Received 17 July 2023, accepted 31 July 2023, date of publication 7 August 2023, date of current version 14 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3302692

RESEARCH ARTICLE

DANS: Deep Attention Network for Single Image Super-Resolution

JAGRATI TALREJA[1], (Graduate Student Member, IEEE), SUPAVADEE ARAMVITH[2], (Senior Member, IEEE), AND TAKAO ONOYE[3], (Senior Member, IEEE)

[1]Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand
[2]Multimedia Data Analytics and Processing Unit, Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand
[3]Graduate School of Information Science and Technology, Osaka University, Suita 565-0871, Japan

Corresponding author: Supavadee Aramvith (supavadee.a@chula.ac.th)

This research is funded by the Second Century Fund (C2F), Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University Bangkok, 10330, Thailand. This research is also funded by Thailand Science Research and Innovation Fund Chulalongkorn University (CU_FRB65_ind (9)_157_21_23), The NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B04G640053] and also funded by Thailand Science research and Innovation Fund Chulalongkorn University (IND66210019).

ABSTRACT The current advancements in image super-resolution have explored different attention mechanisms to achieve better quantitative and perceptual results. The critical challenge recently is to utilize the potential of attention mechanisms to reconstruct high-resolution images from their low-resolution counterparts. This research proposes a novel method that combines inception blocks, non-local sparse attention, and a U-Net network architecture. The network incorporates the non-local sparse attention on the backbone of symmetric encoder-decoder U-Net structure, which helps to identify long-range dependencies and exploits contextual information while preserving global context. By incorporating skip connections, the network can leverage features at different scales, enhancing the reconstruction of high-frequency information. Additionally, we introduce inception blocks allowing the model to capture information at various levels of abstraction to enhance multi-scale representation learning further. Experimental findings show that our suggested approach produces superior quantitative measurements, such as peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), visual information fidelity (VIF), and visually appealing high-resolution image reconstructions.

INDEX TERMS Image super-resolution, inception blocks, non-local sparse attention, U-Net.

I. INTRODUCTION

In convolutional neural network-based image processing, single image super-resolution (SISR) is one of the most important research fields. Super-Resolution (SR) is a classified modern-day problem needing immediate solutions. All the vision task-based devices require embedding fast and less complex Super-Resolution (SR) algorithms for faster and high-quality processing of visuals, i.e., images, videos, or live streaming. Focusing on images, the SR appertains to recuperating high-quality images from given low-quality images, and this process is termed Image Super-Resolution (ISR). SISR is an ill-posed problem because of the generation of multiple High-Resolution (HR) images corresponding to a single Low-Resolution (LR) image. Due to this reason, detailed constraints of images such as high-frequency, low-frequency, spatial characteristics, or a variety of image priors and domains are used as features of an image. These insightful and hierarchical features recover the finer detail of any image, which further helps in security and surveillance and many other vision-based applications such as image segmentation [1], [4], reconstruction [2], estimation [3], object and anomaly detection [5], etc. having varied applicability in security,

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

surveillance [6], medical [7], face [8], [9], satellite imaging [10], remote sensing [11]. Learning-based algorithms have recently shown remarkable performance compared to conventional super-resolution methods. Convolutional neural networks (CNN) dominate the culture for modeling the algorithms in Deep Learning (DL) [12], [13], [14], [15], [16]. The introduction of neural networks (NN) in Super-Resolution Convolution Neural Networks (SRCNN) [17] proposed by Dong et al. has led to many advancements in the field of SISR. After the SRCNN approach, the research community explored ISR regarding model frameworks, up-sampling methods, network design, learning strategies, etc. Introducing CNNs [17], Generative Adversarial networks (GANs) [18], attention mechanisms [19], [20], transformative discriminative networks [21], and residual networks [22] has significantly improved the quality of LR images. To reduce the computational cost and training complexity, Wavelet-SRNet [23] and Deep wavelet super-resolution (DWSR) [24] were the first to super-resolve images in the wavelet domain, which uses the contextual information of an image by using low and high-frequency sub-bands having topological information. Fan et al. [25] concluded the favorability of sparsity [26], [27] constraints by imposing sparsity [28] and proved the increase in efficiency of neural networks [29]. Meanwhile, with network architecture, development turned towards exploring the sparsity constraints using attention mechanisms in sequential models. It emerged as a top-tier standard for Peak Signal-to-Noise ratio (PSNR) or Structural Similarity Index Measure (SSIM). Non-Local Sparse attention (NLSN) [30] is one of the most efficient current examples of a non-local channel attention-based sequential model for getting structural information in dominant or more focused regions. However, there is room for improving computational efficiency and reconstruction of finer details while preserving the natural textures of this sequential model.

Considering the current scenario of fast processing smart edge devices or embedded systems on chip (SoC), the model size must be considered for achieving faster and state-of-the-art (SOTA) comparable outcomes for higher and arbitrary enlargement factors. Several enhancements in ISR are based on sequential model designs. In this experiment, research work showed higher performance gains with reduced model size. Even though the sequential model helps to improve the network performance, they still face some limitations.

(1) Above mentioned methods are computationally expensive and require substantial processing power and time. They also require storing and accessing large affinity matrices demanding significant memory resources. (2) These methods can sometimes introduce unwanted artifacts or distortions into super-resolved images. The artifacts can include checkboard patterns, jagged edges, or blurring in certain regions. (3) When the input image is extremely low-resolution and contains noise or compression artifacts, earlier methods can struggle to recover fine details or remove noise patterns effectively.

An efficient way to overcome all those problems is to combine the sequential models [30] with non-sequential models [4]. The non-sequential methods have been proven to achieve meaningful features and effectively capture regional and global image subjects. Furthermore, this approach [30] helps to address the information loss encountered in deep neural networks. Isola et al. [31] first considered sequential models as encoder-decoder networks [32] and linked them with non-sequential architectures to separate high and low-frequency components of an image. Working with the same approach, we propose a novel approach for image super-resolution that combines non-local sparse attention with a U-Net network architecture and integrates inception blocks. The U-Net architecture serves as the backbone of our proposed network architecture. The skip connections [33] and encoder-decoder [32] structure enable efficient feature learning at various scales. The non-local sparse attention module enables the network to capture long-range dependencies by modeling the correlations between image patches and enhancing information exchange across the image, leading to improved utilization of contextual information for SR reconstruction. The integration of skip connections [34] ensures that the network can leverage features from different layers, enabling the efficient recovery of high-frequency information. To further enhance multi-scale representation learning, we introduce inception blocks into our network architecture, enabling the model to capture information at various levels of abstraction. This enriches the network's capability to capture local and global structures, enhancing super-resolution performance. In summary, the main contributions of our proposed model are listed below: (i) Propose a non-sequential backbone with skip connection to enable efficient feature learning at different scales and reconstruction of fine image details while maintaining global context. (ii) Integrate the non-local sparse attention module to capture long-range dependencies, enhance information exchange across image patches and improve utilization of contextual information for SR reconstruction. (iii) Employ inception block and benefit from parameterized and efficient feature extraction to enhance the network capability to capture local and global structures at different levels of abstraction. The remaining article comprises Section II, which briefly reviews the relevant work of the proposed method, and Section III describes the methodology of the network. Section IV presents the experimental results and comparative analysis with state-of-the-art methods. Discussion and Conclusion with future work are in sections V and VI.

II. RELATED WORK

Image super-resolution (ISR) is a well-studied problem in computer vision, and numerous techniques have been proposed to tackle the challenge of reconstructing high-resolution images from low-resolution inputs. This section provides an extensive overview of the related work in image super-resolution, focusing on deep learning-based methods, attention mechanisms, and network architectures. Deep learning-based approaches have revolutionized the field of image super-resolution by harnessing the representation power of convolutional neural networks (CNNs). Dong et al. introduced the pioneering work of Super-Resolution Convolutional Neural Network (SRCNN) [17], which directly utilized a shallow network to learn the mapping from low-resolution to high-resolution images. SRCNN achieved impressive results and laid the foundation for subsequent research in deep learning-based SR techniques. Deeper and more complex network architectures have been proposed to improve SR performance. For instance, Very Deep Super-Resolution (VDSR) [35] introduced a deeper network using a 20-layer residual network. The residual learning framework allowed VDSR [35] to efficiently capture residual information and achieve state-of-the-art results at that time. After this, many methods were employed, such as Deeply Recursive Network (DRCN) [36], but it shows slow training convergence. To extend this Residual Encoder-Decoder Network (REDNet) [37] was used to enhance model performance and fasten the convergence. To further speed up the training convergence of these models, Denoising Convolutional Neural Network (DnCNN) [38] was introduced by Chen et al. Since these deeper networks could not perform on small edge devices, the concept of introducing lightweight models was introduced [39], [40], [41]. Deep Recursive Residual Network (DRRN) [42] by Tai et al. and Information Multi-Distillation Network (IMDN) [39] by Hui et al. are two examples of lightweight models. Enhanced Deep Super-Resolution (EDSR) [22] enhanced the network architecture by increasing the network depth, utilizing residual blocks, and adopting an improved optimization strategy. EDSR [22] won the New Trends in Image Restoration and Enhancement (NTIRE) 2017 challenge on single image Super-Resolution: Dataset and Study and became a benchmark for subsequent SR models. After this, researchers shifted to making shallower networks and focusing on designing models with less memory consumption and computation time. For this, a Persistent Memory Network for image restoration (MemNet) [43] was introduced, which combined skip connections with CNN layers. This model was shallower compared to EDSR [22] and VDSR [35]. Multi-Scale Residual Network (MSRN) [44] by Li et al. employed adaptive feature extraction and used hierarchical information for image SR. Some methods also improve noise in an image, such as Learning a Single Convolutional Super-Resolution Network for Multiple degradations (SRMDNF) [45]. Furthermore, Deep Recurrent Fusion Network (DRFN) [46] proposed a transposed layer method for scale computation. To reduce the computational cost, Hung et al. suggested the concept of a Super Sampling network (SSNet) [47]. A multiple-cascaded information distillation block was introduced in Fast and Accurate Single Image Super-Resolution via Information Distillation Network (IDN) [48] to construct high-quality residuals in SR. While deep learning-based methods have pulled off remarkable results in ISR, there are still several challenges that researchers are actively addressing. One challenge is the trade-off between computational efficiency and reconstruction quality. Deep network architectures with many parameters are computationally expensive, making them less practical for real-time applications. Addressing this challenge requires exploring network compression techniques, model quantization, and efficient network architectures. Squeeze-and-Excitation Next for Single Image Super-Resolution (SENext) [14] helps address these challenges, balance performance and computational cost, and avoid the risk of overfitting. However, these challenges still need to be explored and addressed carefully. Attention mechanisms have gained significant attention to address these issues in various computer vision tasks, including image super-resolution. These mechanisms aim to capture long-range dependencies and exploit contextual information within images. Figures 1a, 1b and 1c show the different attention mechanisms used in state-of-the-art methods. Channel attention was first introduced in Deep Residual Channel Attention Networks (RCAN) [19]. After this, channel attention has been used in many networks for improving performance, e.g., Cross-Scale Non-Local (CS-NL) [49], Holistic Attention Networks (HAN) [50], and Multi-FusNet of Cross Channel Networks (MFCC) [16]. Further, improvements were made by using the second-order feature statistics of global average pooling (GAP) and introducing the second-order channel attention (SOCA) [51] module for more variational features. The exemplar model using channel attention mechanisms is Image Super-Resolution using RCAN [19], Second Order Attention Network (SAN) [51], and Densely Residual Laplacian Network (DRLN) [52]. After this, many other attention mechanisms, such as the self-attention mechanism introduced by Zhang et al. [53], allow the network to attend to the different spatial locations and capture global correlations. The self-attention module employs learned attention maps to reweight feature responses, enhancing the network's ability to focus on informative image regions. This attention mechanism has been successfully applied in SR models, improving reconstruction quality. Spatial attention creates a spatial map and utilizes the interdependencies of channels and features. It focuses on the informative part by applying average and max-pooling along the axis of the channels and then integrating them to produce efficient feature maps. Spatial attention has been introduced in models such as Image super-resolution via channel attention and spatial attention [54] and Residual Feature Aggregation Networks (RFANet) [55]. Another example of an attention mechanism is non-local attention [56] which has also been explored in the context of ISR. Non-local Neural Networks (NLNN) [29] introduced non-local operations and sparse coding [57] that capture relationships between all possible pairs of positions in an image. These operations allow the network to model the interactions between distant image regions, facilitating the exploitation of long-range dependencies. NLSN [30] extended the non-local mechanism by introducing sparsity [58], [59] to reduce computational complexity while

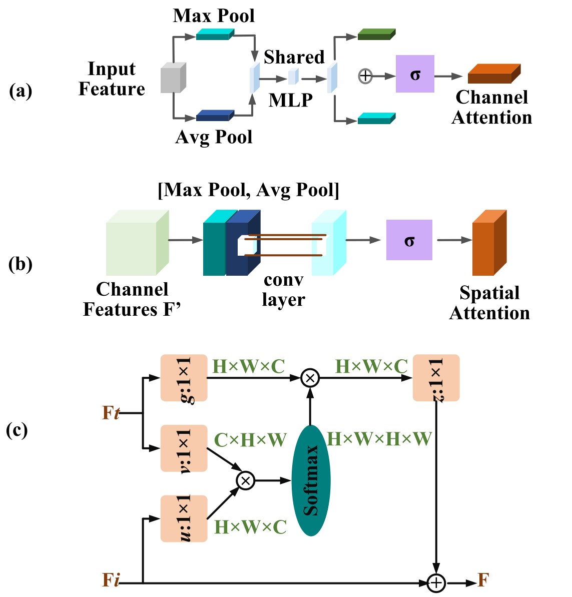

FIGURE 1. Different Attention mechanisms in image super-resolution (a) Channel Attention [19], (b) Spatial Attention [54], and (c) Non-Local Attention [49].

preserving the non-local modeling capability. The sparse attention attends to a subset of non-local positions, effectively capturing relevant contextual information.

The U-Net architecture, proposed by Ronneberger et al. [4], was widely adopted in various image restoration tasks, including image super-resolution. In [4], an encoder-decoder architecture with skip connections facilitates low-level and high-level feature learning. The skip connections enable the network to combine low-level details with high-level semantic information, enhancing the reconstruction of fine image structures. U-Net [4] has been further improved by incorporating dilated convolutions, residual blocks, and other modifications to enhance its performance in SR tasks. Inception blocks, introduced in GoogLeNet [60], have been widely utilized for multi-scale feature extraction in deep networks. Inception blocks consist of parallel convolutional layers with different receptive fields [61], allowing the model to capture information at multiple scales. This multi-scale representation learning enhances the network's ability to capture local and global structures, facilitating improved reconstruction quality. Inception blocks have been incorporated into SR models to capture diverse image features and enable more effective feature extraction. While substantial progress has been made in the field of image super-resolution, there is still a need for techniques that can effectively capture long-range dependencies, exploit contextual information, and enhance the reconstruction of fine details. This work proposes a novel approach that combines non-local sparse attention with a U-Net network architecture augmented with inception blocks. Integrating these components aims to leverage the strengths of attention mechanisms, multi-scale feature extraction, and deep network architectures to enhance image super-resolution performance further. In summary, deep learning-based approaches, attention mechanisms, and network architectures have significantly advanced in image super-resolution. Combining these techniques has led to significant improvements in reconstruction quality, enabling the generation of high-resolution images with enhanced details. Ongoing research addresses challenges such as computational efficiency, generalization across domains, and adapting SR techniques to video super-resolution [5]. These advancements are crucial for realizing the full potential of image super-resolution in various applications, including face image super-resolution [3], medical imaging [7], surveillance [6], remote sensing [11], and digital content creation [8], [62].

III. PROPOSED METHOD

This section presents our proposed novel approach in single image super-resolution by fusion of Non-Local Sparse attention mechanism [30] into U-Net [4] framework Network. Furthermore, we employed the inception block [60] in the network architecture to extract various contextual features at distinct levels of abstraction. Additionally, skip connections are introduced to transmit low-frequency information at each network stage to reduce the required parameters for the computation. The up-sampling and down-sampling are presented in the network to localize the high-resolution features to generate more precise results for the contextual regions in an image.

As shown in Figure 2, our proposed Deep Attention Network for Single Image Super-Resolution (DANS) comprises three staged encoder-decoder frameworks. Each stage consists of Inception Block and NLSA block. The architecture of DANS uses 5 NLSA [30] Blocks with up-sampling and down-sampling. It also contains six inception blocks to better understand fine-grained details and broader contextual details at different levels of abstraction. The image is enlarged by a factor of two at the encoder side and reduced by two at the decoder. The contextual details of an image are stored in high-frequency signals and are refined by propagating it through the encoder-decoder framework. In contrast, the low-frequency information is passed through the skip connections. This symmetric encoder-decoder design produces more precise SR results than a sequential model. Since hierarchical features [63] help to understand global and fine-grained details better, adopting an encoder-decoder framework by introducing up-sampling and down-sampling instead of adopting a sequential framework helps to learn hierarchical features in an image. It is to be noted that the encoder-decoder network is more flexible with variable input and output data [4]. The encoder can accommodate input of varying sizes, and the decoder can generate output with variable sizes. Also, encoder-decoder frameworks are well suited for structured outputs, i.e., processing images containing text. Therefore, each modality of images with text is processed independently by the encoder and integrated to generate joint representations of the output by the decoder.

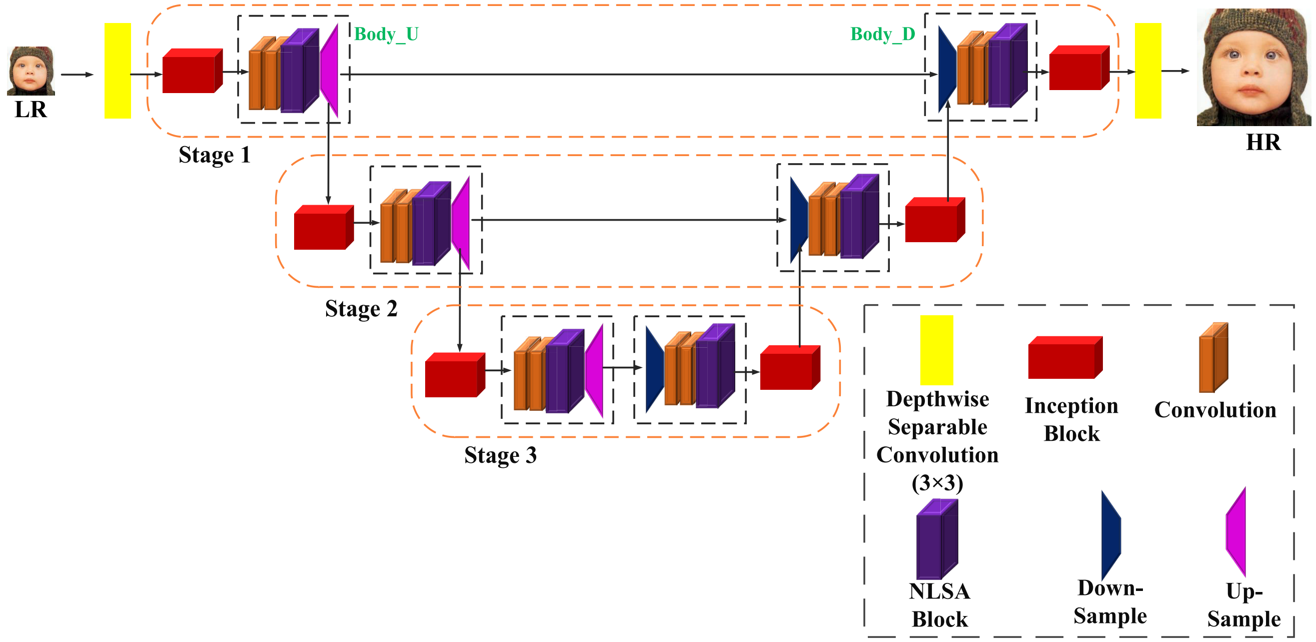

FIGURE 2. The proposed network architecture of Deep Attention Network for Single Image Super-Resolution (DANS).

Instead of residual blocks, we have used inception blocks to extract features at various abstraction levels to capture global and local contextual details better. This is much more useful for object detection or scene understanding. Inception Block helps to achieve diverse and complementary sets of feature maps because of parallel pathways and different receptive fields [61]. The parameter sharing makes the inception block more compact than the residual block. The diversity of features captured benefits tasks requiring a wide range of visual patterns.

The Deep Feature Extraction Block indicated by Body_U and Body_D in Figure 2 consists of two 3 × 3 convolutions, one Non-Local Sparse Attention Block and an up-sample module in Body_U at the encoder side and down-sampler in Body_D at the decoder side at each stage of the framework. Non-local sparse attention has shown brilliant performance in NLSN [30] for extracting contextual details in SISR, but it increases the computational cost when used in a sequential model. Our proposed approach uses NLSA block in an encoder-decoder framework to reduce the number of parameters and ultimately help reduce the computational cost compared to a sequential model framework. Since sparsity constraints in the NLSA block help focus on contextual details, the encoder-decoder framework preserves the spatial information in the model architecture. Therefore, using the NLSA block in this framework allows the decoder to easily access the preserved low-level features from the encoder through skip connections.

A. DEEP FEATURE EXTRACTION (DFE)

Deep feature extraction is used to capture an image's textural and contextual features, also known as deep features. This technique mainly processes and refines the high-frequency information in the image patches. Deep Feature Extraction (DFE) is classified into two categories 1.) Deep Feature Extraction Up-Sampler Block, and 2.) Deep Feature Extraction Down-Sampler Block. As seen in Figure 2, at every stage, for the encoder side, we have used Deep Feature Extraction Up-Sampler Block termed Body_U, and for the decoder, we have used Deep Feature Extraction Down-Sampler Block termed Body_D.

1) DEEP FEATURE EXTRACTION UP-SAMPLER

As seen in Figure 3, the Deep Feature extraction Up-Sampler (DFE_U) block consists of two 3 × 3 convolutional layers, one NLSA Block, and an up-sampling or down-sampling module stacked together. The 3 × 3 convolutional layers help to extract the local features to learn spatial and discriminative features [64] from the input data. The NLSA [30] blocks help to analyze the relationship between different spatial positions of the extracted features, and finally, the up-sampling module enlarges the derived characteristics. The output of the Deep Feature Extraction Up-Sampler Block is shown in Equation 1.

$$H_0 = H_{DFE\_U}(H_{IB}), \tag{1}$$

where $H_{DFE\_U}(.)$ represents deep feature extraction up-sampler operation, and (H_{IB}) is the output of the inception block where the original input LR image is fed after passing through the depthwise separable convolution. After obtaining the up-sampled deep features, H_0 is used as the input of the next stage Inception block and the Deep Feature Extraction Down-Sampler Block DFE_D.

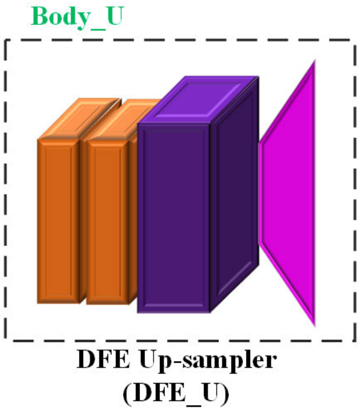

FIGURE 3. The structure of Deep Feature Extraction Up-Sampler Block.

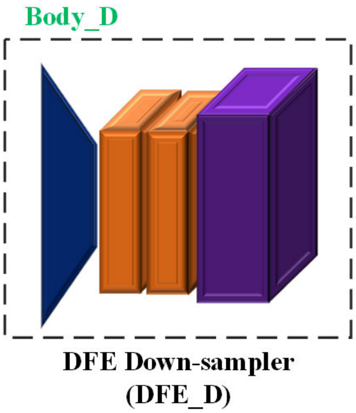

FIGURE 4. The structure of Deep Feature Extraction Down-Sampler Block.

2) DEEP FEATURE EXTRACTION DOWN-SAMPLER

This block is described in Figure 4. It uses a down-sampling module to shrink the spatial resolution of the characteristic maps and aggregate details at a coarse scale. Like DFE_U, in Deep Feature extraction Down-Sampler (DFE_D) also stack two 3 × 3 convolutions with NLSA [30] block to capture deep contextual feature effectively. Still, instead of an up-sampler, we use a down-sampler to help downscale the extracted features to process through the NLSA [30] Block. The output of the Deep Feature Extraction Down-Sampler Block is shown in Equation 2. It is to be noted that the DFE_D excepts two inputs, i.e., one from the DFE_U Block and the other from the inception block.

$$H_I = H_{DFE\_D}(H_{IB} + H_{DFE\_U}), \tag{2}$$

where $H_{DFE\_D}(.)$ represents deep feature extraction down-sampling operation, and (H_{IB}) is the output of the inception block. After obtaining the up-sampled deep features, H_I is used as the input of the next stage Inception block and the Deep Feature Extraction Down-Sampler Block.

This process is repeated at each network stage, and the HR image is obtained after passing through the depth-wise separable convolution operation.

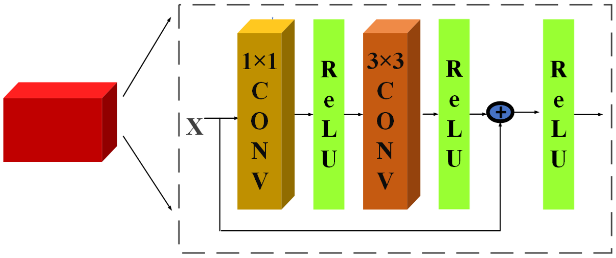

FIGURE 5. Inception Block.

B. INCEPTION BLOCK

The Inception block, as seen in Figure 5, also known as the Inception module, is a building block of deep neural network architectures, originally introduced in the GoogLeNet [60] model for image classification. It is designed to capture information at multiple scales and learn diverse and rich feature representations.

The key idea behind the Inception block is to parallelize and concatenate multiple convolutional operations of different filter sizes, allowing the network to capture local and global information. This enables the model to learn a wide range of features at various levels of abstraction within the same layer.

The Inception block typically combines 1 × 1, 3 × 3, and 5 × 5 convolutional filters, and max-pooling operations. It includes multiple parallel branches, each performing a different convolution operation. The final outputs of the Inception block are generated by concatenating these branches along the channel dimension. The 1 × 1 convolutions are used to make the input less dimensional and control the computational complexity of the block.

One of the major advantages of the Inception block is its ability to efficiently capture regional and global details within a layer and maintain a minimum number of parameters while keeping a large receptive field [61]. This makes the Inception block effective for jobs such as image classification [65], object detection [1], and semantic segmentation [4].

In our proposed model, we used 1 × 1 and 3 × 3 convolutional layers inside the inception block followed by Rectified Linear Unit (ReLU) activation. We further controlled experiments by changing the activations to PReLU and CReLU to check their effect on the model's performance.

The potential applications of our proposed framework in image and computer vision tasks such as image segmentation, classification, and object detection. The main versatile application of our proposed approach is in medical imaging [7], image restoration [32], and video super-resolution [5], where efficient models are highly desirable.

IV. EXPERIMENTAL RESULTS

To show the quantitative and qualitative visual results of our proposed DANS model, several experiments were conducted on benchmark test datasets to verify its performance. Furthermore, the computational cost in terms of

TABLE 1. The total number of images available in the Training and Testing Dataset.

Datasets	Train	Test	Number of Images
DIV2K [66]	✓	×	800
Set5 [67]	×	✓	5
Set14 [68]	×	✓	14
BSD100 [69]	×	✓	100
Urban100 [70]	×	✓	100
Manga109 [71]	×	✓	109

network parameters and execution time are discussed in this section. Our model training performance is evaluated with average PSNR (dB) and loss versus epoch convergence. Additionally, an ablation study has been shown for different rounds of NLSA [30] with the baseline during the testing process. Structural ablation study has been demonstrated by using different activation functions in the inception block and showing performance convergence per epoch. Finally, noise degradation analysis compared with different selective models has been presented to verify that our model also has better quantitative performance on noisy and blurry images.

A. EXPERIMENTAL SETTINGS

This section demonstrates the training details, evaluation metrics, and the datasets used in training and testing the effectiveness of our proposed model on public datasets. It is to be noted that the training and the testing sets are different.

1) DATASETS

We used the DIV2K [66] dataset to train our SR model. It contains 800 high-quality training images. This dataset contains diversified variations in its images. For testing, we used five benchmark datasets for comparison, which are Set5 [67], Set14 [68], BSD100 [69], Manga109 [71], and Urban100 [70]. Table 1 shows the training and testing datasets with the number of images.

2) TRAINING DETAILS

We cropped random low-resolution patches with sizes 48 × 48 to train our model. The low-resolution images for ×2, ×3, ×4, and ×8 were obtained using MATLAB R2022b. The proposed network is trained on NVIDIA GeForce GTX 2080ti GPU with 24GB memory. Python 3.6 programming language with PyTorch 1.1.0 platform has been used for coding the algorithm of the proposed model. Eight hundred samples from DIV2K [66] datasets are obtained for training the model. We select an Adam optimizer with $\beta_1 = 0.90$ and $\beta_2 = 0.99$ for optimization purposes. The learning rate of the proposed model is kept being 10^{-4} and reduced to half every 200 epochs.

3) TESTING DETAILS

Our proposed model has been tested on five standard benchmark datasets, i.e., Set5 [67], Set14 [68], BSD100 [69], Urban100 [70], and Manga109 [71] datasets. The LR image is obtained by downsampling the HR images using bicubic kernels. 48 × 48 patches are randomly cropped from the training samples and divided into mini-batches of 8 images. Data augmentation also creates more samples for the algorithm by flipping and randomly rotating for 90, 180, and 270 degrees. Input is expected to be 100 × 100 spatial with 64 input and output channels. The range of scaled images for LR is set in the range [- 1, 1]. The Mean Squared Error (MSE) has been computed on the image intensity range [-1, 1]. PSNR and SSIM are standard evaluation metrics for quantitatively comparing our model with state-of-the-art methods.

B. QUANTITATIVE COMPARISONS WITH STATE-OF-THE-ART MODELS

Table 2 presents the tabular standard metric comparison of five benchmark test datasets. Quantitative analysis of our proposed DANS shows a comparison with seventeen SOTA methods, such as Bicubic, SRCNN [17], FSRCNN [72], VDSR [35], RDN [73], LapSRN [74], SENext [14], RCAN [19], MemNet [43], RNAN [75], MFCC [16], SRFBN [76], SAN [51], EDSR [22], HAN [50], SwinIR [77], and NLSN [30]. As shown in Table 2, our proposed DANS quantitative results have significantly outperformed the state-of-the-art methods in terms of PSNR and SSIM. Our proposed DANS model is better in performance on all test datasets for scale factors ×2, ×3, ×4, and ×8. Furthermore, our proposed method obtained a higher value of PSNR/SSIM on all averages compared to other SOTA models.

C. COMPARISON ANALYSIS BASED ON THE NUMBER OF MODEL PARAMETERS

In Figure 6, the comparison in parameters versus PSNR has been shown for our proposed DANS model. The performance is evaluated on the Set5 [67] test dataset for our proposed model DANS with an enlargement factor of ×2. A reduction in the number of parameters demonstrates a reduction in computational cost. Compared to other deep learning models, the DANS model helps reduce the model's size better. DANS has parameters about 94% less than EDSR [22], 84% less than RCAN [19], 88% less than RDN [73], 45% less than NLSN [30], 23% less than HAN [50], and 29% less than SRFBN [76]. Figure 6 shows that our proposed method has lesser parameters than six other state-of-the-art methods.

D. COMPARISON ANALYSIS OF PSNR AND SSIM ON THE IMAGE SR DATASETS FOR ENLARGEMENT FACTORS OF ×4 AND ×8

Figure 7, Figure 8, Figure 9, and Figure 10 shows the performance comparison of different existing image SR methods using standard objective measures, i.e., PSNR and SSIM on benchmark datasets (Set5 [67], Set14 [68], BSD100 [69], Urban100 [70], Manga109 [71]) for enlargement factor of ×4 and ×8. The quantitative results reveal that our proposed DANS attains the most effective results as compared to NLSN [30], HAN [50], SwinIR [77], EDSR [22], and

TABLE 2. Quantitative evaluation of our proposed DANS with SR models. Average values of PSNR/SSIM have also been reported on enlargement factors ×2, ×3 ×4, and ×8. The best quantitative value has been recorded as bold with Red color. The second-best quantitative value is shown in blue color with an underline.

Method	Factor	#Param	Set5 [67]		Set14 [68]		BSD100 [69]		Urban100 [70]		Manga109 [71]		Average	
			PSNR↑	SSIM↑	PSNR↑	SSIM ↑	PSNR↑	SSIM ↑	PSNR↑	SSIM ↑	PSNR↑	SSIM ↑	PSNR↑	SSIM ↑
Bicubic	×2	-/-	33.68	0.9304	30.24	0.8691	29.56	0.8435	26.88	0.8405	31.05	0.9349	30.23	0.8832
SRCNN [17]	×2	57K	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.51	0.8946	35.72	0.9680	33.11	0.9219
FSRCNN [72]	×2	12K	36.98	0.9556	32.62	0.9087	31.50	0.8904	29.58	0.9009	36.62	0.9710	33.56	0.9260
VDSR [35]	×2	665K	37.53	0.9587	33.05	0.9127	31.90	0.8960	30.77	0.9141	37.16	0.9740	33.24	0.9314
MemNet [43]	×2	677K	37.78	0.9597	33.28	0.9142	32.08	0.8978	31.31	0.9195	37.72	0.9740	34.43	0.9330
LapSRN [74]	×2	812K	37.52	0.9591	32.99	0.9124	31.80	0.8949	30.41	0.9101	37.53	0.9740	33.87	0.9302
SENext [14]	×2	97K	38.04	0.9608	34.24	0.9181	32.21	0.8997	32.43	0.9287	38.79	0.9774	35.14	0.9369
RDN [73]	×2	21,900K	38.24	0.9614	34.01	0.9212	32.34	0.9017	32.89	0.9353	39.18	0.9780	35.33	0.9395
RCAN [19]	×2	16,000K	38.27	0.9614	34.12	0.9216	32.41	0.9027	33.34	0.9384	39.44	0.9786	35.52	0.9405
RNAN [75]	×2	1,350K	38.17	0.9611	33.87	0.9207	32.32	0.9014	32.73	0.9340	39.23	0.9785	35.26	0.9391
MFCC [16]	×2	1,861K	38.16	0.9606	33.85	0.9195	32.28	0.9010	32.65	0.9331	39.11	0.9780	35.21	0.9384
SRFBN [76]	×2	3,500K	38.11	0.9609	33.82	0.9196	32.29	0.9010	32.62	0.9328	39.08	0.9779	35.18	0.9384
SAN [51]	×2	1,550K	38.31	0.9620	34.07	0.9213	32.42	0.9028	33.10	0.9370	39.32	0.9792	35.44	0.9404
EDSR [22]	×2	43,000K	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9351	39.10	0.9773	35.28	0.9386
HAN [50]	×2	3,230K	38.27	0.9614	34.16	0.9217	32.41	0.9027	33.35	0.9385	39.46	0.9785	35.53	0.9405
NLSN [30]	×2	4,475K	38.34	0.9618	34.08	0.9231	32.43	0.9027	33.42	0.9394	39.59	0.9789	35.57	0.9412
SwinIR [77]	×2	878K	38.38	0.9620	34.24	0.9233	32.47	0.9032	33.51	0.9401	39.70	0.9794	35.66	0.9416
DANS (Ours)	×2	2,456K	**38.49**	**0.9622**	**34.28**	**0.9248**	**32.64**	**0.9039**	**33.58**	**0.9404**	**39.72**	**0.9796**	**35.74**	**0.9422**
Bicubic	×3	-/-	30.40	0.8686	27.54	0.7741	27.21	0.7389	24.46	0.7349	26.95	0.8566	27.31	0.7945
SRCNN [17]	×3	57K	32.75	0.9090	29.29	0.8215	28.41	0.7863	26.24	0.7991	30.48	0.9117	29.44	0.8455
FSRCNN [72]	×3	12K	33.16	0.9140	29.42	0.8242	28.52	0.7893	26.41	0.8064	31.10	0.9210	29.70	0.8516
VDSR [35]	×3	665K	33.66	0.9213	29.78	0.8318	28.83	0.7976	27.14	0.8279	32.01	0.9340	30.28	0.8624
MemNet [43]	×3	677K	34.09	0.9248	30.00	0.8350	28.96	0.8001	27.56	0.8376	32.51	0.9369	30.62	0.8669
LapSRN [74]	×3	812K	33.82	0.9227	29.79	0.8320	28.82	0.7973	27.07	0.8271	32.21	0.9350	30.36	0.8631
SENext [14]	×3	54K	34.32	0.9255	**31.08**	0.8419	29.11	0.8047	28.60	0.8519	33.63	0.9451	31.35	0.8738
RDN [73]	×3	21,900K	34.71	0.9296	30.57	0.8468	29.26	0.8093	28.80	0.8653	34.13	0.9484	31.49	0.8798
RCAN [19]	×3	16,000K	34.74	0.9299	30.65	0.8482	29.32	0.8111	29.09	0.8702	34.44	0.9499	31.64	0.8818
RNAN [75]	×3	1,350K	34.66	0.9290	30.52	0.8462	29.32	0.8090	28.75	0.8646	34.25	0.9483	31.50	0.8794
MFCC [16]	×3	2,230K	34.67	0.9294	30.51	0.8456	29.22	0.8080	28.64	0.8616	34.15	0.9478	31.43	0.8793
SRFBN [76]	×3	3,500K	34.70	0.9292	30.51	0.8461	29.24	0.8084	28.73	0.8641	34.18	0.9481	31.47	0.8791
SAN [51]	×3	1,550K	34.75	0.9300	30.59	0.8476	29.33	0.8112	28.93	0.8671	34.30	0.9494	31.58	0.8810
EDSR [22]	×3	43,000K	34.65	0.9280	30.52	0.8462	29.25	0.8093	28.80	0.8653	34.17	0.9476	31.48	0.8792
HAN [50]	×3	3,230K	34.75	0.9299	30.67	0.8483	29.32	0.8110	29.10	0.8705	34.48	0.9500	31.66	0.8819
NLSN [30]	×3	4,475K	34.85	0.9306	30.70	0.8485	29.34	0.8117	29.25	0.8726	34.57	0.9508	31.74	0.8824
SwimIR [77]	×3	886K	34.89	0.9312	30.77	0.8503	29.37	0.8124	29.29	0.8744	34.74	0.9518	31.81	0.8840
DANS (Ours)	×3	2,456K	**34.96**	**0.9329**	30.88	**0.8512**	**29.42**	**0.8132**	**29.31**	**0.8752**	**34.88**	**0.9519**	**31.89**	**0.8848**
Bicubic	×4	-/-	28.43	0.8109	26.00	0.7023	25.96	0.6678	23.14	0.6574	25.15	0.7890	25.68	0.7250
SRCNN [17]	×4	57K	30.48	0.8628	27.50	0.7513	26.90	0.7103	24.52	0.7226	27.66	0.8580	27.40	0.7785
FSRCNN [72]	×4	12K	30.70	0.8657	27.59	0.7535	26.96	0.7128	24.60	0.7258	27.89	0.8590	27.57	0.7850
VDSR [35]	×4	665K	31.35	0.8838	28.02	0.7678	27.29	0.7252	25.18	0.7525	28.82	0.8860	28.13	0.8031
MemNet [43]	×4	677K	31.74	0.8893	28.26	0.7723	27.40	0.7281	25.50	0.7630	29.42	0.8942	28.46	0.8094
LapSRN [74]	×4	812K	31.54	0.8866	28.09	0.7694	27.32	0.7264	25.21	0.7553	29.09	0.8900	28.27	0.8060
SENext [14]	×4	54K	31.50	0.8947	**28.99**	0.7812	**28.49**	0.7357	26.64	0.7839	30.48	0.9084	29.22	0.8208
RDN [73]	×4	21,900K	32.47	0.8990	28.81	0.7871	27.72	0.7419	26.61	0.8028	31.00	0.9151	29.32	0.8291
RCAN [19]	×4	16,000K	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087	31.22	0.9173	29.46	0.8317
RNAN [75]	×4	1,350K	32.49	0.8982	28.83	0.7878	27.72	0.7421	26.61	0.8023	31.09	0.9149	29.34	0.8291
MFCC [16]	×4	2,157K	32.42	0.8973	28.73	0.7849	27.67	0.7399	26.48	0.7977	30.98	0.9131	29.25	0.8265
SRFBN [76]	×4	3,500K	32.47	0.8983	28.81	0.7868	27.72	0.7409	26.60	0.8015	31.15	0.9160	29.35	0.8287
SAN [51]	×4	1,550K	32.64	0.9003	28.92	0.7888	27.78	0.7436	26.79	0.8068	31.18	0.9169	29.46	0.8312
EDSR [22]	×4	43,000K	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148	29.32	0.8289
HAN [50]	×4	3,230K	32.64	0.9002	28.90	0.7890	27.80	0.7442	26.85	0.8094	31.42	0.9177	29.52	0.8321
NLSN [30]	×4	4,475K	32.59	0.9000	28.87	0.7891	27.78	0.7444	26.96	0.8109	31.27	0.9184	29.49	0.8325
SwinIR [77]	×4	897K	32.72	0.9021	28.94	0.7914	27.83	0.7459	27.07	0.8164	31.67	0.9226	29.64	0.8356
DANS (Ours)	×4	2,456K	**32.78**	**0.9028**	28.98	**0.7928**	27.97	**0.7468**	**27.32**	**0.8189**	**31.74**	**0.9228**	**29.75**	**0.8368**
Bicubic	×8	-/-	24.40	0.6580	23.10	0.5660	23.67	0.5480	20.74	0.5160	21.47	0.6500	22.68	0.5876
SRCNN [17]	×8	57K	25.33	0.6900	23.76	0.5910	24.13	0.5660	21.29	0.5440	22.46	0.6950	23.42	0.5739
FSRCNN [72]	×8	12K	25.60	0.6970	24.00	0.5990	24.31	0.5720	21.45	0.5500	22.72	0.6920	23.46	0.5696
VDSR [35]	×8	665K	25.93	0.7240	24.26	0.6140	24.49	0.5830	21.70	0.5710	23.16	0.7250	23.50	0.5800
MemNet [43]	×8	677K	26.16	0.7414	24.38	0.6199	24.58	0.5842	21.89	0.5825	23.56	0.7387	24.11	0.6529
LapSRN [74]	×8	812K	26.15	0.7380	24.35	0.6200	24.54	0.5860	21.81	0.5810	23.39	0.7350	24.04	0.6520
MSRN [44]	×8	6,226K	26.59	0.7254	24.88	0.5961	24.70	0.5610	22.37	0.6077	24.30	0.7701	24.56	0.6520
SENext [14]	×8	97K	26.87	0.7415	**25.73**	0.6200	**26.79**	0.5847	21.90	0.5829	23.96	0.7389	25.05	0.6536
EDSR [22]	×8	43,000K	26.96	0.7762	24.91	0.6420	24.81	0.5985	22.51	0.6221	24.69	0.7841	24.74	0.6824
AWSRN [78]	×8	2,348K	26.97	0.7747	24.96	0.6414	24.80	0.5967	22.45	0.6174	24.69	0.7842	24.77	0.6828
DBPN [2]	×8	10,000K	26.96	0.7762	24.91	0.6420	24.81	0.5985	22.51	0.6221	24.60	0.7732	24.75	0.6824
MFCC [16]	×8	2,453K	27.07	0.7762	25.01	0.6412	24.84	0.5980	22.54	0.6196	24.63	0.7791	24.81	0.6828
RDN [73]	×8	21,900K	27.21	0.7840	25.13	0.6480	24.88	0.6010	22.73	0.6312	25.14	0.7897	25.02	0.6907
RCAN [19]	×8	16,000K	27.31	0.7878	25.23	0.6511	24.98	0.6058	23.00	0.6452	25.24	0.8029	25.15	0.6985
HAN [50]	×8	3,230K	27.33	0.7884	25.24	0.6510	24.98	0.6059	22.98	0.6437	25.20	0.8011	25.14	0.6980
DANS (Ours)	×8	2,456K	**27.58**	**0.7908**	25.32	**0.6516**	25.12	**0.6066**	**23.18**	**0.6458**	**25.38**	**0.8036**	**25.31**	**0.6996**

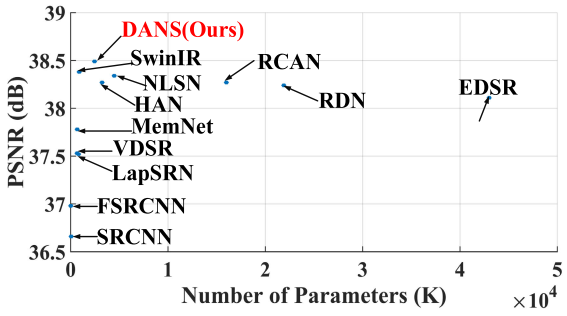

FIGURE 6. Comparison of model parameters versus PSNR on the image dataset of Set5 [67] with enlargement factor ×2.

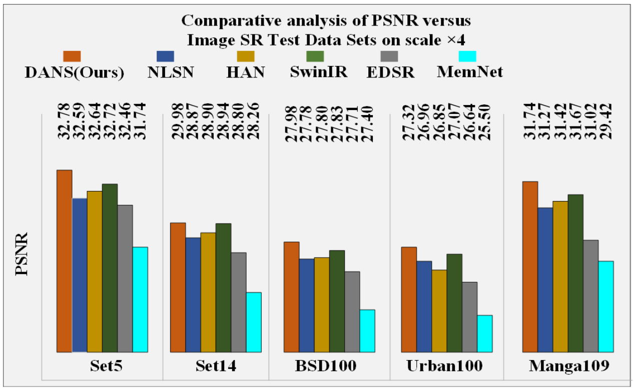

FIGURE 7. Comparative analysis of PSNR versus Image SR Test Data Sets on enlargement factor ×4.

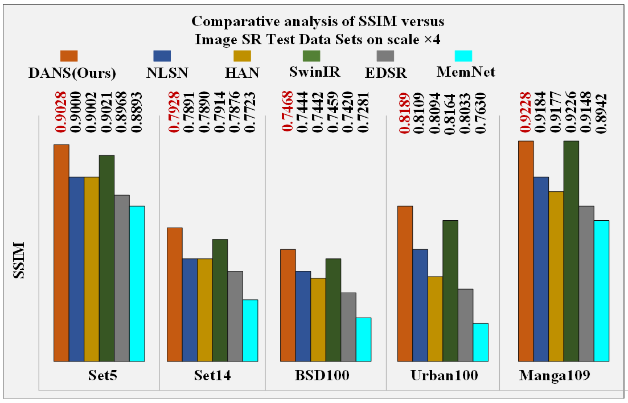

FIGURE 8. Comparative analysis of SSIM versus Image SR Test Data Sets on enlargement factor ×4.

MemNet [43] on enlargement factor of ×4 and RCAN [19], HAN [50], RDN [73], DBPN [2] and AWSRN [78] on enlargement factor of ×8. The best improvement of our model for PSNR is shown in Set5 [67] and Manga109 [71] datasets at scale factor ×4, and for SSIM is shown in Manga109 [71] datasets at scale factor ×8.

E. QUANTITATIVE ANALYSIS OF PSNR VERSUS EXECUTION TIME

This section shows the performance of DANS regarding PSNR versus execution time, as shown in Figure 11. We used NVIDIA GeForce GTX 2080ti GPU with 24GB memory to evaluate the state-of-the-art methods. For evaluation, GitHub codes provided by the research community have been used. Figure 11 shows the trade-off between PSNR versus execution time on Set5 [67] scale factor ×4. Our proposed method gains the highest PSNR of 32.78 and is faster than five state-of-the-art methods (RCAN [19], EDSR [22], NLSN [30], MemNet [43], and VDSR [35]) except the SRCNN [17], FSRCNN [72] and LapSRN [74]. Furthermore, as seen in Figure 12, our proposed DANS has lesser computation costs regarding floating point operations (FLOPs).

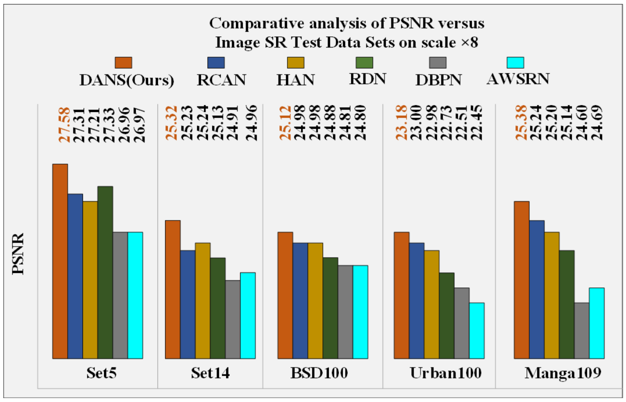

FIGURE 9. Comparative analysis of PSNR versus Image SR Test Data Sets on enlargement factor ×8.

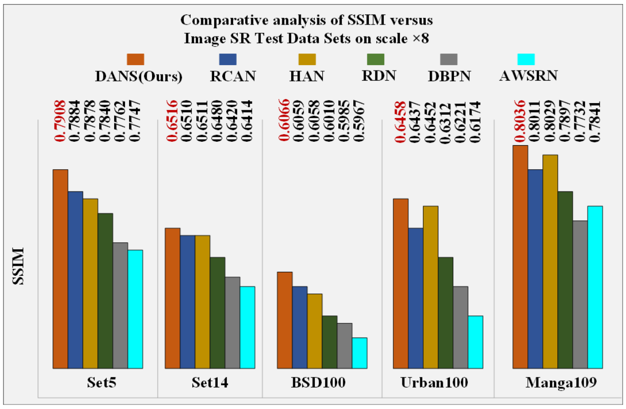

FIGURE 10. Comparative analysis of SSIM versus Image SR Test Data Sets on enlargement factor ×8.

F. PERFORMANCE ANALYSIS OF OUR MODEL DURING TRAINING FROM THE EXISTING SR METHOD

In this subsection, we discuss performance evaluation during the training of our model. The average PSNR (dB) per epoch is shown in Figure 13, demonstrating that our model shows better training convergence than an existing SR model. The training hyperparameters are kept the same for a fair comparison. This evaluation is calculated for the training of enlargement factor of ×4 on the DIV2K [66] Dataset.

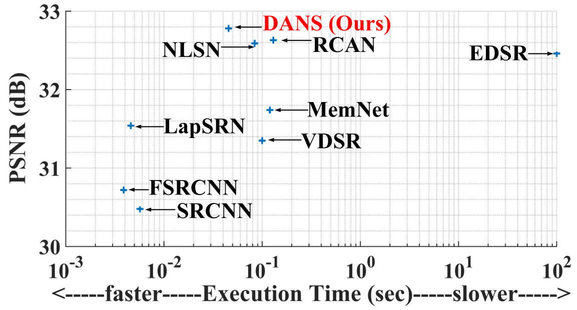

FIGURE 11. Quantitative assessment of running time versus PSNR on Set5 [67] with scale factor ×4.

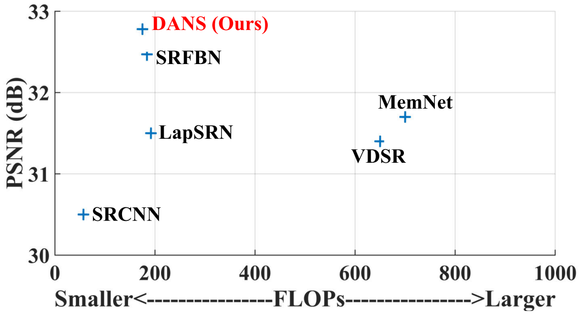

FIGURE 12. Quantitative assessment of GFLOPs versus PSNR on Set5 [67] scale factor ×4.

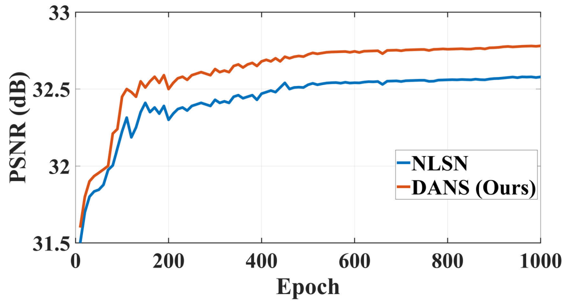

FIGURE 13. Quantitative consideration of PSNR for an existing SR method on a scale factor ×4 on DIV2K [66] Dataset.

G. LOSS ANALYSIS OF OUR MODEL DURING TRAINING FROM THE EXISTING SR METHOD

This section describes the graph of the average training loss of our model. Figure 14 shows that our model shows better loss convergence than an existing SR model NLSN [30]. Our proposed DANS shows better and smoother convergence in an average loss in Figure 14 and average PSNR (dB) in Figure 13 during training. It is noted that average PSNR and loss are calculated for the training of scale ×4 on the DIV2K [66] dataset.

H. SPACE COMPLEXITY ANALYSIS

An amount of memory space is required for an algorithm to operate on a computer. The space complexity of a deep CNN model represents how much memory it requires to run. The performance of a proposed algorithm establishes a balance between space and time (the complexity of space and time). In this section, we evaluate the space complexity on the publicly available Set5 test dataset with an enlargement factor of ×2. The space complexity of Set5 images such as baby, bird, butterfly, head, and woman are calculated with five state-of-the-art methods, including HR and LR images. Figure 15 shows that our proposed method has less space complexity (storage memory) than existing state-of-the-art methods.

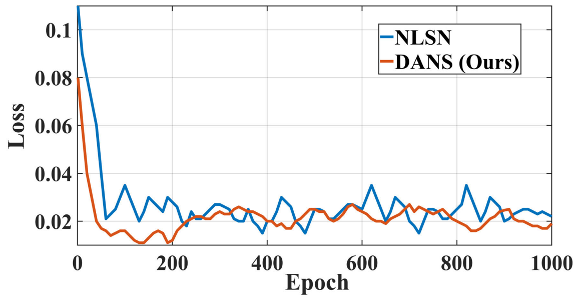

FIGURE 14. Loss versus epoch curve enlargement factor ×4 on DIV2K [66] dataset.

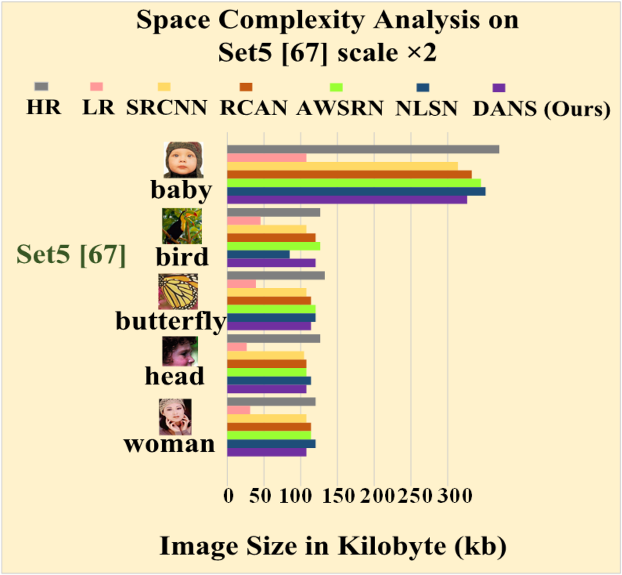

FIGURE 15. Space complexity analysis for Set5 [67] Dataset images on scale factor ×2.

I. TIME COMPLEXITY ANALYSIS

The time required for completing each epoch during the training of a deep learning model demonstrates its complexity in terms of time. This is known as the time complexity of the DL model. In Figure 16, we show the time each epoch takes for 100 training epochs for an existing state-of-the-art method NLSN [30], and our proposed DANS. The curve shows a significant gap, indicating that our proposed DANS takes less training time for each epoch. Hence DANS show lesser time complexity than the baseline model.

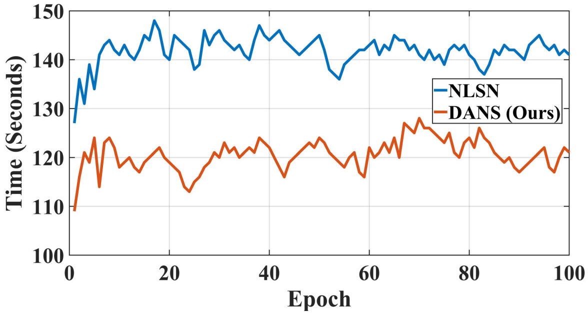

FIGURE 16. Time complexity (Training time per epoch) for 100 epochs on DIV2K [66] Dataset images on scale factor ×4.

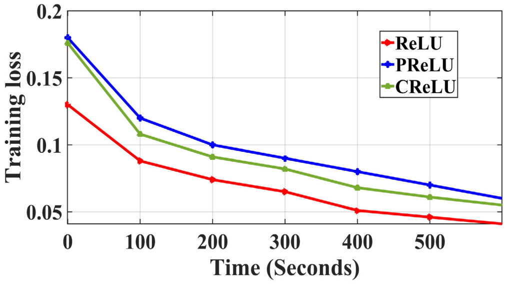

FIGURE 17. Convergence rate analysis of proposed DANS for activactions on Set5 [67] for scale factor ×4.

J. CONVERGENCE RATE ANALYSIS

In deep learning, a model's convergence rate refers to how rapidly it can arrive at the best solution during training. It focuses on comprehending the learning algorithm's speed, time, and effectiveness for minimizing the loss during training. Figure 17 shows the convergence analysis for loss during training of the proposed DANS model for Set5 [67] scale ×4. Loss convergence of the model is calculated with different activations ReLU, PReLU and CReLU. As seen in Figure 17, ReLU shows a lesser loss convergence rate than CReLU and PreLU for our proposed DANS model.

Time complexity, space complexity, and convergence rate analyses are crucial to fully grasp the practical viability of an image super-resolution deep learning approach. These include the trade-off between computational demands, achievable performance, hardware accelerators (such as GPUs) for effective training, inference, and the available computing resources. When evaluating the practicality of an image super-resolution deep learning approach, examining the time complexity, space complexity, and convergence rate is helpful.

K. PERCEPTUAL QUALITY COMPARISON

Figure 18, Figure 19, Figure 20, Figure 21, Figure 22, Figure 23, and Figure 24 presents the visual quality of up-sampling factors ×4 and ×8 for image SR test datasets, including Set5 [67], Set14 [68], BSD100 [69], Urban100 [70] and Manga109 [71]. Blurry results are observed on up-sampling factor ×8 for Bicubic, LapSRN [74], and MSRN [44]. Even though improving an image for an enlargement factor of ×8 is difficult, our proposed DANS favorably reconstructs the fine contextual detail and constructively subdues the artifacts because of our encoder-decoder approach combined with Non-local sparse attention (NLSN) [30]. Non-local sparse attention (NLSN) [30] excels at capturing long-range dependencies and modeling global contextual details. Combining it with the encoder-decoder structure of U-Net [4] leads to increased discriminative [66] and informative feature representation both locally and globally.

For enlargement factor ×4, we used the barbara image from Set14 [68] dataset, Img_148026 from BSD100 [69], Img_092 from Urban100 [70], and TaiyouNiSmash image from Manga109 [71] dataset. For enlargement factor ×8, we used Img_253027 from BSD100 [69], Img_060 from Urban100 [70], and Hamlet image from Manga109 [71] dataset. Our proposed DANS shows visually pleasing patches and better quantitative metrics (PSNR/SSIM) as compared to other state-of-the-art methods such as Bicubic, MSRN [44], EDSR [22], AWSRN [78], RCAN [19], and NLSN [30] for ×4 and Bicubic, LapSRN [74], MSRN [44], DBPN [2], AWSRN [78], and RCAN [19] for enlargement factor of ×8.

L. ABLATION STUDY

In this section, we conduct controlled experiments to analyze our proposed model. The proposed model has six inception and 5 Non-local Sparse Attention (NLSA) blocks. We insert up-sampling and down-sampling with NLSA [30] block to give it an encoder-decoder [4] structure. Finally, we introduce a skip connection to the model to make it lightweight. The ablation study on the proposed model has been done in the following ways: (1) By changing activation functions to ReLU, PReLU, and CReLU in the inception block of the network design, (2) by changing the number of attention rounds on the Local Sensitivity Hashing technique inside the NLSA [30] block, (3) by noise degradation analysis of our proposed model using different activation at noise level 15 and degradation kernel is set to be 0.5, (4) by comparison analysis with traditional denoising techniques, and (5) by calculating Visual Information Fidelity (VIF). We conduct these experiments to check their effect on the performance of the proposed model.

1) ABLATION STUDY WITH DIFFERENT ACTIVATION FUNCTIONS

He et al. [65] first introduced Parametric Rectified Linear Unit (PReLU) and Concatenated Rectified Linear Unit (CReLU). PReLU and CReLU have been introduced as extensions of ReLU to address its limitations. PReLU allows for more flexibility in the network by introducing a learnable parameter to provide a small negative slope for negative inputs. On the other hand, CReLU concatenates ReLU with ReLU-like functions with negative slopes, providing a more complex activation function. Henceforth, we attempt to change the activation function inside the inception block

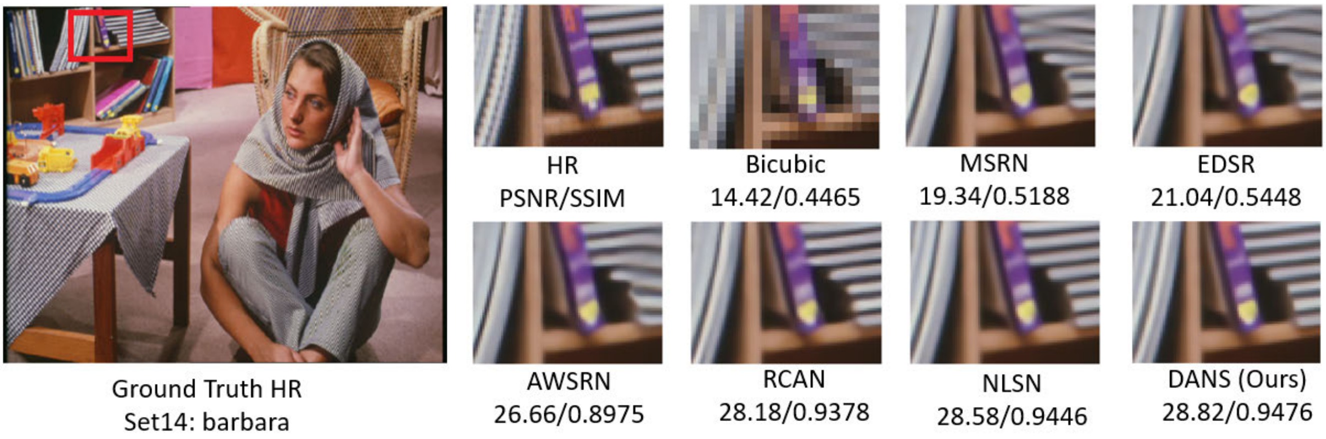

FIGURE 18. Qualitative improvement of Barbara image from Set14 [68] dataset on a scale factor of ×4.

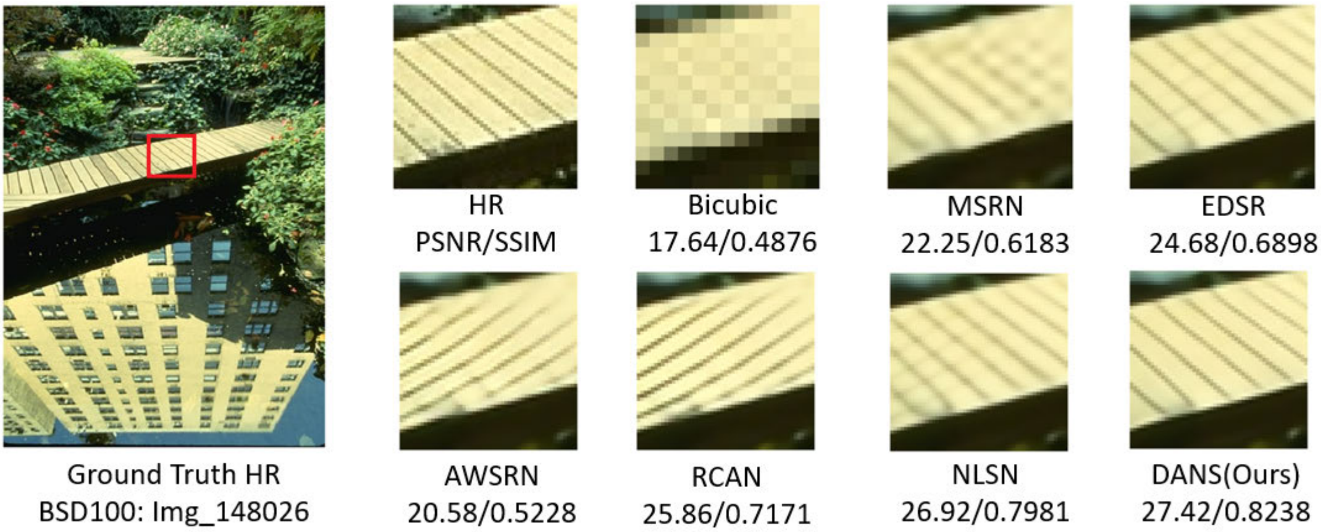

FIGURE 19. Qualitative improvement of image Img_148026 from BSD100 [69] dataset on a scale factor ×4.

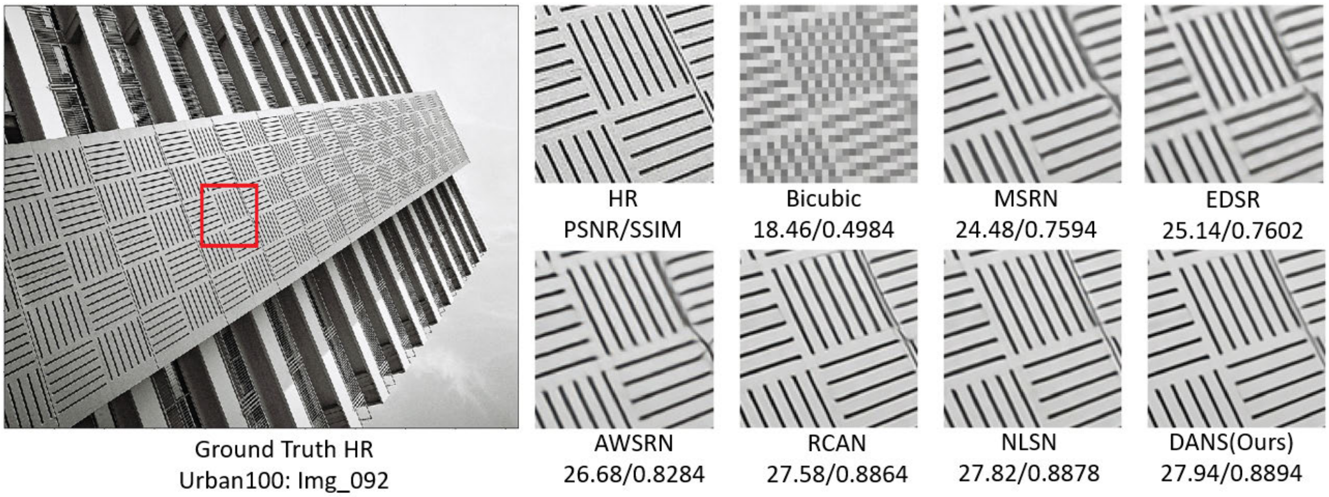

FIGURE 20. Qualitative improvement of image Img_092 from Urban100 [70] dataset on a scale factor ×4.

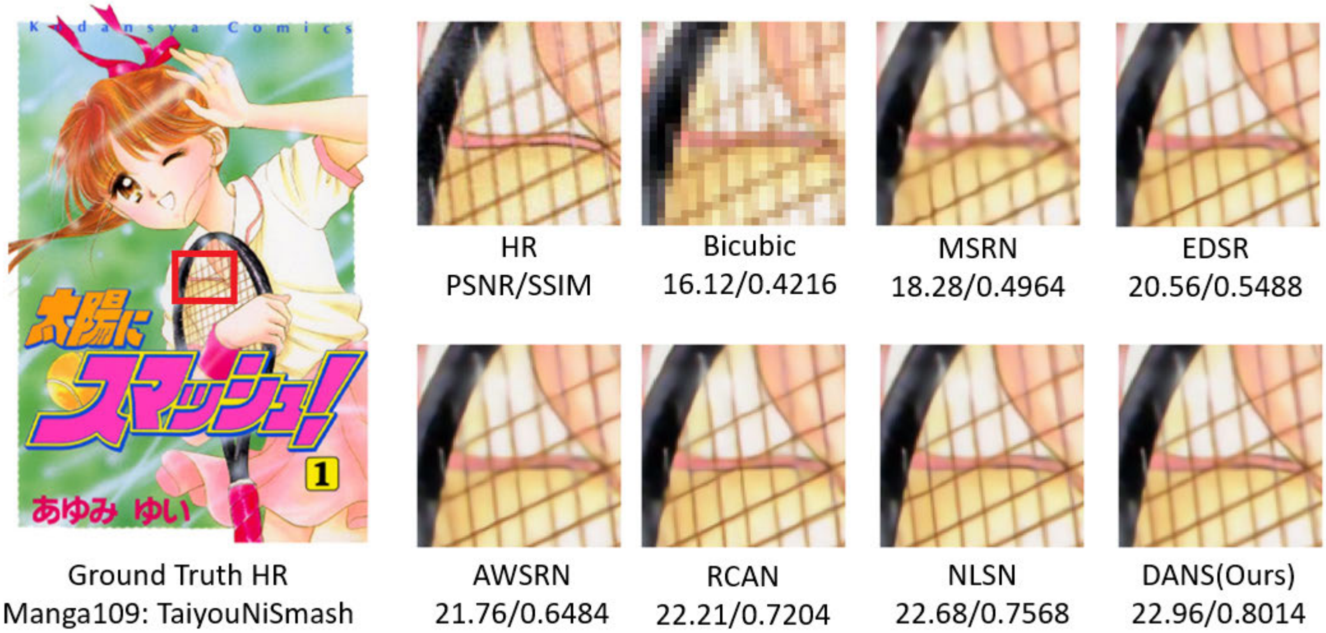

FIGURE 21. Qualitative improvement of TaiyouNiSmash image from Manga109 [71] image dataset on a scale factor ×4.

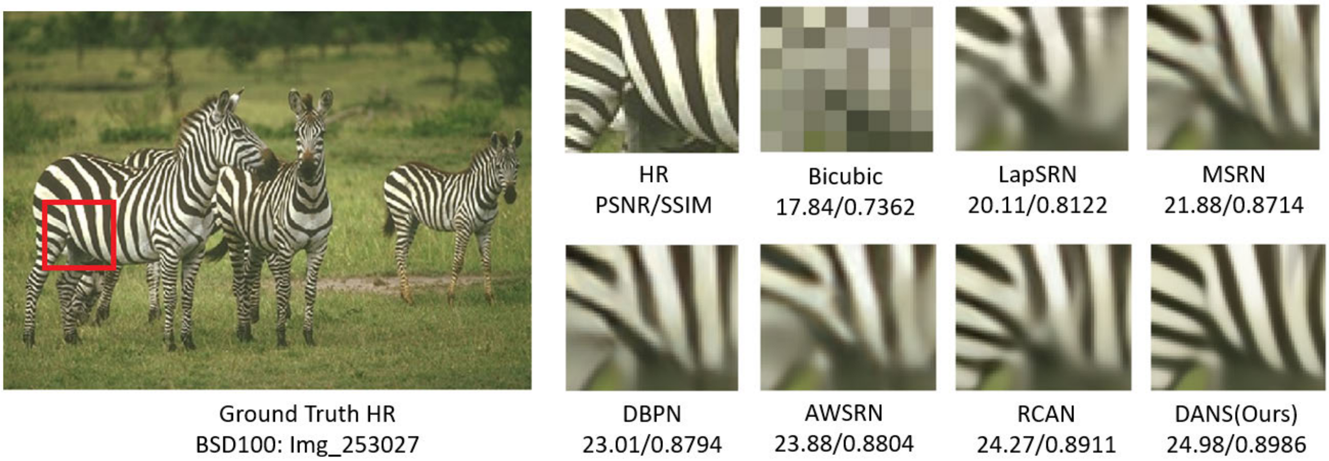

FIGURE 22. Qualitative improvement of Img_253027 image from BSD100 [69] image dataset on scale factor of ×8.

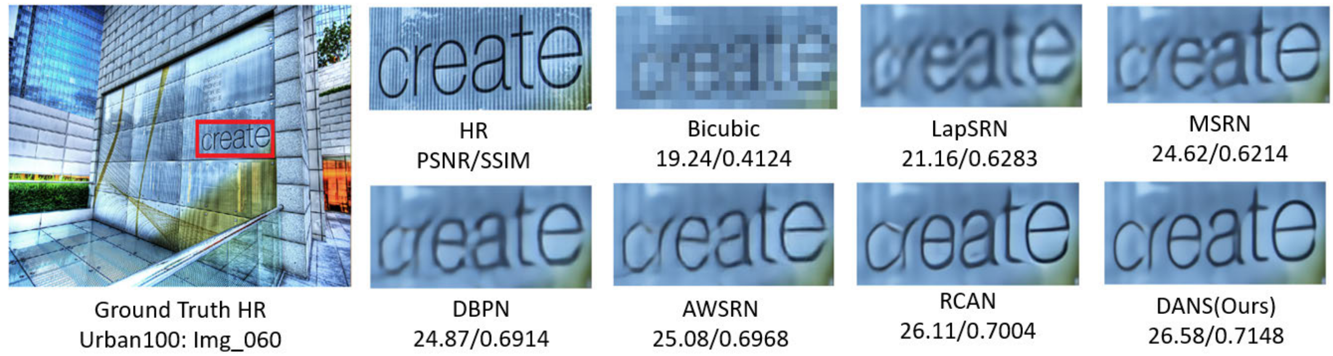

FIGURE 23. Qualitative improvement of Img_060 image from Urban100 [70] image dataset on scale factor of ×8.

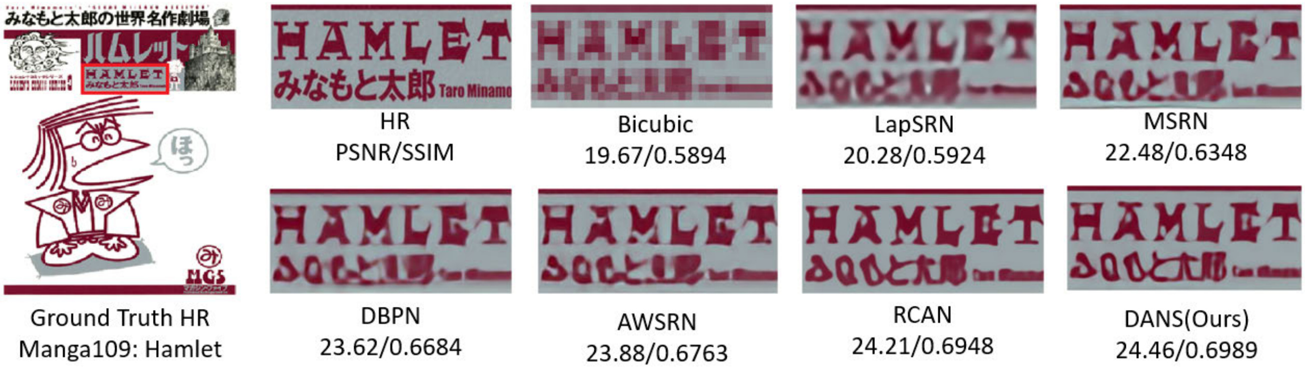

FIGURE 24. Qualitative improvement of Hamlet image from Manga109 [71] image dataset on the scale factor ×8.

TABLE 3. Ablation study of different activations in inception blocks, including ReLU, CReLU, and PReLU. The quantitative value of average PSNR calculated on Set5 [67] enlargement factor ×4 on 50 epochs. The best quantitative value has been recorded as bold with Red color. The second-best quantitative value is shown in blue color with an underline.

Activation Function	Inception Block			Average PSNR
ReLU	✓	×	×	**32.56**
PReLU	×	✓	×	32.50
CReLU	×	×	✓	32.39

in the proposed model to check its effect on performance. Figure 25 shows the activation function has been changed inside the inception block of the model to see the impact on performance.

The ReLU activation function is known for its simplicity, computational efficiency, and better sparsity. Even though PReLU and CReLU are advancements over ReLU, they cannot provide better sparsity than ReLU. This has been demonstrated by evaluating PSNR for different designs of models having inception blocks with ReLU, PReLU, and CReLU in Table 3. The red color demonstrates the best value, and the blue underlined demonstrates the second-best value. It can be observed in Figure 26 that ReLU shows better convergence as compared to PReLU and CReLU when it comes to dealing with sparsity.

Table 3 shows that ReLU gives better PSNR on an average calculated on Set5 [67] for enlargement factor ×4 compared to PReLU and CReLU. Since ReLU is computationally efficient, as shown in Figure 26, the network with ReLU activation converges better. It helps the network train faster than those with PReLU and CReLU.

2) ABLATION STUDY WITH DIFFERENT ROUNDS OF ATTENTION IN NON-LOCAL SPARSE ATTENTION BLOCKS

Since the Local Sensitivity Hashing (LSH) [80] technique used in NLSA [30] works on improving the robustness of the model, its computational cost is furthermore reduced by adjusting the attention rounds r. Table 4 shows the result of the model trained and evaluated for different attention rounds. This result indicates that increasing the number of hashing rounds either at training or evaluation improves the accuracy of the super-resolution model. As a result, the best performance in terms of PSNR is achieved at the highest value, i.e., 8 for hashing rounds. Figure 27 shows the effect of

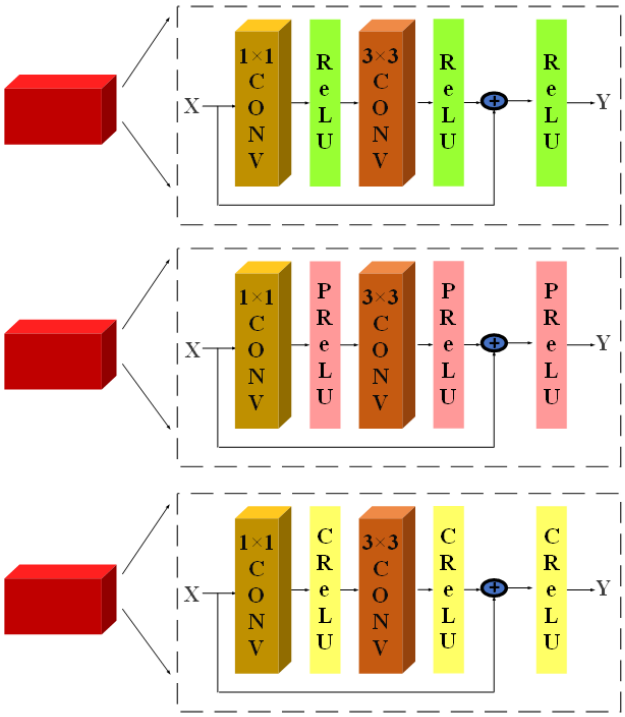

FIGURE 25. Inception Block with different activations.

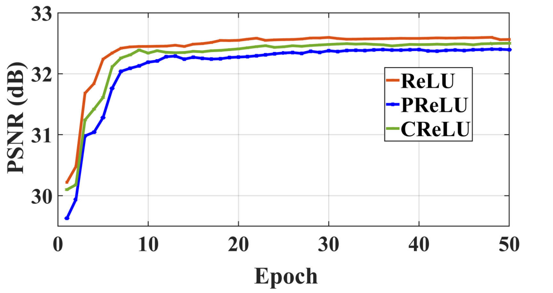

FIGURE 26. Performance assessment of PSNR (dB) versus training epoch for different activations on enlargement factor ×4 on Set5 [67].

TABLE 4. Effect of different attention rounds on performance for enlargement factor of ×2. The best quantitative value has been recorded as bold with Red color. The second best quantitative value is shown in blue color with an underline.

Non-Local attention rounds	Set5 [67]		Set14 [68]		BSD100 [69]		Urban100 [70]		Manga109 [71]	
	NLSN	DANS	NLSN	DANS	NLSN	DANS	NLSN	DANS	NLSN	DANS
r=1	37.87	37.92	33.76	33.81	31.38	31.44	32.26	32.28	38.36	38.40
r=2	37.90	37.96	33.80	33.84	31.42	31.48	32.30	32.33	38.37	38.42
r=4	37.92	<u>37.99</u>	33.86	<u>33.90</u>	31.47	<u>31.52</u>	32.32	<u>32.36</u>	38.40	<u>38.44</u>
r=8	37.93	**38.08**	33.89	**33.94**	31.50	**31.53**	32.35	**32.38**	38.43	**38.46**

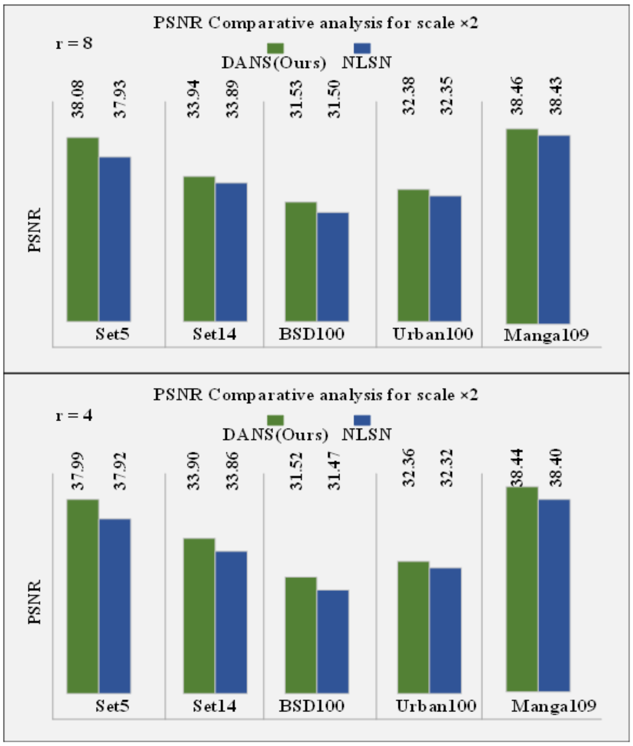

FIGURE 27. Performance assessment of PSNR (dB) versus image dataset for different rounds of activations on enlargement factor ×2.

attention rounds for r = 4 and r = 8 on the benchmark dataset for an enlargement factor of × 2.

3) NOISE DEGRADATION ANALYSIS OF THE PROPOSED MODEL

Performance evaluation using different activations, i.e., ReLU, PReLU, and CReLU, has been demonstrated in Table 5 on Set5 [67] for enlargement factors ×2 and ×4, respectively. Gaussian noise has been added to the image keeping the noise level at 15, and the degradation kernel is set to 0.5. Table 5 shows that our proposed model performs better for noise-degraded images and can be used as a denoiser [79].

Table 5 shows that ReLU performs better in terms of PSNR and SSIM for noise-degraded images than PReLU and CReLU activation. Since ReLU promotes sparsity in activations, it shows the best performance. The red color indicates the best performance, and blue with an underline indicates the second-best. The high PSNR and SSIM show the robustness of ReLU against noise degradation in the proposed model.

TABLE 5. Quantitative evaluation of different activations on noise degradation of images on Set5 [67] for enlargement factors of ×2 and ×4. The best quantitative value has been recorded as bold with Red color. The second best quantitative value is shown in blue color with an underline.

Methods	Factor	Set5 [67]	
		PSNR	SSIM
DANS (with PReLU)	×2	28.61	0.6804
DANS (with CReLU)	×2	<u>28.86</u>	<u>0.6838</u>
DANS (with ReLU)	×2	**28.94**	**0.6842**
DANS (with PReLU)	×4	23.34	0.4694
DANS (with CReLU)	×4	<u>23.52</u>	<u>0.4712</u>
DANS (with ReLU)	×4	**23.68**	**0.4742**

4) COMPARISON ANALYSIS WITH TRADITIONAL DENOISING TECHNIQUES

In this section, we show the comparison of our proposed DANS model on Set14 [68] Dataset on scale ×4 with classical denoising methods such as Block Matching and 3D Filtering (BM3D) [81], Weighted Nuclear Norm Minimization with Application to Image Denoising (WNNM) [82], Denoising Convolutional Neural Network (DnCNN) [38], Fast and Flexible Solution for CNN-Based Image Denoising (FFDNet) [83] and Nonlocally centralized sparse representation for image restoration (NCSR) [84]. Performance comparison in terms of PSNR is shown using Gaussian noise keeping noise level (σ), i.e., $\sigma = 5$, $\sigma = 10$ and $\sigma = 15$ in Table 6. It can be observed from Table 6 that our proposed DANS model shows better performance at noise level $\sigma = 5$.

5) VISUAL INFORMATION FIDELITY (VIF)

Visual Information Fidelity (VIF) is a statistic used to evaluate how well a processed or compressed image maintains the integrity of the original image. It measures the degree of visual perception and visual information preservation similarity. Luminance, contrast, structure, and texture are just a few variables that VIF considers when assessing the fidelity of the processed image. It provides a thorough evaluation by considering both local and global visual data.

Table 7 shows the VIF calculation for our proposed method. The Visual Information Fidelity is calculated in Luma (Y) Chroma Blue (Cb) Chroma Red (Cr) (YCbCr) color space since the Human Visual System (HVS) is very sensitive to high-frequency details in the Luma component. Hence, the luma component in YCbCr color space shows better detection of textual information. The VIF values are

TABLE 6. Performance evaluation for noise degradation of images on Set14 [68] for scale factor ×4. The best quantitative value has been recorded as bold with Red color. The second best quantitative value is shown in blue color with an underline.

Methods / Noise Level	Factor	BM3D [81]	WNNM [82]	DnCNN [38]	FFDNet [83]	NCSR [84]	DANS (Our)
$\sigma = 5$	×4	29.72	29.88	29.93	30.06	30.13	**30.34**
$\sigma = 10$	×4	28.64	28.82	29.46	29.38	29.42	29.68
$\sigma = 15$	×4	27.81	27.94	28.24	28.44	28.68	28.76

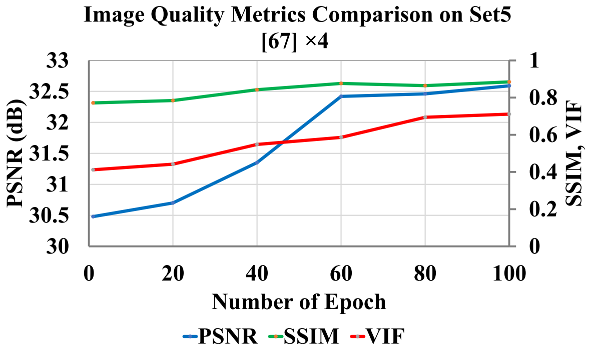

FIGURE 28. PSNR (dB), SSIM and VIF assessment on Set5 [67] image test dataset for enlargement factor ×4.

TABLE 7. Quantitative evaluation of our proposed DANS model in terms of VIFon Set5 [67] test dataset with enlargement factor ×4. The best quantitative value has been recorded as bold with Red color.

Methods	Factor	VIF
SRCNN [17]	×4	0.561
FSRCNN [72]	×4	0.613
VDSR [35]	×4	0.683
DANS (Our)	×4	**0.711**

calculated for Set5 [67] on scale factor ×4. Figure 28 compares of PSNR, SSIM, and VIF on Set5 [67] image test dataset for scale factor ×4.

V. DISCUSSION

In the article being discussed, non-local sparse attention (NLSN) and the U-Net framework are combined to propose a novel method for SISR. The model's effectiveness, computational cost, and perceptual quality of the reconstructed high-resolution images are all improved by the adding the inception block, skip connections, and Depth-wise Separable Convolution. Results from comparative evaluations and carefully monitored trials show that the suggested DANS model efficiently enhances the quantitative and perceptual quality of the reconstructed images. The model can effectively capture both local and global information. It is because of the inclusion of non-local sparse attention, which improves reconstruction results when dealing with different up-sampling factors.

The five benchmark test datasets analysis further demonstrates the DANS model's success in terms of quantitative and qualitative performance measures. The suggested method performs better than current approaches, demonstrating its capacity to produce high-quality reconstructed images under various up-sampling factors. These findings indicate the DANS model's potential as a useful tool for various image improvement applications. One noteworthy feature is the proposed method's reduced computational expense, attained using skip connections and parameter-effective Depth-wise Separable Convolution. The model becomes more computationally efficient without compromising performance by lowering the number of parameters to mitigate the vanishing gradient problem during the training. This is very useful in actual situations with constrained processing resources.

Even though the results presented are encouraging, it is vital to recognize some of the study's limitations. The performance of the DANS model was evaluated using benchmark datasets; however, how well it performs in tough or real-world circumstances is still unknown. Testing the model on a wider variety of photos, especially those with complicated properties or diverse visual content, should be a part of future studies. The authors also provide an overview of their future research goals, which include improving the model for real-time and video super-resolution applications. Using the DANS model in such situations is anticipated to offer insightful information and significantly advance image super-resolution.

Deep Attention Network for SISR (DANS) is a variant of an encoder-decoder network and attention mechanism used in Deep Learning models. This model introduces the concept of the encoder-decoder framework with skip connections and sparsity to reduce computational requirements while improving its performance and encouraging parallelism and scalability in deep learning models. As computational complexity reduces, parallelizing computations over many processors or devices becomes easier. The training and inference procedures may be accelerated, making it easier to deal with larger models or process data in real-time.

DANS can give deep learning models interpretability and explainability, which help to understand the areas that contribute the most to the model's predictions by seeing the attention maps the model produces. These features can help with model debugging, highlighting key components, or understanding how the model makes decisions. It can be applied to various domains and tasks within deep learning, such as natural language processing, computer vision, machine translation, image recognition, and video understanding.

VI. CONCLUSION AND FUTURE WORK

This paper presents a novel fusion of non-local sparse attention NLSN and U-Net framework for single image super-resolution to improve the efficiency and reduce the computational cost of the model. Furthermore, integration of the inception block in the network makes it flexible for the computation of diversified data of the image, which in turn improves the perceptual quality of the image. Additionally, skip connections introduced into the network reduce the parameters used in the network model and ultimately help reduce the model's computational cost. It further helps to reduce the vanishing gradient problem during training. The generalization and convergence of our model are enhanced using Depth-wise Separable Convolution. Depth-wise separable convolution also helps to enhance parameter, computation efficiency and encourages channel-wise feature reuse. Using non-local sparse attention NLSN in the encoder-decoder framework results in efficient local and global modeling. Furthermore, the relative assessment and controlled experiment show the DANS model's effectiveness in improving the reconstructed HR image's quantitative and perceptual quality. A thorough analysis of five benchmark datasets revealed that the proposed DANS model also enhances reconstruction outcomes in terms of quantitative and qualitative norms for up-sampling factors of ×2, ×3, ×4, and ×8. In the future, we will advance our model to introduce real-time and video super-resolution applications under complex scenarios.

REFERENCES

[1] J. Gupta, "Learning rich features from RGB-D images for object detection and segmentation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 345–360.

[2] 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.

[3] Z. Zhang, "Facial landmark detection by deep multi-task learning," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 94–108.

[4] Ronneberger, O., P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[5] A. Basharat, A. Gritai, and M. Shah, "Learning object motion patterns for anomaly detection and improved object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8, doi: 10.1109/CVPR.2008.4587510.

[6] L. Zhang, H. Zhang, H. Shen, and P. Li, "A super-resolution reconstruction algorithm for surveillance images," *Signal Process.*, vol. 90, no. 3, pp. 848–859, Mar. 2010.

[7] H. Greenspan, "Super-resolution in medical imaging," *Comput. J.*, vol. 52, no. 1, pp. 43–63, Jan. 2009.

[8] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "FSRNet: End-to-end learning face super-resolution with facial priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2492–2501.

[9] C. Wang, J. Jiang, Z. Zhong, and X. Liu, "Propagating facial prior knowledge for multitask learning in face super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7317–7331, Nov. 2022.

[10] S. Shakya, S. Kumar, and M. Goswami, "Deep learning algorithm for satellite imaging based cyclone detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 827–839, 2020.

[11] H. Zhang, Z. Yang, L. Zhang, and H. Shen, "Super-resolution reconstruction for multi-angle remote sensing images considering resolution differences," *Remote Sens.*, vol. 6, no. 1, pp. 637–657, Jan. 2014.

[12] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, Oct. 2021.

[13] N. Barla. *A Gentle Introduction to Deep Learning—The ELI5 Way*. Accessed: Aug. 7, 2023. [Online]. Available: https://www.v7labs.com/blog/deep-learning-guide

[14] W. Muhammad, S. Aramvith, and T. Onoye, "SENext: Squeeze-and-ExcitationNext for single image super-resolution," *IEEE Access*, vol. 11, pp. 45989–46003, 2023.

[15] A. Hajian and S. Aramvith, "Fusion objective function on progressive super-resolution network," *J. Sensor Actuator Netw.*, vol. 12, no. 2, p. 26, Mar. 2023.

[16] W. Ruangsang, S. Aramvith, and T. Onoye, "Multi-FusNet of cross channel network for image super-resolution," *IEEE Access*, vol. 11, pp. 56287–56299, 2023.

[17] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[18] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.

[19] Y. Zhang, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 286–301.

[20] J.-H. Kim, J.-H. Choi, M. Cheon, and J.-S. Lee, "MAMNet: Multi-path adaptive modulation network for image super-resolution," 2018, *arXiv:1811.12043*.

[21] X. Yu and F. Porikli, "Face hallucination with tiny unaligned images by transformative discriminative neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 1–19.

[22] 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. 1132–1140.

[23] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-SRNet: A wavelet-based CNN for multi-scale face super resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1698–1706.

[24] T. Guo, H. S. Mousavi, T. H. Vu, and V. Monga, "Deep wavelet prediction for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1100–1109.

[25] Y. Fan, "Neural sparse representation for image restoration," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 15394–15404.

[26] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[27] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 370–378.

[28] K. In Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, Jun. 2010.

[29] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.

[30] 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. 3516–3525.

[31] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.

[32] X. Mao, C. Shen and Y.-B. Yang, "Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2802–2810.

[33] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4809–4817.

[34] J. Yamanaka, S. Kuwashima, and T. Kurita, "Fast and accurate image super-resolution by deep CNN with skip connection and network in network," in *Proc. ICNIP*, 2017, pp. 217–225.

[35] 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.

[36] 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.

[37] J. Jiang, L. Zheng, F. Luo, and Z. Zhang, "RedNet: Residual encoder–decoder network for indoor RGB-D semantic segmentation," 2018, *arXiv:1806.01054*.

[38] J. Chen and F. Li, "Denoising convolutional neural network with mask for salt and pepper noise," *IET Image Process.*, vol. 13, no. 13, pp. 2604–2613, Nov. 2019.

[39] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2024–2032.

[40] N. Ahn, B. Kang, and K.-A. Sohn, "Fast accurate and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 252–268.

[41] J.-H. Choi, J.-H. Kim, M. Cheon, and J.-S. Lee, "Lightweight and efficient image super-resolution with block state-based recursive network," 2018, *arXiv:1811.12546*.

[42] 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. 2790–2798.

[43] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4539–4547.

[44] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 517–532.

[45] 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.

[46] X. Yang, H. Mei, J. Zhang, K. Xu, B. Yin, Q. Zhang, and X. Wei, "DRFN: Deep recurrent fusion network for single-image super-resolution with large factors," *IEEE Trans. Multimedia*, vol. 21, no. 2, pp. 328–337, Feb. 2019.

[47] K.-W. Hung, Z. Zhang, and J. Jiang, "Real-time image super-resolution using recursive depthwise separable convolution network," *IEEE Access*, vol. 7, pp. 99804–99816, 2019.

[48] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 723–731.

[49] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T. S. Huang, and H. Shi, "Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5689–5698.

[50] B. Niu, W. Wen, W. Ren, X. Zhang, L. Yang, S. Wang, K. Zhang, X. Cao, and H. Shen, "Single image super-resolution via a holistic attention network," in *Proc. Eur. Conf. Comput. Vis.* Glasgow, U.K.: Springer, Aug. 2020, pp. 191–207.

[51] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11057–11066.

[52] S. Anwar and N. Barnes, "Densely residual Laplacian super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1192–1204, Mar. 2022.

[53] X. Zhang, H. Zeng, S. Guo, and L. Zhang, "Efficient long-range attention network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Tel Aviv, Israel: Springer, Oct. 2022, pp. 649–667.

[54] E. Lu and X. Hu, "Image super-resolution via channel attention and spatial attention," *Appl. Intell.*, vol. 52, no. 2, pp. 2260–2268, Jan. 2022, doi: 10.1007/s10489-021-02464-6.

[55] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2356–2365.

[56] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," 2019, *arXiv:1903.10082*.

[57] Z. Wang, D. Liu, Z. Wang, J. Yang, and T. Huang, "Deeply improved sparse coding for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 370–378.

[58] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2004, pp. 1.

[59] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[60] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[61] R. Wang, M. Gong, and D. Tao, "Receptive field size versus model depth for single image super-resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 1669–1682, 2020.

[62] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.

[63] M. Su, S. Lai, Z. Chai, X. Wei, and Y. Liu, "Hierarchical recursive network for single image super resolution," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops (ICMEW)*, Jul. 2019, pp. 595–598.

[64] S.-J. Park, "SRFeat: Single image super-resolution with feature discrimination," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 439–455.

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

[66] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1122–1131.

[67] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L.-A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, p. 135.

[68] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.*, Jun. 2012, pp. 711–730.

[69] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vision (ICCV)*, Jul. 2001, pp. 416–423.

[70] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5197–5206.

[71] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, "Sketch-based Manga retrieval using manga109 dataset," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21811–21838, Oct. 2017.

[72] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 391–407.

[73] 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.

[74] 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. 5835–5843.

[75] W. Ai, X. Tu, S. Cheng, and M. Xie, "Single image super-resolution via residual neuron attention networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 1586–1590.

[76] 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. 3862–3871.

[77] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1833–1844.

[78] C. Wang, Z. Li, and J. Shi, "Lightweight image super-resolution with adaptive weighted learning network," 2019, *arXiv:1904.02358*.

[79] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2808–2817.

[80] A. Frome and J. Malik, "Object recognition using locality-sensitive hashing of shape contexts," in *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice*. Cambridge, MA, USA: MIT Press, 2005, pp. 221–247.

[81] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2007, p. 313.

[82] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.
[83] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.
[84] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

JAGRATI TALREJA (Graduate Student Member, IEEE) received the B.Tech. degree in electronics and communication engineering networks from the Pranveer Singh Institute of Technology, Kanpur, Uttar Pradesh, India, in 2019. She is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, Chulalongkorn University Bangkok, Thailand. Her research interests include electrical engineering, neural networks, and machine learning, specifically in deep learning image super-resolution.

SUPAVADEE ARAMVITH (Senior Member, IEEE) received the B.S. degree (Hons.) in computer science from Mahidol University, in 1993, and the M.S. and Ph.D. degrees in electrical engineering from the University of Washington, Seattle, USA, in 1996 and 2001, respectively. In June 2001, she joined Chulalongkorn University, where she is currently an Associate Professor with the Department of Electrical Engineering, specializing in video technology. She has successfully advised 32 bachelor's, 27 master's, and ten Ph.D. graduates. She has published over 130 articles in international conference proceedings and journals with four international book chapters. She has rich project management experiences as a project leader and a former technical committee chairs to the Thailand government bodies in telecommunications and ICT. She is very active in the international arena with the leadership positions in the international network, such as JICA Project for AUN/SEEDNet, and the professional organizations, such as the IEEE, IEICE, APSIPA, and ITU.

TAKAO ONOYE (Senior Member, IEEE) received the B.E. and M.E. degrees in electronic engineering and the Dr.Eng. degree in information systems engineering from Osaka University, Osaka, Japan, in 1991, 1993, and 1997, respectively. He was an Associate Professor with the Department of Communications and Computer Engineering, Kyoto University, Kyoto, Japan. Since 2003, he has been a Professor with the Department of Information Systems Engineering, Osaka University. He has published over 200 research papers in VLSI design and multimedia signal processing in reputed journals and proceedings of international conferences. His research interests include media-centric low-power architecture and its SoC implementation. He has been served as a member of the CAS Society Board of Governors, since 2008. He is also a member of IEICE, IPSJ, and ITE-J.