MPRNet: Multi-Path Residual Network for Lightweight Image Super Resolution

Armin Mehri
*Computer Vision Center,
Edifici O, Campus UAB,
08193, Bellaterra,
Barcelona, Spain
amehri@cvc.uab.es

Parichehr B.Ardakani
*Computer Vision Center,
Edifici O, Campus UAB,
08193, Bellaterra,
Barcelona, Spain
pbehjati@cvc.uab.es

Angel D. Sappa
+ESPOL Polytechnic University,
Guayaquil, Ecuador
*Computer Vision Center,
08193, Bellaterra, Barcelona, Spain
sappa@ieee.org

Abstract

Lightweight super resolution networks have extremely importance for real-world applications. In recent years several SR deep learning approaches with outstanding achievement have been introduced by sacrificing memory and computational cost. To overcome this problem, a novel lightweight super resolution network is proposed, which improves the SOTA performance in lightweight SR and performs roughly similar to computationally expensive networks. Multi-Path Residual Network designs with a set of Residual concatenation Blocks stacked with Adaptive Residual Blocks: (i) to adaptively extract informative features and learn more expressive spatial context information; (ii) to better leverage multi-level representations before upsampling stage; and (iii) to allow an efficient information and gradient flow within the network. The proposed architecture also contains a new attention mechanism, Two-Fold Attention Module, to maximize the representation ability of the model. Extensive experiments show the superiority of our model against other SOTA SR approaches.

1. Introduction

Single Image Super Resolution (SISR) targets to recover a high-resolution (HR) image from its degraded low-resolution (LR) one with a high visual quality and enhanced details. SISR is still an active yet challenging topic to research due to its complex nature and high practical values in improving image details and textures. SR is also critical for many devices such as HD TVs, computer displays and portable devices like cameras, smartphones, tablets, just to mention a few. Moreover, it leads to improvements in various computer vision tasks, such as object detection [9], medical imaging [10], security and surveillance imaging [49], face recognition [31], astronomical images [28] and many other domains [25, 41, 43]. Image super-resolution is challenging due to the following reasons: i) SR is an ill-posed inverse problem, since instead of a single unique solution, there exist multiple solutions for the same low-resolution image; and ii) as the up-scaling factor increases, the complexity of the problem increases [7]. The retrieval of missing scene details becomes even more complicated with greater factors, which often leads to the reproduction of incorrect information. Due to the rapid development of deep learning methods, recent years have witnessed an explosive spread of CNN models to perform SISR. The obtained performance has been consistently improved by designing new architectures or introducing new loss functions. Though significant advances have been made, most of the works in SR were dedicated to achieve higher PSNR with the design of a very deep network, which causes the increase in the numbers of computational operations.

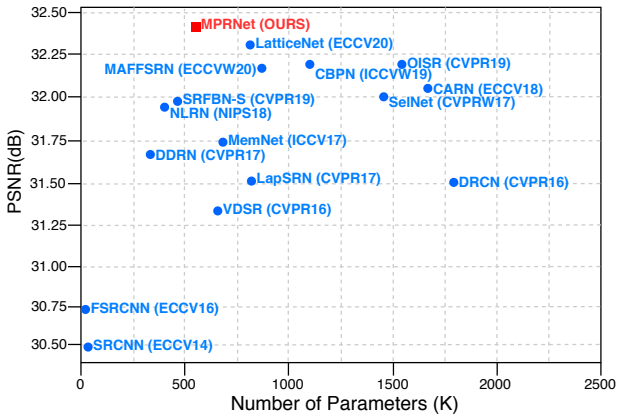

Figure 1: PSNR vs. Parameters trade-off on Set5 ($\times 4$). MPRNet achieves superior performance among all lightweight models.

In this paper, to design a practical network for real-world applications and tackle with mentioned downsides, a novel lightweight architecture is introduced, referred to as Multi-Path Residual Network (MPRNet), to adaptively learn most valuable features and construct the network to focus on learning high-frequency information. Additionally, to seek a better trade-off between performance and

applicability, we introduce a novel module, referred to as Residual Module (RM), which contains Residual Concatenation Blocks that are connected to each other with a Global Residual connection; build with a set of Adaptive Residual Blocks (ARB) with a Local Residual Connection (LRC). Each ARB is defined as a divers residual pathways learning to make use of all kind of information form LR image space, which the main parts of the network can access to more rich information. So, our MPRNet design has the benefits of multi-level learning connections and also takes advantage of propagating information throughout the network. As a result, each block has access to information of the precedent block via local and global residual connections and passes on information that needs to be preserved. By concatenating different blocks followed by 1×1 convolutional layer the network can reach to both intermediate and high-frequency information, resulting in a better image reconstruction. Finally, in order to enhance the representation of the model and even make it robust against challenging datasets and noise, we propose a lightweight and efficient attention mechanism, Two-Fold Attention Mechanism (TFAM). TFAM is working by considering both the inner channel and spatial information to highlight the important information. This TFAM helps to adaptively preserve essential information and overpower the useless ones. The proposed model is illustrated in Fig. 2. In brief, the main contributions are in three-fold:

- An efficient Adaptive Residual Block (ARB) is proposed by well-focusing on spatial information via a multi-path residual learning to enhance the performance at a negligible computational cost. Comprehensive study shows the excellent performance of ARB.

- A new attention mechanism (TFAM) is proposed to adaptively re-scale feature maps in order to maximize the representation power of the network. Since its low-cost, it can be easily applied to other networks, and has the better performance than other Attention Mechanisms.

- A lightweight network (MPRNet) is proposed to effectively enhance the performance via multi-level representation and multiple learning connections. The MPRNet is built by fusing the proposed ARB with the robust TFAM to generate more accurate SR image. MPRNet achieves the excellent performance among all the lightweight state-of-the-art approaches with lower model size and computational cost (Fig. 1).

2. Related Work

In this section, recent state-of-the-art SR deep learning approaches are detailed. In section 2.2, SR lightweight models, which focus on compressing the number of parameters and operations are reviewed. Finally, an overview of Attention Mechanisms is given in section 2.3.

2.1. Deep Learning Based Image Super-Resolution

Dong et al. [7] present one of the first work using CNN to tackle the SR task (i.e., SRCNN). The SRCNN receives an upsampled image as an input that cost extra computation. Later on, to address this drawback, FSRCNN [8] and ESPCN [35] have been proposed to reduce the large computational and run time cost by upsampling the features near to the output of the network. This tactic leads results in efficient approaches with low memory compared to SRCNN. However, the entire performance could be reduced if there are not enough layers after the upsampling process. In addition, they cannot manage multi-scale training, as the size of the input image differs for each upsampling scale.

Even though the strength of deep learning shows up from deep layers, the above-mentioned methods are referred to as shallow network due to the training difficulties. Therefore, Kim et al. [18] use residual learning to ease the training challenges and increase the depth of their network by adding 20 convolutional layers. Then, [37] has proposed memory block in MemNet for deeper networks and solve the problem of long-term dependency with 84 layers. Thus, CNN-based SR approaches demonstrate that deeper networks with various types of skip connections show better performance. Thereby, Lim et al. [24] introduce EDSR by expanding the network size and enhancing the residual block by omitting the batch normalization from residual block. Zhang et al. [47] propose RDN with residual and dense skip connections to fully use hierarchical features. Li et al. [22] propose a network with more than 160 layers plus improved residual units. Despite of the fact that they achieve higher PSNR values, the number of parameters and operations are increased, which leads to high risk of overfitting and limits for real-world applications.

2.2. Deep Learning Lightweights Super Resolution

In recent years the interest of building lightweight and efficient models has been increased in SISR to reduce the computational cost. Several lightweight networks have been introduced, such as SRCNN [7], FSRCNN[8], ESPCN[35], which were the first attempts, but they could not perform well. Later, Ahn et al. [1] design a network that is suitable in the mobile scenario by implementing a cascade mechanism beyond a residual network (CARN), in order to obtain lightweight and improve reconstruction but it is at the cost of reduction of PSNR. Then, a neural architecture search (NAS)-based strategy has been also proposed in SISR to construct efficient networks—MoreMNAS [4] and FALSR [3]. But due to limitation in strategy, the performance of these models are limited. Later, [32] introduces MAFFSRN by proposing multi-attention blocks to

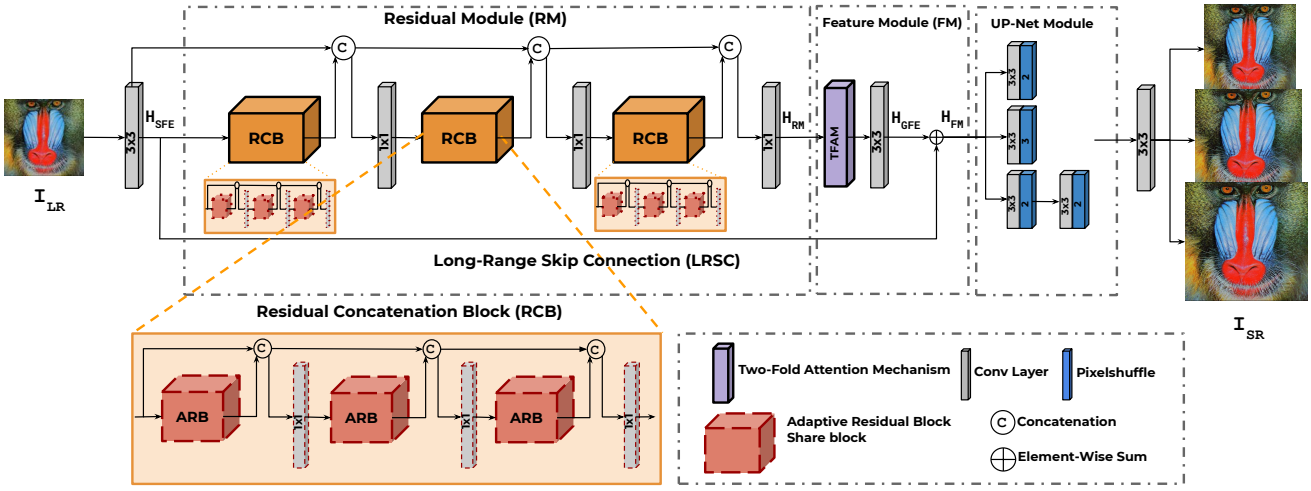

Figure 2: The overall network architecture of the proposed Multi-Path Residual Network (MPRNet).

improve the performance. Recently, LatticeNet [27] introduces an economical structure to adaptively combine Residual Blocks, which achieve good results. All these works suggest that the lightweight SR networks can keep a good trade-off between PSNR and parameters.

2.3. Attention Mechanism

Attention can be described as a guide to bias the allocation of available computer resources to the most important informative elements of an input. Recently, some works have focused on attention mechanism for deep neural networks. Hu et al. [15] introduce squeeze-and-excitation (SE) block, a compact module to leverage the relationship between channels. Also, Woo et al. [42] propose a Convolutional Block Attention Module (CBAM) to exploit the inner-spatial and inner-channel relationship of features to achieve a performance improvement in image classification.

Recently, RCAN [46] designs a very deep network with a channel attention mechanism to enhance the reconstruction results by only considering inner-channel information, which call first-order statistics. In contrast, Dai et al. [5] introduce the second-order attention network in order to explore more powerful feature expression. More recently, Li et al., [27] propose enhanced spatial attention (ESA) to make the residual features to be more focused on critical spatial contents. In the current work, motivated by attention mechanisms and considering that there are different types of information within and across feature space, which have different contributions for image SR, a Two-Fold Attention Mechanism is proposed that adaptively highlight the important information by considering both channel and spatial information to boost the performance of the network.

3. Multi-Path Residual Network

3.1. Network Structure

The proposed model (MPRNet – Fig 2) consists of four different modules, namely, Shallow Feature Extraction (SFE); Residual Module that contains Residual Concatenation Blocks (RCBs); Feature Module that includes a Two-Fold Attention Mechanism (TFAM) and a Global Feature Extractor with a Long-Range Skip Connection; and the multi-scale UP-Net module at the end of network. Let's consider $\{I_{LR}, I_{SR}\}$ as the input and output of the network respectively. The SFE is a Conv layer with a kernel size of 3×3, which can be formulated as follow:

$$\boldsymbol{H}_{SFE} = f_{SFE}(\boldsymbol{I}_{LR}; W_c), \tag{1}$$

where $f_{SFE}(\cdot)$ and W_c indicates Conv operation and parameters applied on I_{LR}. $\boldsymbol{H}_{SFE}$ denotes the output of SFE, which later is used as the input to Residual Module. Lets $\boldsymbol{H}^{i,j}_{RM}$ be the output from the i-th Residual Concatenation Block (RCB) that has j-th inner Adaptive Residual Blocks (ARBs). The Residual Module can be defined as:

$$\boldsymbol{H}_{RM} = f([\boldsymbol{H}_{SFE}, ..., \boldsymbol{H}^{i-1}_{RCB}(\boldsymbol{H}^{j-1,R}_{ARB}; W^j_c), \boldsymbol{H}^i_{RCB}]; W^i_c), \tag{2}$$

where $\boldsymbol{H}_{RM}$ is the output of the Residual Module. Note that our RM contains multi-level learning connections followed by a 1×1 Conv layer to control the output after each block, which helps our model to quickly propagate information all over the network (lower to higher layers and vice-versa in term of back propagation) and also let the network to learn multi-level representations. So, i-th RCB can be defined as:

$$\boldsymbol{H}^i_{RCB} = f([\boldsymbol{H}^{j,R}_{ARB}, ..., \boldsymbol{H}^{j-1,R}_{ARB}(\boldsymbol{H}^{i-1}; W^i_c)]; W^j_c). \tag{3}$$

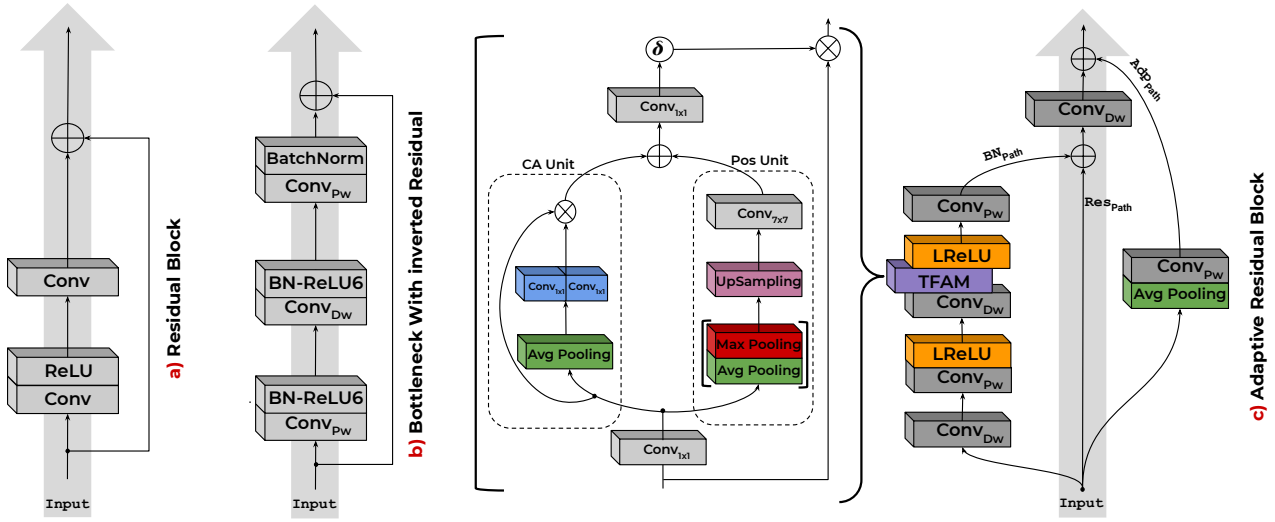

Figure 3: Illustrations of different structure of residual blocks: a) Residual block in EDSR [24]; b) Bottleneck with inverted residual from [14]; c) Proposed Adaptive Residual Block and Two-Fold Attetion Module.

Then, the output of RM feed to the Feature Module by firstly refining the feature maps (i.e., re-calibrate) throughout the TFAM and then extracting more abstract features. Later, accumulate with LRSC to efficiently alleviate the gradient vanishing/exploding problems and make sure that network has access to unmodified information before UP-Net:

$$\boldsymbol{H}_{FM} = f_{GFE}(\boldsymbol{H}_{TFAM}(\boldsymbol{H}_{RM}; W_c); W_c) + \boldsymbol{H}_{LRC}, \tag{4}$$

where $\boldsymbol{H}_{TFAM}$ denotes our TFAM and $\boldsymbol{H}_{LRC}$ is Long-Range Residual Connection. The last stage is the Multi-Scale Up-Net Module to reconstruct the image from obtained feature-maps. The upsampling module is inspired by [1] and followed by a Conv layer:

$$\boldsymbol{H}_{UP} = f^{\uparrow}_{pix}(\boldsymbol{H}_{FM}), \tag{5}$$

where $f^{\uparrow}_{pix}(\cdot)$ indicates the Up-net module function and $\boldsymbol{H}_{FM}$ is the output of FM. The upsampled features are reconstructed with a Conv layer:

$$\boldsymbol{I}_{SR} = f_{REC}(\boldsymbol{H}_{up}) = \boldsymbol{H}_{MPRNet}(\boldsymbol{I}_{LR}), \tag{6}$$

where $f_{REC}(\cdot)$ and $\boldsymbol{H}_{MPRNet}(\cdot)$ denote the reconstruction layer and function of our MPRNet. In the next subsections, more details about the Adaptive Residual Block and Two Fold Attention Mechanism are given.

3.2. Adaptive Residual Block

This research focuses on designing a efficient and effective Residual Block based on Depthwise (Dw) and Pointwise (Pw) Convolutions for SISR. [34] introduced linear bottleneck with an inverted residual structure. However, this structure deliver chances of losing information and weaken the propagation capability of gradients across layers, due to gradient confusion arising from the narrowed feature space [6, 21]. Thus, we propose a novel Residual Block that mitigates the aforementioned issues; it is well-optimized especially for the SR tasks, called Adaptive Residual Block (ARB). Unlike [34], ARB introduces new features and operations by proposing a multi learning pathways with a completely new structure. Each learning path is responsible to extract different kind of information before aggregation. So, the main part of network can have access to more rich information and performs notably well in noisy LR and generates more accurate SR image. The ARB consists of three different learning pathways that are detailed below. Fig 3 shows each of the ARB components.

Bottleneck Path: We design our Bottleneck path (BN) based on the following insights: i) Extract richer spatial information since spatial information is key importance in SR tasks; ii) prevent very wide feature maps in the middle of the building block, which unavoidably growing the computational load of relevant layers; iii) preserve the BN path low-cost and efficient. Thus, Dw Convolutions with small kernel size (3×3) are chosen since they are lightweight and they can learn expressive features when conducted to the high dimensional space. So, we initiate the BN path by using a Dw convolution with kernel size 3×3 towards the high dimensional features space to richer spatial information to be encoded and generate meaningful representations. Also, a Pw convolution is used after each Dw convolution in our design to produce new features by encoding the inter-channel information and reduce the computational cost. We shared the same number of channels and resolution along the BN path to prevent of sudden rise of computational burden in middle of the path. Furthermore, we conjunct our TFAM into the BN path after the second Dw convolution

to spotlight the informative features along the channel and spatial axes. By doing so, the BN path is working with high dimensional features space, which makes the pathway efficient, low-cost, and well-focused on spatial context information compared to [34].

Adaptive Path: It is proposed by taking the advantages of global average pooling accompanied by a 1×1 Pw convolution. Average Pooling layers have been employed to take the average value of the features from the feature space to smooth and eliminate the noise from the LR image and reduce the dimensionality of each feature map but retains the important information to help the network to generate robust feature maps in challenging situations—noisy LR image. So, the network can generate a sharper and well-detailed SR image.

Residual Path: Unlike [34] that puts the residual path between narrowed feature space that cause gradient confusion, in our ARB, we place the residual path on the high dimensional representations to transfer more information from the bottom- to top-layers. Such structure facilitate the gradient propagation between multiple layers and help the network to optimize better during training.

Thus, the information from BN- and Res-paths aggregate together, followed by another Dw convolution. We found out adding the Dw convolution before final aggregation with Adp path is essential for performance improvement since Dw encourage the network to learn more meaningful spatial information. Extensive experiments show that, our ARB is more beneficial than the existed ones for SISR tasks and improved the results with a large margin.

3.3. Two-Fold Attention Module

A novel Attention Mechanism (TFAM) has been proposed to boost the performance of our Adaptive Residual Block and refine the high-level information in the Feature Module (FM) by focusing on both channel and spatial information. The best way to amplify efficiency of ARB is through the union of the channel and spatial attention mechanism, since the residual features need to be well-focused on both information. In detail, TFAM is designed to focus on the important features on the channel information via CA unit and spotlight on the region of interest via Pos unit. Thus, each unit can learn 'what' and 'where' to attend in the channel and spatial axes respectively to recover edges and textures more accurately. As a result, TFAM works better than other attention mechanism [15, 16, 27, 42] by emphasising informative features and reducing worthless ones.

Channel Unit. CA unit starts with an average pooling to exploit first-order statistics of features followed by two Conv layer, which they work side by side, each seeing half of the input channels, and producing half the output channels, and both subsequently concatenated to even have more low-cost unit. Thus, CA unit modulates features globally, where the summary statistics per channel are computed. Then, used to emphasize meaningful feature maps while redundant useless features are diminished. Especially, CA unit focuses on 'what' is meaningful given an input image.

Positional Unit. Pos unit designed as a complementary unit to our CA unit. The feature map information is varied over spatial positions therefore, Pos unit concerns about the position of the informative part of the image and focuses on that region. Pos unit requires a large receptive field to work perfectly in SR tasks unlike the classification task. Thus, Average- and Max pooling operations with a large kernel size have been employed and then concatenated them to generate an efficient feature descriptor. afterward, an Up-Sampling layer is used to retrieve the spatial dimensions, which is followed by a Conv layer to generate a spatial attention map.

Finally, highlighted information from both units aggregated together followed a 1×1 Conv layer and a sigmoid operation to firstly, recover the channel dimensions and then generate the final attention mask. Also, a residual connection used to transfer HR features to the end of module.

4. Experimental Results

4.1. Setting

Datasets & Evaluation Protocol. Following previous works [5, 27], we use $DIV2K$ [38] dataset to train (800 images) and validate (100 images) our model. The proposed model is evaluated with the standard benchmark datasets, namely, $Set5$ [2], $Set14$ [44], $B100$ [30], and $Urban100$ [17]. Two widely used quantitative metrics have been considered to measure its performance: PSNR and SSIM [40], computed between the obtained images and the corresponding ground truths. Both metrics are computed on the Y channel in the $YCbCr$ space.

Degradation Models. Following the work of [47], three different degradation models created to simulate LR images and make fair comparisons with available methods. Firstly, a bicubic (BI) down-sampling dataset with scaling factors [$\times2$, $\times3$, $\times4$] has been created. Blur-Down-sampled (BD) is the second one to blur and down-sample HR images with a Gaussian kernel 7×7, and $\sigma = 1.6$. Then, images are down-sampled with scaling factor $\times3$. Aside from the BD, a more challenging model has been created, referred to as (DN). DN degradation model is down-sampling HR images with bicubic followed by adding 30% Gaussian noise.

Training Details. In the training stage, RGB input patches are used with size of 64×64 from each of the randomly selected 64 LR training images. Patches are augmented by random horizontally flips and 90 degree rotation. AdamP [13] optimizer has been employed. The initial learning rate set to 10^{-3} and its halved every 4×10^5 steps. $L1$ is used as loss function to optimize the model. The PyTorch framework is used.

Table 1: Comparison with lightweight SOTA methods on the Bicubic (**BI**) degradation for scale factors $[\times 2, \times 3, \times 4]$. **Red** is the Best and **Blue** is the second best performance. We assume that the generated SR image is $720P$ to calculate Multi-Adds (MAC).

Params-MAC / Dataset	Methods / Scale	VDSR [18] $655K-612.6G$ PSNR/SSIM	LapSRN[20] $813K-149.6G$ PSNR/SSIM	MemNet[37] $677K-2662.4G$ PSNR/SSIM	NLRN[26] $350K-32.5$ PSNR/SSIM	SRFBN_S[23] $483K-119G$ PSNR/SSIM	CARN[1] $1592K-90.9G$ PSNR/SSIM	CBPN [48] $1197K-97.9G$ PSNR/SSIM	OISR_RK2_s[12] $1540K-114.2G$ PSNR/SSIM	MAFFSRN-L[32] $830K-38.6G$ PSNR/SSIM	LatticeNet[29] 777K-43.6G PSNR/SSIM	MPRNet [Ours] **538K-31.3G** PSNR/SSIM
Set5	×2	37.53/0.9587	37.52/0.9590	37.87/0.9597	38.00/0.9603	37.78/0.9597	37.76/0.9590	37.90/0.9590	37.90/0.9600	38.07/0.9607	**38.15/0.9610**	38.08/0.9608
	×3	33.66/0.9213	——	34.09/0.9248	34.27/0.9266	34.20/0.9255	34.29/0.9255	——	34.39/0.9273	34.45/0.9277	34.53/0.9281	**34.57/0.9285**
	×4	31.35/0.8838	31.54/0.8850	31.74/0.8893	31.92/0.8916	31.98/0.9594	32.13/0.8937	32.21/0.8944	32.21/0.8903	32.20/0.8953	32.30/0.8962	**32.38/0.8969**
Set14	×2	33.03/09124	33.08/0.9130	33.28/0.9142	33.46/0.9159	33.35/0.9156	33.52/0.9166	33.60/0.9171	33.58/0.9172	33.59/0.9177	33.78/0.9193	**33.79/0.9196**
	×3	29.77/0.8314	——	30.00/0.8350	30.16/0.8374	30.10/0.8350	30.29/0.8407	——	30.33/0.8420	30.40/0.8432	30.39/0.8424	**30.42/0.8441**
	×4	28.01/0.7674	28.19/0.7720	28.26/0.7723	28.36/0.7745	28.45/0.7779	28.60/0.7806	28.63/0.7813	28.63/0.7822	28.62/0.7822	28.68/0.7830	**28.69/0.7841**
B100	×2	31.90/0.8960	31.80/0.8950	32.08/0.8978	32.19/0.8992	32.00/0.8970	32.09/0.8978	32.17/0.8989	32.18/0.8996	32.23/**0.9005**	32.25/**0.9005**	**32.25**/0.9004
	×3	28.82/0.7976	——	28.96/0.8001	29.06/0.8026	28.96/0.8010	29.06/0.8434	——	29.10/0.8083	29.13/0.8061	29.15/0.8059	**29.17/0.8073**
	×4	27.29/0.7251	27.32/0.7280	27.40/0.7281	27.48/0.7306	27.44/0.7313	27.58/0.7349	27.58/0.7356	27.58/0.7364	27.59/0.7370	27.62/0.7367	**27.63/0.7385**
Urban100	×2	30.76/0.9140	30.41/0.9100	31.31/0.9195	31.81/0.9249	31.41/0.9207	31.92/0.9256	32.14/0.9279	32.21/0.8950	32.38/0.9308	32.43/0.9302	**32.52/0.9317**
	×3	27.14/0.8279	——	27.56/0.8376	27.93/0.8453	27.66/0.8415	28.06/0.8493	——	28.03/0.8544	28.26/0.8552	28.33/0.8538	**28.42/0.8578**
	×4	25.18/0.7524	25.21/0.7560	25.50/0.7630	25.79/0.7729	25.71/0.7719	26.07/0.7837	26.14/0.7869	26.14/0.7874	26.16/0.7887	26.25/0.7873	**26.31/0.7921**

Table 2: Comparison with SOTA methods on challenging datasets ("**BD**" and "**DN**") for scale factor $\times 3$. **Red** is the Best and **Blue** is the second best performance.

Dataset	Methods / Degradation	Bicubic PSNR/SSIM	SPMSR[33] PSNR/SSIM	SRCNN[7] PSNR/SSIM	FSRCNN[8] PSNR/SSIM	VDSR[18] PSNR/SSIM	IRCNN_G[45] PSNR/SSIM	IRCNN_C[45] PSNR/SSIM	SRMD(NF)[39] PSNR/SSIM	RDN[47] PSNR/SSIM	MPRNet [Ours] PSNR/SSIM
Set5	**BD**	28.34/0.8161	32.21/0.9001	31.75/0.8899	26.58/0.8224	33.29/0.9139	33.38/0.9182	29.55/0.8246	34.09/0.9242	**34.57/0.9280**	**34.57**/0.9278
	DN	24.14/0.5445	—-	27.04/0.7638	24.28/0.7124	27.42/0.7372	24.85/0.7205	26.18/0.7430	27.74/0.8026	28.46/0.8151	**28.54/0.8175**
Set14	**BD**	26.12/0.7106	28.89/0.8105	28.64/0.7997	24.86/0.7246	29.58/0.8259	29.73/0.8292	27.33/0.7135	30.11/0.8364	**30.53/0.8447**	30.47/0.8427
	DN	23.14/0.4828	—-	25.56/0.6592	23.25/0.5956	25.60/0.6706	23.84/0.6091	24.68/0.6300	26.13/0.6974	**26.60/0.7101**	26.25/0.6954
B100	**BD**	26.02/0.6733	28.13/0.7740	27.33/0.7500	24.15/0.6728	28.61/0.7900	28.65/0.7922	26.46/0.6572	28.98/0.8009	**29.23/0.8079**	29.19/0.8062
	DN	22.94/0.4461	—-	25.45/0.6198	23.95/0.5695	25.22/0.6271	23.89/0.5688	24.52/0.5850	25.64/0.6495	25.93/0.6573	**25.95/0.6616**
Urban100	**BD**	23.20/0.6661	25.84/0.7856	25.19/0.7591	22.95/0.6836	26.68/0.8019	26.77/0.8154	24.89/0.7172	27.50/0.8370	**28.46/0.8581**	28.31/0.8538
	DN	21.63/0.4701	—-	23.59/0.6580	21.74/0.5724	23.33/0.6579	21.96/0.6018	22.63/0.6205	24.28/0.7092	24.92/0.7362	**25.00/0.7406**

4.2. Comparison with state-of-the-art Methods

4.2.1 Results with BI Degradation

Table 1 presents comparisons between the proposed MPRNet and 10 most recent lightweight SOTA models on BI degradation model for scale factor $[\times 2, \times 3, \text{and } \times 4]$ to verify the effectiveness of our MPRNet (we exclude some lightweight methods [7, 8, 19, 35, 36, 41] from table 1 since their results are worse than MemNet). Table 1 also contains the number of parameters and operations to show the model complexity. In almost all the cases, our MPRNet achieves superior results among all the aforementioned approaches. MPRNet performs especially well on Urban100. This is particularly because the Urban100 includes rich structured contents and our model can consistently accumulate these hierarchical features to form of more representative features and well-focused on spatial context information. This characteristic can be confirmed by our MPRNet SSIM scores, which focuses on the visible structures in the image. In Fig. 4 a couple of qualitative results on scale factor $\times 4$ are depicted. The proposed MPRNet can generally yield to more precise details. In both images in Fig. 4, the texture direction of the reconstructed images from all compared methods is completely wrong. However, results from the proposed MPRNet makes full use of the abstract features and recover images accurately similar to ground truth texture.

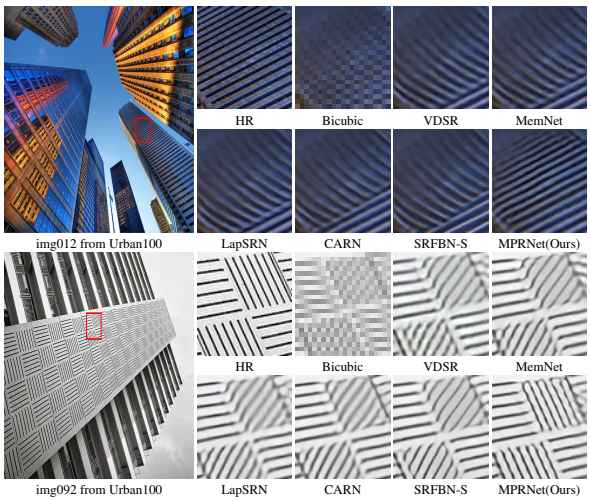

Figure 4: Qualitative results on **BI** degradation dataset with scale factor $\times 4$.

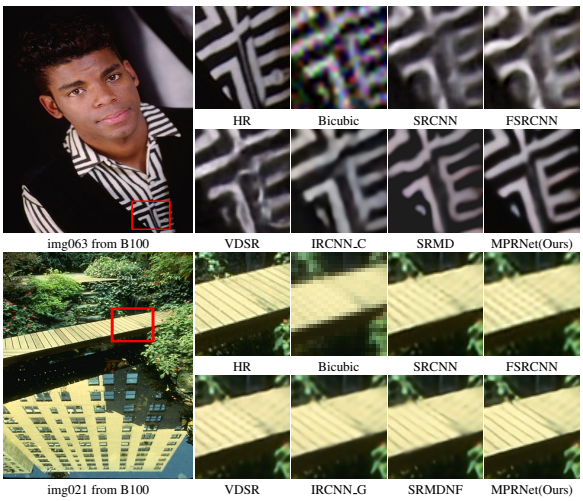

Figure 5: Qualitative results on **DN** and **BD** degradation datasets with a scale factor $\times 3$.

4.2.2 Results with BD and DN Degradation Models

In Table 2, the performance of MPRNet on BD and DN benchmark datasets, together with SOTA methods, are presented. Due to degradation mismatch, SRCNN, FSRCNN, and VDSR for both BD and DN have been re-trained. As can be appreciated, MPRNet achieves remarkable results over all the lightweight SOTA models on challenging benchmark datasets. RDN [47] also listed as a high-capability model to show the superior performance of MPRNet compared to very costly model in the BD and DN datasets. RDN performs sightly better in some BD datasets but not in DN datasets. Obviously, this result was expected since RDN is very expensive compared to low-cost MPRNet (it is almost $\times 44$ more costly). Fig 5 depicts some visual results on both challenging BD and DN benchmark datasets. As can be appreciated the MPRNet with the help of the proposed TFAM performs better in comparison with SOTA methods in terms of producing more convincing results by cleaning off noise and blurred regions from SR images, which results in a sharper SR image with fine details.[1]

4.3. Ablation Study

To further investigate the performance of the proposed model, a deep analysis on the Two-Fold Attention Module, the Adaptive Residual Block, and Residual Learning Connections is performed via an extensive ablation study.

Two-Fold Attention Module. In this section, Deep investigation of the impacts of our proposed TFAM on SOTA SR models are provided. The performance of image SR has improved greatly with the application of Attention Mechanism (AM). Table 3 shows the performance of applying recent AMs including Channel and spatial attention residual (CSAR) [16], Enhanced Spatial Attention (ESA)[27], and our Two-Fold Attention Module (TFAM) on EDSR, RCAN, and MSRN. For a fair comparison, all the models were re-trained with their default setting and AMs are added to the end of their Block, and replaced in the same place as RCAN's Channel Attention placed. As can be seen, by using the aforementioned attention module, the performance of the baseline models are increased that shows the importance of AM in SR tasks. By applying the CSAR to the mentioned approaches, PSNR improves in EDSR and MSRN but does not show enough improvement in RCAN. In contrast, ESA is enhanced version of CASR, which combine both the channel and spatial information, improves all the baseline models. However ESA cannot completely boost the power of the networks due to lack of highlighting informative feature in spatial information. For this propose, we introduce Two-Fold Attention Module, which consider both channel and spatial information and maximize the performance of the networks. TFAM extracts the channel and spatial statistic among channels and spatial axis to further enhance the discriminative ability of the network. As a results, TFAM shows better performance than all the aforementioned ones and boosted the baseline SOTA.

Furthermore, Table 4 contains the study on impact of recent AMs on our MPRNet. Namely, SE[15], CBAM [42], CSAR [16], ESA [27], and TFAM. We apply all the aforementioned AM to our ARB blocks and Feature Module, and provide the performance. The proposed MPRNet with CBAM, could not achieve better results than baseline or SE due to losing channel information and applying the Max-pooling in CA unit which shows harm the performance. Unlike, MPRNet with CASR achieves better results than CBAM and SE because of considering both channel and spatial information but not better than ESA. However, our TFAM performs better among all the AMs by calculating the first order statistics on CA unit and applying Avg- and Max-pooling operations along the channel axis, which is effective in highlighting informative regions and extracts the most important features like edges.

Table 3 also shows the efficiency of our ARB with conjunction of TFAM when it is applied to other SOTA models. As indicated, ARB with TFAM together can improve the PSNR of SOTA models with a large margin.

Adaptive Residual Block. Table 5 presents the impact of different Residual Blocks and the proposed Adaptive Residual Block (ARB) on our MPRNet. In this work, three different structures of residual blocks from SOTA models are considered to compare with our proposed ARB, namely, MobileNet-BottleneckBlock, EDSR-ResBlock, RCAN-ResidualChannelBlock. All the models were trained with the same settings. As can be seen, MobileNet-BottleneckBlock could not perform well in SR tasks due to difficulty of extracting high-frequency information and gradient confusion. EDSR-ResBlock is the ResNet without batch normalization layer, but still could not achieve good results due to the lack of extracting rich feature maps and eliminating noises from LR feature space. RCAN-ResidualChannelBlock performs better than aforementioned ResBlock due to channel attention in their structure. However RCAN-ResidualChannelBlock did not show better results than our proposed ARB since our ARB can learn more expressive spatial information, have access to high-dimensional information and also with the help of TFAM can maximize the whole performance of block.

Additionally, effect of each learning pathways of ARB on the performance is provided. ARB_B, ARB_{BA} and ARB_R are Adaptive Residual Block with bottleneck path; ARB with bottleneck and adaptive paths; ARB with bottleneck and residual path respectively. As shown in Table 5, MPRNet with all learning pathways (ARB) achieves the best performance among all the mentioned ResBlock and

[1] Additional analyses (such as Inference time, Memory consumption, and etc.) and more visual results can be found in supplementary material.

Table 3: Effect of Attention Mechanisms and proposed Adaptive Residual Block on SOTA models. The best **PSNR** (dB) are highlighted.

Name	EDSR					RCAN					MSRN				
	Baseline	Attention Modules			ResBlock	Baseline	Attention Modules			ResBlock	Baseline	Attention Modules			ResBlock
Channel and spatial attention residual[16]		✓					✓					✓			
Enhanced Spatial Attention[27]			✓					✓					✓		
Two-Fold Attention Module[Ours]				✓	✓				✓	✓				✓	✓
Adaptive Residual Block[Ours]					✓					✓					✓
PSNR on Set5 (×4)	32.46	32.48	32.51	32.54	**32.65**	32.63	32.64	32.67	32.70	**32.78**	32.25	32.27	32.30	32.34	**32.39**
PSNR on Urban100 (×4)	26.64	26.66	26.69	26.71	**26.79**	26.82	39.84	26.86	26.89	**26.96**	26.22	26.25	26.29	26.32	**26.41**

Table 4: Impact of different Attention Mechanisms on MPRNet.

Dataset	Baseline	SE	CBAM	CSAR	ESA	TFAM
Set14 (×4)	28.57	28.59	28.54	28.61	28.64	**28.67**
Urban100 (×4)	26.19	26.21	26.18	26.23	26.25	**26.29**

Table 5: Effect of different configs of Residual Block and each learning pathway of the Adaptive Residual Block

Configs	MobileNet BnBlock	EDSR ResBlock	RCAN ResBlock	ARB_B	ARB_{BA}	ARB_R	ARB
BN_p				✓	✓	✓	✓
Adp_p					✓		✓
Res_p						✓	✓
B100 (×4)	27.24	27.44	27.52	27.46	27.58	27.55	**27.63**
Urban100 (×4)	25.79	25.96	26.08	26.05	26.15	26.11	**26.31**

Table 6: Study on combining different Residual Connections.

Options		Baseline	$1st$	$2nd$	$3rd$	$4th$	$5th$
Residual Learning Connections	LRC	✗	✓			✓	✓
	GRC	✗		✓		✓	✓
	LRSC	✗			✓		✓
PSNR on Set5 (×3)		34.42	34.40	34.47	34.45	34.52	**34.57**
PSNR on Urban100 (×3)		28.30	28.29	28.35	28.33	28.38	**28.42**

combinations of different ARB learning pathways. This is caused by allowing the main parts of network to focus on more informative components of the LR features and force the network to focus more on abstract features, which are important in SR tasks. Furthermore, the proposed pathways helps the model to converge better and performs better than all the baseline models. In a nutshell, information propagates locally via residual path, adaptively extract the informative features via adaptive path, and learn more meaningful spatial information by Bottleneck path. By doing so, information is transmitted by multiple pathways inside of ARB and main parts of network access to more expressive and richer feature maps, resulting in superior PSNR.

Effect of Residual Learning Connections. Table 6 shows the extensive study of the impact of Residual Learning Connections on our design of MPRNet, i.e. Local Residual Connection (LRC), Global Residual Connection (GRC), and Long Range Skip Connection (LRSC). In this work, residual connections except LRSC comprise concatenation followed by a 1×1 Conv layer. As we can see, MPRNet without any residual connection performs relatively low (i.e. baseline). However, MPRNet with only GRC in Residual Module shows better performance than baseline since GRC transports the information from mid- to high-layers and helps the model to better leverage multi-level representations by collecting all information before the next module.

On the contrary, MPRNet with only LRC inside Residual Concatenation Block could not perform better than the MPRNet with GRC. This behavior was expected as mentioned in [11] that 1×1 Conv layer on the residual connection can confuse optimization and prevent information propagation due to multiplicative manipulations. However, MPRNet can show better performance by using both connections ($4th$ col.). This is due to GRC eases the information propagation issue that LRC suffers from.

To end this, LRSC also added to the MPRNet to carry the shallow information to high-level layers. Thus, information is transferred by multiple connections, which mitigates the vanishing gradient problem and network has access to multi-level representation. As a results, MPRNet with all connections ($5th$ col.) can performs greatly better.

Model Complexity Analysis. Figure 1 indicates the comparison regard to the model size and PSNR with 15 recent state-of-the-art SR models. Our MPRNet achieves the best performance among all the lightweight SR approaches with much fewer parameters and achieves better or comparable results when compared with computationally expansive models. This shows that our MPRNet is well-balanced in terms of model size and reconstruction results.

5. Conclusions

This paper proposes a novel lightweight network (MPRNet) that achieves the best performance against all existing lightweight SOTA approaches. The main idea behind of this work is to design an advanced lightweight network to deliver almost similar results to heavy computational networks. A novel Residual Module is proposed to let abundant low-level information to be avoided through multiple connections. In addition, an efficient Adaptive Residual Block is proposed to allows MPRNet achieves more rich feature-maps through the multi-path learning. Furthermore, to maximize the power of the network a Two-Fold Attention Module is proposed, which refine the extracted information along channel and spatial axes to further enhance the discriminative ability of the network. Extensive evaluations and comparisons are provided.

Acknowledgment

This work has been partially supported by the Spanish Government under Project TIN2017-89723-P; the "CERCA Programme / Generalitat de Catalunya"; and the ESPOL project CIDIS-56-2020.

References

[1] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 252–268, 2018.

[2] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012.

[3] Xiangxiang Chu, Bo Zhang, Hailong Ma, Ruijun Xu, Jixiang Li, and Qingyuan Li. Fast, accurate and lightweight super-resolution with neural architecture search. *arXiv preprint arXiv:1901.07261*, 2019.

[4] Xiangxiang Chu, Bo Zhang, Ruijun Xu, and Hailong Ma. Multi-objective reinforced evolution in mobile neural architecture search. *arXiv preprint arXiv:1901.01074*, 2019.

[5] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 11065–11074, 2019.

[6] Zhou Daquan, Qibin Hou, Yunpeng Chen, Jiashi Feng, and Shuicheng Yan. Rethinking bottleneck structure for efficient mobile network design. *arXiv preprint arXiv:2007.02269*, 2020.

[7] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014.

[8] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016.

[9] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):142–158, 2015.

[10] Hayit Greenspan. Super-resolution in medical imaging. *The Computer Journal*, 52(1):43–63, 2008.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.

[12] Xiangyu He, Zitao Mo, Peisong Wang, Yang Liu, Mingyuan Yang, and Jian Cheng. Ode-inspired network design for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1732–1741, 2019.

[13] Byeongho Heo, Sanghyuk Chun, Seong Joon Oh, Dongyoon Han, Sangdoo Yun, Youngjung Uh, and Jung-Woo Ha. Slowing down the weight norm increase in momentum-based optimizers. *arXiv preprint arXiv:2006.08217*, 2020.

[14] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. *arXiv preprint arXiv:1905.02244*, 2019.

[15] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[16] Yanting Hu, Jie Li, Yuanfei Huang, and Xinbo Gao. Channel-wise and spatial feature modulation network for single image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[17] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015.

[18] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[19] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016.

[20] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017.

[21] Duo Li, Aojun Zhou, and Anbang Yao. Hbonet: Harmonious bottleneck on two orthogonal dimensions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3316–3325, 2019.

[22] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 517–532, 2018.

[23] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2019.

[24] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[25] Ding Liu, Zhaowen Wang, Yuchen Fan, Xianming Liu, Zhangyang Wang, Shiyu Chang, and Thomas Huang. Robust video super-resolution with learned temporal dynamics. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2507–2515, 2017.

[26] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. In *Advances in Neural Information Processing Systems*, pages 1673–1682, 2018.

[27] Jie Liu, Wenjie Zhang, Yuting Tang, Jie Tang, and Gangshan Wu. Residual feature aggregation network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2359–2368, 2020.

[28] Andrei P Lobanov. Resolution limits in astronomical images. *arXiv preprint astro-ph/0503225*, 2005.

[29] Xiaotong Luo, Yuan Xie, Yulun Zhang, Yanyun Qu, Cuihua Li, and Yun Fu. Latticenet: Towards lightweight image super-resolution with lattice block.

[30] David Martin, Charless Fowlkes, Doron Tal, Jitendra Malik, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Iccv Vancouver:, 2001.

[31] Sivaram Prasad Mudunuri and Soma Biswas. Low resolution face recognition across variations in pose and illumination. *IEEE transactions on pattern analysis and machine intelligence*, 38(5):1034–1040, 2015.

[32] Abdul Muqeet, Jiwon Hwang, Subin Yang, Jung Heum Kang, Yongwoo Kim, and Sung-Ho Bae. Ultra lightweight image super-resolution with multi-attention layers. *arXiv preprint arXiv:2008.12912*, 2020.

[33] Tomer Peleg and Michael Elad. A statistical prediction model based on sparse representations for single image super-resolution. *IEEE transactions on image processing*, 23(6):2569–2582, 2014.

[34] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[35] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[36] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155, 2017.

[37] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE international conference on computer vision*, pages 4539–4547, 2017.

[38] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 114–125, 2017.

[39] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4799–4807, 2017.

[40] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[41] Zhangyang Wang, Shiyu Chang, Yingzhen Yang, Ding Liu, and Thomas S Huang. Studying very low resolution recognition using deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4792–4800, 2016.

[42] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–19, 2018.

[43] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5505–5514, 2018.

[44] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010.

[45] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938, 2017.

[46] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018.

[47] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018.

[48] Feiyang Zhu and Qijun Zhao. Efficient single image super-resolution via hybrid residual feature learning with compact back-projection network. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[49] Wilman WW Zou and Pong C Yuen. Very low resolution face recognition problem. *IEEE Transactions on image processing*, 21(1):327–340, 2011.

SUPPLEMENTARY MATERIAL

MPRNet: Multi-Path Residual Network for Lightweight Image Super Resolution

Armin Mehri
*Computer Vision Center, Edifici O, Campus UAB, 08193, Bellaterra, Barcelona, Spain
amehri@cvc.uab.es

Parichehr B.Ardakani
*Computer Vision Center, Edifici O, Campus UAB, 08193, Bellaterra, Barcelona, Spain
pbehjati@cvc.uab.es

Angel D. Sappa
+ESPOL Polytechnic University, Guayaquil, Ecuador
*Computer Vision Center, 08193, Bellaterra, Barcelona, Spain
sappa@ieee.org

Summary

The following items are contained in the supplementary material:

1) Memory Complexity Analysis
2) Inference Time and Memory Consumption
3) Width Multiplier
4) Additional Qualitative Results

1. Memory Complexity Analysis

In this section, we compare the proposed MPRNet with the most recent lightweight and expensive networks: LapSRN, VDSR, DRCN, SelNet, DRRN, MemNet, SRFBN, CARN, MSRN, OISR, CBPN, MAFFSRN, and LatticeNet in term of number of MAC operations (Multi-Adds) and reconstruction results (PSNR) to show the efficiency of the purposed MPRNet. In Fig. 1, reconstruction results (PSNR) and MAC (G), which shows the number of multiply-accumulate operations, are illustrated. As we can see, our MPRNet can achieve better results with a large gap among all the recent networks with less needed MAC operations; and even perform better than MSRN, which has more than 160 layers by only 13% of the total number of MSRN multiply-accumulate operations ($1365.4G$).

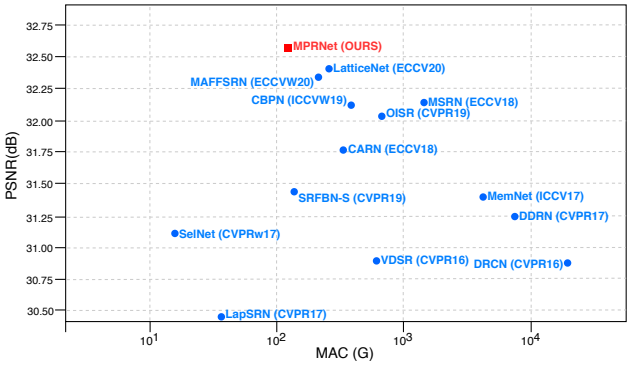

Figure 1: PSNR $vs.$ MAC on Urban100 for scale factor $\times 2$.

2. Inference Time and Memory Consumption

Table 1 illustrates the superiority of the proposed MPRNet in terms of Inference Time (s) and Memory Consumption (MB) when it compares with the recent light- and heavy-weight state of the art approaches on Urban100 for scale factor $\times 4$, namely MemNet, SRFBN, CARN, RCAN, RDN, EDSR. We consider the pyTorch version of MemNet instead of Caffe version due to large memory consumption in Caffe. The inference time and memory consumption of each approach is evaluated using their official code on the same environment. The MPRNet has the fastest inference time while using less memory compared to other approaches, which reflect the efficiency of the proposed method.

Table 1: Average Inference Time (s) and Memory Consumption (MB) comparisons with other SOTA models on Urban100 for scale factor $\times 4$.

Model	Params.	Time	Memory	PSNR
MemNet	$667K$	0.543	3,170	25.54
SRFBN-S	$483K$	0.0069	2,960	25.71
CARN	$1592K$	0.0047	3,015	26.07
RCAN	$16000K$	0.5927	2,731	26.82
RDN	$22000K$	0.0294	3,835	26.61
EDSR	$43000K$	0.0841	8,263	26.64
MPRNet [Ours]	$\mathbf{538K}$	**0.0095**	**2,154**	**26.31**

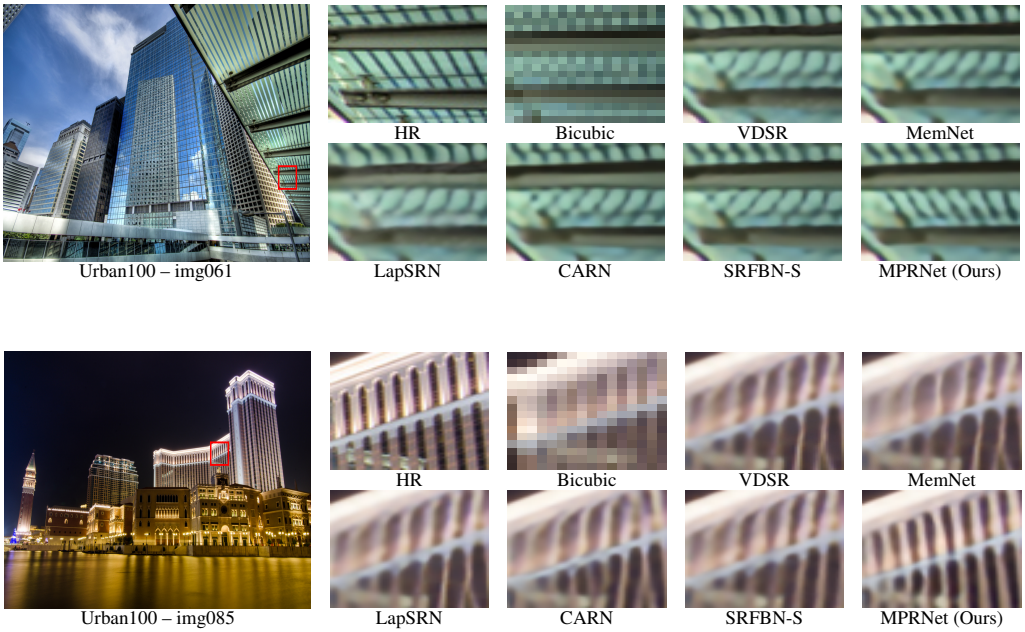

Figure 2: Qualitative results on **BI** degradation model with a scale factor $\times 4$ on $Urban100$ dataset.

3. Depth Multiplier

In Table 2, the effect of depth multiplier on model size and reconstruction results are illustrated. Similar to MobileNetV2, we employed depth multiplier ($alpha$) to make our MPRNet even more light cost with small reduction in performance. Depth multiplier is a float number between 0 and 1 that controls the depth of input layer. $\alpha = 1$ is the baseline model. By decreasing α, model size and computational cost are reduced. As can be seen, the proposed MPRNet with $372.7K$ ($\alpha = 0.25$), can achieve a good performance among the lightweight SOTA methods.

Table 2: Impact of Depth Multiplier on MPRNet

Depth Multiplier	1.0	0.75	0.5	0.25
# Parameters	$538.2K$	$470.4K$	$416.2K$	$372.7K$
Set114 ($\times 4$)	32.38	32.23	32.01	31.84
Urban100 ($\times 4$)	26.31	26.16	25.99	25.83

As we can see, by analyzing the number of parameters and MAC operations vs PSNR, inference time, memory consumption, and reconstruction result, the proposed MPRNet can prove that it is well-balanced in terms of speed, accuracy and computation cost.

4. Additional Qualitative Results

In this section, additional results are provided showing the superiority of the SR images obtained with the proposed model. Qualitative results with all degradation models (i.e., **BI**, **BD**, and **DN**) are presented below.

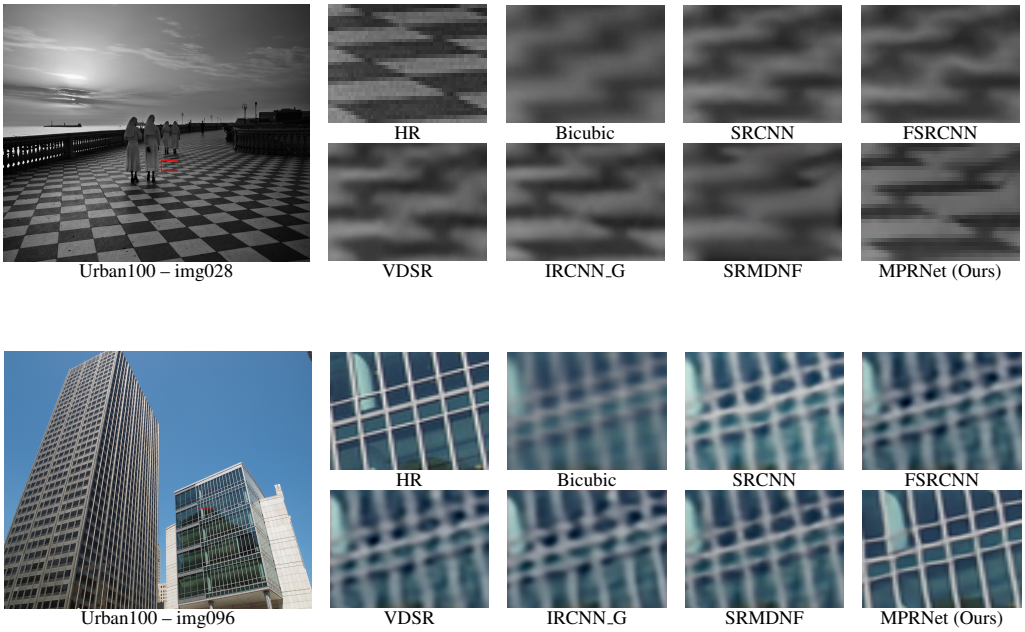

Figure 3: Qualitative results on **BN** degradation model with a scale factor $\times 3$.

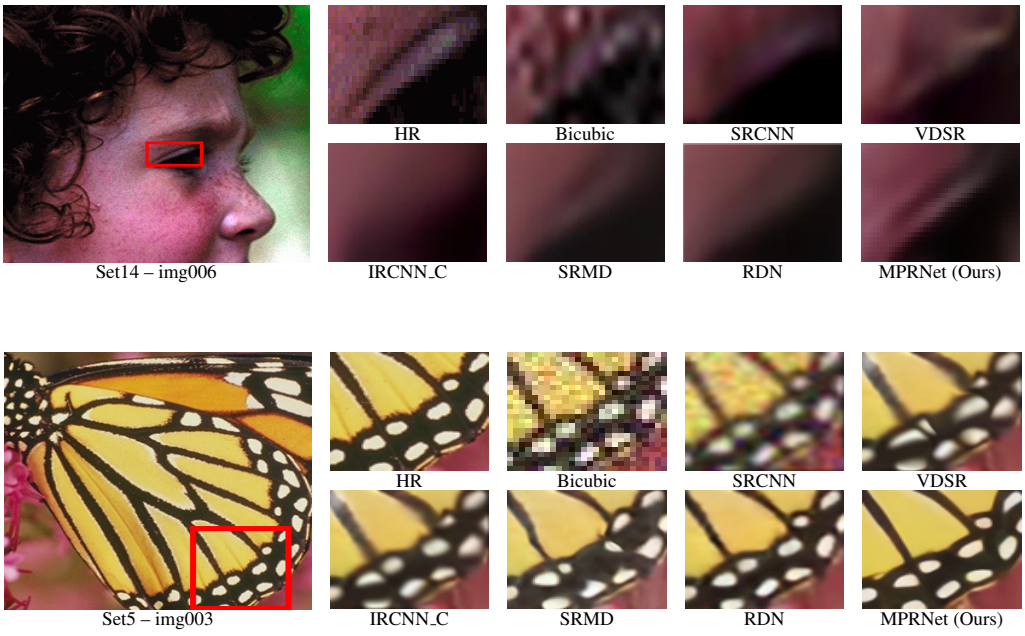

Figure 4: Qualitative results on **DN** degradation model with a scale factor $\times 3$.