IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 May 2023, accepted 25 May 2023, date of publication 2 June 2023, date of current version 9 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3282571

RESEARCH ARTICLE

Multi-FusNet of Cross Channel Network for Image Super-Resolution

WATCHARA RUANGSANG[1], (Graduate Student Member, IEEE), SUPAVADEE ARAMVITH[1], (Senior Member, IEEE), AND TAKAO ONOYE[2], (Senior Member, IEEE)

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

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

This research is funded by The 90th Anniversary of Chulalongkorn University Fund (Ratchadaphiseksomphot Endowment Fund), 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 Thailand Science research and Innovation Fund Chulalongkorn University (IND66210019).

ABSTRACT Image Super-resolution (SR) has gained considerable attention in artificial intelligence (AI) research and image-based applications. Recent deep learning-based SR models have demonstrated remarkable accuracy and perceptual quality in the resulting images. However, the computational cost and model parameters are the most challenging limitations in real-world applications. Additionally, designing an efficient and lightweight SR algorithm to improve the perceptual quality of the SR images is a critical issue. According to these considerations, we propose a Multi-FusNet of Cross Channel Network (MFCC) network by modeling a multipath residual network, named multi-RG, with cross-filtering fusion. Additionally, a pixel shuffling fusion technique is used to fuse low-level features into the up-sampled features of the multi-RG. The experimental results show the comparison of the proposed MFCC to the state-of-the-art SR models. The proposed method significantly reduces the number of network parameters (8.4 times compared to RCAN) while preserving the visual quality of the result and achieving the best PSNR value compared to the other state-of-the-art methods.

INDEX TERMS Residual network, super-resolution, multipath residual.

I. INTRODUCTION

The image Super-resolution (SR) technique aims to reconstruct a high-resolution (HR) image from the low-resolution (LR) input. The SR is considered an ill-posed problem. Therefore, many SR techniques have been developed for producing HR images. However, the perceptual quality and execution time of the SR model are two critical factors in designing an effective and robust SR model for real-world applications. A real-world problem in utilizing closed-circuit television (CCTV) [1] for intelligent monitoring [2], such as traffic congestion and accidents [3], home management security [4], face recognition [5], and social identification [6], has appeared because of the low-resolution (LR) images captured by CCTV cameras [7].

Many research topics, such as feature preservation in video coding [8], [9] and super-resolution [10], [11], [12], have been published to solve the low-resolution problem. Deep Convolutional Neural Networks (CNN) have recently become a powerful model for solving the ill-posed SR problem [13]. Dong et al. [14], [15] proposed a super-resolution algorithm using a three-layer deep convolutional neural network (SRCNN) that was linearly stacked together. In this shallow and straightforward architecture, the first layer is designed to extract the features from the LR input, the second layer is used for non-linear mapping from low-dimensional to high-dimensional features, and the final layer is responsible for aggregating the feature maps of the earlier layer to the final HR result. This SR network is trained using an

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

end-to-end approach to minimize the Mean Squared Error (MSE) between the ground truth (GT) image and the reconstructed (SR) image. Following the SRCNN model, other models, such as Very Deep Super-resolution [16] (VDSR) and Deeply Recursive Convolutional Network (DRCN), achieved significant improvements by increasing the depth of their CNN architectures.

Following the DRCN and VDSR models, various architectures based on different image reconstruction approaches, including the residual network [12], [17], [18], [19], [20], [21], [22], [23], [24], Recursive Network [25], [26], [27], progressive reconstruction [28], dense connection networks [29], [30], multi-branch architectures [31], [32], and attention-based mechanisms [33], [34], have been proposed to improve the results of SR models by increasing the data flow in the deep learning architecture.

Inspired by the Residual Network architecture [17] (ResNet), some SR models have attempted to design deeper architectures. Lim et al. proposed a deep SR architecture called Enhanced Deep Residual Network [35] (EDSR) based on the residual concept. This SR model stacks residual blocks to design deeper networks (almost 165 convolutional layers) and achieves considerable improvement compared with earlier SR models. However, training such a deep trainable network, such as EDSR [35] is challenging, and the large number of parameters in a deep network is a significant obstacle to fast execution in real-world applications and hardware implementations.

On the other hand, some recent CNN-based models [23], [25], [28], [31], [32] utilize a multipath network architecture to solve the limitations of deep networks. In this approach, rather than using a single-path network, the multipath network operates in parallel. Based on this concept, the depth of the network decreases, whereas the performance of the SR model increases. Based on the multipath structure, Hui et al. proposed an Information Distillation Network (IDN) by designing cascaded network paths operating in parallel. This implies that the layers of the network do not require waiting for the calculations of previous layers. Additionally, the different types of extracted features from each path are mixed, which improves the SR model's operation time and the perceptual quality of the result. Although the SR models based on the multipath approach achieved acceptable performance and execution time, their results achieved low PSNR and SSIM values.

Additionally, the low-level feature-sharing approach is used in some SR models [20], [21], [24], [33] to enhance the low-frequency information flow in the SR network structure. This low-level feature-sharing approach attempts to transfer the low-level features of the early CNN layer, to the latest layers and fuse the low and high-level features. Hence, this technique improves reconstruction quality by enhancing the sharpness of the SR image. Motivated by this technique, SR models such as the Feedback Network [20] (SRFBN), Adaptive Weighted Super-Resolution Network [24] (AWSRN), SelNet [33], and Attentive Auxiliary Features [21] (A2F) utilize the feature sharing approach. This idea demonstrates more effective enhancement of the lightweight architecture.

To address these issues, we propose an efficient lightweight SR model for image enlargement, with the following main contributions:

1) Proposing the multi-depth cross-channel network to obtain local pixel attention features from low-resolution images.

2) Investigating the doubling stage of the residual identity connection to retrieve merged features represented by low-level features.

3) Exploring low-level feature sharing to fuse low-level information with unsampled features enhances the model's reconstruction ability.

The remainder of this paper is organized as follows. Section II presents a review of related works. Section III explains the proposed method. Section IV discusses the experimental results, and Section V presents the conclusions.

II. RELATED WORKS

Numerous image SR models have been studied in computer vision. Single Image Super-Resolution (SISR) [13] extensively utilizes deep Convolutional Neural Networks (CNN). CNN uses the LR image as the input and reconstructs the SR image. Excellent networks [36] based on CNN can be categorized into six groups: Linear Networks [14], [15], [27], [37], [38], Residual Networks [16], [25], [28], [29], [30], [34], [35], Recursive Network [25], [26], [27], Progressive Reconstructions Network [28], Multi-Branch Nets [31], [32], and Attention Based Networks [33], [34].

A. LINEAR NETWORKS

Linear networks have simple architectures with a single path that linearly stacks the convolution layers to allow the information to flow in the network. These linear models are further categorized into early and late up-sampling designs. The early up-sampling design was inspired by SRCNN [14], [15], which operates based on up-samples of the LR image in the first stage and, then reconstructs the HR image. This model uses an early up-sampling design. This model [16] obtains a larger region of contextual information to improve the results. Additionally, they increased the depth of the VDSR network by stacking 20 convolutional layers. The FSRCNN model [37] uses the late up-sampling framework. This framework performs an up-sampling operation toward the end of the network to improve the computational cost. The architecture of the FSRCNN contains four convolution layers and one deconvolution layer at the end of the model to produce the upscaled image. The output result was reconstructed by combining residual learning with bicubic interpolation. This model was designed for rapid processing in real-time applications.

B. RESIDUAL NETWORKS

The Residual Network introduces skip connections into the neural network architecture. This idea attempts to focus on high-frequency information in a very deep network. The residual network concept is further categorized into single-stage and multi-stage residual networks. Inspired by the Residual Network architecture [17] (ResNet), the Enhanced Deep Super-Resolution (EDSR) was modified by removing Batch Normalization (BN) layers and ReLU activation. This model decreases the number of parameters while simultaneously improving the performance of the SR model. This method [18] proposes a cascading mechanism to improve the performance of the model and weight trade-offs. CARN uses the ResNet [17] architecture, and a cascading mechanism at the local and global levels is used to include features from all layers. The multi-scale residual network (MSRN) [19] has been proposed to address feature utilization and the adaptation of arbitrary scaling factor problems. This model can fuse the image features at different scales. This is the first multi-scale module based on a residual structure, which is very easy to train. The model shows superior performance compared to other state-of-the-art models on various benchmark datasets.

The Adaptive Weight SR Network (AWSRN) [24] was designed to resolve the heavy computational cost problem. This model consists of Local Fusion Blocks (LFB) designed with residual learning-based embryonic adaptive voluminous residual units (ARWU) and a local residual fusion entity (LRFU). Apart from the LFB, it also contains an adaptive weight multi-scale module (AWMS) to enhance the reconstruction layer. The AWMS is an important contributor to the design of lightweight network structures. The Inception Network [12] proposed an asymmetric residual architecture to reduce the number of parameters. They were inspired by the Inception network concept, Muhammad et al. [12] proposed Multi-Scale Inception Based Super-Resolution (MSISRD). In this SR model, the short and long feature information is directly extracted using a locally residual asymmetric convolutional block and an inception-based asymmetric convolutional block architecture by the model.

The A2F model [21] utilizes additional features and a channel attention mechanism to improve the model's performance while reducing the weight of model. This study has proven that having fewer auxiliary features results in less high-frequency information and consequently decreases the accuracy of the SR model. In addition, the A2F model outperforms other state-of-the-art models on all scales and has a faster execution time. The FALSR method [23], Fast and Lightweight SR with Neural Architecture Search contributes to maximizing the balance between the image restoration and the models' weight. In the proposed model, an elastic search approach is used, which is based on a hybrid controller at both the micro and macro levels.

SFFN [39] proposed an efficient feature fusion block, along with lightweight and shallow residual blocks. This model efficiently fuses the features of different blocks and improves the model's performance and execution time. Moreover, they introduced an attention mechanism for reinforcing the useful cross-layer features of each channel. This lightweight SR model outperformed other state-of-the-art methods.

C. RECURSIVE NETWORKS

This design focuses on breaking the more significant SR problem into a simple smaller entity. The contributions of this network design are as follows. The DRCN model [25] is based on a recursive CNN containing almost 16 layers of recursion. This method improves the performance without increasing the parameters. The only drawback, that is the learning difficulty of this method, can be solved by recursive supervision or skipping connections. This model reduces the weight of the network by introducing recursion and skip connections. This reduced the training difficulty of the model.

D. PROGRESSIVE RECONSTRUCTION NETWORKS

The progressive reconstruction approach suggests a progressive network in the SR area to improve the SR results with larger scaling factors. Another benefit of the progressive approach is that the predictions are made in multiple sub steps. The Laplacian Pyramid Framework [28] (LapSRN) uses progressive up-sampling to reconstruct fast and accurate residuals of HR images. Some important limitations of previous state-of-the-art models, such as high computational cost, blurry images, and learning difficulty, were overcome by the LapSRN model because of the progressive approach in the architecture of the SR model. This method uses cascaded CNNs to predict the sub-band surplus in a rough-to-fine texture. The LapSRN method has 27 layers overall, takes LR as input, uses residual learning, and performs progressive reconstruction with a char bonnier loss function. The proposed method constructs high-quality HR images faster than other state-of-the-art methods. It also helps to remove the blurred kernels. The only problem with this model is that it does not hallucinate fine details over large scales.

E. MULTI-BRANCH NETS

The multi-branch architecture proposed a successful model for increasing the information flows between the network layers. To obtain diverse information and features from multiple scales, a multi-branch network architecture was used. This architecture obtains complementary information and merges them for better HR reconstruction. Information Multi-distillation Network [32] (IMDN) proposed a lightweight multi-branch architecture to solve the learning complexity limitation caused by the numerous convolutional layers. In this multi-branch architecture, a distillation block is designed to extract hierarchical features and combine them using cascaded Information Multi-distillation Blocks (IMDB). The IMDB blocks are formed from distillation blocks, and the fusion module extracts features at a coarse level, retaining partial information. It then aggregates them

using the channel attention mechanism to improve the refined information (edges, corners, and textures).

F. ATTENTION BASED NETWORKS

To improve the performance of learning-based SR models, an attention-based technique was designed as an enhancement module to pay attention to specific varying features. A deep CNN with Selection Units [33] (SelNet) is motivated by CNN's linear mapping techniques. The Rectified Linear Unit (RLU) was used for linearly mapping the LR images, inspiring the creation of a non-linear unit known as the Selection Unit (SU). Because SU combines identity mapping and a sigmoid switching function, it has better control over the data passed through than ReLU [40]. The results show that the proposed network has a much lower computational complexity and outperforms the baseline model with only ReLU and state-of-the-art SR methods. Very Deep Residual Channel Attention Networks [34] (RCAN) utilizes the Residual Channel Attention architecture for their SR model. They designed residual in residual (RIR) architecture block consisting of Residual Group (RG) and a Residual Channel Attention (RCA) block. The RG structure uses a short skip connection as a residual component, whereas the RCA utilizes a long skip connection to target the LR feature components. Additionally, channel attention (CA) was introduced to affect the feature rescaling channel. Although RCAN produces high-quality SR results, the complexity of the architecture increases processing time [13], [36]. Channel Split Image Super-Resolution (CSISR) [41] improves the learning capability of the SR model with a novel channel attention mechanism. The proposed attention mechanism utilizes a combination of global average and standard deviation pooling along with the non-linear mapping layers. In addition, CSISR demonstrated an efficient and lightweight architecture to enhance computational complexity problems and outperformed other state-of-the-art models. Based on the dynamic residual attention (DRA) approach [42], the dynamic residual self-attention network (DRSAN) proposed a lightweight SR model. The proper weights for each residual path statistical investigation of the input image, and interrelation between residual paths boost the reconstruction capability of this model. Additionally, a residual self-attention (RSA) block was proposed to generate 3-D attention maps without additional parameters. In [43], the Information-Growth Attention Network (IGAN) has introduced a new type of attention mechanism called the ''information-growth attention.'' This attention mechanism focuses on features that have the potential for large information-growth capacity by analyzing the differences between the current features and the previous features within the network. The Context Reasoning Attention Network (CRAN) [44] adoptively adjusts the convolution kernel based on the global context. This model first extracts global context descriptors and, then introduces channel and spatial interactions to produce a context reasoning attention mask. In [45], a second-order attention network (SAN) used a trainable second-order channel attention (SOCA) module to rescale channel-wise features with second-order feature statistics. This approach results in more discriminative representations.

III. MULTI-FusNet OF CROSS CHANNEL NETWORK

The proposed SR network is designed based on a multipath residual architecture that provides a wider network rather than a deeper one, resulting in more efficient and faster execution. The architecture, named Multi-FusNet of Cross Channel Network (MFCC), consists of four main modules: feature extraction, Residual Group (Multi-RG), enlargement, and low-level fusing, as illustrated in Figure 1. The Multi-RG architecture has been designed by integrating the RCAN [34] with a multi-identical residual link. Due to the increase in the number of multiplicities (possible paths from the input to the output layer) in the proposed architecture, the information flow between RG blocks gradually increases, which helps reduce the computational complexity. The first convolution layer is the feature extraction module, which feeds low-level features to the RG blocks and low-level fusion module. The proposed cascading topology in the MFCC network is composed of three different paths, that form a multipath residual configuration.

Each Residual Group (RG) block consists of N stacked Residual Channel Attention (RCA) blocks and a short residual skip connection within the block. The first path consists of two stacked Residual Group blocks, whereas the second path has one Residual Group block cascaded with the first path. The third path of the model bypasses the low-level details of the earlier layer and fuses them with the up-sampled features of multi-RGs, as shown in Figure 1. The proposed SR model begins by feeding the LR input to a convolutional layer. The resulting features are then passed through three different paths, with a kernel size of $k \times k$. To perform image enlargement, we utilize the pixel shuffle technique, which transforms low-level feature maps into different channels and shuffles the features to enlarge them.

Although the RCAN model offers high accuracy, its large number of parameters results in slow execution times which makes its implementation in real-time applications challenging. To create a lightweight architecture, we reduce the number of residual groups in our model and incorporate a multipath residual network architecture. This design improves both accuracy and processing speed, resulting in a more efficient model compared to those that use a non-cascading (deep) architecture. To further improve our model's capability to extract sharp attribute details, we exploit the low-level features of early CNN layers and share them with the up-sampled features of our multipath residual network. Many SR models suffer from over-smooth degradation due to the lack of high-frequency details in the latest layer of CNN, leading to perceptually unpleasant images at large scales. By incorporating a pixel shuffling fusion technique, we can overcome this limitation and produce high-quality results. The following section describes the residual group block, multipath residual configuration, and pixel shuffle fusion method.

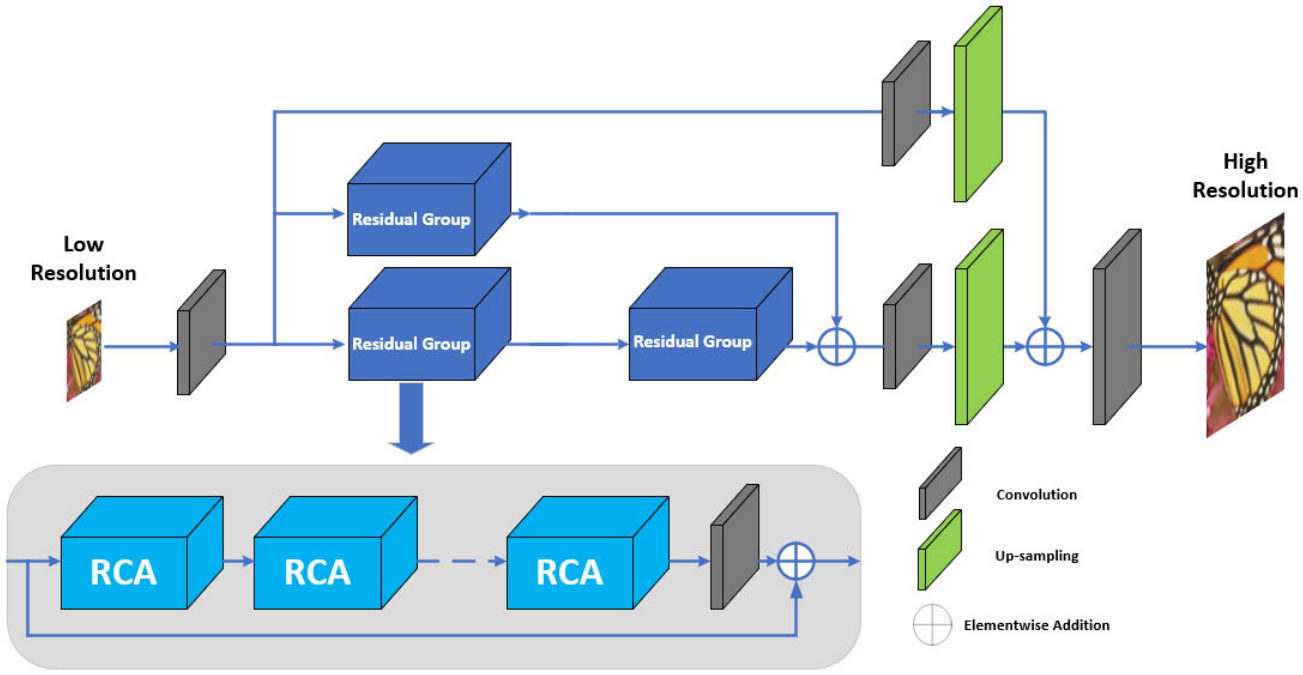

FIGURE 1. Multi-FusNet of cross channel network.

A. RESIDUAL GROUP

In this section, the Residual Group (RG), which is a robust feature extraction of low-resolution, is presented. The proposed Residual Group is constructed by applying identity residual connection at the edge of N sequences of the statistical Channel Attention network in [34], where $N > 0$. The output of RG is denoted by O_{RG}, and defines as in Eq. (1).

$$O_{RG} = I_{RG} + W_{RG}(F_{RCA_N}) \tag{1}$$

where I_{RG} represents RG input. W_{RG} denotes the weight parameter in the Residual Group block, and F_{RCA_N} is the channel-wise feature of the Residual Channel Attention (RCA) block. The channel-wise feature from the RCA can be determined by Eq. (2) and (3).

$$F_{RCA_N} = H^N_{RCA}\left(F_{RCA_{N-1}}\right) \tag{2}$$

$$F_{RCA_0} = H^0_{RCA}\left(I_{RCA}\right) \tag{3}$$

where F_{RCA_N} and $F_{RCA_{N-1}}$ denote the outputs of N^{th} and $(N-1)^{th}$ Residual Channel Attention blocks, respectively. H_{RCA} shows the corresponding operation function of the RCA. F_{RCA_0} is the output of the first Residual Channel Attention block. I_{RG} shows the input of the first Residual Channel Attention block. The Channel Attention (CA) mechanism is a technique that uses the interdependencies among the feature channels. The CA technique leads to more focus on informative features in the SR model and consequently improves the image reconstruction capability of the model. More details on the CA mechanism and the corresponding operation function in the Residual Channel Attention block can be found in RCAN [34].

B. MULTIPATH RESIDUAL

As shown in Figure 1, the combination of Residual Group blocks under the multipath-residual architecture [46] is used in our model. Based on multipath residual evidence [47], a wider residual architecture significantly improves the accuracy and computation speed of the model compared with a deeper residual architecture. These improvements are related to the increasing multiplicity of the wider residual network. Multiplicity implies the number of possible paths from the input layer to the output layer. A sequence of two Residual Group blocks is utilized in the first path of our model, and one Residual Group block is employed in the second path. Eq.4 defines the multipath output of the proposed model.

$$O_{MR} = O^2_{RG} + O^1_{RG} \tag{4}$$

where O^2_{RG} and O^1_{RG} denote the outputs of two Residual Group blocks in the first path and a Residual Group block in the second path, respectively. O_{MR} denotes the output of the proposed multipath residual architecture.

C. PIXEL SHUFFLE FUSION

A low-level feature-sharing approach is employed to improve the sharpness of reconstructed results. Because the low-level features of the early layer contain more high-frequency information, sharing them improves the challenging weakness of the SR model in recovering the sharp attributes of the lines and edges. Simultaneously, it preserves our model against over-smoothing degradation.

Our model utilizes the pixel shuffle fusion approach to bypass the low-frequency features of the early layer of the

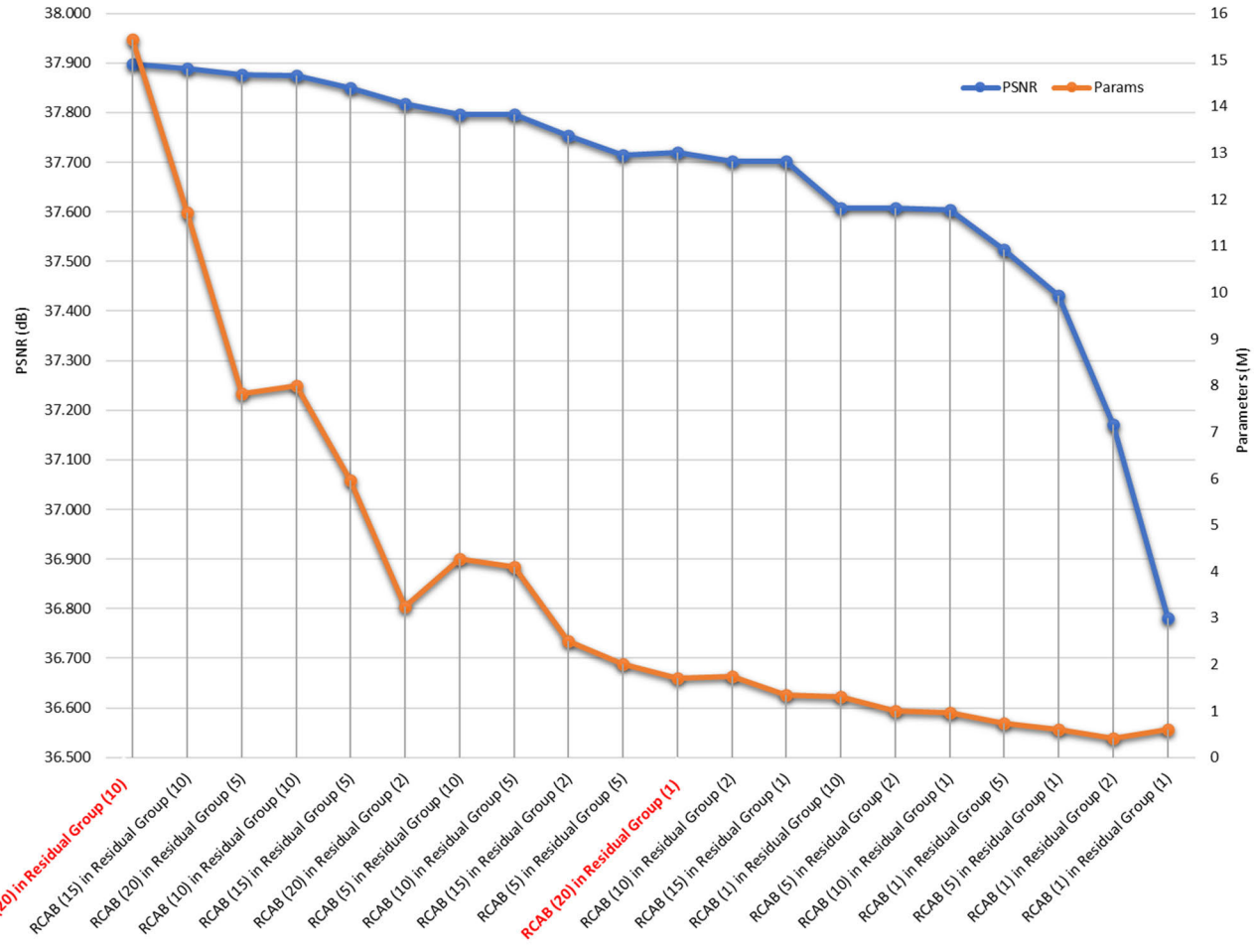

FIGURE 2. Comparison PSNR vs. parameter of number RCA with residual group.

SR network to up-sampled features. The proposed model employs the pixel shuffle [48] to up-sample the image. Based on the feature sharing concept, the features of the early layer are also up-sampled by the pixel shuffle model and fused with the up-sampled features of the multipath residual network, as shown in Figure 1.

$$PS_{Fus} = PS_1\left(W_{MR}\left(O_{MR}\right)\right) + PS_2(W_p\left(F_L\right)) \tag{5}$$

where, PS_1 and PS_2 represent the pixel shuffle up-sampling on the multipath residual network and low-level features sharing, respectively. W_{MR} and W_p represent the convolution operations of the multipath residual output and up-sampled low-level features, respectively. The pixel shuffle can be mathematically expressed as Eq.6.

$$\begin{aligned} PS(U)_{i,j,c} = U_{\left[\frac{i}{a}\right],\left[\frac{j}{a}\right],} C \cdot a \cdot mod\ (i,a) \\ + C \cdot a \cdot mod\ (i,a) + c \end{aligned} \tag{6}$$

where $PS(U)$ is the output, a is the scale factor, i, j, show pixel coordinates. c is the channel position. To modernize the final SR image, the up-sampled high-frequency details of the early layer are fused with the up-sampled features of the multipath residual model.

$$SR = W_{SR}(PS_{Fus}) \tag{7}$$

where PS_{Fus} denotes the fusing pixel shuffle result and, W_{SR} defines the last convolution operation to produce the final SR result. Utilizing the proposed fused approach improves the capability of our model to recover the sharp attributes of images and improves the perceptual quality of results by preventing an over-smoothing problem.

$$L_1\left(\emptyset\right) = \frac{1}{n \times m} \sum_{i=1}^{n} \sum_{j=1}^{m} \|SR(i,j) - y(i,j)\| \tag{8}$$

where SR and y are the result and the reference image, respectively, n and m are parameters related to the training dataset.

IV. EXPERIMENTAL RESULTS

In this section, several experiments were conducted to validate the performance of our model. First, the hyper-parameter settings of the proposed model are explained. The experimental results and analysis are then demonstrated.

TABLE 1. The extension of residual in based line model.

Method			Params	Set5 (×2)	Best Epoch	All Epoch
Residual Channel Attention (RCA)	Residual Group	Upscale		PSNR (dB)		
(20)	(10)	N	15,444,667	38.25	524	600
2 Residual Network of branch (20)	(1)	N	1,713,835	38.06	556	600
(20)	2 Residual Networks	N	1,713,835	38.14	576	600
(20)	2 Residual Networks	Y	1,861,547	38.16	417	600

A. PARAMETER SETTINGS

To train our SR model, we utilized the DIV2K [49] dataset, which included 800 images. For testing our SR model, five standard benchmark datasets including Set5 [49], Set14 [50], B100 [51], Urban100, and Manga109 were considered. For the degradation models, we apply Bicubic Interpolation (BI) was used in our experiments. Y- PSNR, and Y-SSIM were used to evaluate SR model accuracy. For data augmentation, 800 training images were randomly applied with three rotations such as 90, 180, 120 degrees, and horizontal flipping. We extracted 16 in 48 $\times$ 48 test LR color patches to get the input for each training batch. We train our model with an ADAM optimizer, and the parameter values of β_1, β_2, and ϵ are 0.9, 0.999, and 10^{-8}, respectively. Initially, we set the learning rate to 10^{-4} and reduced it to half every 2×10^5 iterations. The models were implemented using PyTorch on a Titan Xp GPU.

B. EXPERIMENTAL RESULTS AND ANALYSIS

Figure 2 shows the performance of the SR model using various RCA and Residual Group (RG) combinations. The results of various SR configurations on the Set5 dataset with a scale factor of two, which were trained under a fixed seed for 50 epochs with the DIV2K dataset, are depicted in this graph. The horizontal axis of the graph displays 20 distinct RCA and Residual Group combinations, arranged in descending order from the highest number of parameters to the lowest. The baseline model (RCAN architecture) comprises 20 RCA configurations and 10 Residual Groups, with 15 million parameters showing the highest performance.

The other combinations differed based on the number of RCA and Residual Groups. Regarding lightweight architecture, the Residual technique provides a wider network architecture and optimizes network parameters and accuracy. Residual Networks [52] optimize residual blocks by expanding the residual information to a broader network architecture. This network architecture improves processing speed by removing some parts of the sequential block from a parallel block. It was also proven that the results showed better accuracy and convergence.

According to the graph, RG has a negligible impact on quality. Therefore, only one RG block is used to reduce the number of parameters. However, if the number of RCAs reduce, there is a considerable decrease in the PSNR value. Regarding our contribution to the design of a lightweight SR architecture, it is important to ensure that the network parameters should be constrained to less than 2 million.

As shown in Figure 2, the PSNR decreases gradually as we move from the baseline architecture to lighter network architectures. Considering the constraint of network parameters for a lightweight model, combining 20 RCA with one RG indicates the optimum trade-off between the minimum number of parameters and the maximum PSNR for our lightweight network. In addition, by checking the effectiveness of RCA and RG, we found that the PSNR drops faster than decreasing the number of RG layers if we reduce the number of RCA layers. Moreover, by reducing the number of RCA layers, the performance of the parameter is reduced less than by decreasing the number of RG layers. Figure 2 presents a comparison between the baseline model and various extensions of the RCA and Residual Group (20 RCA with one RG). The PSNR values, which were calculated using the Set5 dataset with a scale factor of two, are shown. The models were trained for 600 epochs by using a fixed seed. The baseline architecture, with 20 RCA and 10 Residual Group (RG) configuration has a total of 15 million parameters and, achieved the highest PSNR. Among the other lightweight extensions of RCA and RG, a configuration with 20 RCAs and two Residual Network achieved 38.16 PSNR.

It's worth noting that this network configuration (shown in Figure 1) reduced the network parameters by 88%, while only decreasing the accuracy by only 0.09 dB compared to the baseline. Another benefit of this extension is that the 20 RCAs

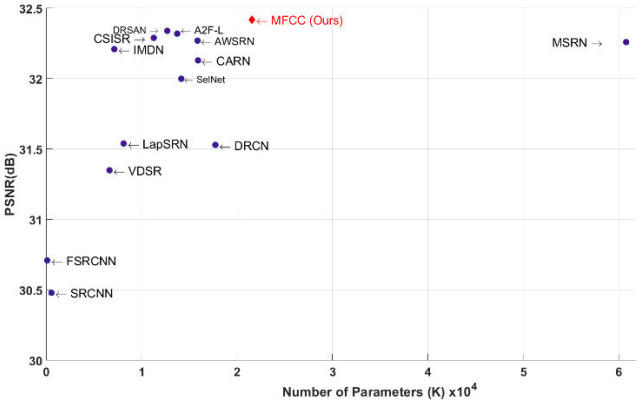

FIGURE 3. Performance and number of parameters evaluated on Set5 dataset at scale ×4.

TABLE 2. The extension of residual in based line model.

Scale	Model	Param	Multi-Adds	Set5		Set14		B100		Urban100		Manga109	
				PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×2	FSRCNN [37]	12K	6.0G	37.00	0.9558	32.63	0.9088	31.53	0.8920	29.88	0.9020	36.67	0.9694
	SRCNN [15]	57K	52.7G	36.66	0.9542	32.42	0.9063	31.36	0.8879	29.50	0.8946	35.74	0.9661
	VDSR [16]	665K	612.6G	37.53	0.9587	33.03	0.9124	31.90	0.8960	30.76	0.9140	37.22	0.9729
	DRCN [25]	1774K	17,974.3G	37.63	0.9588	33.04	0.9118	31.85	0.8942	30.75	0.9133	37.63	0.9723
	LapSRN [28]	813K	29.9G	37.52	0.9590	33.08	0.9130	31.80	0.8950	30.41	0.9100	-	-
	SelNet [33]	974K	225.7G	37.89	0.9598	33.61	0.9160	32.08	0.8984	-	-	-	-
	CARN [18]	1592K	222.8G	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256	-	-
	MSRN [19]	5930K	13654G	38.08	0.9607	33.70	0.9186	32.23	0.9002	32.29	0.9303	38.69	0.9772
	IMDN [32]	694K	158.8G	38.00	0.9605	33.63	0.9177	32.19	0.8996	32.17	0.9283	38.88	0.9774
	AWSRN [24]	1397K	320.5G	38.11	0.9608	33.78	0.9189	32.26	0.9006	32.49	0.9316	38.87	0.9776
	A2F-L [21]	1363K	306.0G	38.09	0.9607	33.78	0.9192	32.23	0.9002	32.46	0.9313	38.95	0.9772
	CSISR [41]	1017K	233.1G	38.06	0.9606	33.73	0.9193	32.24	0.9004	32.27	0.9298	-	-
	FALSR-A [23]	1021K	234.7G	37.82	0.9595	33.55	0.9168	32.12	0.8987	31.93	0.9256	-	-
	DRSAN [42]	1190K	274.6G	38.14	0.9611	33.75	0.9188	32.25	0.9010	32.46	0.9317	-	-
	SFFN [39]	912K	138.7G	38.02	0.9606	33.59	0.9177	32.20	0.9000	32.34	0.9298	-	-
	MFCC (Ours)	1861K	364.5G	38.16	0.9611	33.85	0.9195	32.28	0.9010	32.65	0.9331	39.11	0.9780
×3	FSRCNN [37]	12K	5.0G	33.16	0.9140	29.43	0.8242	28.53	0.7910	26.43	0.8080	30.98	0.9212
	SRCNN [15]	57K	52.7G	32.75	0.9090	29.28	0.8209	28.41	0.7863	26.24	0.7989	30.59	0.9107
	VDSR [16]	665K	612.6G	33.66	0.9213	29.77	0.8314	28.82	0.7976	27.14	0.8279	32.01	0.9310
	DRCN [25]	1774K	17,974.3G	33.82	0.9226	29.76	0.8311	28.80	0.7963	27.15	0.8276	32.31	0.9328
	SelNet [33]	1159K	120.0G	34.27	0.9257	30.30	0.8399	28.97	0.8025	-	-	-	-
	MSRN [19]	6114K	150.6G	34.46	0.9278	30.41	0.8437	29.15	0.8064	28.33	0.8561	33.67	0.9456
	CARN [18]	1592K	118.8G	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493	-	-
	IMDN [32]	703K	71.5G	34.36	0.9270	30.32	0.8417	29.09	0.8046	28.17	0.8519	33.61	0.9445
	AWSRN [24]	1476K	150.6	34.52	0.9281	30.38	0.8426	29.16	0.8069	28.42	0.8580	33.85	0.9463
	A2F-L [21]	1367K	136.3G	34.54	0.9283	30.41	0.8436	29.14	0.8062	28.40	0.8574	33.83	0.9463
	CSISR [41]	1063K	108.5G	34.46	0.9277	30.41	0.8435	29.15	0.8061	28.35	0.8563	-	-
	DRSAN [42]	1290K	133.4G	34.59	0.9286	30.42	0.8443	29.18	0.8079	28.52	0.8593	-	-
	MFCC (Ours)	2230K	374.1G	34.67	0.9294	30.51	0.8456	29.22	0.8080	28.64	0.8616	34.15	0.9478
×4	FSRCNN [37]	12K	4.6G	30.71	0.8657	27.59	0.7535	26.98	0.7150	24.62	0.7280	27.90	0.8517
	SRCNN [15]	57K	52.7G	30.48	0.8628	27.49	0.7503	26.90	0.7101	24.52	0.7221	27.66	0.8505
	VDSR [16]	665K	612.6G	31.35	0.8838	28.01	0.7674	27.29	0.7251	25.18	0.7524	28.83	0.8809
	LapSRN [28]	813K	149.4G	31.54	0.8850	28.19	0.7720	27.32	0.7280	25.21	0.7560	29.09	0.8845
	SelNet [33]	1417K	83.1G	32.00	0.8931	28.49	0.7783	27.44	0.7325	-	-	-	-
	CARN [18]	1592K	90.9G	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837	-	-
	MSRN [19]	6078K	89.6G	32.26	0.8960	28.63	0.7836	27.61	0.7380	26.22	0.7911	30.57	0.9103
	IMDN [32]	715K	71.99G	32.21	0.8948	28.58	0.7811	27.56	0.7353	26.04	0.7838	30.45	0.9075
	AWSRN [24]	1587K	33.7G	32.27	0.8960	28.69	0.7843	27.64	0.7385	26.29	0.7930	30.72	0.9109
	A2F-L [21]	1374K	77.2	32.32	0.8964	28.67	0.7839	27.62	0.7379	26.32	0.7931	30.72	0.9115
	CSISR [41]	1128K	65.13G	32.29	0.8967	28.67	0.7835	27.64	0.7378	26.32	0.7931	-	-
	DRSAN [42]	1270K	88.7G	32.34	0.8960	28.65	0.7841	27.63	0.7390	26.33	0.7936	-	-
	SFFN [39]	923K	34.6G	32.23	0.8950	28.58	0.7813	27.56	0.7813	26.15	0.7877	-	-
	MFCC (Ours)	2157K	395.2G	32.42	0.8973	28.73	0.7849	27.67	0.7399	26.48	0.7977	30.98	0.9131
×8	FSRCNN [37]	12K	4.6G	25.42	0.6440	25.34	0.5482	24.21	0.5112	21.32	0.5090	22.39	0.6357
	SRCNN [15]	57K	52.7G	25.34	0.6471	23.86	0.5443	24.14	0.5043	21.29	0.5133	22.46	0.6606
	VDSR [16]	665K	612.6G	25.73	0.6743	23.20	0.5110	24.34	0.5169	21.48	0.5229	22.73	0.6688
	DRCN [25]	1,774K	17,974G	25.93	0.6740	24.25	0.5510	24.49	0.5170	21.71	0.5230	23.20	0.6690
	LapSRN [28]	813K	--	26.15	0.7028	24.45	0.5792	24.54	0.5293	21.81	0.5555	23.39	0.7068
	MSRN [19]	6,226K	89.6G	26.59	0.7254	24.88	0.5961	24.70	0.5410	22.37	0.5977	24.28	0.7517
	AWSRN[24]	2,348K	33.7G	26.97	0.7747	24.99	0.6414	24.80	0.5967	22.45	0.6174	24.60	0.7782
	MFCC (Ours)	2,453K	518.2G	27.07	0.7762	25.01	0.6412	24.84	0.5980	22.54	0.6196	24.63	0.7791

and two Residual Networks with upscaling converge faster than the other extensions, reaching convergence at the 417th epoch. It is noticeable that the upscale parameter in this table refers to the up sampled low-level feature sharing, as demonstrated in Figure 1. Our experiment shows that the edges of the SR images recovered more clearly using this configuration. The proposed network architecture reduces the number of parameters by up to eight times compared to RCAN, with comparable quality.

Table 2 presents comparisons of our proposed model with other state-of-the-art models on scaling factors of ×2, ×3, ×4, and ×8. The evaluation was performed using five benchmark datasets: Set5, Set14, B100, Urban100, and Manga109. Moreover, the comparison includes the number of parameters and multi-adds of each model. The best performance is denoted by red numbers and the second-best performance is represented by blue numbers. Based on Table 2, it can be seen that our model outperforms other in terms of

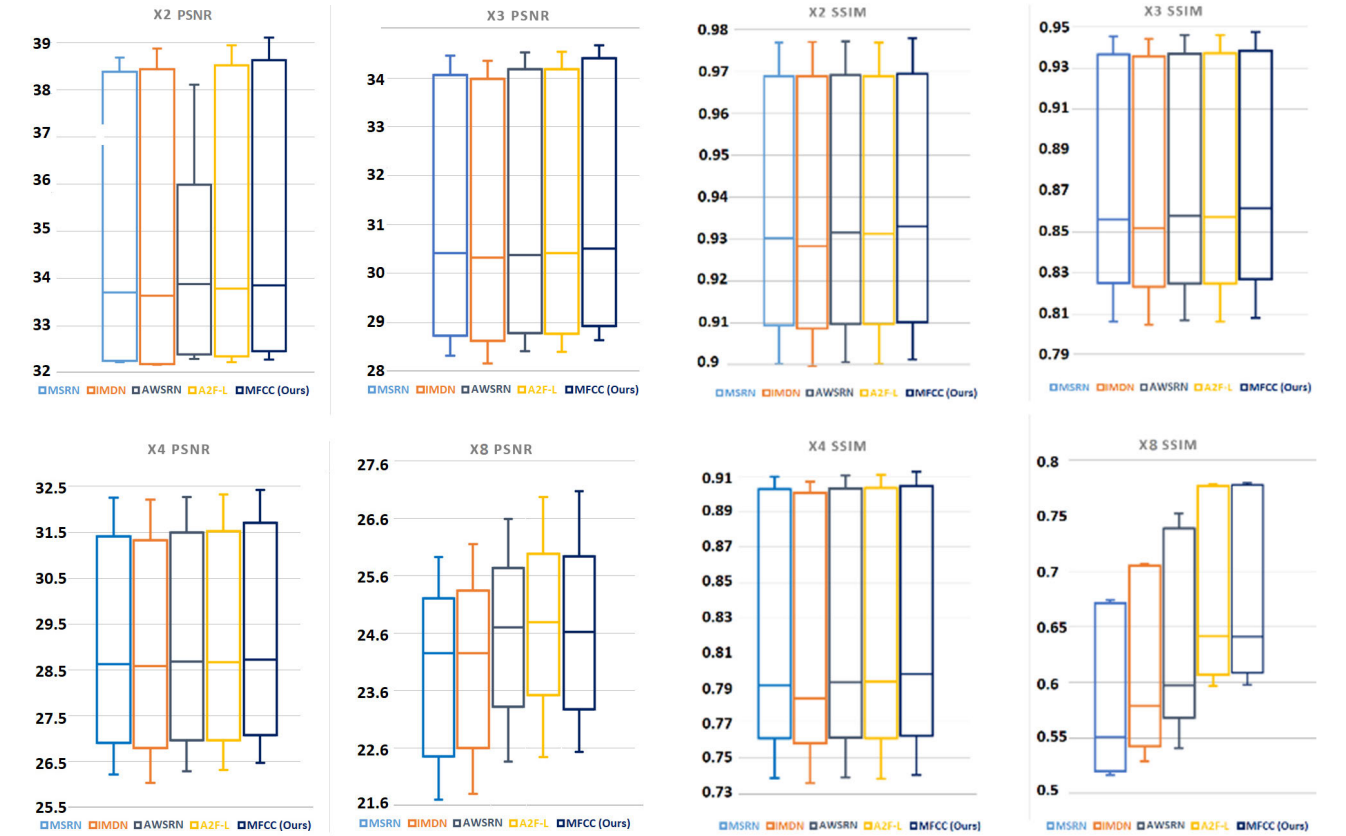

FIGURE 4. Statistical analysis of average of PSNR and SSIM on All dataset.

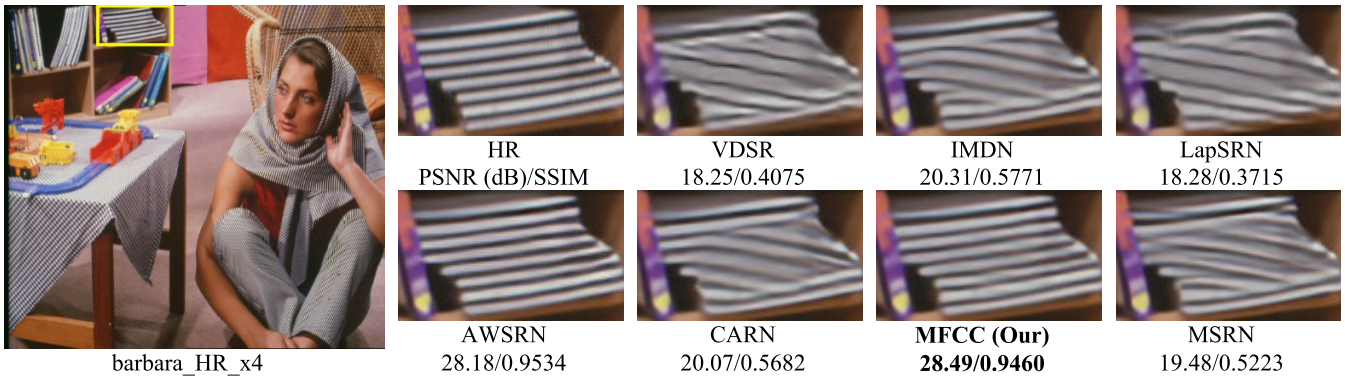

FIGURE 5. Qualitative comparison of barbara image belongs to Set14 dataset for scale ×4.

achieving higher performance with a reasonable number of parameters and multi-adds. Specifically, in comparison to the second-best PSNR at scale factor of 2, our model provides an improvement of 0.05 dB, 0.07 dB, 0.02 dB, 0.16 dB, and 0.16 dB for Set5, Set14, B100, Urban100, and Mango109, respectively.

Our model demonstrates a notable improvement in PSNR for upscale factors of 3× and 4×, with an accuracy of 0.3 dB and 0.26 dB compared to the second best on the Manga109 dataset, respectively. For the most challenging scale of 8×, our model exhibits superior performance on all datasets, except for the SSIM of the Set14 dataset. Nonetheless, our performance on SSIM is the second-best and differs only slightly (0.0002) from the AWSRN [24] model. Figure 4 shows the statistical analysis of our model, focusing on the PSNR and SSIM performance on all datasets across different scaling factors. The charts also provide a comparison with four other lightweight SR models: MSRN [19], IMDN [32], AWSRN [24], and A2F-L [21]. According to the statistical analysis of PSNR and SSIM, our model outperforms

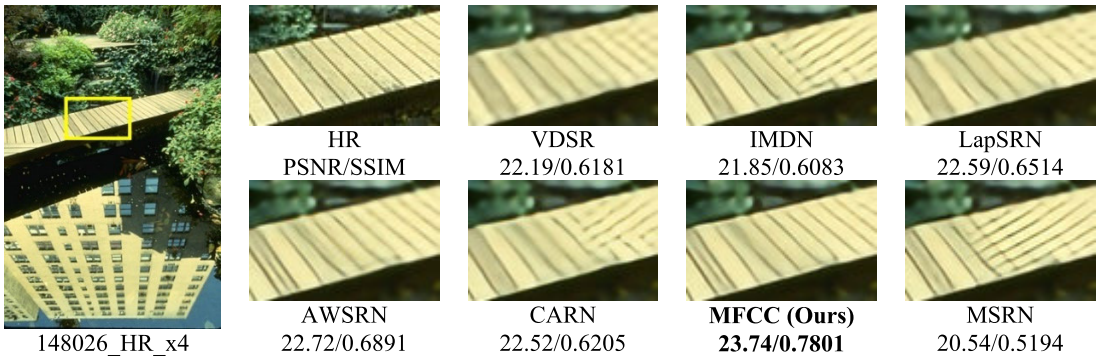

FIGURE 6. Qualitative comparison of image 148026 belongs to B100 dataset for scale ×4.

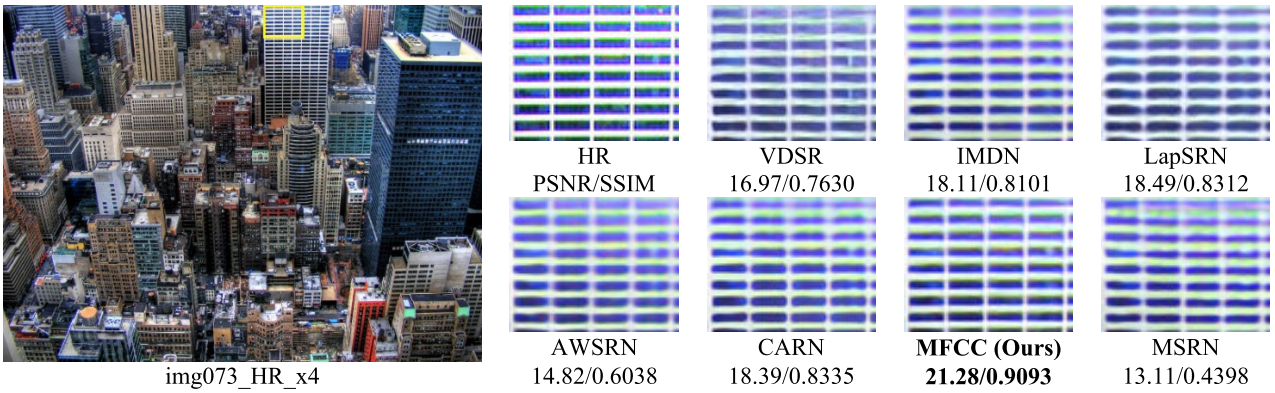

FIGURE 7. Qualitative comparison of img073 belongs to Urban100 dataset for scale ×4.

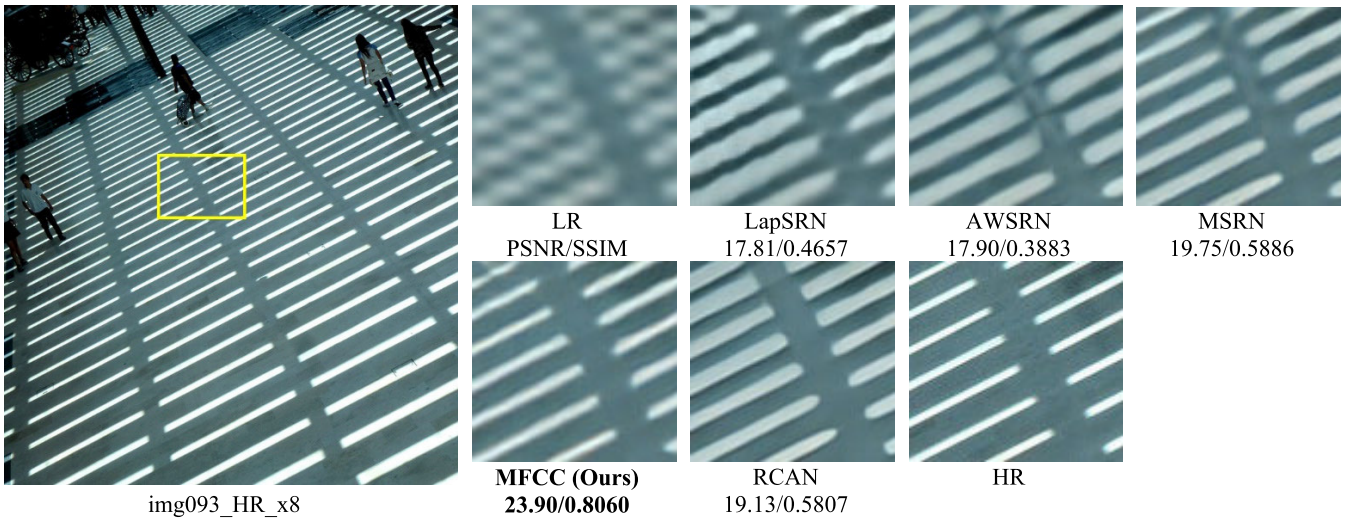

FIGURE 8. Qualitative comparison of img093 belongs to Urban100 dataset for scale ×8.

other lightweight state-of-the-art models for all scale factors. Figure 5, Figure 6, and Figure 7 demonstrate the visual comparisons of our proposed model at scale ×4 with other state-of-the-art models, including VDSR [16], IMDN [32], LapSRN [28], AWSRN [24], CARN [18], and MSRN [19]. The demonstrated patches were added according to the network parameters of the models.

Figure 5 illustrates a visual comparison of "Barbara" image belonging to the Set14 dataset at scale ×4. From the observation of the demonstration, it is apparent that our lightweight model demonstrates a strong capability to accurately reconstruct parallel lines. Additionally, the MRCC model achieved the highest PSNR and SSIM, indicating its superiority over the other models.

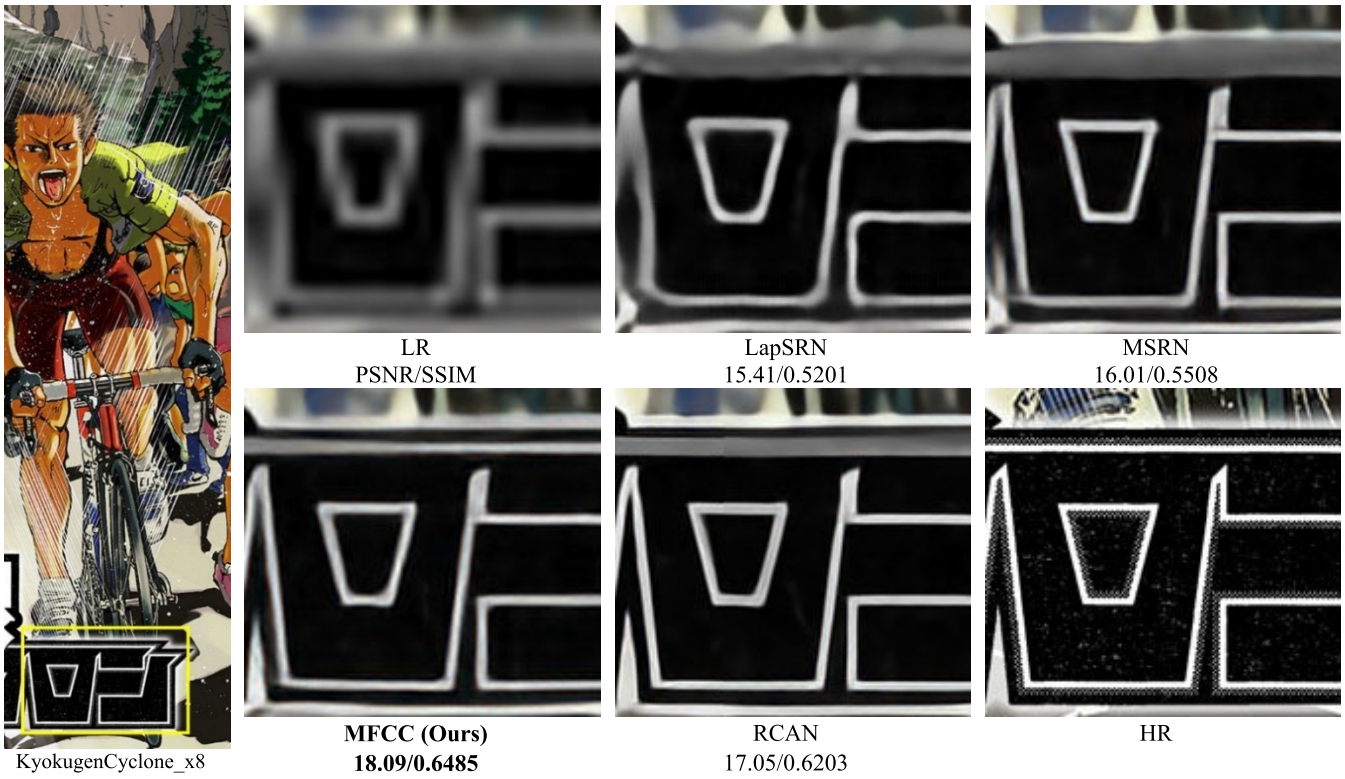

FIGURE 9. Qualitative comparison of an image belongs to Manga109 dataset for scale ×8.

Figure 6 shows the visual comparison of "148026" belonging to the B100 dataset at scale ×4. The MFCC model achieved the highest PSNR and SSIM compared with other state-of-the-art models. Additionally, the high-frequency details of the tiny lines were recovered more effectively without any over-smoothing degradation. Figure 7 shows a visual comparison of "img037" belonging to the Urban100 dataset at scale ×4. Our lightweight MFCC model effectively produces an SR image that closely resembles the ground-truth (GT) image. Conversely, the results obtained from the other models exhibit shortcomings when reconstructing a sharp image. Our model achieved the highest PSNR and SSIM values.

The values (PSNR and SSIM) indicate our model's superiority over the other models. Figure 8, and Figure 9 display the visual comparisons of our proposed model at scale ×8 with other state-of-the-art models, including LapSRN [28], MSRN [19], and RCAN [34]. Figure 8 illustrates a visual comparison of "img093", which belongs to the Urban100 dataset. Compared to the other models, our lightweight model reconstructs the high-frequency details of lines, similar to the GT image. The PSNR and SSIM of our resultant image show significant improvement compared to other state-of-the-art models. Figure 8 compares the results of "Kyokugen-Cyclone" image which belongs to Manga109 dataset. The PSNR and SSIM of the proposed model are the highest. The other models could not reconstruct the parallel lines located at the top of the selected patch and merged the lines. In contrast, our model shows a robust ability to produce tiny edges at this scale.

Figure 3 compares the number of parameters and the performance of different SR models at a scale of ×4 on the Set5 dataset. Several state-of-the-art approaches including VDSR [16], LapSRN [18], DRCN [25], SelNet [33], CARN [18], IMDN [32], A2F-L [21] and AWSRN [24] were chosen to analyze our model's performance. As demonstrated in the graph, our MFCC model has the highest PSNR (32.42 dB), while the number of parameters in our model is 2.15 million.

V. CONCLUSION

This study proposed a lightweight single-image super-resolution model based on constructing Residual Group blocks on a multipath residual architecture (MFCC). Utilizing a multipath residual network increases the efficiency of the proposed lightweight model. In addition, we addressed the lack of low-frequency details by employing the pixel-shuffle fusion method. Based on this approach, the low-frequency details of the early layer are up-sampled and bypassed into the up-sampled features of the multipath residual network. The high and low-frequency information of these layers are fused, which improves the line and edge reconstruction capability of the proposed model. The experimental results on five benchmark datasets demonstrate that our lightweight MFCC model outperforms other state-of-the-art models, particularly on a scale of ×8.

In future work, the proposed model will be tuned to achieve the optimum parameters and then implemented on FPGAs to support real-world applications.

REFERENCES

[1] J. Trimek, "Public confidence in CCTV and fear of crime in Bangkok, Thailand," *Int. J. Criminal Justice Sci.*, vol. 11, no. 1, pp. 7–29, 2016.

[2] S. Taweesaengsakulthai, S. Laochankham, P. Kamnuansilpa, and S. Wongthanavasu, "Thailand smart cities: What is the path to success?" *Asian Politics Policy*, vol. 11, no. 1, pp. 144–156, Jan. 2019.

[3] P. Vichian, T. Prachuab, D. Ubon, P. Jitrapun, and A. C. P.-A. Mark, "The design of real estate housing project in Bangkok and the metropolitan, Thailand for the year 2021," *I-Manager's J. Manag.*, vol. 15, no. 1, p. 19, 2020.

[4] C. Sirirattanapol, M. Nagai, A. Witayangkurn, S. Pravinvongvuth, and M. Ekpanyapong, "Bangkok CCTV image through a road environment extraction system using multi-label convolutional neural network classification," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 3, p. 128, Mar. 2019.

[5] S. Saypadith and S. Aramvith, "Real-time multiple face recognition using deep learning on embedded GPU system," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2018, pp. 1318–1324.

[6] T. Ganokratanaa, S. Aramvith, and N. Sebe, "Unsupervised anomaly detection and localization based on deep spatiotemporal translation network," *IEEE Access*, vol. 8, pp. 50312–50329, 2020.

[7] S. Aramvith, S. Pumrin, T. Chalidabhongse, and S. Siddhichai, "Video processing and analysis for surveillance applications," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Jan. 2009, pp. 607–610.

[8] S. Chen, H. M. Maung, and S. Aramvith, "Improving feature preservation in high efficiency video coding standard," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2016, pp. 1–5.

[9] S. Chen, S. Aramvith, and Y. Miyanaga, "Encoder control enhancement in HEVC based on R-Lambda coefficient distribution," in *Proc. Int. Symp. Multimedia Commun. Technol. (ISMAC)*, Aug. 2019, pp. 1–4.

[10] W. Ruangsang and S. Aramvith, "Super-resolution for HD to 4K using analysis K-SVD dictionary and adaptive elastic-net," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2015, pp. 1076–1080.

[11] W. Ruangsang and S. Aramvith, "Efficient super-resolution algorithm using overlapping bicubic interpolation," in *Proc. IEEE 6th Global Conf. Consum. Electron. (GCCE)*, Oct. 2017, pp. 1–2.

[12] W. Muhammad and S. Aramvith, "Multi-scale inception based super-resolution using deep learning approach," *Electronics*, vol. 8, no. 8, p. 892, Aug. 2019.

[13] 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.

[14] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Computer Vision—ECCV*, vol. 8689. Cham, Switzerland: Springer, 2014, pp. 184–199.

[15] 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.

[16] 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.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[18] 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)*, 2018, pp. 252–268.

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

[20] 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.

[21] X. Wang, Q. Wang, Y. Zhao, J. Yan, L. Fan, and L. Chen, "Lightweight single-image super-resolution network with attentive auxiliary feature learning," in *Proc. Asian Conf. Comput. Vis.*, 2020, pp. 1–17.

[22] X. Chu, B. Zhang, and R. Xu, "Multi-objective reinforced evolution in mobile neural architecture search," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 99–113.

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

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

[25] 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.

[26] 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.

[27] 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.

[28] W. Lai, J. Huang, N. Ahuja, and M. 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] J. Choi and M. Kim, "A deep convolutional neural network with selection units for super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1150–1156.

[34] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.

[35] 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.

[36] S. Anwar, S. Khan, and N. Barnes, "A deep journey into super-resolution: A survey," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, May 2021.

[37] 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.

[38] 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.

[39] Z. Wang, Y. Liu, R. Zhu, W. Yang, and Q. Liao, "Lightweight single image super-resolution with similar feature fusion block," *IEEE Access*, vol. 10, pp. 30974–30981, 2022.

[40] W. Shang, K. Sohn, D. Almeida, and H. Lee, "Understanding and improving convolutional neural networks via concatenated rectified linear units," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2217–2225.

[41] K. Prajapati, V. Chudasama, K. Upla, K. Raia, R. Ramachandra, and C. Busch, "Channel split convolutional neural network for single image super-resolution (CSISR)," in *Proc. 16th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Dec. 2021, pp. 1–8.

[42] K. Park, J. W. Soh, and N. I. Cho, "A dynamic residual self-attention network for lightweight single image super-resolution," *IEEE Trans. Multimedia*, vol. 25, pp. 907–918, 2023.

[43] Z. Li, G. Li, T. Li, S. Liu, and W. Gao, "Information-growth attention network for image super-resolution," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 544–552.

[44] Y. Zhang, D. Wei, C. Qin, H. Wang, H. Pfister, and Y. Fu, "Context reasoning attention network for image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4258–4267.

[45] T. Dai, J. Cai, Y. Zhang, S. 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.

[46] A. Veit, M. J. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, vol. 29, pp. 1–9.
[47] Z. Meng, L. Li, X. Tang, Z. Feng, L. Jiao, and M. Liang, "Multipath residual network for spectral–spatial hyperspectral image classification," *Remote Sens.*, vol. 11, no. 16, p. 1896, Aug. 2019.
[48] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
[49] 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.
[50] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.* Cham, Switzerland: Springer 2010, pp. 711–730.
[51] J. 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.
[52] M. Abdi and S. Nahavandi, "Multi-residual networks: Improving the speed and accuracy of residual networks," 2016, *arXiv:1609.05672*.

WATCHARA RUANGSANG (Graduate Student Member, IEEE) received the B.Eng. degree in electrical and electronic engineering from the King Mongkut's University of Technology North Bangkok, Thailand, in 2011, and the M.Eng. degree in electrical engineering from Chulalongkorn University, Bangkok, Thailand, in 2015, where he is currently pursuing the Ph.D. degree in electrical engineering. His current research interests include image super resolution, computer vision and machine learning, and deep learning applications.

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 nine Ph.D. graduates. She has published over 130 papers 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 chair to the Thailand Government bodies in telecommunications and ICT. She is very active in the international arena with 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, Japan, in 1991, 1993, and 1997, respectively. He is currently a Professor and the Dean of the Graduate School of Information Science and Technology, Osaka University. His research interests include media-centric low-power system architecture and implementation. He has also taken various volunteer positions in academic societies, such as the Editor-in-Chief of *IEICE Transactions on Fundamentals* (Japanese Edition), the IEEE CAS Society Board of Governors, the IEEE Region 10 Treasurer, the IEEE Region 10 Vice Chair, and the IEEE Japan Council Chair. He is a member of IEICE, IPSJ, and ITE-J.