ITC 3/55
Information Technology and Control
Vol. 55 / No. 3/ 2026
pp. 1038-1051
DOI 10.5755/j01.itc.55.3.44625

SR-HDRNet: A Unified Network for Single-Image Super-Resolution and HDR Enhancement

Received 2026/03/23 | Accepted after revision 2026/06/28

HOW TO CITE: Chen, H., Mu, S. (2026) SR-HDRNet: A Unified Network for Single-Image Super-Resolution and HDR Enhancement. *Information Technology and Control*, 55(3), 1035-1048. https://doi.org/10.5755/j01.itc.55.3.44625

SR-HDRNet: A Unified Network for Single-Image Super-Resolution and HDR Enhancement

Huafeng Chen, Shaoshuo Mu*

School of Media Engineering, Communication University of Zhejiang, Key Lab of Film and TV Media Technology of Zhejiang Province, Hangzhou, China

Corresponding author: ciomphit@cuz.edu.cn

Super-resolution (SR) and high-dynamic-range (HDR) enhancement are different image restoration tasks and require specialized models to be trained for each task. This work aims to tackle the all-in-one enhancement task that handle both super-resolution and HDR enhancement with a single model. To facilitate research, we construct a new dataset, namely SR-HDRSet, containing over 5,000 training samples. Inspired by the conventional 3D lookup table (3D-LUT) theory, deep network, and guided filter, we further propose a unified network for joint super-resolution and HDR image restoration model called SR-HDRNet. First, we integrated neural networks with 3D-LUT to create an image-guided 3D-LUT learnable module that expanded the color gamut and dynamic range. Next, we introduced a novel deep multi-scale feature improvement module designed to extract multi-scale texture features and efficiently expand the receptive field. Finally, we developed a learning-guided filter (LGF) module that enhanced the edge details, leading to the final reconstructed HR-HDR image. On the 2× and 3× experiments, the mean PSNR values of the proposed model surpassed the other state-of-the-arts methods by 0.61-1.43 dB and by 0.26-1.71 dB, respectively. Meanwhile, the SSIM also achieved competitive values. The comparisons validate that our proposed framework outperforms the current state of the art.

KEYWORDS: All-in-one image restoration, High dynamic range, Super-resolution, 3D-LUT, Multi-scale enhancement, Guided filter.

1. Introduction

Image restoration aims to reconstruct high-quality images from degraded ones. It has been widely applied in unmanned aerial vehicle (UAV) remote sensing, multimedia, and other fields, where displays are increasingly equipped with 4K ultra-high-definition (UHD) and high dynamic range (HDR) capabilities to deliver superior visual experiences. However, the high cost of producing UHD content and lengthy production cycles remain significant barriers. Consequently, high-definition (HD) content with lower resolution (e.g., 2K) and standard dynamic range (SDR) is still widely used. Therefore, converting low-resolution standard dynamic range (LR-SDR) images into high-resolution high dynamic range (HR-HDR) images with intelligent computational methods has become essential at the content production end.

In this paper, we aim to tackle the joint super-resolution and HDR restoration task, which can reconstruct one HR-HDR image directly from a single LR-SDR image. However, such joint SR-HDR is a complex problem. To achieve the goal, we need solve two problems simultaneously: super-resolution converting from LR to HR and HDR enhancement converting from SDR to HDR. Super resolution focuses on solving the problem of spatial detail loss, paying attention to the clarity, edge integrity, and texture authenticity of images [1, 27]. HDR algorithms aim to restore richer brightness levels and color ranges in the original scene, compensating for the lost visual information in SDR images due to dynamic range limitations [21, 11]. Therefore, they are different image restoration tasks and require specialized models to be trained for each task. Many researchers have attempted to cascade SR and HDR algorithms to generate HR-HDR images from LR-SDR ones [9, 19]. However, it is inefficient, as this step-by-step approach presents shortcomings, including error accumulation, high system complexity, and inconsistent training objectives among modules, making achieving optimal performance complex [34]. For the joint SR and HDR task, it is essential to simultaneously restore fine-detail and contrast while increasing the spatial resolution, dynamic and color gamut range when predicting the HR-HDR image from the LR-SDR input. To solve these drawbacks, we address the all-in-one image restoration task to handle multiple tasks of super-resolution and HDR enhancement with a single model. In our proposed architecture, called SR-HDRNet, the LR-SDR images can be directly reconstructed into HR-HDR format on the device. This not only brings benefits to the terminal and content production ends, but also reduces the required bandwidth.

Drawing inspiration from the conventional HDR algorithm [16] and SR theory [29], we introduce a novel all-in-one framework for joint super-resolution and HDR image restoration. A traditional 3D lookup table (3D-LUT) is an effective structure that is a color correction technique primarily employed in luminance-predicting-ITM [20]. It facilitates efficient mapping by sparsely sampling input ranges and recording the corresponding output values within the 3D lattice [43]. The framework simulates the optimal color transformation at the low level but also adapts widely varying image content [36]. This study integrates 3D-LUT, enhancement network, and guide filter to construct a unified SR-HDRNet model.

Our contributions are as follows:

1. We introduce a unified network for joint super-resolution and HDR image restoration, called SR-HDRNet that offers low complexity and enables end-to-end synchronized enhancement of resolution and dynamic range.
2. Learning from color correction technique, we first introduce the 3D-LUT into the SR network while expanding the brightness range and color spectrum to compensate for the lost visual information in SDR images due to dynamic range limitations.
3. We also develop a novel deep multi-scale feature improvement module, which extracts multi-scale texture features and efficiently enhance spatial resolution with fine-detail. And the model introduces learnable guide filter to further enhance the quality of the output images.

2. Related Work

This section briefly reviews the SR and HDR imaging techniques.

2.1. Super-resolution Method

The field of SR began with Dong's [10] pioneering work, which introduced a convolutional neural network (CNN)-based model that demonstrated impressive performance. Following this, rapid development in deep learning and graphics processing unit technology has spurred extensive research in deep-learning SR, which can be categorized into two types of research: image SR and video SR. Many scholars have proposed various innovative deep CNN network structures for image SR. Some notable examples include deep networks with residual learning [17, 18], residual Laplacian pyramid network structures [44], recursive learning and densely connected network structures [32, 33], deep back-projection structures [12], and residual dense network structures [45]. More recent approaches also incorporate transformer structures [25], each contributing to significant advancements in visual quality. For instance, Wang et al. [37] introduced the spatial feature transformation network structure and enhanced the super-resolution generative adversarial network (SRGAN) structure, resulting in a more realistic and natural visual quality. Hu et al. [14] proposed a new method called Meta-SR, which is the first to achieve super-resolution with arbitrary scale factors by a single model, including non-integer scale factors. Chen et al. [7] proposed a novel hybrid attention transformer model, demonstrating robust performance in activating more input pixels for enhanced reconstruction. Sun et al. [31] proposed a hybrid CNN-Transformer image super-resolution network, introducing the spatial pyramid swin transformer and gated convolution modules to capture long-range dependencies and multi-scale features efficiently; their method outperforms SwinIR with higher PSNR, fewer parameters, and lower computational costs.

Unlike image SR models, video SR focuses on leveraging spatio-temporal information from consecutive frames to generate coherent videos. This process comprises several essential steps, including motion estimation, motion compensation, information fusion, and up-sampling [8, 42]. Li et al. [24] presented an innovative loss function and developed a video SR model based on generative adversarial networks, effectively removing temporal artifacts without compromising visual quality. Liang et al. [26] proposed a novel video super-resolution reconstruction algorithm based on deep learning and spatio-temporal feature similarity to address noise, over-smoothing, and visual artifacts. Additionally, Zhu et al. [46] pioneered using spiking neural networks (SNNs) for video SR reconstruction based on event cameras. SNNs are well-suited for large-scale regression tasks, as their spiking neurons can extract temporal information, presenting promising opportunities for low-power applications.

2.2. HDR Method

HDR significantly enhances the display capabilities of SDR on TV by offering an expanded color range, higher brightness upper limit, and lower brightness lower limit. This model enhances image quality in terms of contrast and resolution, providing a more immersive experience for viewers. The process of converting SDR to HDR involves transforming dynamic range. Recent advancements in deep learning-based ITM technology have become a research hot-spot [15, 28], proposing various network structures, such as deep HDR imaging models based on CNN [13, 22], U-net models [6], deep reciprocating HDR imaging network [40] and HDR model based on large foreground motions [39]. Inverse tone mapping presents an inverse ill-posed problem, typically categorized into multi-exposure and single-image HDR reconstruction. Multi-exposure HDR reconstructs high dynamic range images by merging a series of SDR images generated by an optical camera under different exposures [30, 3]. This approach requires alignment before fusion. Steven et al. [35] developed an innovative alignment-free network utilizing dynamic and static context, integrating spatial and channel attention modules to produce remarkable results. However, these methods rely on multiple exposure images of static scenes and require additional registration for dynamic scenes to avoid blur or ghosts, increasing the model's complexity. Consequently, single-image HDR reconstruction is still the focus of current research [5], aiming to bypass the registration and directly use a single SDR image to reconstruct the HDR image. Zou et al. [47] proposed an HDR model from a single raw image that leverages the unique attributes of raw data to facilitate effective mapping from raw data to HDR, successfully restoring HDR imaging details in the case of underexposure.

The contemporary approach to converting LR-SDR to HR-HDR in a single image generally follows a step-by-step process, enhancing resolution through SR techniques and dynamic range enhancement via HDR methods. Kim et al. [19] was the first to introduce a cascaded network capable of jointly training and designing a multi-purpose convolution network that performed SR, inverse tone mapping, and joint SR inverse tone mapping. In contrast, some scholars have researched SR-HDR enhancement of multi-exposure images of the same scene. Deng et al. [9] developed a deep coupled feedback network that effectively achieved high-dynamic-range enhancement and SR, demonstrating excellent SR and fusion accuracy. Tan et al. [34] created a deep neural network combining SR and HDR techniques, successfully reconstructing HR-HDR images from multiple LR-SDR images with varying exposures. Unlike the above mentioned approaches, this study emphasizes single-frame images and proposes a novel joint SR-HDRNet model. Our method reconstructs the HR-HDR image directly from a single LR-SDR image and demonstrates strong generalization capabilities across diverse images.

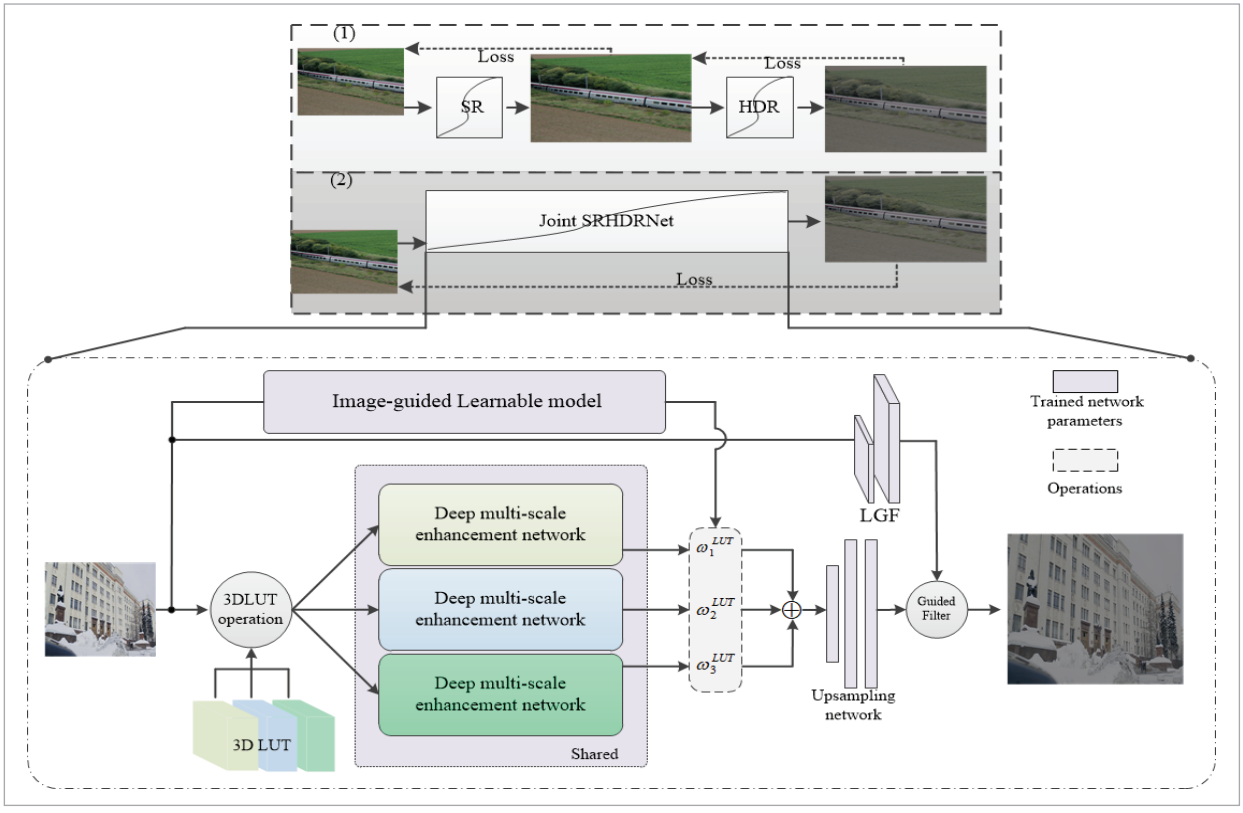
3. Approach

3.1. Overall Framework

The challenge of joint learning SR-HDR arises from the need to simultaneously enhance dynamic range and resolution, which are inherently different tasks. The traditional method is conducted by cascading SR and HDR algorithms, allowing for synchronized enhancement of resolution and dynamic range. Figure 1(1) illustrates that SR and HDR models are constructed for SR, with the HDR model realizing dynamic range enhancement. Despite its advantages, solving a complex multi-task restoration problem through a step-by-step cascading approach has notable drawbacks, including error accumulation and

Figure 1

Framework of proposed SR-HDRNet. LGF: Learned guided filter; Shared: Network parameter sharing; Operations: Mathematical calculation without learned weight parameters.



high system complexity, making achieving optimal performance difficult. In contrast to these methods, this study introduces an all-in-one image restoration model, called SR-HDRNet, as shown in Figure 1(2). This model integrates image-guided learnable 3D-LUT, enhancement network, and a guide filter. By incorporating these components, SR-HDRNet enables a seamless end-to-end process for synchronized enhancement of both resolution and dynamic range.

As shown in Figure 1, the main strategies included in our model can be illustrated from aspects of image-guided 3D-LUT learnable network, deep multi-scale enhancement network(DMEN) and learned guided filter module. The first network designs three trained 3D-LUTs to enhance and expand color gamut and dynamic range, which generates three feature images from the input LR-SDR image through 3D-LUT operations. Passing through the main design of DMEN, these three feature images undergo fusion and achieve rough HR-HDR output by applying a linear operation with parameters. This module can extract multi-scale texture features and reconstruct the global information and multi-scale local feature in input LR-SDR image by expanding the receptive field. To further fine-tune and enhance the rough output's edge detail, learned guided filter module is finally applied, from which the true HR-HDR image can be accomplished and shown as the final output.

3.2. Image-guided 3D-LUT Learnable Network

The commonly-used enhancement network cannot simultaneously improve both resolution and dynamic range, which would result in color distortion. Though many works have attempted to cascade SR and HDR algorithms to generate HR-HDR images from LR-SDR ones. However, these methods suffer from heavy computational burdens due to complicated optimization processes or neural architecture designs [34]. In fact, most of the commonly-used enhancement operations are pixel-independent. They are equivalent to three-dimensional color transformation functions, which map an input color point to another location in the color space. To overcome the computational complexity and reconstruct color gamut, the 3D-LUT represents a viable data structure for achieving efficient pixel mapping, which cannot be achieved by the current enhancement models. In this work, we are the first to introduce the learnable image-guided 3D-LUT module into the super-resolution enhancement network to enhance and expand color gamut and dynamic range. A 3D-LUT defines a 3D lattice that consists of P^3 elements, where P is the number of bins in each color channel. Every element of 3D-LUT is defined by pixel-to-pixel transformation $H(i,j,k)$ between the input RGB color and output RGB color. The 3D LUT operation is defined mathematically in (1), where O signifies the output for 3D LUT with r, g, and b channels. The learnable parameters of our method are encapsulated within the 3D LUT, consisting of $3P^3$ parameters.

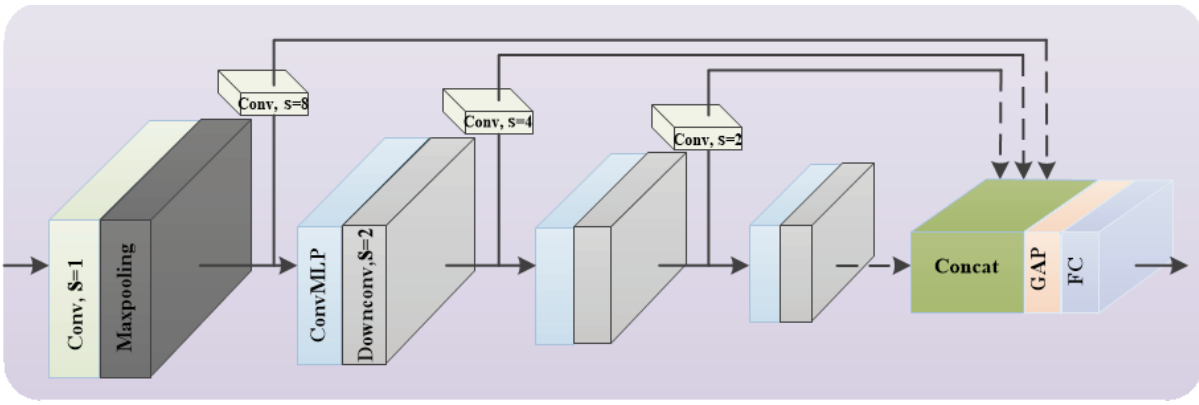
$$O^r_{i,j,k}, O^g_{i,j,k}, O^b_{i,j,k} = H(I^r_{i,j,k}, I^g_{i,j,k}, I^b_{i,j,k}). \quad (1)$$

where, I defines an indexing RGB color. $i,j,k = 0,1,...,M - 1$, and M is set to 33 which is the number of bins in each color channel. When the 3D LUT lattices are learned, an input pixel looks up its nearest sampling points according to its color and computes its transformed output via interpolation [36].

Image-guided learnable network. The weight parameters (ω_1^{LUT}, ω_2^{LUT}, ω_3^{LUT}) of 3D-LUT are crucial. The image quality assessment network is an effective method with a significant capacity for image comprehension [41]. Its hidden layers are designed to automatically learn the image's abstract information layer by layer using numerous data, yielding detailed features with rich semantic information. The multi-layer perceptron (MLP) architecture has demonstrated state-of-the-art performance and has been widely applied to visual tasks [4]. Li et al. [23] combined MLP and convolution to suggest a ConvMLP for vision, achieving advanced performance with fewer parameters and computations. A new image-guided learnable model is introduced using the ConvMLP and skip connection, following the image-quality assessment network principle, as shown in Figure 2. The network consists of one conv+max-pooling, three convMLP+downconv, global average pooling, and fully connected layers. The convMLP consists of MLP, and Depthwise conv introduces locality features and performs local information interaction by adding a convolutional layer. The skip connection propagates the upper feature information to the lower feature, obtaining higher abstraction and stronger generalization ability. This architectural design ultimately enables the model to achieve a higher level of abstraction and better generalization.

Figure 2

Image-guided learnable network. (S: stride; GAP: global average pooling; FC: fully connected layer; ConvMLP: Convolutional MLP; Concat: dimension concatenation)



3.3. Deep multi-scale Enhancement Network

During reconstruction, the enhancement network must not only capture strong global information but also retain the locality of multi-scale features, thereby significantly restoring the rich diversity of the original HR-HDR images. However, the commonly-used U-Net [2] architecture, which employs a pyramid structure with skip connections between the upsampling and downsampling layers, may still compromise multi-scale representation capabilities. Inspired by the U-net and multi-scale convolution kernel, a new deep multi-scale enhancement network (DMEN) is proposed, which is shared and designed to optimize three channels of the overall model in Figure 1, minimizing the overhead by producing the three different outputs. This operation can reconstruct the global information and multi-scale local feature in input LR-SDR image by expanding the receptive field.

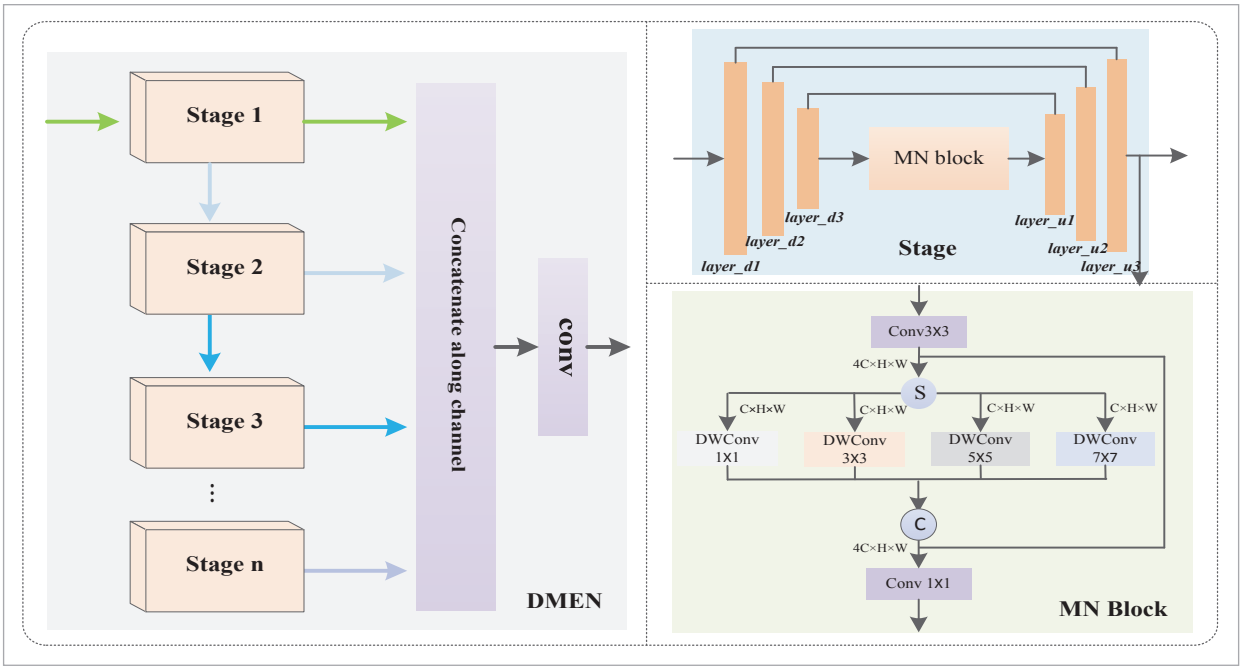
$$
\begin{aligned}
&layer_d1 = conv(x)\\
&layer_d2 = downsample(conv(layer_d1))\\
&layer_d3 = downsample(conv(layer_d2))\\
&layer_u1 = conv(MN(layer_d3)) + layer_d3\\
&layer_u2 = upsample(conv(layer_u1)) + layer_d2\\
&layer_u3 = upsample(conv(layer_u2)) + layer_d1
\end{aligned}
\tag{2}
$$

Figure 3 shows the expanded view of a DMEN consisting of n-stage modules, one concatenation along a channel, and one convolution (Conv). The stage modules share the same convolution structure, consisting of a three-layer upsampling convolution structure (*Layer_u*1, *Layer_u*2, *Layer_u*3), a three-layer downsampling convolution structure (*Layer_d*1, *Layer_d*2, *Layer_d*3), and one multi-scale convolution block. *Layer_u* and *Layer_d* are given by Formula (2), where MN (·) denotes a multi-scale convolution block of three layers. The first layer used 3×3 conv to extract global features. The second layer split four parts along the channel, and each part was operated by DWConv with different scale kernels (1×1, 3×3, 5×5, 7×7), extracting multi-scale texture features of different receptive fields. The last layer employs 1×1 convolution to fuse multi-scale features, streamlining the model by reducing the number of parameters and network complexity.

3.4. Learned Guided Filter Module

Considering the efficiency and performance, the edge enhancement is used in our method to improve the reconstructed image's edge detail. Many recent approaches to high-level visual tasks cascade a guided filter behind their original architecture to improve the results. The guided filter is an effective type of edge protection filter. It is able to suppress the gradient inversion artifacts, preserving the edge smoothing properties with higher levels of detail. A learnable guided filter module is created, as illustrated in Figure 1. The model comprises a learnable guided filter (LGF), an upsampling network, and guided filter operation. The LGF features a three-layer convolution network to learn the guided image I_i. Moreover, the

Figure 3
Deep multi-scale enhancement network (© denotes concatenation along the channel dimension, Ⓢ denotes split along the channel).



upsampling network improves the resolution by deconvolution to obtain the input p_i of the guided filter. Moreover, I_i and q_i will still faithfully guide the final output with guided filter operation. The guided filter operation, driven by a local linear model, is as follows:

$$q_i = a_k I_i + b_k, \forall i \in \omega_k, \tag{3}$$

where I_i and q_i are the guided image and filtered output, respectively, i is the pixel index, ω_k is the sliding window, and k is the index of the sliding window. The a_k and b_k compute the coefficients from the guided image and the input image. The operation (3) is based on the calculation process in literature [38].

4. Loss

The following three loss terms are used to train the proposed model:

$$L_{MSE} = \sum \left\| \hat{y}_i - y_i \right\|^2 \tag{4}$$

$$L_{TM} = \sum \left\| T(\hat{y}_i) - T(y_i) \right\|^2 \tag{5}$$

$$L_S = \sum_n \left\| \omega_n \right\|^2. \tag{6}$$

L_{MSE} is MSE loss, ensuring content consistency of generated image. L_{TM} indicates HDR loss, often used for the HDR imaging field, where $T(y) = \frac{\log(1+\mu y)}{\log(1+\mu)}$ (μ=5000). Moreover, y_i indicates the frame i of the ground-truth; $\hat{y}_i$ represents the SR-HDR image generated by the proposed model. L_s is introduced to ensure LUTs' smoothness and reduce artifacts, and ω_n is the image-guided 3D-LUT learnable network parameters.

Finally, the new joint loss for training is established as follows:

$$L = L_{MSE} + \lambda_D L_{TM} + \lambda L_s. \tag{7}$$

The proposed network is end-to-end trainable. Therefore, the loss L_{MSE}, L_{TM}, and L_S terms are simulta-

neously optimized when training our framework. We do not introduce any other constraint or loss function to the learning of 3D-LUT learnable network, willing that it can be image-adaptive for SR-HDRNet. λ and λ_D are the coefficients that balance the different loss terms. Following [35], We set λ_D=0.0001 and λ=0.0001 based on the experiment and research[36].

5. Experiment

5.1. SR-HDRSet Datasets

In this part, we first collected and built a super-resolution and HDR image dataset, namely SR-HDRSet, with more than five thousand samples. For the SR-HDRSet, the images are from 700 4K-HDR (BT.2020 with PQ-OETF) videos and the corresponding 4K-SDR (BT.709) videos, ensuring data generality and diversity. All the videos are publicly available to users provided by Tencent AI Lab. Among them, the 4K-HDR videos were captured by professional cameras, while the corresponding 4K-SDR videos were obtained by professional engineers using the Reinhard tone mapping method to convert the HDR videos into SDR videos. We randomly built the HDR datasets from 4K-HDR videos and randomly sampled eight frames of every 4K-HDR video. In total, 4K-HDR datasets contain 5600 images. Moreover, we used the same method to build the SDR datasets from 4K-SDR videos containing 5600 4K-SDR images corresponding to HDR datasets. Among these, 100 4K-HDR and 4K-SDR image pairs were extracted as the validation set, and the rest was used as the training set. For the purpose of saving storage resources, we further processed these 4K frames with a resolution of 3840×2160 into 1080×1080 using a random crop. Through the above process, the SR-HDRSet was constructed, which can be used as public datasets of joint SR-HDR task and is available on website we provide. The proposed and comparative algorithms were trained using the same datasets of SR-HDRSet.

During our training, 4K-HDR images are used as the HR-HDR label. And the corresponding 4K-SDR images are down-scaled by 2× or 3×, forming the LR-SDR images as the input. The proposed and comparative algorithms were trained using the same datasets. Meanwhile, for the test datasets in the comparative experiment, we randomly selected 20 image pairs with a resolution of 3840×2160 from 10 videos, which was extracted additionally by the Tencent AI Lab and different from training set videos. The SDR frames are further bilinearly downsampled with 2× and 3×, which are considered to compose the LR-SDR inputs. Our dataset is stored in the RGB format instead of YUV for convenience.

5.2. Experimental Setup

All experiments were conducted in the RGB channel. During the training phase, we randomly cropped a 480×480 patch from the HR-HDR image as ground truth. The corresponding patch in the LR-SDR image was identically cropped and used as the input. Data augmentation was performed through random rotation and reflection to improve the network's generalization ability. The mini-batch size was set to 8. Our framework was trained on an NVIDIA RTX 3090 GPU. The model was optimized using the Adam optimizer with $\beta 1 = 0.9$ and $\beta 2 = 0.999$. Training proceeded for a total of 800 epochs. The initial learning rate was set to 1×10^{-4} and was subsequently downscaled by a factor of 0.5 every 100 epochs until 600 epoch. The model achieving the best performance on the validation set was saved for final testing.

5.3. Results

Qualitative and quantitative evaluations were conducted to analyze the advantages of this algorithm. To address joint SR and HDR enhancement tasks, we computed the mean quantitative metrics PSNR and SSIM, which were used to evaluate the quality of SR-HDR images.

5.3.1. Ablation Study

Against our final model, we trained the joint SR-HDR model for the ablation study to examine the impact of stages' number in a DMEN. We designed three comparative experiments: Ours_x1, Ours_x3, and Ours_x5. In this context, "Ours_x1" signifies that the DMEN module utilizes one stage; "Ours_x3" employs three stages, and "Ours_x5" incorporates five stages. We focused on a subset of the ablation study's validation set, comparing the running time, the validation PSNR, and the SSIM performance for the joint SR-HDR model with different DMEN modules. Figure 4 illustrates the SR-HDRNet results of three comparative trials. Meanwhile, qualitative evaluation was performed, as shown in Table 1.

Figure 4
Ablation comparative experiments. (a) 1920×1080SDR_Input. (b) 3840×2160HDR_GT. (c) Image details of SR-HDRNet with Ours_x1. (d) Image details of SR-HDRNet with Ours_x3. (e) Image details of SR-HDRNet with Ours_x5. (f) Image details of Ground truth.

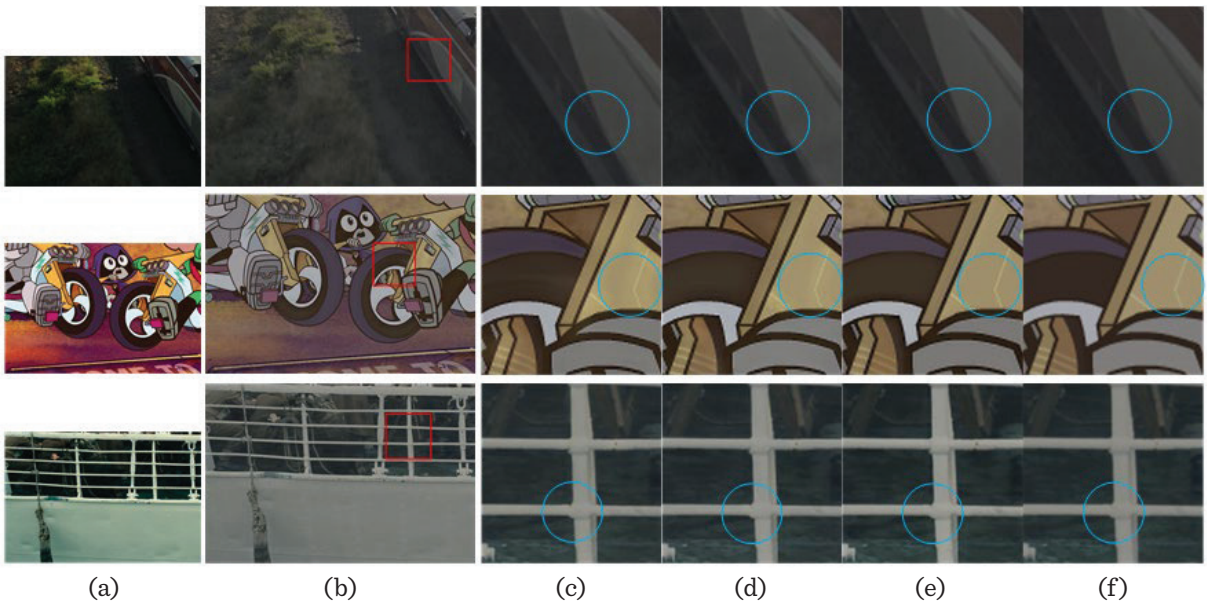


Table 1
Quantitative comparison (PSNR/SSIM) for ablation study.

Comparisons	PSNR	SSIM	Running time (ms)
Ours_x1	34.42	0.9740	2.873
Ours_x3	35.08	0.9774	4.707
Ours_x5	35.58	0.9797	6.751

Table 2
Quantitative evaluation (PSNR/SSIM).

Methods	×2		×3	
	PSNR	SSIM	PSNR	SSIM
SwinIR+HDR	34.73	0.9709	32.65	0.9266
SPNet+HDR	34.84	0.9734	33.18	0.9373
HAT+HDR	34.69	0.9697	32.53	0.9253
SPNet+SHDR	35.50	0.9706	33.98	0.9461
SwinIR+SHDR	35.51	0.9717	33.77	0.9420
HAT+SHDR	35.48	0.9712	33.74	0.9423
SR-HDRNet(ours)	**36.12**	**0.9723**	**34.24**	**0.9494**

As illustrated in Figure 4 and Table 1, increasing the number of stages substantially enhances visual effects and higher quantitative comparison values (PSNR and SSIM). For example, the PSNR and SSIM values of Ours_x5 improved by 0.5 and 0.0025, respectively, compared to Ours_x3 and by 1.16 and 0.0057 compared to Ours_x1. Visually, with more stages, the SR-HDR results increasingly resemble the ground truth, as demonstrated in Figures 4 (c)–(f). However, adding more stages deepens the network structure, increases the model's time complexity, and demands additional device memory and GPU computing power, which will limit the processing of large images. Consequently, we used the SR-HDRNet model with the DMEN module of Ours_x3 to achieve the best trade-off between SR-HDR performance and training time.

5.3.2. Comparisons with State-of-the-Art Methods

Visual comparison. This section compares our method with the latest state-of-the-art methods for SR and HDR tasks on the testing datasets after training with baseline SR-HDRSet. We selected 20 image pairs from the testing datasets for comparison. Fig-

ures 5-8 show the visual comparison of several restored images of testing datasets on 2× and 3×. The classical SR algorithms and HDR algorithms are cascaded as the state-of-the-art methods: SPNet_SHDR method [31, 22], HAT_SHDR method [7, 22], SwinIR_SHDR method [25, 22], SPNet_HDR method [31, 6], HAT_HDR method [7, 6], and SwinIR_HDR method [25, 6]. For all these, we used the source code provided by the authors with all the comparative experiments using the same training set. The reconstructed images were magnified locally to compare the details of different methods. The qualitative assessments reveal that the output of our proposed SR-HDRNet algorithm (Figures 5–8(g)) closely resembles the ground truth (Figures 5–8(h)) in terms of image detail and color. Notably, the edge lines in our results appear sharper and more refined than in other results. As highlighted by the green circles in Figure 5, the image edges of Figure 5(g) are more delicate and align more closely with the ground truth in Figure 5(h). However, the edges of Figure 5(a–f) were not fragile enough. The detailed comparison of Figures 6-8 further corroborates the advantages of our approach.

Figure 5

2×SR-HDR—visual comparison 1. (a) SwinIR_HDR. (b) SPNet_HDR. (c) HAT_HDR. (d) SwinIR_SHDR. (e) HAT_SHDR. (f) SPNet_SHDR. (g) Proposed method. (h) Ground truth.

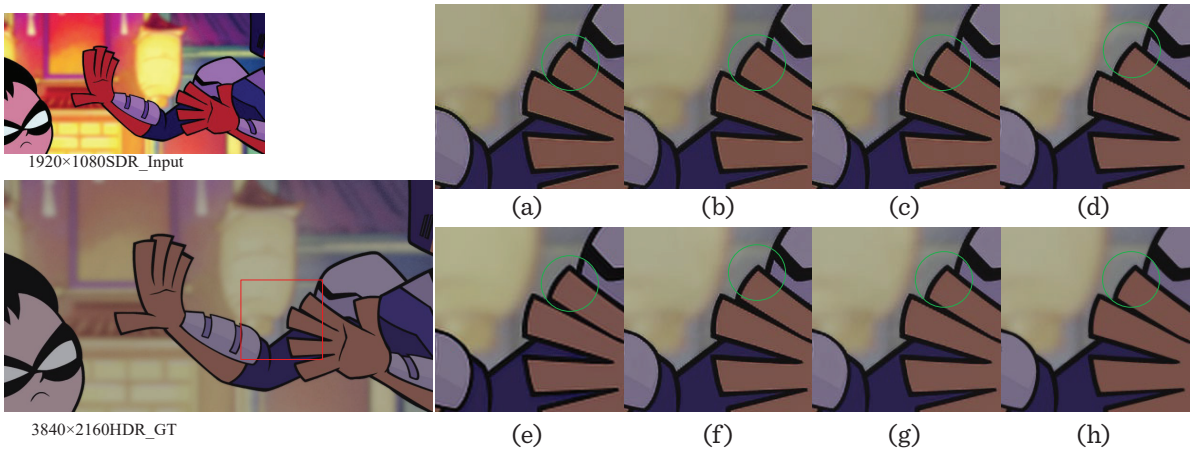


Figure 6

2×SR-HDR—visual comparison 2. (a) SwinIR_HDR. (b) SPNet_HDR. (c) HAT_HDR. (d) SwinIR_SHDR. (e) HAT_SHDR. (f) SPNet_SHDR. (g) Proposed method. (h) Ground truth.

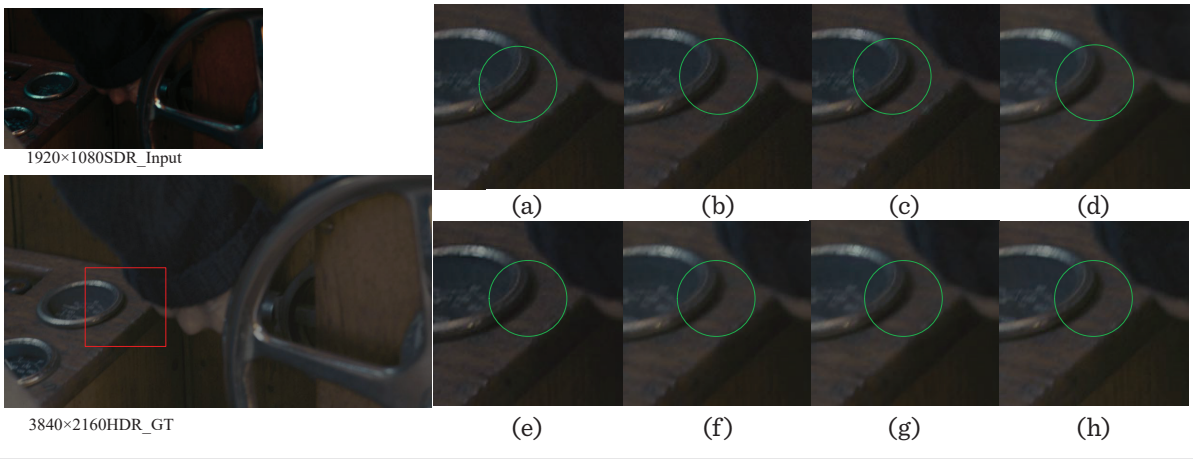


Figure 7

3×SR-HDR—visual comparison 1. (a) SwinIR_HDR. (b) SPNet_HDR. (c) HAT_HDR. (d) SwinIR_SHDR. (e) HAT_SHDR. (f) SPNet_SHDR. (g) Proposed method. (h) Ground truth.

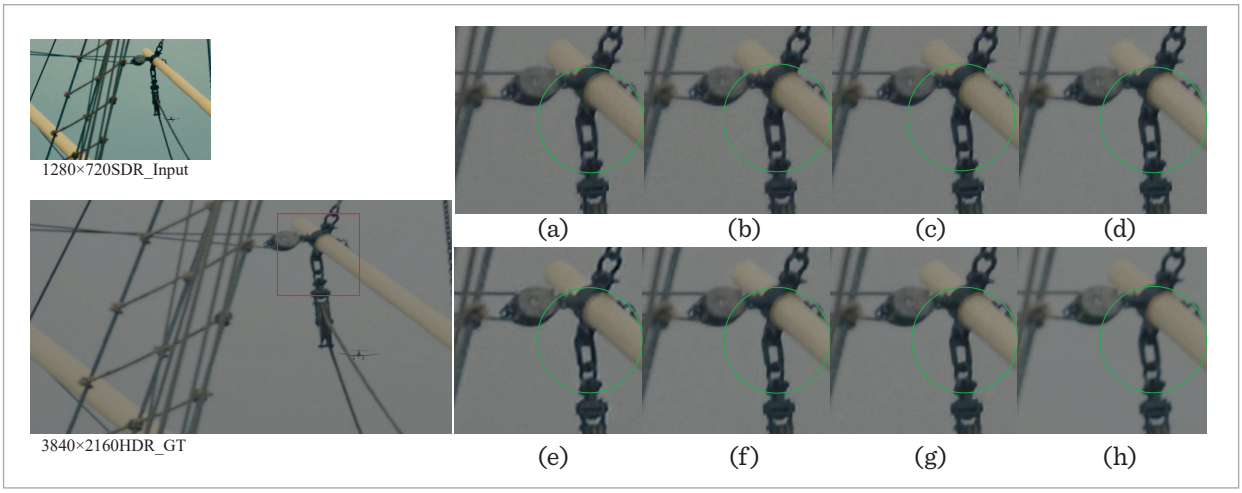
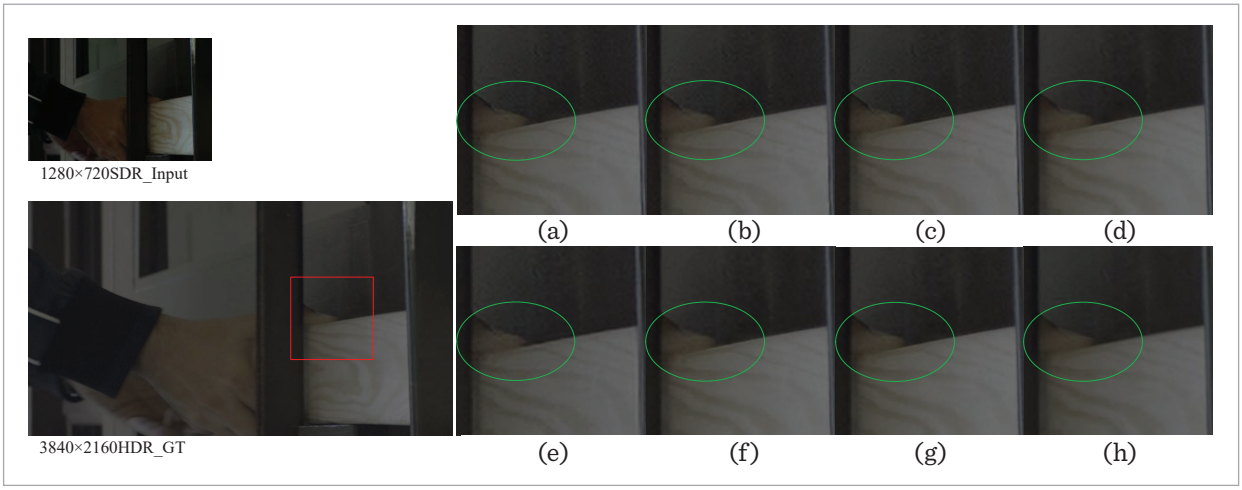


Figure 8

3×SR-HDR—visual comparison 2. (a) SwinIR_HDR. (b) SPNet_HDR. (c) HAT_HDR. (d) SwinIR_SHDR. (e) HAT_SHDR. (f) SPNet_SHDR. (g) Proposed method. (h) Ground truth.



Quantitative results. Table 2 presents the quantitative comparison of our method against other state-of-the-art techniques. Our approach demonstrates competitive results under all scaling factors. The mean PSNR values of the proposed SR-HDRNet model surpass those of other methods by 0.61–1.43 dB on 2× and by 0.26–1.71 dB on 3×. We also observed superior mean SSIM values, reinforcing our method's effectiveness. Thus, the visual and quantitative evaluations confirm that our proposed algorithm achieves exceptional results.

6. Conclusion

In this paper, we presented SR-HDRNet, an end-to-end framework for joint super-resolution and HDR enhancement. First, it introduced the three learned 3D-LUT lattices into a unified network to enhance spatial resolution and dynamic range. Then, the novel (deep multi-scale enhancement network) DMEN module advanced the understanding of feature reconstruction through a receptive field expansion mechanism that captures both global information

and multi-scale local features. In addition, considering the SR-HDR task is an issue of high-frequency reconstruction and local high dynamic range enhancement, we designed a learned guided filter module that combines the enhanced output and the original input as the final output. Experiments demonstrate that SR-HDRNet surpassed the other state-of-the-art methods by 0.61–1.43 dB on the 2× factor and by 0.26–1.71 dB on the 3× factor, with stable performance across diverse image types. Future research could focus on further optimizing the model for real-time performance and exploring various tasks in remote sensing and multimedia fields such as super-resolution, motion deblurring, dehazing and underwater/low-light enhancement.

Acknowledgement

This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No.ZCLQN26D0102, and the "Pioneer" and "Leading Goose" R&D Program of Zhejiang under Grant No. 2023C01212.

References

1. AI-Mekhlafi, H., Liu, S. Single Image Super-Resolution: A Comprehensive Review and Recent Insight. Frontiers of Computer Science, 2024, 18, 181702. https://doi.org/10.1007/s11704-023-2588-9
2. Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., Merhof, D. Medical Image Segmentation Review: The Success of U-Net. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12), 10076-10095. https://doi.org/10.1109/TPAMI.2024.3435571
3. Catley-Chandar, S., Tanay, T., Vandroux, L., Leonardis, A., Slabaugh, G., Pérez-Pellitero, E. FlexHDR: Modeling Alignment and Exposure Uncertainties for Flexible HDR Imaging. IEEE Transactions on Image Processing, 2022, 31, 5923-5935. https://doi.org/10.1109/TIP.2022.3203562
4. Chen, S., Xie, E., Ge, C., Chen, R., Liang, D., Luo, P. CycleMLP: A MLP-Like Architecture for Dense Visual Predictions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(12), 14284-14300. https://doi.org/10.1109/TPAMI.2023.3303397
5. Chen, S. K., Yen, H. L., Liu, Y. L., Chen, M. H., Hu, H. N., Peng, W. H., Lin, Y. Y. Learning Continuous Exposure Value Representations for Single-Image HDR Reconstruction. In IEEE International Conference on Computer Vision (ICCV), Paris, France, 2023, 12944-12954. https://doi.org/10.1109/ICCV51070.2023.01194
6. Chen, X., Liu, Y., Zhang, Z., Qiao, Y., Dong, C. HDRUNet: Single Image HDR Reconstruction with Denoising and Dequantization. In IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, USA, 2021, 354-363. https://doi.org/10.1109/CVPRW53098.2021.00045
7. Chen, X. Y., Wang, X. T., Zhou, J., Qiao, Y., Dong, C. Activating More Pixels in Image Super-Resolution Transformer. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, 2023, 22367-22377. https://doi.org/10.1109/CVPR52729.2023.02142
8. Chu, M., Xie, Y., Laura, L. T. Temporally Coherent GANs for Video Super-Resolution (TecoGAN). arXiv:1811.09393, 2020. https://doi.org/10.1145/3386569.3392457
9. Deng, X., Zhang, Y., Xu, M., Gu, S., Duan, Y. Deep Coupled Feedback Network for Joint Exposure Fusion and Image Super-Resolution. IEEE Transactions on Image Processing, 2021, 30, 3098-3112. https://doi.org/10.1109/TIP.2021.3058764
10. Dong, C., Loy, C. C., He, K., Tang, X. Learning a Deep Convolutional Network for Image Super-Resolution. In European Conference on Computer Vision (ECCV), Zurich, Switzerland, 2014, 184-199. https://doi.org/10.1007/978-3-319-10593-2_13
11. Guan, Y., Xu, R., Yao, M., Gao, R., Wang, L., Xiong, Z. Diffusion-Promoted HDR Video Reconstruction. In European Conference on Computer Vision (ECCV), Milan, Italy, 2024, 20-38. https://doi.org/10.1007/978-3-031-91838-4_2
12. Haris, M., Shakhnarovich, G., Ukita, N. Deep Back-Projection Networks for Super-Resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, USA, 2018, 1664-1673. https://doi.org/10.1109/CVPR.2018.00179
13. Hu, L., Chen, H., Allebach, J. P. Joint Multi-Scale Tone Mapping and Denoising for HDR Image Enhancement. In IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW), Waikoloa,

HI, USA, 2022, 729-738. https://doi.org/10.1109/WACVW54805.2022.00080

14. Hu, X. C., Mu, H. Y., Zhang, X., Wang, Z., Tan, T., Sun, J. Meta-SR: A Magnification-Arbitrary Network for Super-Resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019, 1575-1584. https://doi.org/10.1109/CVPR.2019.00167
15. Junghee, K., Siyeong, L., Suk-Ju, K. End-to-End Differentiable Learning to HDR Image Synthesis for Multi-Exposure Images. In Association for the Advancement of Artificial Intelligence (AAAI), 2021, 1780-1788. https://doi.org/10.1609/aaai.v35i2.16272
16. Kalantari, N. K., Ramamoorthi, R. Deep High Dynamic Range Imaging of Dynamic Scenes. ACM Transactions on Graphics, 2017, 36(4), 144:1-144:12. https://doi.org/10.1145/3072959.3073609
17. Kim, J., Lee, J. K., Lee, K. M. Deeply-Recursive Convolutional Network for Image Super-Resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, 1637-1645. https://doi.org/10.1109/CVPR.2016.181
18. Kim, J., Lee, J. K., Lee, K. M. Accurate Image Super-Resolution Using Very Deep Convolutional Network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, 1646-1654. https://doi.org/10.1109/CVPR.2016.182
19. Kim, S. Y., Kim, M. A Multi-Purpose Convolutional Neural Network for Simultaneous Super-Resolution and High Dynamic Range Image Reconstruction. In Asian Conference on Computer Vision (ACCV), Perth, Australia, 2018, 379-394. https://doi.org/10.1007/978-3-030-20893-6_24
20. Kosugi, S., Yamasaki, T. Unpaired Image Enhancement Featuring Reinforcement-Learning-Controlled Image Editing Software. In Association for the Advancement of Artificial Intelligence (AAAI), New York, USA, 2020, 11296-11303. https://doi.org/10.1609/aaai.v34i07.6790
21. Lazri, Z. M., Su, G. M. Iterative Graph-Based HDR Image Enhancement. In The Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, USA, 2023, 1090-1094. https://doi.org/10.1109/IEEECONF59524.2023.10476898
22. Le, P. H., Le, Q., Nguyen, R., Hua, B. S. Single-Image HDR Reconstruction by Multi-Exposure Generation. In IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, USA, 2023, 4063-4072. https://doi.org/10.1109/WACV56688.2023.00405
23. Li, J., Hassani, A., Walton, S., Shi, H. ConvMLP: Hierarchical Convolutional MLPs for Vision. In IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Vancouver, Canada, 2023, 6307-6316. https://doi.org/10.1109/CVPRW59228.2023.00671
24. Li, S., He, F. X., Du, B., Zhang, L., Xu, Y., Tao, D. Fast Spatio-Temporal Residual Network for Video Super-Resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019, 10522-10531. https://doi.org/10.1109/CVPR.2019.01077
25. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R. SwinIR: Image Restoration Using Swin Transformer. In IEEE International Conference on Computer Vision Workshops (ICCVW), Montreal, Canada, 2021, 1833-1844. https://doi.org/10.1109/ICCVW54120.2021.00210
26. Liang, M., Du, J. P., Li, L. H., Xue, Z., Wang, X., Kou, F., Wang, X. Video Super-Resolution Reconstruction Based on Deep Learning and Spatio-Temporal Feature Self-Similarity. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(9), 4538-4553. https://doi.org/10.1109/TKDE.2020.3034261
27. Liu, Z., Liu, H. CMCNet: Enhancing Face Image Super-Resolution Through CNN-Mamba Collaboration. The Visual Computer, 2025, 41, 12367-12382. https://doi.org/10.1007/s00371-025-04162-7
28. Marcel, S. S., Tsang, I. R., Khademi Kalantari, N. Single Image HDR Reconstruction Using a CNN with Masked Features and Perceptual Loss. ACM Transactions on Graphics, 2020, 39(4), 80:1-80:10. https://doi.org/10.1145/3386569.3392403
29. Mu, S. S., Zhang, Y., Jiang, Y. DRN-VideoSR: A Deep Recursive Network for Video Super-Resolution Based on a Deformable Convolution Shared-Assignment Network. Multimedia Tools and Applications, 2023, 82, 14019-14035. https://doi.org/10.1007/s11042-022-13818-8
30. Niu, J., Wu, J., Liu, W., Guo, W., Lau, R. W. H. HDR-GAN: HDR Image Reconstruction from Multi-Exposed LDR Images with Large Motions. IEEE Transactions on Image Processing, 2021, 30, 3885-3896. https://doi.org/10.1109/TIP.2021.3064433
31. Sun, Y., Wang, J., Yang, Y., Zhang, Y. SPSTNet: Image Super-Resolution Using Spatial Pyramid Swin Transformer Network. Signal, Image and Video Processing, 2025, 19, 278. https://doi.org/10.1007/s11760-025-03880-8
32. Tai, Y., Yang, J., Liu, X. Image Super-Resolution Via Deep Recursive Residual Network. In IEEE Confer-

ence on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017, 3147-3155. https://doi.org/10.1109/CVPR.2017.298

33. Tai, Y., Yang, J., Liu, X. Memnet: A Persistent Memory Network for Image Restoration. In IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, 4549-4557. https://doi.org/10.1109/ICCV.2017.486
34. Tan, X., Chen, H., Xu, K., Jin, Y., Zhu, C. Deep SR-HDR: Joint Learning of Super-Resolution and High Dynamic Range Imaging for Dynamic Scenes. IEEE Transactions on Multimedia, 2023, 25, 750-763. https://doi.org/10.1109/TMM.2021.3132165
35. Tel, S., Wu, Z. W., Zhang, Y. L., Heyrman, B., Demonceaux, C., Timofte, R., Ginhac, D. Alignment-Free HDR Deghosting with Semantics Consistent Transformer. In IEEE International Conference on Computer Vision (ICCV), Paris, France, 2023, 12790-12799. https://doi.org/10.1109/ICCV51070.2023.01179
36. Wang, T., Li, Y., Peng, J., Ma, Y., Wang, X., Song, F., Yan, Y. Real-Time Image Enhancer via Learnable Spatial-Aware 3D Lookup Tables. In IEEE International Conference on Computer Vision (ICCV), Montreal, Canada, 2021, 2451-2460. https://doi.org/10.1109/ICCV48922.2021.00247
37. Wang, X. T., Yu, K., Wu, S. X., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., Tang, X. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In European Conference on Computer Vision (ECCV), Munich, Germany, 2018, 1-16. https://doi.org/10.1007/978-3-030-11021-5_5
38. Wu, H., Zheng, S., Zhang, J., Huang, K. Fast End-to-End Trainable Guided Filter. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, USA, 2018, 1838-1847. https://doi.org/10.1109/CVPR.2018.00197
39. Wu, S., Xu, J., Tai, Y. W., Tang, C. K. Deep High Dynamic Range Imaging with Large Foreground Motions. In European Conference on Computer Vision (ECCV), Munich, Germany, 2018, 120-135. https://doi.org/10.1007/978-3-030-01216-8_8
40. Yang, X., Xu, K., Song, Y., Zhang, Q., Wei, X., Lau. R. Image Correction via Deep Reciprocating HDR Transformation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, USA, 2018, 1798-1807. https://doi.org/10.1109/CVPR.2018.00193
41. Yang, Y., Xu, L., Li, L., Qie, N., Li, Y., Zhang, P., Guo, Y. Personalized Image Aesthetics Assessment with Rich Attributes. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022, 19829-19837. https://doi.org/10.1109/CVPR52688.2022.01924
42. Yi, P., Wang, Z. Y., Jiang, K., Jiang, J., Ma, J. Progressive Fusion Video Super-Resolution Network via Exploiting Non-Local Spatio-Temporal Correlations. In IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 2019, 3106-3115. https://doi.org/10.1109/ICCV.2019.00320
43. Zeng, H., Cai, J. R., Li, L., Cao, Z., Zhang, L. Learning Image-Adaptive 3D Lookup Tables for High-Performance Photo Enhancement in Real Time. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4), 2058-2073. https://doi.org/10.1109/TPAMI.2020.3026740
44. Zhang, S., Lin, Y., Sheng, H. Residual Networks for Light Field Image Super-Resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019, 11046-11055. https://doi.org/10.1109/CVPR.2019.01130
45. Zhang, Y. L., Tian, Y. P., Kong, Y., Zhong, B., Fu, Y. Residual Dense Network for Image Restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(7), 2480-2495. https://doi.org/10.1109/TPAMI.2020.2968521
46. Zhu, L., Wang, X., Chang, Y., Li, J., Huang, T., Tian, Y. Event-Based Video Reconstruction via Potential-Assisted Spiking Neural Network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022, 3584-3594. https://doi.org/10.1109/CVPR52688.2022.00358
47. Zou, Y. H., Yan, C. G., Fu, Y. RawHDR: High Dynamic Range Image Reconstruction from a Single Raw Image. In IEEE International Conference on Computer Vision (ICCV), Paris, France, 2023, 12300-12310. https://doi.org/10.1109/ICCV51070.2023.01133

This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).