IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 May 2024, accepted 16 May 2024, date of publication 20 May 2024, date of current version 30 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3403096

RESEARCH ARTICLE

Semantic-Aware Guided Low-Light Image Super-Resolution

SHENG REN[1], RUI CAO[2], WENXUE TAN[1], AND YAYUAN TANG[3]
[1]School of Computer and Electrical Engineering, Hunan University of Arts and Science, Changde 415000, China
[2]Information and Network Center, Central South University, Changsha 410083, China
[3]School of Information Engineering, Hunan University of Science and Engineering, Yongzhou 425199, China

Corresponding author: Rui Cao (ruicao@csu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62102147; in part by the National Science Foundation of Hunan Province under Grant 2022JJ30424, Grant 2022JJ50253, and Grant 2022JJ30275; in part by the Scientific Research Project of Hunan Provincial Department of Education under Grant 21B0616 and Grant 21B0738; and in part by Hunan University of Arts and Sciences Doctoral Research Initiation Project under Grant 22BSQD02 and Grant 20BSQD13.

ABSTRACT The single image super-resolution based on deep learning has achieved extraordinary performance. However, due to inevitable environmental or technological limitations, some images not only have low resolution but also low brightness. The existing super-resolution methods for restoring images through low-light input may encounter issues such as low brightness and many missing details. In this paper, we propose a semantic-aware guided low-light image super-resolution method. Initially, we present a semantic perception guided super-resolution framework that utilizes the rich semantic prior knowledge of the semantic network module. Through the semantic-aware guidance module, reference semantic features and target image features are fused in a quantitative attention manner, guiding low-light image features to maintain semantic consistency during the reconstruction process. Second, we design a self-calibrated light adjustment module to constrain the convergence consistency of each illumination estimation block by self-calibrated block, improving the stability and robustness of output brightness enhancement features. Third, we design a lightweight super resolution module based on spatial and channel reconstruction convolution, which uses the attention module to further enhances the super-resolution reconstruction capability. Our proposed model surpasses methods such as RDN, RCAN, and NLSN in both qualitative and quantitative analysis of low-light image super-resolution reconstruction. The experiment proves the efficiency and effectiveness of our method.

INDEX TERMS Low-light image, semantic-aware, super-resolution.

I. INTRODUCTION

Among the external information obtained by human beings, visual information accounts for about 63% to 83%. Fast and efficient processing of visual information is an important core of human intelligence. In recent years, with the rapid development of artificial intelligence (AI) technology, new technologies such as visual AI, decision-making AI, speech semantic AI and AI robot have an unprecedented impact in medical care, technology, education, transportation, finance, entertainment and other fields. In the AI market, visual AI has the highest share of more than 40%. The efficient and rapid processing of visual information has become a crucial component of AI technology. The emergence of computer vision technology based on deep neural network has greatly promoted the development of visual AI in terms of quality, efficiency and practical application.

Visual AI is based on high-quality images or video data. However, due to the influence of hardware cost, equipment process, shooting time and angle, some of the image or video data acquired by hardware sensors have low resolution and brightness. Low quality data seriously affects the accuracy of visual AI, such as target detection, image segmentation, facial recognition, pattern recognition, image

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

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

classification and other upstream visual AI. Super-resolution technology belongs to low-level vision AI technology, which can reconstruct low-resolution images into high-resolution data without increasing the hardware cost, effectively improving the accuracy and robustness of upstream vision AI tasks.

Dong Cao et al. first applied deep neural network to super-resolution reconstruction, and the proposed Super-Resolution Convolutional Neural Networks (SRCNN) greatly improves the quality of reconstruction [1]. Subsequently, Dong Cao et al. proposed Fast SRCNN (FSRCNN), which further improves the efficiency of super-resolution reconstruction [2]. After He et al. presented the residual neural network, Kim et al. applied the idea of residual learning to super-resolution reconstruction, and proposed method such as Super-Resolution using Very Deep convolutional networks (VDSR), Deeply-Recursive Convolutional Network for Image Super-Resolution (DRCN) [3], [4]. The super-resolution reconstruction uses residual neural networks as the backbone network, which enables the super-resolution model to focus on learning high-frequency information. The low frequency information is transmitted directly from the input end by using the fast connection, which can greatly reduce the computation and improve the quality of super-resolution reconstruction, and become the basis of constructing super-resolution networks. Subsequently, the introduction of attention mechanisms, generative adversarial networks, prior knowledge and other technologies continuously improved the quality and efficiency of super-resolution reconstruction.

The super-resolution technology based on deep learning has achieved extraordinary performance. However, existing super-resolution methods mainly rely on deep neural networks to learn the nonlinear mapping relationship between pixels in low-resolution (LR) images and high-resolution (HR) images, lacking understanding and utilization of overall semantic knowledge of images. In addition, due to inevitable environmental or technological limitations, some images not only have low resolution but also low brightness. The existing super-resolution methods for restoring images through low-light input may encounter issues such as low brightness and many missing details. To address the above issues, we propose a semantic-aware guided low-light image super-resolution method, which uses a semantic segmentation network pre-trained on a large-scale dataset PASCAL-Context to provide semantic prior guidance knowledge for super-resolution reconstruction. and we build an adaptive brightness enhancement module to learn the illumination component to enhance the brightness of the input low resolution image. The main contributions of this paper are summarized as follows:

(1) We propose a semantic-aware guided low-light image super-resolution framework based on the rich semantic prior knowledge provided by semantic segmentation network, including intermediate features and semantic mapping. The framework uses semantic-awareness guidance module (SAGM) to guide the low-light super-resolution module to have the region awareness ability and restrict the reconstruction authenticity of detail-rich regions.

(2) We design an adaptive lighting adjustment module (LAM) for learning illumination components, which constrains the convergence consistency of each illumination estimation block through self-calibrated blocks. This module improves the stability and robustness of the output brightness enhancement feature, and improves the brightness of the input low resolution image.

(3) We design a lightweight super-resolution module (SRM) based on spatial and channel reconstruction convolution. On the basis of semantic guidance and LAM, the ability of SRM is enhanced through feature distillation, feature aggregation, and feature optimization, as well as the introduction of attention module. The experimental results demonstrate large performance improvements by our semantic-aware guided super-resolution (SAGSR).

II. RELATED WORK

Early super-resolution algorithms were mostly based on interpolation, while traditional interpolation algorithms used pixels around sampling points to recover target pixels, including nearest neighbor interpolation, bilinear interpolation, bicubic interpolation, etc. Before the application of deep learning technology to super-resolution reconstruction, most advanced super-resolution algorithms were example-based. According to different learning databases, example-based super-resolution algorithms can be divided into internal example-based super-resolution algorithms and external example-based super-resolution algorithms. In 2014, Dong et al. from the Chinese University of Hong Kong first applied deep learning to super-resolution reconstruction and proposed the super-resolution algorithm SRCNN based on convolutional neural networks [1]. Subsequently, deep learning based super-resolution algorithms became a research hotspot, and researchers continuously improved super-resolution algorithms from aspects such as network structure, loss function, and learning mechanism, leading to the development of numerous outstanding super-resolution algorithms.

A. SINGLE IMAGE SUPER-RESOLUTION

Ma et al. proposed a method of using gradient maps to constrain the structure of super-resolution images [5]. The dynamic convolutional network proposed by Xu et al. solves the super-resolution restoration problem for multiple combinations of fuzzy kernels and noisy images [6]. Yoo et al. presented a super-resolution reconstruction network based on data augmentation [7]. Hussein et al. solved the problem of mismatch between test data and training data in practical scenarios using the correction filter for super-resolution [8]. Chu et al. introduced a lightweight and accurate super-resolution method that leverages Neural Architecture Search (NAS) for efficient and rapid performance [9]. Wang et al. designed a degenerated perceptual super-resolution network to solve the blind super-resolution problem of unsupervised degenerated

representation learning [10]. The super-resolution method based on additive neural networks adopted by Song et al., which used addition operations in the output layer calculation to avoid the significant computational power consumption of traditional convolutional neural network convolution kernel multiplication [11]. Wang et al. introduced a approach rooted in sparse mask networks. This method utilized sparse masks to differentiate between crucial regions (e.g., edges and textures) and less significant areas, thereby concentrating computational resources on the important regions to curtail redundant computations [12]. Kong et al. combined classification network and super-resolution network into a unified framework, using classification network to classify sub images into different categories based on the difficulty of restoration, and then using super-resolution modules to perform different super-resolution algorithms for different categories [13]. Deng et al. introduced a Laplacian multi-level separation framework for achieving super-resolution reconstruction of panoramic images across various latitude bands [14]. Liang et al. proposed a method for constructing a kernel prior modeling approach based on normalized flows by learning the invertible mapping between anisotropic gaussian kernel distributions and controllable hidden distributions, aiming to address the challenge of blind super-resolution reconstruction [15]. Son et al. developed a super-resolution network capable of achieving arbitrary image transformations [16]. Kim et al. [17] proposed a kernel adaptive locally adjusted blind super-resolution network framework based on super-resolution features [17]. Bhat et al. designed a multi-frame super-resolution network framework by composing information from various actual image signals [18]. Hui et al. introduced an adaptive modulation super-resolution reconstruction network tailored for handling multiple degradations [19]. Jo et al. proposed a search-based super-resolution reconstruction method for applications with limited computational power such as smartphones and televisions [20].

B. GUIDED SUPER-RESOLUTION

Wang et al. proposed a super-resolution method based on spatial feature transformation, which uses prior category information to solve the problem of unreal super-resolution textures [21]. Zhang et al. designed an end-to-end depth model that adaptively transfers textures from reference images based on texture similarity to enrich the detailed information of high-resolution images [22]. Yang et al. proposed a Transformer-based reference-based super-resolution method, which utilizes attention mechanisms to discover the deep feature correspondence between the low-resolution image and the reference image, thereby transferring accurate texture features during the super-resolution reconstruction process and improving the quality of the reconstruction [23]. Zhou et al. proposed a reference-based super-resolution network that takes the underlying scene structure as a clue. This network incorporates of a plane-aware attention mechanism, a multi-scale guided upsampling module, and a super-resolution synthesis module, collectively enabling high-fidelity super-resolution reconstruction even at high magnification [24]. Jiang et al. introduced a C2-Matching algorithm with clear robust matching to solve the conversion and resolution gap between input image and reference image. To bridge the conversion gap, the comparative correspondence network leverages an enhanced representation of the input image to learn robust conversion correspondences. For the resolution gap, knowledge distillation is used to guide low resolution to high resolution matching from easier high resolution matching through the teacher-student correlation distillation method, which effectively improves the quality of reference-based super resolution reconstruction [25]. Lu et al. introduced a reference image super-resolution technique that addresses the limitations of current reference-based super-resolution methods by accounting for potential distribution disparities between the low-resolution image and the reference image. The method constructs coarse-to-fine matching schemes by matching and extracting modules, and learns the distribution difference between low-resolution images and reference images by spatial adaptive modules, and remaps the distribution of reference image features to the distribution of low-resolution image features in a spatially adaptive way [26]. Hayat et al. proposed a combined channel and spatial attention block to extract features incorporated with a specific but very strong parallax attention module for endoscopic image super-resolution [27]. Subsequently, a multi-stage network with a pioneering stereo endoscopic attention module was proposed to integrate edge-guided stereoscopic attention mechanisms into every interaction of stereoscopic features [28].

C. SEMANTIC SEGMENTATION

Semantic segmentation involves labeling each pixel in the input image with its corresponding category, assigning an accurate semantic label in the process. Since image segmentation necessitates the consideration of location information for each pixel, encoder-decoder structures are frequently employed. The encoder generates a low-resolution feature map by reducing the spatial resolution of the input image through downsampling, and the decoder recovers the low-resolution feature map into a high-resolution segmentation map through upsampling. Vijay Badrinarayanan et al. introduced a novel deep convolutional encoder-decoder architecture for image segmentation (SegNet). In this architecture, the decoder adopts the pooled index nonlinear upsampling of the encoder, which not only solves the loss of position information caused by multiple pooling, but also effectively reduces the computational load and improves the efficiency of semantic segmentation [29]. Lin et al. introduced a Multi-path Refinement Networks for High-Resolution Semantic Segmentation (RefineNet). By employing residual convolution, multi-resolution fusion, and chain residual pooling, it has effectively improve the quantitative indexes of semantic segmentation [30]. Zhao et al. proposed the Pyramid Scene

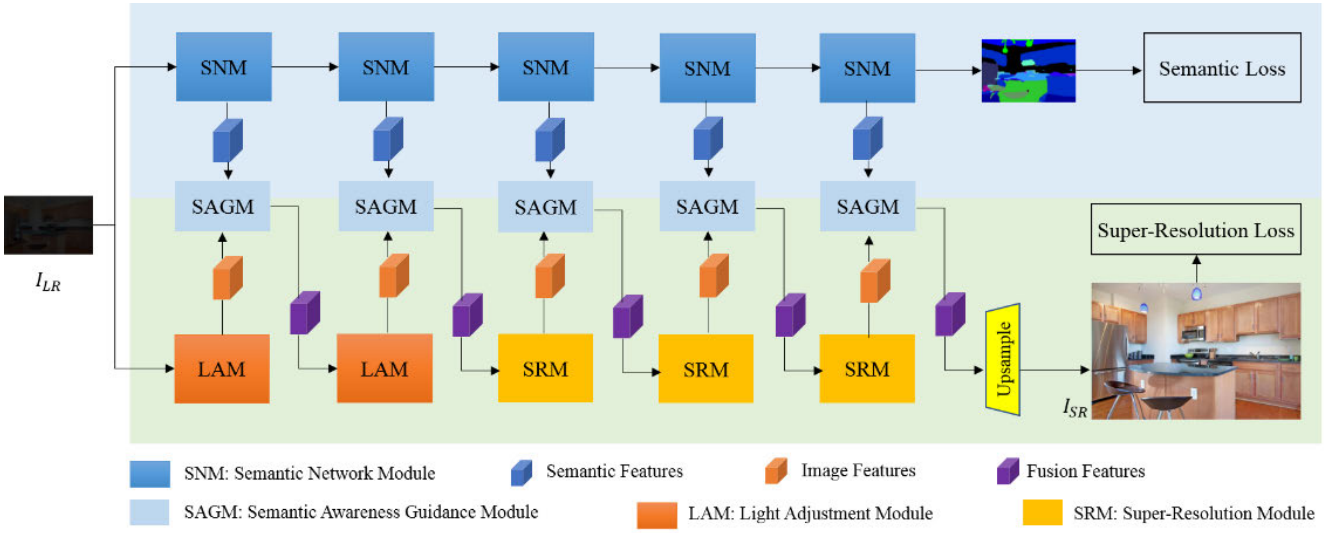

FIGURE 1. Overview of our semantic-aware guided super-resolution framework (SAGSR).

Parsing network (PSPnet), which using the spatial pyramid pooling to generate feature maps possessing varying receptive field sizes, which are then merged to harness the global scene category cues, ultimately enabling multi-level semantic feature fusion [31]. Chen et al. introduced Semantic image segmentation with deep convolutional nets and fully connected CRFs (DeepLab) in their work, and subsequently enhanced the semantic segmentation capabilities of DeepLab, culminating in the development of four distinct versions: DeepLabV1, DeepLabV2, DeepLabV3, and DeepLabV3+ [32], [33], [34], [35]. Wang et al. introduced Deep High-Resolution Representation Learning for Visual Recognition (HRNet), utilizing parallel connections of convolutional streams ranging from high to low resolutions to enhance the accuracy of semantic segmentation by facilitating continuous information exchange among various resolutions [36]. Aakerberg et al. proposed a novel framework, semantic segmentation guided real-world super-resolution, which uses an auxiliary semantic segmentation network to guide the SR learning [37]. Wu et al. designed a semantics-aware approach to better preserve the semantic fidelity of generative real-world image super-resolution [38]. Park et al. proposed a novel semantic SR method that is based on the generative adversarial network framework and self-distillation [39]. Li et al. proposed the simple and effective semantic-aware discriminator to excavate the semantics of images from a well-trained semantic extractor [40]. We make a comparison between the strength and weakness of SR methods in recent years, as shown in Table 1.

Our solution explores the challenges of the previously proposed approach by building a semantically aware guided SR framework, integrating LAM and SRM. Low-light image SR guided by semantic perception can be widely used in public security, criminal investigation and other fields, and it is of great significance to actively explore related solutions.

TABLE 1. Comparison of super-resolution methods.

Method	Year	Strength	Weakness
HRNet [34]	2020	High semantic segmentation accuracy	Lack of SR reconstruction
MASA-SR[26]	2021	Robust to handle reference images.	Requires reference images, and lacks semantic guidance
SSG-RWSR [37]	2022	Semantic Segmentation Guided	Lack of low-light processing
SeeSR [38]	2023	Reproduce more realistic image details	Texture distortion and lack of low-light processing
semantic SR [39]	2024	Learn implicit category-specific semantic priors	Reconstruct image texture distortion
SED [40]	2024	Semantic-aware discriminator	Lack of low-light processing

III. THE PROPOSED METHOD

A. PROPOSED SUPER-RESOLUTION FRAMEWORK

Single-image super-resolution refers to the reconstruction of a high-resolution image based on a low-resolution image. In existing research, it is commonly presumed that low-resolution images originate from degraded high-resolution images. Equation (1) illustrates the degradation model, with k denoting the blur kernel, s indicating the down-sampling scale factor, and n representing noise.

To facilitate the procurement of paired images for training super-resolution models, numerous super-resolution techniques usually overlook the impact of blur and noise, opting instead to utilize bicubic down-sampling of high-resolution images for the construction of training and testing datasets.

$$I_{LR} = (I_{HR} * k) \downarrow_s + n \tag{1}$$

In reality, the degradation of images is significantly more intricate than bicubic downsampling. For instance, owing to unavoidable environmental or technical constraints, the

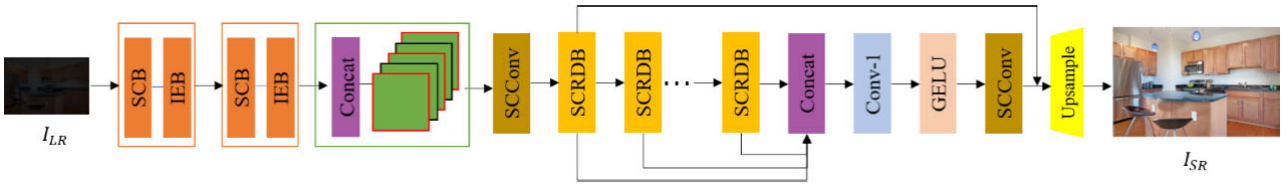

FIGURE 2. The architecture of low-light image super-resolution network (LISR).

brightness and resolution of the image may be low, leading to a compromised visual perception quality. To recover the concealed finer points during the super-resolution reconstruction process, we introduce a semantically perception-guided super-resolution method, the architecture of which is depicted in Figure 1 Semantic segmentation is to segment the object in the image on the basis of understanding the image scene through the intensive marking of pixels in the image. We firmly believe that the abundant semantic prior knowledge offered by semantic segmentation networks serves a crucial purpose in bolstering the effectiveness of low-light image super-resolution networks and elevating the authenticity of reconstructed imagery. The Semantic Network Module (SNM), which has been pre-trained on extensive datasets, can furnish abundant semantic prior knowledge encompassing intermediate features and semantic mappings. SNM uses semantic awareness guidance module to integrate semantic features and image features of target features in a quantitative attention way, and guides the features of low-light images to maintain semantic consistency. Light adjustment module including self-calibrated block (SCB) and illumination estimation block (IEB). By leveraging the SCB, it ensures convergence consistency among the output results of each IEB, enhancing the stability of brightness features. Furthermore, it facilitates rapid, adaptable, and robust illumination enhancement. The Super-Resolution Module primarily comprises a spatial and channel reconstruction distillation module that replaces redundant convolution operations with spatial and channel reconstruction convolutions. Additionally, an attention module is employed to further enhance the reconstruction capabilities of the SRM.

B. SEMANTIC AWARENESS BASED ON TRANSPOSED-ATTENTION MECHANISM

Current super-resolution methods primarily focus on learning the mapping between low-resolution image and high-resolution image pixels, neglecting the comprehension and utilization of the comprehensive semantic knowledge within the image. Our designed semantic segmentation module is capable of providing abundant prior knowledge for super-resolution reconstruction of low-light images, guiding the low-light super-resolution module to possess regional awareness, limiting the reconstruction authenticity of detailed regions, and ensuring color consistency in SR images [41]. To construct the SNM, we leverage the pre-trained High-Resolution Network (HRNet) [36] on the extensive PASCAL-Context dataset, providing semantic prior guidance for the low-light image super-resolution reconstruction module.

In SAGSR, the input image is a low-resolution, low-light image $I_{LR} = \mathcal{R}^{W\times H\times 3}$ with a height of H and a width of W, which is limited by the exposure time of the hardware equipment and the shooting environment. Guided by the semantic segmentation module, the process of super-resolution reconstruction for low-light images can be formulated as a function, as demonstrated in Equations (2) and Equations (3).

$$SP = F_{ss}(I_{LR}; \theta_{ss}) \tag{2}$$

$$\widetilde{I_{HR}} = F_{sr}(I_{LR}, SP; \theta_{sr}) \tag{3}$$

$$\widetilde{\theta_{sr}} = argmin\mathcal{L}(\widetilde{I_{HR}}, I_{HR}, SP) \tag{4}$$

In Equation (2), semantic prior knowledge is represented SP, encompassing multi-scale, multi-dimensional intermediate features along with semantic segmentation outcomes. F_{ss} represents a fully trained semantic segmentation network, and θ_{ss} represents the network parameters for semantic segmentation, which is frozen during the training process. In Equation (3), semantic prior knowledge SP and the low-resolution image I_{LR} serve as inputs to the super-resolution network F_{sr}, with $\widetilde{I_{HR}}$ representing the resulting super-resolution reconstruction. During the training phase, θ_{ss} will be guided by semantic prior knowledge SP to minimize the objective function, with the update process outlined in Equation (4). Among then, I_{HR} represents the ground truth, and $\mathcal{L}(\widetilde{I_{HR}}, I_{HR}, SP)$ represents the objective function of SAGSR.

To leverage the semantic priors from SNM for enhancing the features of the Low-Illumination Super-Resolution Network (LISRM), we design a SAGM based on the transposed-attention mechanism. This mechanism fusions intermediate features from different scales of SNM and LISRM to obtain refined feature maps. SAGM serves as a bridge between SNM and LISRM, establishing a connection between the two heterogeneous tasks.

We define f_{ss}^i as the intermediate feature of SNM, and f_{sr}^i as the intermediate feature of LISRM. The SAGM performs pixel-level interactions between f_{ss}^i 和 f_{sr}^i to obtain refined features f_{ro}^i, where $i = 1,2,3\ldots n$. The SAGM employs a transposed-attention mechanism to compute the cross-modal similarity between features f_{ss}^i and f_{sr}^i [42], resulting in a semantically aware attention map AM^i. This computation process is outlined in Equation (5).

$$AM^i = Softmax(LN(CL\left(f_{ss}^i\right)) \times LN(CL\left(f_{sr}^i\right))/\sqrt{C}) \tag{5}$$

$$f_{ro}^{i} = FN\left(LN(CL\left(f_{sr}^{i}\right))\right) \times AM^{i} + f_{sr}^{i} \tag{6}$$

In Equation (5), CL denotes convolution layers, LN represents layer normalization, and C refers to the number of channels in the feature. To begin, we employ CL to transform f_{ss}^{i} and f_{sr}^{i} into the same dimension and utilize LN to constrain gradient variations. Then, f_{ss}^{i} and f_{sr}^{i} of the same dimension are computed using matrix multiplication, divided by the number of channels C of the feature to generate a semantic-aware graph. Finally, the semantic-aware attention map AM^{i} for the i-th SAGM is obtained through the softmax function, representing the interrelationship between f_{ss}^{i} and f_{sr}^{i}. Using Equation (6), with AM^{i}as the input, refined feature f_{ro}^{i} is obtained. Among them, FN represents the feed forward network, and f_{ro}^{i} represents the final refined feature of the i-th SAGM.

C. SELF-CALIBRATED LIGHT ADJUSTMENT MODULE

The adaptive lighting adjustment module is designed to learn the illumination component and enhance the brightness of low-resolution input images. Based on the retinex theory, there exists a relationship between the low-light input image I_{LR} and the desired sharp image I_{SR}, given by $I_{LR} = I_{SR} \otimes IC$, where IC represents the illumination component. Using the stage-wise optimization strategy [43], [44], we construct a self-calibrated light adjustment module with parameters θ and ϑ. This LAM includes self-calibrated block and illumination estimation block, as show in Figure 2.

The IEB is modeled using a stage-wise optimization strategy, with its fundamental unit described by Equation (7). The illumination component and residual term at the i-th stage are denoted as IC^{i} and $\mathcal{H}_{\theta}(IC^{i})$, and $i = 1,2,3,\ldots n$. As there exist linear connections between illumination and low-light observation in most regions, the parameterized operator $\mathcal{H}_{\theta}$ in the IEB learns residual mappings from these two factors, enabling significant computational savings while enhancing the quality and stability of illumination.

$$IC^{i+1} = IC^{i} + \mathcal{H}_{\theta}(IC^{i}) \tag{7}$$

$$SC^{i} = IC^{0} + \mathcal{M}_{\vartheta}(IC^{0} \oslash IC^{i}) \tag{8}$$

$$IC^{i+1} = SC^{i} + \mathcal{H}_{\theta}(SC^{i}) \tag{9}$$

SCB is utilized to constrain the convergence state of each LAM, and its computational process can be described using Equation (8). The initial input of LAM for low-light observation, denoted as IC^{0}, undergoes a series of transformations mediated by parameterized operators, denoted as $\mathcal{M}_{\vartheta}$, at each stage of computation. We have linked the input IC^{i} and the initial low-light observation IC^{0} from each stage as inputs to $\mathcal{M}_{\vartheta}$, and utilize IC^{0} to construct residual connections for the SCB, resulting in stage-wise constrained outputs SC^{i} for the SCB. By integrating SCB and IEB, the LAM computation process is detailed in Equation (9). LAM exhibits superior robustness and adaptability to unknown and complex scenarios through the convergence constraints offered by SCB.

D. SUPER-RESOLUTION MODULE

To address the challenges of high computational costs and inefficient reconstruction in single-image super-resolution models, we employ spatial and channel reconstruction convolution, which replaces redundant convolution operations. Furthermore, we introduce the attention module CCA to enhance the capabilities of the super-resolution module. The architecture of our low-light super-resolution reconstruction network, as depicted in Figure 2, comprises three stages: feature extraction, feature fusion, and reconstruction.

$$I_{LR}^{n} = Concat_{n}(LAM\ (I_{LR})) \tag{10}$$

$$F_{i} = H_{i}\left(F_{i-1}\right), i = 1, 2 \ldots, n \tag{11}$$

$$F_{fused} = H_{fusion}\left(Concat(F_{1}, \ldots, F_{i-1})\right) \tag{12}$$

$$I_{SR} = H_{rec}\left(F_{fused} + F_{0}\right) \tag{13}$$

As shown in Equation (10), initially, the output of LAM is replicated n times and concatenated together to serve as the input for SRM. Where $Concat(\cdot)$ denotes the operation of connection along the channel dimension, and n represents the quantity of $LAM\ (I_{LR})$ to be interconnected. During the feature extraction stage, we utilize SCConv for the initial extraction of features, followed by multiple SCRDBs for deep feature extraction. This progressive optimization of extracted features is described by Equation (11). The structure of SRCDB, as illustrated in Figure 3, encompasses three stages: feature distillation, feature aggregation, and feature optimization. In the feature distillation stage, the input features are categorized into distillation features and refinement features. The distillation features undergo feature aggregation via Conv-1, while the refinement features are gradually refined through SRCN and undergo feature aggregation using SCConv in the fourth layer. Subsequently, the aggregated features are optimized after undergoing dimension reduction through Conv-1. By introducing CCA, we aim to enhance the representational power of SRM while maintaining its efficiency.

Fully utilizing features of different layers within SRM can significantly enhance reconstruction quality. As shown in Equation (12), we fuse F_{i} at different layers to obtain F_{fused} and then map it using Conv-1 and the GELU activation function. Subsequently, an SCConv is employed to refine the features. To fully leverage residual learning, we employ a long skip connection, connecting F_{0} to the output of SCConv, and employ pixel shuffle upsampling to achieve image super-resolution reconstruction. The reconstructed features encompass both F_{fused} and F_{0}, as demonstrated in Equation (13).

IV. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL SETTINGS

1) DATASETS AND METRICS

In qualitative evaluation, we tested on datasets from different scenarios, including the MIT dataset [45] and the LSRW dataset [46]. In the quantitative evaluation, we tested on the five datasets most commonly used for super resolution,

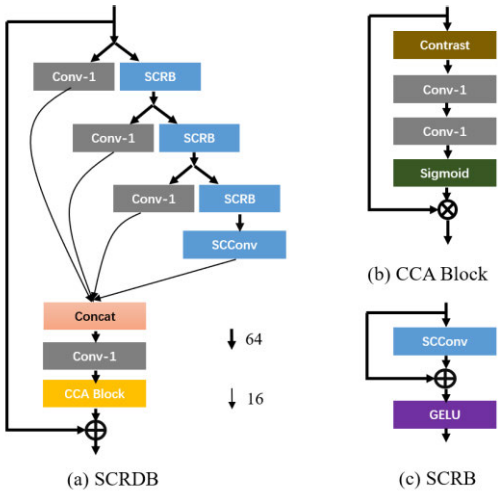

FIGURE 3. (a)The architecture of SCRDB. (b) The architecture of CCA Block, consists of a contrast layer, two Conv-1, and a Sigmoid. (c) The architecture of SCRB, consists of a SCConv and GELU.

including Set5, Set14, B100, Urban100, and Manga109. The evaluation metrics employed were peak signal to noise ratio (PSNR) and structural similarity (SSIM).

2) COMPARED METHODS

We trained SR network models, such as SRCNN, FSRCNN, VDSR, EDSR, DPBN, RDN, RCAN, NLSN, LKDN etc., for comparison in the same software and hardware environment, using the same dataset, setting the same training rounds, and using Set5, Set14, B100, Urban100, and Manga109 as test sets to compare semantic perception SR models with comparative SR models in quantitative evaluation. In qualitative analysis, we mainly selected VDSR, EDSR,, DBPN, RDN, RCAN, NLSN and LKDN as comparative methods. In quantitative analysis, we mainly selected lightweight SR methods, including SRCNN, FSRCNN, VDSR, LapSRN, DRRN, RFDN, VAPSR, LKDN-S, LKDN as well as some classic SR methods, including D-DBPN, RDN, RCAN, etc.

3) IMPLEMENTATION DETAILS

In the training process, we use random rotation and horizontal flipping for data augmentation and use the ADAM optimizer to set the momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, with the initial learning rate is set to 1×10^{-4} and halved every 1×10^2 iterations. Besides, all the experiments are implemented on a common operating platform (an NVIDIA RTX 3090ti 24G GPU), using Python 3.6.2 for encoding.

B. COMPARISON WITH STATE-OF-THE-ART METHODS

1) QUALITATIVE EVALUATION

SAGSR benefits from the rich prior semantic knowledge provided by SNM, which can ensure semantic consistency of low-light image features during the reconstruction process. We have selected three different living scene images, including outdoor rest areas, walls, and corridors. These images have low brightness and resolution due to issues with shooting time and hardware craftsmanship. From Figure 4, it can be seen that our proposed SAGSR reconstruction method performs significantly better than the comparison methods. The rest area image reconstructed by SAGSR clearly shows three sofas, two coffee tables, and promotional posters on the wall. The wall image reconstructed by SAGSR can clearly show a blue and a green triangle decoration, while other comparison methods can only vaguely see the two triangle decorations, making it difficult to distinguish colors. The corridor image reconstructed by SAGSR shows the text presented on the promotional poster, but other comparison methods cannot recognize the text on the promotional poster.

In order to clearly compare the reconstruction effects of various methods, we selected some detailed areas of LR images and displayed the 4 × reconstruction effect of details in three life scene images. From Figure 5, it can be seen that our proposed SAGSR is significantly superior to the comparison methods VDSR, EDSR, DBPN, RDN, RCAN, NLSN and LKDN. Taking corridor promotional posters as an example, the images reconstructed by SAGSR can recognize most of the text on the poster, but the comparison methods is difficult to recognize the text information on the poster. It should be noted that considering the hardware computing power of practical application scenarios, our LAM and SRM are both lightweight modules. The time for SAGSR reconstruction (4×) of a low resolution image with a resolution of 280 × 210 is approximately 0.07 seconds. In this comparison, the reconstruction time for a low-resolution image (280 × 210) using the DBPN method (4x) is approximately 34.63 seconds, while EDSR takes around 1.12 seconds, RCAN and NLSN approximately 1.25 seconds, RDN around 1.34 seconds, and VDSR approximately 1.07 seconds. SAGSR not only outperforms comparison methods in reconstructing visual effects, but also outperforms contrast methods in reconstruction efficiency.

C. QUANTITATIVE EVALUATION

In quantitative analysis, in order to fairly compare the indicators PSNR and SSIM of various super-resolution methods, we removed LAM from SAGSR and set the number of SRM to 8, and the number of channels is set to 64. LAM can improve the brightness of images, but in PSNR and SSIM, the adjustment of pixel brightness by LAM can lead to inaccurate (much lower) PSNR and SSIM values. It is worth noting that SNM provides rich semantic prior knowledge composed of intermediate features and semantic mappings during the training phase, and integrates semantic features and reconstructed features through SAGM in a quantitative attention manner. In the inference stage, we froze the SNM to ensure the lightweight of the SR model. The parameter size of lightweight SAGSR is 372K, and we can further reduce the parameter size of SAGSR to 186K by setting the number of SRM to 4. As shown in Table 2, the PSNR and SSIM of SAGSR outperform most comparison methods on five datasets including Set5. Among all the comparison methods, the parameter size of RFDN is 550K, which is closest

FIGURE 4. Visual comparison of SAGSR with the state-of-the-art methods on ×4 SR.

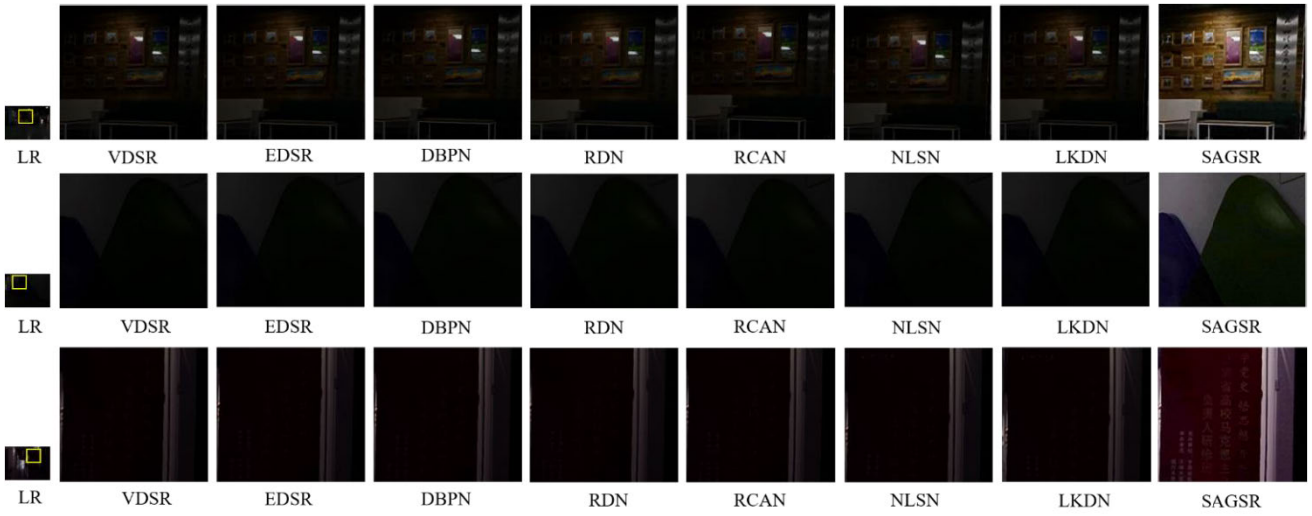

FIGURE 5. Visual comparison (details) of SAGSR with the state-of-the-art methods on ×4 SR.

to SAGSR. The PSNR value of SAGSR is 0.26dB-1.14dB higher than RFDN, and the SSIM is 0.037-0.088 higher. The PSNR and SSIM values of RCAN are closest to SAGSR, with the PSNR value of SAGSR increasing by 0.02dB-0.07dB and SSIM increasing by 0.0-0.002 compared to RCAN. However, the parameter quantity of RCAN is about 16000K, far exceeding SAGSR. Ensuring the lightweight of the SR model during the inference phase is beneficial for its practical application.

D. ABLATION STUDY

1) EFFECTIVENESS OF SNM

To verify the role of prior semantic knowledge in super-resolution reconstruction, we tested the PSNR and SSIM values with and without SNM on five datasets, referring to the quantitative analysis settings. As Table 3 demonstrates, the removal of SNM led to a decrease in both PSNR and SSIM values across all datasets.

The removal of SNM resulted in the largest decrease in PSNR value on Set5 dataset, reaching 0.52dB, and the largest decrease in SSIM value on B100 dataset, reaching 0.088. On the whole, the removal of SNM resulted in a decrease of 0.289dB in PSNR and 0.0578dB in SSIM. These findings clearly highlight the importance of prior semantic knowledge in super-resolution reconstruction.

2) EFFECTIVENESS OF LAM

To verify the role of LAM in SAGSR, we selected 6 images from different life scenes and reconstructed them 4× with and without LAM, respectively. From Figure 6, it can be seen that after removing LAM, many details of the SAGSR reconstructed image are hidden in the dark and difficult to recognize. For example, the umbrella in (a), the promotional poster on the wall in (b), the table in (c), and the cups in (e) and (f) are all hidden in the darkness, and even after SR reconstruction, it is still difficult to clearly observe their detailed information. Guided by prior semantic knowledge of SNM, SAGSR can effectively process low-light and low resolution images by integrating LAM and SRM. When SAGSR

TABLE 2. Quantitative comparison with state-of-the-art methods on benchmark datasets.

SR algorithms	Scale	Set5		Set14		B100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	4	28.42	0.810	26.10	0.702	25.96	0.667	23.15	0.657	24.92	0.789
SRCNN [1]	4	30.48	0.863	27.49	0.750	26.90	0.710	24.52	0.722	27.66	0.851
FSRCNN [2]	4	30.71	0.866	27.59	0.754	26.98	0.715	24.62	0.728	27.90	0.852
VDSR [3]	4	31.35	0.884	28.01	0.767	27.29	0.725	25.18	0.752	28.83	0.881
EDSR [47]	4	32.46	0.897	28.80	0.788	27.71	0.742	26.64	0.803	31.02	0.915
MDSR [47]	4	32.50	0.897	28.72	0.786	27.72	0.742	26.67	0.804	31.03	0.915
DRCN [4]	4	31.53	0.885	28.02	0.767	27.23	0.723	25.14	0.751	28.98	0.882
LapSRN [48]	4	31.54	0.885	28.09	0.770	27.32	0.726	25.21	0.756	29.09	0.890
DRRN [49]	4	31.68	0.889	28.21	0.772	27.38	0.728	25.44	0.764	29.46	0.896
D-DBPN [50]	4	32.47	0.898	28.82	0.786	27.72	0.740	26.38	0.795	30.91	0.914
ESPCN [51]	4	29.21	0.851	26.40	0.744	25.50	0.696	24.02	0.726	23.55	0.795
SRMDNF [52]	4	31.96	0.893	28.35	0.777	27.49	0.734	25.68	0.773	30.12	0.902
RFDN [21]	4	31.96	0.890	28.44	0.780	27.49	0.735	25.77	0.775	30.08	0.900
RDN [53]	4	32.47	0.899	28.81	0.787	27.72	0.742	26.61	0.803	31.00	0.915
SRFBN [54]	4	32.47	0.898	28.81	0.787	27.72	0.741	26.60	0.802	31.15	0.916
RCAN [55]	4	32.43	0.931	28.79	0.848	27.71	0.822	26.65	0.854	31.15	0.935
NLSN [56]	4	32.39	0.900	28.79	0.848	27.72	0.822	26.71	0.854	31.20	0.935
VAPSR [57]	4	32.38	0.898	28.77	0.785	27.68	0.740	26.35	0.794	30.89	0.913
LKDN-S [58]	4	32.10	0.894	28.62	0.782	27.59	0.737	26.07	0.785	30.50	0.908
LKDN [58]	4	32.39	0.898	28.79	0.786	27.69	0.740	26.42	0.797	30.97	0.914
SAGSR(ours)	4	32.45	0.931	28.81	0.849	27.75	0.823	26.72	0.855	31.22	0.937

TABLE 3. Ablation study of SNM.

SR algorithms	Scale	Set5		Set14		B100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SAGSR-woSNM	4	31.93	0.893	28.32	0.781	27.52	0.735	26.47	0.797	31.07	0.900
SAGSR	4	32.45	0.931	28.81	0.849	27.75	0.823	26.72	0.855	31.22	0.937

includes LAM, the image reconstructed by 4x SR can not only improve the resolution of the image, but also recognize many details hidden in the dark. For example, after including LAM reconstruction, it can recognize (b) promotional posters on the wall, and also recognize some of the text and face images on the promotional posters. In figure (e) and (f), not only the cup but also the text on the cup can be recognized.

In order to further validate the effectiveness of LAM, inspired by [59] and [60], we designed LAM-F and LAM-K, and combined them with SRM to 4x reconstruct the same 6 life scene images in Figure 6. From Figure 7, it can be seen that the reconstruction effect of combining LAM-F and LAM-K is superior to direct SR reconstruction of low-light images. However, the images reconstructed by combining LAM-F and LAM-K with SRM, the text in (b), the table in (c), and the text on (e) and (f) cups, were not as clear as the images reconstructed by combining LAM with SRM.

3) EFFECTIVENESS OF SCConv AND CCA IN SRM

To investigate the role of SCConv and CCA in SRM, we referred to the quantitative analysis settings and tested the PSNR and SSIM values of SRM with and without SCConv and CCA on five datasets. From Table 4, it can be seen that removing both SCConv and CCA leads to varying degrees of decrease in PSNR and SSIM values across the five datasets. Among them, removing SCConv resulted in the maximum decrease in PSNR on Set5 and Set14, reaching 0.74dB, and the maximum decrease in SSIM on B100, reaching 0.093dB. Removing CCA resulted in the maximum decrease in PSNR on B100, reaching 0.08dB, and the maximum decrease in SSIM on Urban100, reaching 0.055dB. Overall, removing SCConv resulted in an average decrease of 0.56dB in PSNR and 0.067 in SSIM across the five datasets. Removing CCA resulted in an average decrease of 0.08dB in PSNR and 0.031 in SSIM across 5 datasets.

The conclusion can be drawn from the above three ablation experiments. Guided by SNM semantic knowledge, the SAGSR formed by the fusion of LAM and SRM can reconstruct low-light and low resolution images into visually perceived clear images. From a quantitative analysis perspective, SNM can effectively improve the PSNR and SSIM values of SRM, and SCConv and CCA in SRM are also key

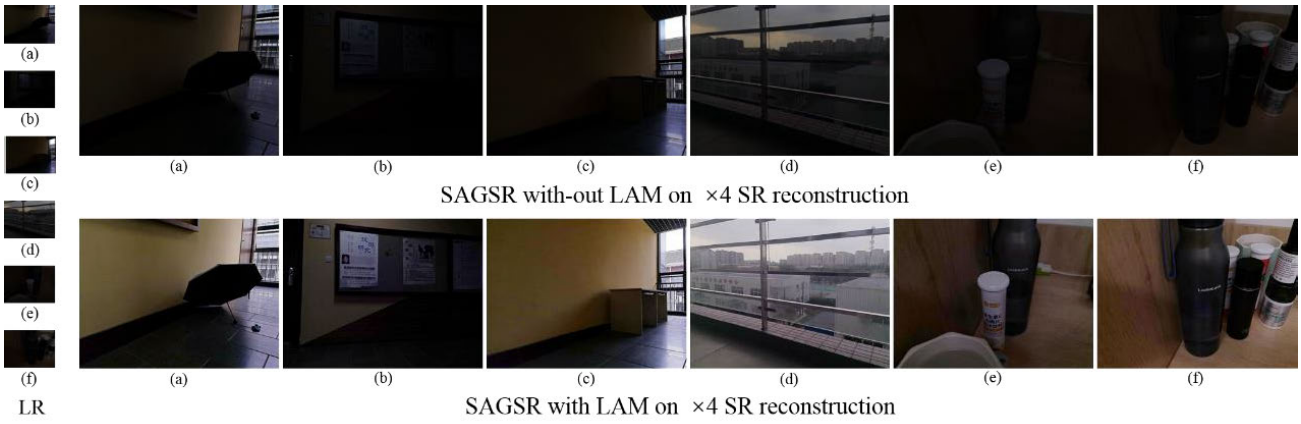

FIGURE 6. Visual comparison of SAGSR with-out and with LAM on ×4 SR.

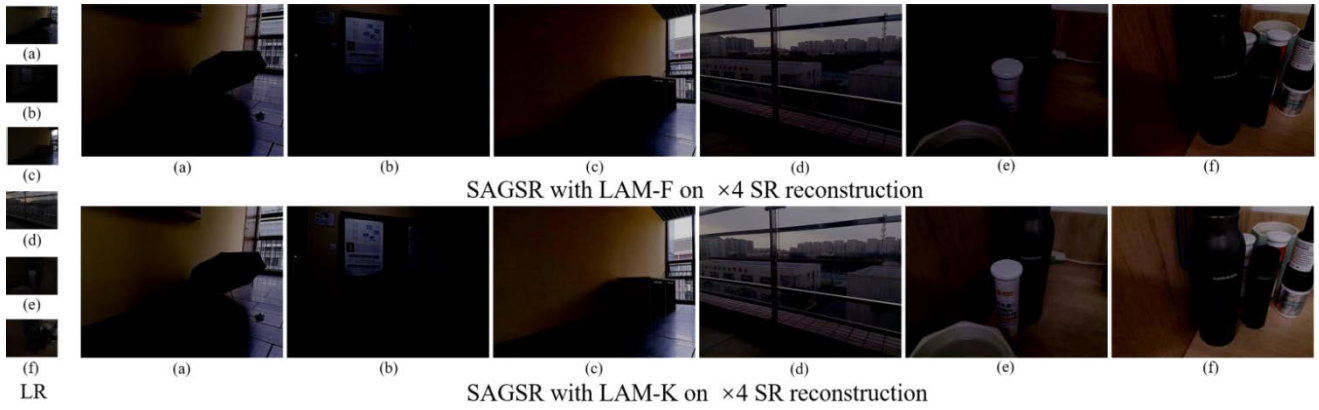

FIGURE 7. Visual comparison of SAGSR with LAM-F and LAM-K on ×4 SR.

TABLE 4. Ablation study of SCConv and CCA in SRM.

SR algorithms	Scale	Set5		Set14		B100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SAGSR-woSCConv	4	31.71	0.887	28.12	0.779	27.22	0.730	26.27	0.786	30.81	0.877
SAGSR-woCCA	4	32.40	0.917	28.76	0.807	27.67	0.803	26.55	0.800	31.15	0.913
SAGSR	4	32.45	0.931	28.81	0.849	27.75	0.823	26.72	0.855	31.22	0.937

factors affecting PSNR and SSIM values. From a qualitative analysis perspective, LAM can effectively enhance the brightness of low-light images, making SR reconstructed images more in line with human visual perception.

V. LIMITATIONS AND FUTURE WORK

Our proposed semantic-aware guided super-resolution model provides a novel solution for SR reconstruction in low-light images. However, our model currently does not support real-time super-resolution due to computational constraints and the lack of resources for real-time application [28]. Future improvements include the use of knowledge distillation, reparameterization technique, and updated optimizers to improve the quality of super-resolution reconstruction while making SR models more lightweight. Additionally, hard-ware improvements could significantly augment real-time processing capabilities [28]. Although our model is currently used for SR reconstruction of low-light images, applying it to SR reconstruction of foggy and rainy days under the framework of semantic guidance is an attractive direction for our future research. These studies are of great significance to public security, criminal investigation and other fields.

VI. CONCLUSION

To solve the problems of low image resolution and brightness caused by environmental and hardware limitations, this paper propose a semantic-aware guided low-light image super-resolution method. Utilizing the rich semantic prior knowledge of the Semantic Network Module, it guides the low-light image features to maintain semantic consistency

during the reconstruction process. SAGM integrates reference Semantic Features and Target Image Features in a quantitative attention approach, guiding low-light image features to maintain semantic consistency during the reconstruction process. LAM constrains the convergence consistency of each illumination estimation block through self-calibrated blocks, improving the stability and robustness of output brightness enhancement features. SRM uses convolution and attention modules based on spatial and channel reconstruction, which can ensure the lightweight of the model on the basis of high-quality reconstruction.

ACKNOWLEDGMENT

The authors declare that they have no conflict of interests.

REFERENCES

[1] C. Dong, C. C. Loy, and K. He, "Learning a deep convolutional network for image super-resolution," in *Proc. 13th Eur. Conf. Comput. Vis. (ECCV)*, Zurich, Switzerland, vol. 13, 2014, pp. 184–199.

[2] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, vol. 14, 2016, pp. 391–407.

[3] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[4] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.

[5] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super resolution with gradient guidance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7769–7778.

[6] Y.-S. Xu, S. R. Tseng, Y. Tseng, H.-K. Kuo, and Y.-M. Tsai, "Unified dynamic convolutional network for super-resolution with variational degradations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12496–12505.

[7] J. Yoo, N. Ahn, and K.-A. Sohn, "Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8375–8384.

[8] S. A. Hussein, T. Tirer, and R. Giryes, "Correction filter for single image super-resolution: Robustifying off-the-shelf deep super-resolvers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1428–1437.

[9] X. Chu, B. Zhang, H. Ma, R. Xu, and Q. Li, "Fast, accurate and lightweight super-resolution with neural architecture search," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 59–64.

[10] L. Wang, Y. Wang, X. Dong, Q. Xu, J. Yang, W. An, and Y. Guo, "Unsupervised degradation representation learning for blind super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10581–10590.

[11] D. Song, Y. Wang, H. Chen, C. Xu, C. Xu, and D. Tao, "AdderSR: Towards energy efficient image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15648–15657.

[12] L. Wang, X. Dong, Y. Wang, X. Ying, Z. Lin, W. An, and Y. Guo, "Exploring sparsity in image super-resolution for efficient inference," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4917–4926.

[13] X. Kong, H. Zhao, Y. Qiao, and C. Dong, "ClassSR: A general framework to accelerate super-resolution networks by data characteristic," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12016–12025.

[14] X. Deng, H. Wang, M. Xu, Y. Guo, Y. Song, and L. Yang, "LAU-net: Latitude adaptive upscaling network for omnidirectional image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9189–9198.

[15] J. Liang, K. Zhang, S. Gu, L. V. Gool, and R. Timofte, "Flow-based kernel prior with application to blind super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10601–10610.

[16] S. Son and K. M. Lee, "SRWarp: Generalized image super-resolution under arbitrary transformation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7782–7791.

[17] S. Y. Kim, H. Sim, and M. Kim, "KOALAnet: Blind super-resolution using kernel-oriented adaptive local adjustment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10611–10620.

[18] G. Bhat, M. Danelljan, L. Van Gool, and R. Timofte, "Deep burst super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9209–9218.

[19] Z. Hui, J. Li, X. Wang, and X. Gao, "Learning the non-differentiable optimization for blind super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2093–2102.

[20] Y. Jo and S. Joo Kim, "Practical single-image super-resolution using look-up table," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 691–700.

[21] X. Wang, K. Yu, C. Dong, and C. Change Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 606–615.

[22] Z. Zhang, Z. Wang, Z. Lin, and H. Qi, "Image super-resolution by neural texture transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7982–7991.

[23] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5791–5800.

[24] Y. Zhou, G. Wu, Y. Fu, K. Li, and Y. Liu, "Cross-MPI: Cross-scale stereo for image super-resolution using multiplane images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14842–14851.

[25] Y. Jiang, K. C. K. Chan, X. Wang, C. C. Loy, and Z. Liu, "Robust reference-based super-resolution via C2-matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2103–2112.

[26] L. Lu, W. Li, X. Tao, J. Lu, and J. Jia, "MASA-SR: Matching acceleration and spatial adaptation for reference-based image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6368–6377.

[27] M. Hayat, S. Armvith, and T. Achakulvisut, "Combined channel and spatial attention-based stereo endoscopic image super-resolution," in *Proc. IEEE Region Conf. (TENCON)*, Oct./Nov. 2023, pp. 920–925.

[28] M. Hayat and S. Aramvith, "E-SEVSR—Edge guided stereo endoscopic video super-resolution," *IEEE Access*, vol. 12, pp. 30893–30906, 2024.

[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[30] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1925–1934.

[31] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

[32] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*.

[33] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[34] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[35] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[36] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021.

[37] A. Aakerberg, A. S. Johansen, K. Nasrollahi, and T. B. Moeslund, "Semantic segmentation guided real-world super-resolution," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 449–458.

[38] R. Wu, T. Yang, L. Sun, Z. Zhang, S. Li, and L. Zhang, "SeeSR: Towards semantics-aware real-world image super-resolution," 2023, *arXiv:2311.16518*.

[39] H. Park, "Semantic super-resolution via self-distillation and adversarial learning," *IEEE Access*, vol. 12, pp. 2361–2370, 2024.
[40] B. Li, X. Li, H. Zhu, Y. Jin, R. Feng, Z. Zhang, and Z. Chen, "SeD: Semantic-aware discriminator for image super-resolution," 2024, *arXiv:2402.19387*.
[41] C. Li, C. Guo, L. Han, J. Jiang, M.-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.
[42] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5728–5739.
[43] 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.
[44] 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.*, Jun. 2021, pp. 10561–10570.
[45] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proc. CVPR*, Jun. 2011, pp. 97–104.
[46] J. Hai, Z. Xuan, S. Han, R. Yang, Y. Hao, F. Zou, and F. Lin, "R2RNet: Low-light image enhancement via real-low to real-normal network," 2021, *arXiv:2106.14501*.
[47] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.
[48] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.
[49] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3147–3155.
[50] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.
[51] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
[52] 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.*, Jun. 2018, pp. 3262–3271.
[53] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.
[54] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3867–3876.
[55] 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)*, 2018, pp. 286–301.
[56] Y. Mei, Y. Fan, and Y. Zhou, "Image super-resolution with non-local sparse attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3517–3526.
[57] L. Zhou, H. Cai, J. Gu, Z. Li, Y. Liu, X. Chen, Y. Qiao, and C. Dong, "Efficient image super-resolution using vast-receptive-field attention," in *Proc. Eur. Conf. Comput. Vis.* Cham: Springer Nature Switzerland, 2022, pp. 256–272.
[58] C. Xie, X. Zhang, L. Li, H. Meng, T. Zhang, T. Li, and X. Zhao, "Large kernel distillation network for efficient single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 1283–1292.
[59] K. Xu, X. Yang, B. C. Yin, and R. W. H. Lau, "Learning to restore low-light images via decomposition-and-enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 2281–2290.
[60] 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.

SHENG REN received the Ph.D. degree in computer science and engineering from Central South University, China, in 2022. He is currently an Associate Professor with Hunan University of Arts and Sciences. His research interests include big data, image and video super-resolution, and video analysis and understanding.

RUI CAO received the M.S. degree in computer science and technology from Central South University, Changsha, China, in 2020. She is currently with the Information and Network Center, Central South University. Her research interests include educational informatization and image super-resolution.

WENXUE TAN received the Ph.D. degree from the College of Computer Science, Beijing University of Technology, in 2016. He is currently a Professor and a Senior Engineer with Hunan University of Art and Science. He has published more than 38 research articles, 19 of which were indexed by EI Compendex or SCI database, and eight of which were referred by the Chinese Science Citation Database. His current research interests include agriculture information technology, artificial intelligence, and cloud information security.

YAYUAN TANG received the Ph.D. degree in computer science and technology from Central South University. She is currently an Associate Professor with the School of Information Engineering, Hunan University of Science and Engineering. Her research interests include intelligent computing, resource scheduling, and big data retrieval.