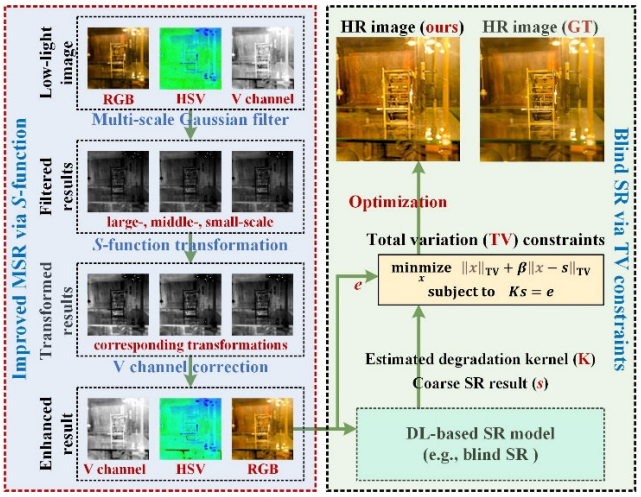

FIGURE 2. Framework of PESiT, which comprises an improved MSR module via S-function and a blind SR module via TV constraints.

Subsequently, the SR module initially uses a blind SR tactic [13] to estimate degradation kernels and acquire super-resolved outcomes, which are considered as coarse SR results. This step harnesses strengths of learning-based methodologies. However, it may not adequately ensure consistencies between SR images and original inputs (See Fig. 2). To address this, the module employs regularization optimization techniques to ensure the consistencies between the input and output pairs, leveraging those benefits of reconstruction-based methods to cater to various degradations and scaling factors. Accordingly, the SR module combines the advantages of both learning- and reconstruction-based strategies, offering excellent adaptability, flexibility, and measurement consistency.

B. IMPROVED MSR VIA S-FUNCTION

For an image I, the mechanism of SSR is described as [6],

$$r_i(x) = \log(R_i(x)) = \log(\frac{I_i(x)}{G(x) * I_i(x)}),$$
$$G(x) = \frac{1}{2\pi\sigma^2} e^{\frac{-x^2}{2\sigma^2}}, \quad (1)$$

where r_i is the output, R_i is the reflectance components of the i^{th} channel map (For an RGB image, $i = 3$), $*$ is the convolution, and G is the Gaussian filter with standard derivation σ (named scale) that defines a scope of convolution kernel. If σ is smaller, image details are effectively improved, but the color of output is easy to be distorted. On the other hand, when σ is larger, the color fidelity of output is better, while

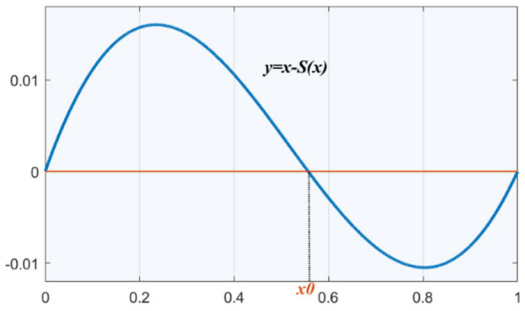

FIGURE 3. The graph of y = x-S(x), if x ∈ [0, 1].

local details are blurrier, and obvious ''halos'' occur at strong edges [7].

$$r_i(x) = \sum_{j=1}^{N} \omega_j \log\left(\frac{I_i(x)}{G_j(x) * I_i(x)}\right), \quad \text{where} \sum_{j=1}^{N} \omega_j = 1, \tag{2}$$

where N is the number of scale parameters, G_j represents the j^{th} Gaussian filtering function with standard derivation σ_j, and ω_j denotes the weighting of j^{th} scale parameter, respectively.

However, the logarithm that employed in (1) and (2) causes some troubles in RGB color space [9], e.g., a potential of color inversion that can be found in those image pixels with 0 values (may jump to 255) or image pixels with 255 values (may drop to 0) [33]. Besides, for a function $h = x\text{-}\ln(x)$, it has a minimum of (1, 1), and is monotonically decreasing in the range of (0, 1] and monotonically increasing in the range of [1, $+\infty$). In this case, those pixels with low magnitude in high magnitude areas dominate variations in regularization term, potentially causing color distortion. Therefore, an S-shaped function is advanced to replace the logarithm in (2), which is defined as,

$$S(x) = e^{1-\frac{2}{e^{\alpha x}+1}} - 1, x \in [0, 1], \quad \text{where } \alpha = ln\frac{1+ln2}{1-ln2}, \tag{3}$$

The weighting parameter α guarantees that the codomain of $S(x)$ is in the interval [0, 1], where $S(0) = 0$ and $S(1) = 1$. The $S(x)$ is monotonically increasing in the range of [0, 1], which tackles effectively the appearance of ambiguity (e.g., value inversion) and/or negative values in the transformation. In addition, the continuity and smoothness of S-function ensure the superiority over other form of function such as piecewise functions.

The graph of $y = x\text{-}S(x)$, $x \in [0, 1]$ is showed in Fig. 3, where 0, x_0 and 1 are three zeros. It can be seen that $S(x)<x$, if $x \in (0, x_0)$, and $S(x)>x$, if $x \in (x_0,1)$. Under the case, when the intensity of an image is normalized to [0, 1], in those pixels with low magnitude ($< x_0$), the intensities are expanded. In contrast, high intensities ($> x_0$) are suppressed. This property can be applied to effectively stretch the intensity distribution of image, which is favorable to improve image contrast.

Inspired by previous pioneering works [8], [9] that employ enhancement implementations on the V channel in HSV space, we operate the enhancement via S-function on the V channel. Considering (3), equation (2) becomes as,

$$r_v(x) = \sum_{j=1}^{N} \omega_j S\left(norm\left(\frac{I_v(x)}{G_j(x) * I_v(x)}\right)\right), \quad \sum_{j=1}^{N} \omega_j = 1, \tag{4}$$

where I_v denotes the V channel image and *norm* is a min-max normalization operation.

To balance enhancement performance against computation complexity of Gaussian filter, those parameters σ_j in (4) are opted as the large-, middle- and small-scale, i.e.,

$$\sigma_j = a_j \cdot max(I_v), \quad j = 1, 2, 3, \tag{5}$$

where a_j is a positive constant. According to the experiments, those parameters a_j are empirically set to 0.8, 0.5, and 0.2 in this section, which guarantees favorable results.

Afterwards, the reflectance estimation acquired from (4) is requisite to be adjusted through the V channel correction, for further bettering the visual quality. According to the histogram of V channel image, the V channel correction is,

$$r_v' = \begin{cases} 0, & \text{if } r_v \le r^{\text{low}} \\ \left(\dfrac{r_v - r^{\text{low}}}{r^{\text{up}} - r^{\text{low}}}\right)^{\gamma}, & \text{if } r^{\text{low}} \le r_v \le r^{\text{up}} \\ 1, & \text{if } r_v \ge r^{\text{up}}, \end{cases} \tag{6}$$

where r^{up} and r^{low} denote high and low shearing points of V channel image r_v, and γ is a constant. The larger γ may cause overexposed, and the smaller value may result in insufficient enhancement. In this study, γ is adjusted experimentally within the range of [0.6,1.2], bringing favorable outcomes. The upper and lower bounds of confidence intervals, set as a confidence level of 0.99, are determined from the cumulative histogram of V channel images, defining the thresholds for high and low shearing points. Finally, the adjusted image is converted back to RGB space, and gain final enhancement results.

C. BLIND SR VIA TV CONSTRAINTS

In view of traditional degradation models, most existing SR models assume that the low-resolution image is a blurred and down-scaled high-resolution image with an additional white Gaussian noise n. The degradation process is formulated as,

$$I_{LR} = (I_{HR} \odot K) \downarrow_s + n, \tag{7}$$

where K, $\odot$ and $\downarrow_s$ denote the blur kernel, convolution and down-sampling with scaling factor of s, respectively. Despite the works gain state-of-the-art performance, they suffer from two major limitations [16]: 1) The trained network is operated to a unique scaling factor and blur kernel. 2) Such approach is difficult to guarantee the consistencies between low- and high-resolution images during testing.

To resolve the first limitation, we focus on the blind SR with multiple degradations, e.g., the blur, noise and down-sampling, which may simultaneously happen in real-world cases [11]. A well-principled blind SR algorithm [13] is applied in this work. In that strategy, the degradation module is firstly reformulated for disentangling the blur kernel estimation and up-sampling. Secondly, a dynamic deep linear kernel is used to further better the kernel estimation. Thirdly, a deep constrained least squares deconvolution module is adopted in the feature domain, which betters robustness to noise, providing a theoretical/principled guidance to get clean features from blurred inputs. Lastly, both clean and primitive features are fed into a dual-path network to yield SR images. Thus, we obtain the estimated degradation kernel and the output result of SR.

To tackle the second limitation, TV constraints are imposed on blind SR output. Similar to [16], the consistency is imposed on low- and high-resolution pairs, that is, reconstruct a high-resolution image x^* from a low-resolution image l through the following optimization strategy,

$$\min_{x} \text{mize } \|x\|_{\text{TV}} + \beta\|x - s\|_{\text{TV}}, \text{ s.t. } \text{K}x = l, \tag{8}$$

where K and s denote the estimated degradation kernel and SR result obtained from the SR model, and β is a regularization parameter. The first term defines a transition among different objects, the second term measures the distance between x^* and s, and the constraint enforces the measurement consistency.

It has been proved that TV minimization is associated with l_1 minimization [16]. Thus, equation (8) can be converted to l_1 minimization,

$$\min_{x} \text{mize } \|x\|_1 + \beta\|x - \hat{s}\|_1, \text{ s.t. } \text{K}x = l, \tag{9}$$

An alternating direction method of multipliers (ADMM) [16] is applied to solve (9). Supplying two auxiliary variables p and q, and defining $\hat{s} = Ds$, where D represents the vertical concatenation of differences, equation (9) is equivalent to the following formula, that is,

$$\begin{aligned} &\min_{p,x,q} \text{mize } \|p\|_1 + \beta\|p - \hat{s}\|_1, \\ &\text{s.t. } \text{K}x = l, Dq = p, q = x, \end{aligned} \tag{10}$$

After that, the correspondence between (10) and a standard operator of ADMM is constructed, that is,

$$\begin{aligned} f(p, x) = ||p||_1 + \beta||p - \hat{s}||_1 + i_{Kx=e}(x), g(q) = 0 \\ F = \begin{bmatrix} -I_{2n} & 0 \\ 0 & I_n \end{bmatrix}, \ G = \begin{bmatrix} D \\ -I_n \end{bmatrix}, \end{aligned} \tag{11}$$

where $\boldsymbol{i}_{\text{K}x=e}(x)$ is the indicator function of $Kx = l$, and I_n is a $n \times n$ identity matrix. This correspondence results in a closed-form solution.

IV. EXPERIMENTAL RESULTS

This section supplies datasets, baseline methods and metrics for comparisons, and then applies the data to demonstrate the effectiveness of the proposed method.

A. DATASETS

The LOL dataset [10] is a benchmark designed to address the real-world challenge of low-light image enhancement, which is usually applied in diverse enhancement strategies [24], [34]. Besides, following [15], [17], [30], [31], we apply commonly-used benchmark data for SR comparisons, including DIV2K [35] and Flickr2K [36]. Thus, three public datasets are chosen for qualitative and quantitative comparisons in this section.

The LOL data comprises paired images captured under low- and normal-light conditions, with normal-light images serving as high-quality GT references for SR comparisons. In order to synthesize low-resolution images, both low- and normal-light images are down-sampled by a scale factor of $\times 4$ using bicubic interpolation. Those down-sampled low-light images are then input into PESiT, while the downscaled normal-light images are denoted as GT for enhancement comparisons. This process results in the creation of LOL-LR data, which consists of 900 image pairs. After, the outputs from the enhancement module are further processed through the SR module (see Fig. 2). This sequence of operations gives rise to the LOL-SR data. To facilitate training and evaluation, LOL-SR dataset is randomly split into three subsets: 80% for training, 10% for validation, and 10% for testing.

DIV2K and Fickr2k datasets provide high-quality images along with corresponding versions at $\times 2$, $\times 3$ and $\times 4$ scales. For the purpose of evaluation, we select the images down-sampled by a factor of $\times 4$ to serve as GT for enhancement comparisons. To simulate low-light conditions, $\times 4$ downscaled images are transformed into HSV color space. The Value (V) component is then dimmed to 0.2 of its original level, but the Hue (H) and Saturation (S) components remain unaltered. The images are subsequently converted back to the RGB color space, resulting in the creation of the DIV2K-LR and Flickr2K-LR datasets for low-light enhancement comparisons. Following this, outputs generated by the enhancement module are fed into SR module, as illustrated in Fig. 2. This process produces DIV2K-SR and Flickr2K-SR datasets. These datasets are divided into subsets for training, validation, and testing, with 80% of the samples allocated for training, using 10% for validation, and 10% for testing purposes.

B. BASELINE METHODS AND METRICS

Since the proposed PESiT contains enhancement and blind SR modules, we adopt nine enhancement methods as baseline for enhancement comparisons and four SR tactics as baseline for SR comparisons. For enhancement comparisons, we use MSR [7], WVmSE [20], Zero-DCE [21], LiCENt [22], BIMEF (bio-inspired multi-exposure fusion) [23], MTFE (multiple transformation function estimation) [28], LWFI (lightweight fast illumination) [33], QDCHE (quadrant dynamic clipped HE) [37], and AGCWHD (adaptive gamma correction with weighted histogram distribution) [38] as baseline approaches. Among, MSR, WVmSE,

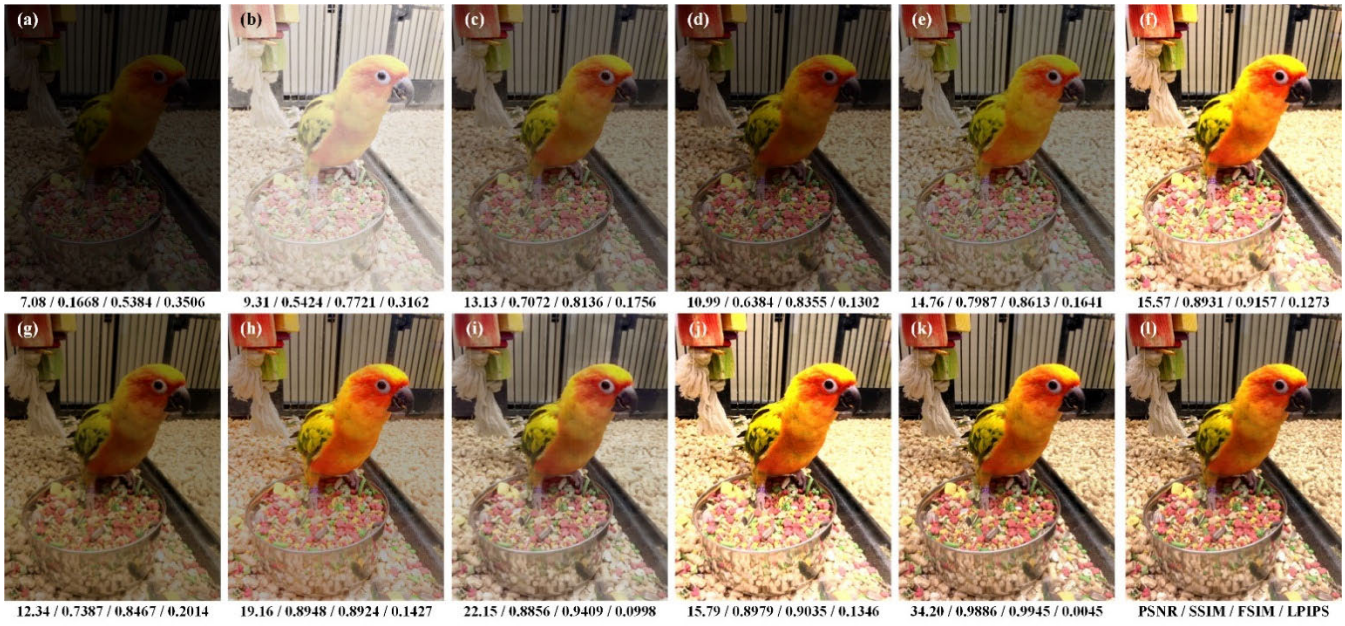

FIGURE 4. A low-light and low-resolution image with a bird (a), enhanced results obtained from MSR (b), BIMEF (c), WVmSE (d), Zero-DCE (e), AGCWHD (f), LWFI (g), LiCNEt (h), MTFE (i), QDCHE (j), and proposed enhancement module (k), and GT (l), respectively.

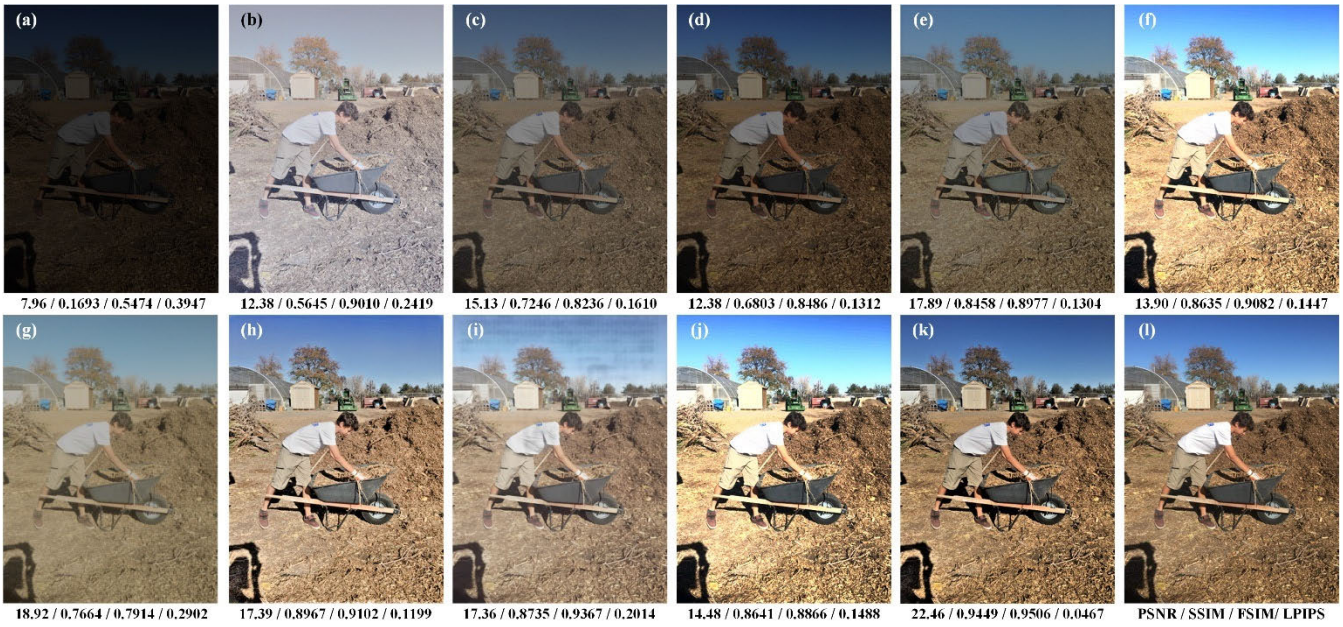

FIGURE 5. A representative low-light and low-resolution image (a), enhanced results obtained from MSR (b), BIMEF (c), WVmSE (d), Zero-DCE (e), AGCWHD (f), LWFI (g), LiCNEt (h), MTFE (i), QDCHE (j), and proposed enhancement module (k), and GT (l), respectively.

QDCHE, AGCWHD, and BIMEF are conventional strategies, which are operated by MATLAB R2022a (MathWorks, Natick, MA). Meanwhile, MTFE, Zero-DCE, LiCENt, and LWFI belong to DL-based tactics, which are used the respective training model provided by the authors.

For the purpose of SR comparisons, we choose DCLS (deep constrained least squares model) [13], DAV [14], Cria-CL [17], and RCAN (residual channel attention network) [31] as baselines. RCAN, Cria-CL represent non-blind SR algorithms, while DAN, DCLS are blind ones. All baseline SR strategies are DL-based, and we utilize the pre-trained models provided by the authors are used in this section.

Four measures are used to assess the enhancement and/or SR performance of different algorithms, such as PSNR, SSIM [39], FSIM [40], and LPIPS [41]. The measures assess the distance between reconstruction results and

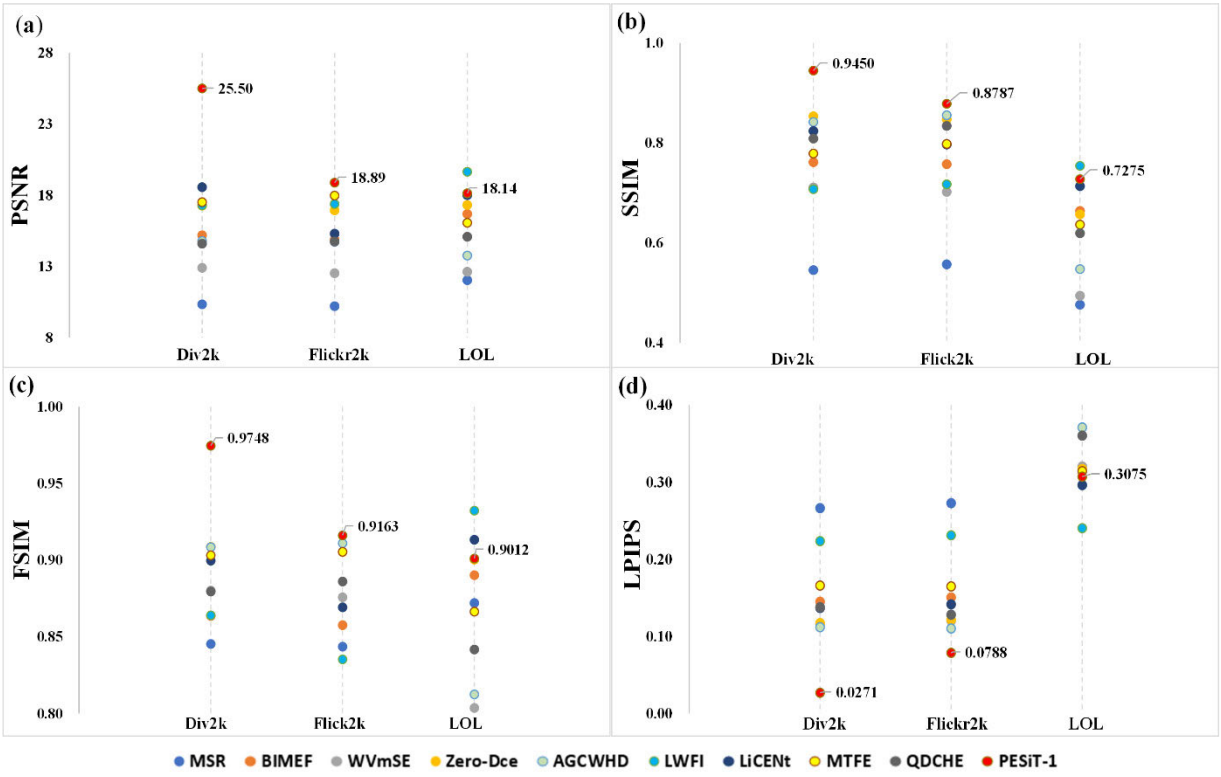

FIGURE 6. Quantitative enhancement comparisons regarding PSNR (a), SSIM (b), FSIM (c), and LPIPS (d) obtained from the baseline approaches and proposed enhancement module on DIV2K, Flickr2K, and LOL datasets.

corresponding GTs, which are widely employed in the assessment of image quality. For example, SSIM is applied to measure structural similarity between paired images, and LPIPS is to judge their perceptual similarity. Higher PSNR, SSIM and FSIM scores, and lower LPIPS scores imply better perceptual quality.

C. ENHANCEMENT COMPARISONS

Fig. 4(a) lists a low-light, low-resolution image depicting a bird scene, characterized by PSNR of 7.08, SSIM of 0.1688, FSIM of 0.5384, and LPIPS of 0.3506. Low image contrast makes it challenging to identify object of interest. Following enhancement with various algorithms (namely MSR, BIMEF, WVmSE, Zero-DCE, AGCWHD, LiCENt, MTFE, QDCHE, and in the proposed module), the contrast is notable improved. The corresponding enhanced results shown in Figs. 4(b)- (k), respectively, with quantitative measures improved and listed below each image, making the object of interest prominent.

When compared to GT, represented by normal-light image in Fig. 4(l), enhanced results produced by the nine baseline approaches exhibit varying degrees of distortion. For instance, MSR yields over-enhancement, and such phenomenon is also observed in outputs of AGCWHD and QDCHE. Conversely, other baselines may cause under-enhancement, e.g., Figs. 4(c)-(e), and (g). However, the proposed enhancement module effectively mitigates those risks of both over- and/or under-enhancement, closely approximating the quality of GT image.

Similarly, a low-light people image is displayed in Fig. 5(a), with PSNR of 7.96, SSIM of 0.1693, FSIM of 0.5474, LPIPS of 0.3947, and Figs. 5(b)-(l) display enhanced results obtained from the baselines and proposed enhancement modules. It can be seen that the over-enhancement phenomena happen in Figs. 5(b), (f), (h), and (j), but the under-enhancement phenomena occurs in Figs. 5(c)-(e), (g), and (i). In this case, those enhanced results impact the visual quality assessment and subsequently affects the subsequent SR process. However, these issues are successfully mitigated in Fig. 5(k), which validates superior performance among the comparisons.

The PSNR, SSIM, FSIM, and LPIPS values shown beneath each image reveal that the low-light and low-resolution image initially has the lowest PSNR, SSIM, and FSIM scores, as well as the highest LPIPS score, which could pose challenges for image analysis. While all the baseline strategies contribute to an improvement in these measures, there are notable variations among them. For instance, MTFE achieves the second-highest improvements in both PSNR and FSIM for Figs. 4 and 5, and LiCENt gains the second-best results for SSIM and LPIPS in Fig. 4. In contrast, proposed enhancement module consistently obtains the highest PSNR, SSIM, and FSIM terms scores, as well as the lowest LPIPS scores across both figures, indicating superior visual quality.

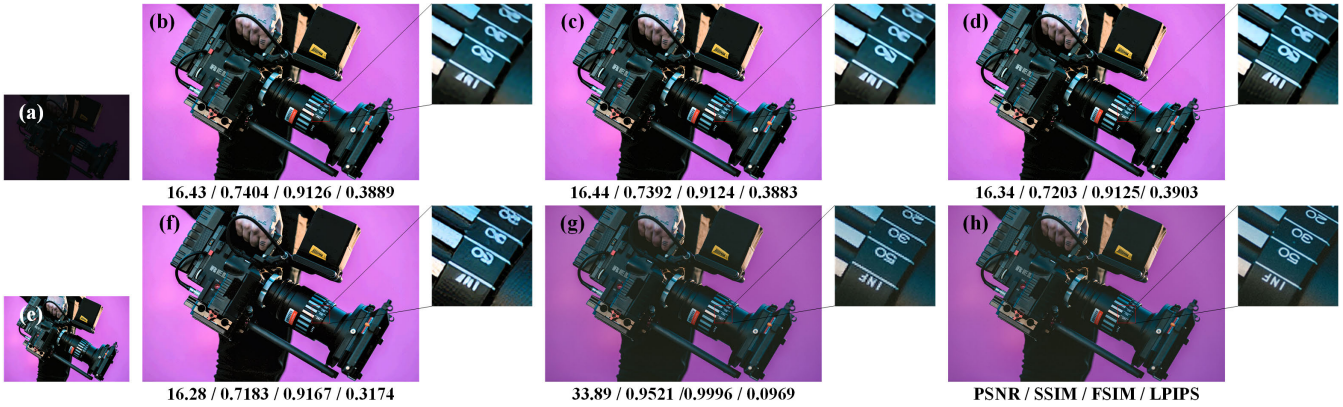

FIGURE 7. *Img* 67 from Div2k, a low-light image (a), enhanced result (e), ×4 SR results of DAV (b), DCLS (c), RCAN (d), Cria-CL (f), and PESiT (g), and GT with high resolution (h), respectively.

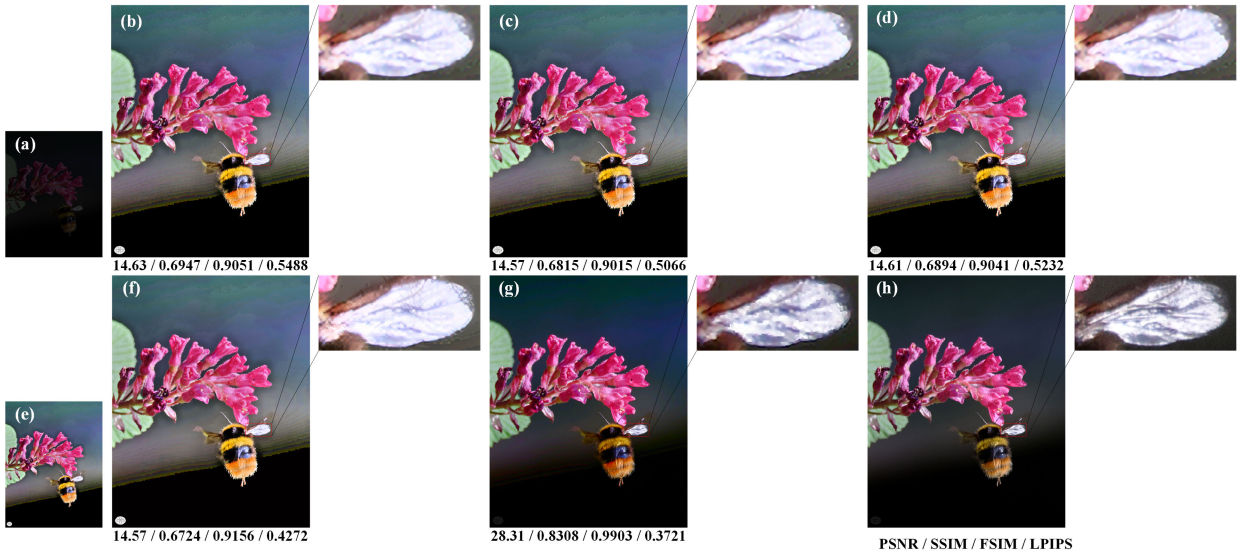

FIGURE 8. *Img* 2414 from Flickr2k, a low-light image (a), enhanced result (e), ×4 SR results of DAV (b), DCLS (c), RCAN (d), Cria-CL (f), and PESiT (g), and GT with high resolution (h), respectively.

Fig. 6 presents quantitative comparisons across the MSR, BIMEF, WVmSE, Zero-DCE, AGCWHD, LWFI, LiCENt, MTFE, QDCHE, and our proposed enhancement module for the three datasets (using the same parameters for testing), focusing on PSNR, SSIM, FSIM and LPIPS metrics. Since the four metrics measure the distance between the reference and enhanced results (PSNR focuses on fidelity of an image, SSIM addresses its structure similarity, FSIM concentrates on its feature similarity, and LPIPS focuses on its perceptual similarity), higher PSNR, SSIM, and FSIM, and lower LPIPS scores indicate better enhancement performance.

As listed in Fig. 6, the proposed enhancement module gains consistently the highest PSNR, SSIM, and FSIM values, along with the lowest LPIPS scores, with average values of 20.14, 0.8586, 0.9206, and 0.1113, respectively. This prove that our enhancement strategy surpasses the baseline methods in terms of each measure in alignment with observations from Fig. 3 and 4. Thus, PESiT exhibits high reliability and robustness in enhancement. This is clearly instrument in enhancing objects of interest, elucidating context, and preserving structural detail information in images.

D. SUPER-RESOLUTION COMPARISONS

Section IV-C reveals that the proposed enhancement module produces enhanced results with impressive performance, such as high fidelity, robustness, and reliability. This guarantees the effectiveness and trustworthiness of subsequent SR process. Fig. 7 presents an illustration applying *Img* 67 from DIV2K dataset, where Fig. 7(a) shows a low-light image with low resolution, and Fig. 7(e) supplies the enhanced outcome gained by the proposed enhancement module. The figure also includes ×4 SR results from DAV Fig. 7(b), DCLS Fig. 7(c), RCAN Fig. 7(d), Cria-CL Fig. 7(f) and PESiT Fig. 7(g), as well as high-resolution GT Fig. 7(h). It is

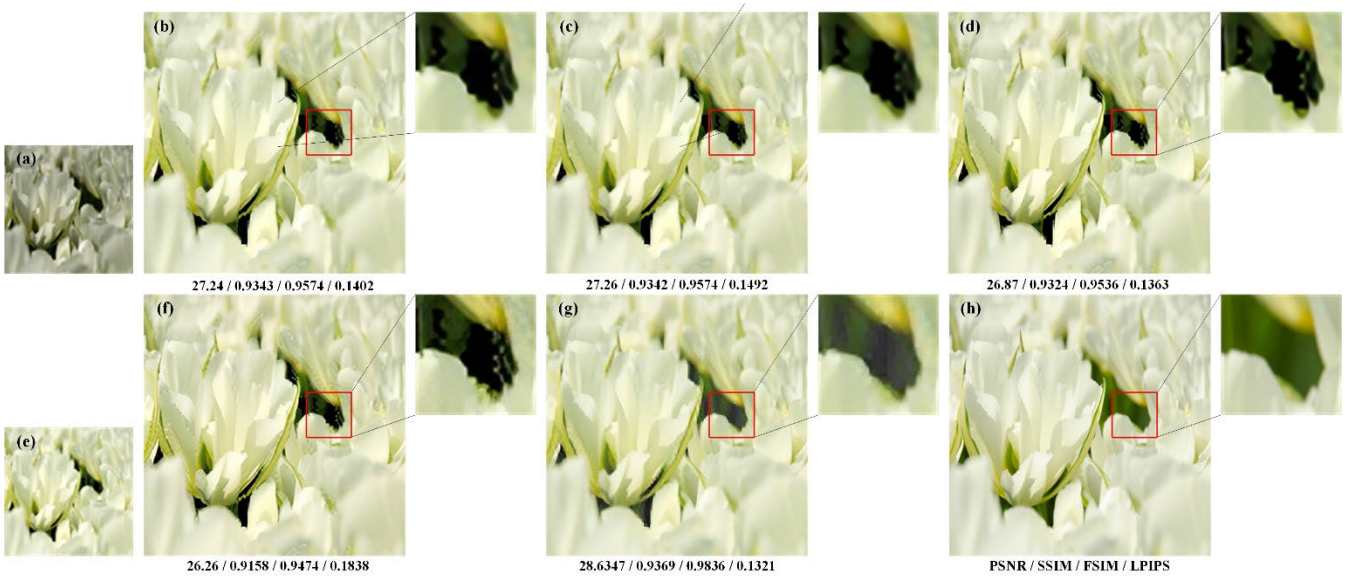

FIGURE 9. *Img* 469 from LOL, a low-light image (a), enhanced result (e), ×4 SR results of DAV (b), DCLS (c), RCAN (d), Cria-CL (f), and PESiT (g), and GT with high resolution (h), respectively.

evident that details e.g., the camera aperture count, which are indiscernible in the GT image, are challenging to recognize in those results from other algorithms like DAV Fig. 7(b) and DCLS Fig. 7(c). This problem is widespread among the other algorithms as well. Howbeit, the image gained by the proposed tactic overcomes this challenge, validating its ability to preserve details more effectively. The quantitative measures supplied below each result substantiate the superior performance of PESiT.

Fig. 8 supplies ×4 SR results for *Img* 2414 from Flickr2K dataset, using the baseline and proposed algorithms, alongside the high-resolution GT. It is evident that the texture on the bee's wings is sharply delineated with clear features in Fig. 8(h). In contrast, the texture detail is indistinct in the results from Figs. 8(b)-(d) and (f), whereas it is notably sharper in Fig. 8(g). Moreover, the background colors in Figs. 8(b)-(d) and (f) differ from that in Fig. 8(g) and Fig. 8(h), suggesting that these methods deviate from GT under the given lighting conditions. As with Fig. 7, the details that are vague in the baseline results are well-defined in Fig. 8(g), indicating that PESiT is good at preserving intricate detail information. Compared to the four baseline tactics, our SR module delivers superior output with optimal performance, as evidenced by the highest PSNR, SSIM, and FSIM scores, as well as the lowest LPIPS value.

Fig. 9 displays *Img* 469 from LOL dataset, including the original low-light image Fig. 9(a), the enhanced image produced by our module (e), and ×4 SR results from DAV Fig. 9(b), DCLS Fig. 9(c), RCAN Fig. 9(d), Cria-CL Fig. 9(f), and PESiT Fig. 9(g), along with the high-resolution GT Fig. 9(h). When compared to the GT, the baseline tactics introduce inaccuracies into the image, particularly noticeable in the RCAN Fig. 9(d) and Cria-CL Fig. 9(f) results. However, such issues are absent in PESiT outcome Fig. 9(g),

TABLE 1. Quantitative SR comparisons on the DIV2K. The best two results are marked in RED and BLUE colors.

Method	PSNR ↑	SSIM ↑	FSIM ↑	LPIPS ↓
DAV	22.40	0.8455	0.9717	0.3269
DCLS	22.40	0.8459	0.9718	0.3391
RCAN	22.46	0.8473	0.9711	0.3163
Cria-CL	21.85	0.8078	0.9727	0.2043
PESiT (ours)	27.36	0.8803	0.9957	0.2676

which also achieves the best auabest quantitative metrics in terms of PSNR, SSIM, FSIM, and the lowest LPIPS, reinforcing the superiority of proposed method.

As for the quantitative measures, Table 1 lists comparisons of PSNR, SSIM, FSIM, and LPIPS metrics among the DAV, DCLS, RCAN, Cria-CL, and PESiT for DIV2K dataset. The results indicate that PESiT achieves the top scores in PSNR, SSIM, and FSIM, and ranks second in LPIPS. RCAN attains the second-highest scores for PSNR and SSIM, while Cria-CL secures the second-highest FSIM and the lowest LPIPS scores. The insights availably confirm that PESiT outperforms those baseline methods, guaranteeing high fidelity, robustness, and reliability in the SR outcomes.

Table 2 provides a quantitative analysis of each metric for Flickr2K data. Unlike Table 1, PESiT achieves the best overall performance across all measures. DAV acquires the second-highest PSNR, and DCLS attains the second-best SSIM score. Regarding LPIPS, Cria-CL achieves the second-lowest score. These results imply that proposed strategy excels in fidelity, structural similarity, and feature similarity for Flickr2K.

Table 3 presents the quantitative comparisons for LOLdata. It is observed that DAV and DCLS achieve the second-highest scores for PSNR, with DAV also acquiring the second rank

TABLE 2. Quantitative SR comparisons on the Flickr2k. The best two results are marked in RED and BLUE colors.

Method	PSNR ↑	SSIM ↑	FSIM ↑	LPIPS ↓
DAV	17.69	0.7876	0.9154	0.3632
DCLS	17.74	0.7898	0.9157	0.3718
RCAN	17.52	0.7816	0.9123	0.3560
Cria-CL	17.26	0.7340	0.9176	0.2592
PESiT (ours)	25.51	0.8417	0.9928	0.2725

TABLE 3. Quantitative SR comparisons on the lol. The best two results are marked in RED and BLUE colors.

Method	PSNR ↑	SSIM ↑	FSIM ↑	LPIPS ↓
DAV	17.30	0.7660	0.8561	0.4780
DCLS	17.30	0.7672	0.8562	0.4885
RCAN	17.17	0.7639	0.8518	0.4616
Cria-CL	16.89	0.6269	0.8418	0.3423
PESiT (ours)	23.55	0.7366	0.9335	0.3366

in SSIM, but DCLS attains the highest SSIM score. Additionally, DCLS achieves the second-highest FSIM result, and Cria-CL obtains the second-highest value for LPIPS term. Importantly, PESiT obtains the best results across PSNR, FSIM, and LPIPS, thereby demonstrating the efficacy of the proposed method.

V. CONCLUSION

To simultaneously enhance the contrast, details, and resolution of low-light and low-resolution images, this work proposes a novel progressive joint approach that integrates enhancement and blind SR strategies, named PESiT. This method addresses those limitations of standalone enhancement or SR approaches, supplying improved robustness, reliability, and effectiveness. Experimental results on DIV2K, Flickr2K, and LOL datasets clearly imply that PESiT not only works more robustly across different low-light / low-resolution cases but also outperforms state-of-the-art methods in terms of key metrics, including the highest PSNR, SSIM, and FSIM scores, and the lowest LPIPS one. This suggest that PESiT effectively preserves fidelity and structural, perceptual, and feature similarity to GT. Therefore, PESiT displays great potentials in the SR of low-light and low-resolution image, aiding in the advancement of image analysis, interpretation and recognition. This lays a solid foundation for our future work in this domain.

REFERENCES

[1] C. Tian, Y. Yuan, S. Zhang, C.-W. Lin, W. Zuo, and D. Zhang, "Image super-resolution with an enhanced group convolutional neural network," *Neural Netw.*, vol. 153, pp. 373–385, Sep. 2022.

[2] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, "MADNet: A fast and lightweight network for single-image super resolution," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1443–1453, Mar. 2021.

[3] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Müller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. 1st Conf. Visualizat. Biomed. Comput.*, Atlanta, GA, USA, 1990, pp. 337–345.

[4] L. Li, R. Wang, W. Wang, and W. Gao, "A low-light image enhancement method for both denoising and contrast enlarging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, Sep. 2015, pp. 3730–3734.

[5] X. Wang, J. Ma, J. Jiang, and X.-P. Zhang, "Dilated projection correction network based on autoencoder for hyperspectral image super-resolution," *Neural Netw.*, vol. 146, pp. 107–119, Feb. 2022.

[6] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[7] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[8] K.-F. Yang, X.-S. Zhang, and Y.-J. Li, "A biological vision inspired framework for image enhancement in poor visibility conditions," *IEEE Trans. Image Process.*, vol. 29, pp. 1493–1506, 2020.

[9] 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.

[10] W. Yang, W. Wang, H. Huang, S. Wang, and J. Liu, "Sparse gradient regularized deep retinex network for robust low-light image enhancement," *IEEE Trans. Image Process.*, vol. 30, pp. 2072–2086, 2021.

[11] G. Yin, W. Wang, Z. Yuan, W. Ji, D. Yu, S. Sun, T.-S. Chua, and C. Wang, "Conditional hyper-network for blind super-resolution with multiple degradations," *IEEE Trans. Image Process.*, vol. 31, pp. 3949–3960, 2022.

[12] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 3262–3271.

[13] Z. Luo, H. Huang, L. Yu, Y. Li, H. Fan, and S. Liu, "Deep constrained least squares for blind image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 17621–17631.

[14] Z. Zhang, K. Gao, J. Wang, S. Ji, X. Gu, and S. Li, "Switching the iteration in alternating optimization for blind super-resolution," *Proc. SPIE*, vol. 12557, pp. 315–322, Jan. 2023.

[15] J. Fu, H. Wang, Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "KXNet: A model-driven deep neural network for blind super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2022, pp. 235–253.

[16] M. Vella and J. F. C. Mota, "Robust single-image super-resolution via CNNs and TV-TV minimization," *IEEE Trans. Image Process.*, vol. 30, pp. 7830–7841, 2021.

[17] Y. Shi, H. Li, S. Zhang, Z. Yang, and X. Wang, "Criteria comparative learning for real-scene image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8476–8485, Dec. 2022.

[18] K. Lyu, S. Pan, Y. Li, and Z. Zhang, "JSENet: A deep convolutional neural network for joint image super-resolution and enhancement," *Neurocomputing*, vol. 489, pp. 570–583, Jun. 2022.

[19] M. T. Rasheed and D. Shi, "LSR: Lightening super-resolution deep network for low-light image enhancement," *Neurocomputing*, vol. 505, pp. 263–275, Sep. 2022.

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

[21] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1777–1786.

[22] A. Garg, X.-W. Pan, and L.-R. Dung, "LiCENt: Low-light image enhancement using the light channel of HSL," *IEEE Access*, vol. 10, pp. 33547–33560, 2022.

[23] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," 2017, *arXiv:1711.05591*.

[24] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, and Y.-Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 6306–6314.

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

[26] X. Guo and Q. Hu, "Low-light image enhancement via breaking down the darkness," *Int. J. Comput. Vis.*, vol. 131, no. 1, pp. 48–66, Jan. 2023.

[27] Y. Guo, Y. Lu, and R. W. Liu, "Lightweight deep network-enabled real-time low-visibility enhancement for promoting vessel detection in maritime video surveillance," *J. Navigat.*, vol. 75, no. 1, pp. 230–250, Jan. 2022.
[28] J. Park, A. G. Vien, M. Cha, T. T. Pham, H. Kim, and C. Lee, "Multiple transformation function estimation for image enhancement," *J. Vis. Commun. Image Represent.*, vol. 95, Sep. 2023, Art. no. 103863.
[29] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[30] Y. Wang, "Edge-enhanced feature distillation network for efficient super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, New Orleans, LA, USA, Jun. 2022, pp. 776–784.
[31] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 286–301.
[32] Y. Zhang, X. Di, B. Zhang, R. Ji, and C. Wang, "Better than reference in low-light image enhancement: Conditional re-enhancement network," *IEEE Trans. Image Process.*, vol. 31, pp. 759–772, 2022.
[33] F. Matin, Y. Jeong, K. Kim, and K. Park, "Color image enhancement using multiscale retinex based on particle swarm optimization method," *J. Phys., Conf.*, vol. 960, Jan. 2018, Art. no. 012026.
[34] Z. Cui, K. Li, L. Gu, S. Su, P. Gao. Z. Jiang, Y. Qiao, and T. Harada, "You only need 90K parameters to adapt light: A light weight transformer for image enhancement and exposure correction," in *Proc. Brit. Mach. Vis. Conf.*, London, U.K., 2022, p. 238.
[35] B. Subramani and M. Veluchamy, "Quadrant dynamic clipped histogram equalization with gamma correction for color image enhancement," *Color Res. Appl.*, vol. 45, no. 4, pp. 644–655, Aug. 2020.
[36] M. Veluchamy and B. Subramani, "Image contrast and color enhancement using adaptive gamma correction and histogram equalization," *Optik*, vol. 183, pp. 329–337, Apr. 2019.
[37] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 1122–1131.
[38] R. Timofte, E. Agustsson, L. V. Gool, M. Yang, and L. Zhang, "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 1110–1121.
[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[40] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[41] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 586–595.

HE DENG received the Ph.D. degree in control science and engineering from Huazhong University of Science and Technology, Wuhan, China, in 2011. He is currently a Professor with Wuhan University of Science and Technology, Wuhan. His current research interests include artificial intelligence and medical image processing.

KAI CHENG received the B.S. degree from Wuhan University of Science and Technology, Wuhan, China, in 2021, where he is currently pursuing the Ph.D. degree. His current research interest includes medical image processing.

YUQING LI received the B.S. degree from Wuhan University of Science and Technology, Wuhan, China, in 2023, where she is currently pursuing the Ph.D. degree. Her current research interest includes artificial intelligence.