IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Real-time Image Super-Resolution using Recursive Depthwise Separable Convolution Network

Kwok-Wai Hung[1], Zhikai Zhang[1], and Jianmin Jiang[1]
[1]College of Computer Science and Software Engineering, Shenzhen University, China

Corresponding author: Jianmin Jiang (e-mail: jianmin.jiang@szu.edu.cn).

This work was supported in part by (i) Natural Science Foundation China (NSFC) under the Grant No. 61602312, 61620106008; and (ii) Shenzhen Commission for Scientific Research & Innovations under the Grant No. JCYJ20160226191842793.

ABSTRACT In recent years, deep convolution neural networks (CNNs) have been widely used for image super-resolution (SR) to achieve a range of sophisticated performances. Despite of the significant advanced made on CNNs, it is still difficult to apply CNNs to practical SR applications due to enormous computations of deep convolutions. In this paper, we propose two lightweight deep neural networks using depthwise separable convolution for real-time image super-resolution. Specifically, depthwise separable convolution divides the standard convolution into depthwise convolution and pointwise convolution to significantly reduce the number of model parameters and multiplication operations. Moreover, recursive learning is adopted to increase the depth and receptive field of the network, in order to improve the super-resolution quality without increasing the model parameters. Finally, we propose a novel technique called Super-Sampling (SS) to learn more abundant high-resolution information by over-sampling the output image followed by adaptive down-sampling. The proposed two models named as SSNet-M and SSNet outperform the existing state-of-the-art real-time image super-resolution networks, including SRCNN, FSRCNN, ESPCN, VDSR, etc, in terms of model complexity, and subjective and PSNR/SSIM evaluations on Set5, Set14, B100, Urban100, and Manga109.

INDEX TERMS Super-resolution; depthwise separable convolution; super-sampling; recursive convolution

I. INTRODUCTION

Image super-resolution (SR) is to convert an observed low resolution (LR) into a high resolution image (HR) by adding plausible high-frequency information to improve the visual quality according to perception of the human visual system. Since converting from LR image to HR image is an ill-posed problem, there are infinite solutions of estimated HR image, such that image SR is a non-trivial task [1].

Conventionally, Bicubic interpolation has been widely adopted as the traditional method for image SR by using cubic polynomials to model the image signals, in order to interpolate the missing HR pixels [2]. However, the abrupt signal changes in natural images can hardly be modeled by 3rd order polynomials, such that adaptive edge-directed interpolation methods were proposed to model the edge characteristics to better reconstruct the edges [2-4].

Due to the rapid development of computer hardware, the processing power of computer has reached a certain level that learning-based approaches were widely developed to explicitly learn the relationship of the LR and HR images using the handcrafted models, which often involve well-designed components, including feature extraction, non-linear mapping and reconstruction, using large external database or sparse representations, etc, as the source of information for offline supervised learning [5-7].

Recently, with the increasing adaptability of neural networks, the generic deep learning models have been proven to provide end-to-end learning ability. Convolution neural networks (CNNs) can be trained for mapping the LR image into the HR image with plausible high-frequency information estimated from the deep convolution layers. Specifically, many super-resolution models based on deep CNNs have achieved obviously better results than previous approaches with expenses of higher computations [8-10].

On the other hand, the deep CNNs have been proposed for image restoration and denoising for remote sensing image [32] and hyperspectral Image [33], which utilizes the spatial-temporal-spectral deep residual CNNs to exploit the multi-modal information for handling image processing applications for significant performance improvements.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2929223, IEEE Access

Dong et al. proposed the first CNN model for image super-resolution called as SRCNN which pre-processes the LR image by Bicubic interpolation before feeding into the neural network [11]. Based on SRCNN, Dong et al. proposed FSRCNN to remove the Bicubic up-sampling of SRCNN by using the deconvolution layer to reconstruct the HR image at the last network layer [12]. Specifically, FSRCNN uses a funnel-shaped network structure, which adopts shrinking, mapping and expanding as non-linear mapping components of the network model, in order to reduce the number of model parameters. FSRCNN uses deconvolution layer at the end of the model, in order to achieve real-time image SR in some scenarios.

Shi et al. proposed a CNN model called as efficient sub-pixel convolutional neural network (ESPCN) which processes the image in the LR space to reconstruct the HR image at the last network layer [13]. However, different from FSRCNN, ESPCN reconstructs the HR image using the sub-pixel layer to perform pixel shuffling using features from the preceding convolution layer, instead of using deconvolution layer for HR image reconstruction to reduce the computation. Shi et al. proved in the paper that sub-pixel layer is an order of magnitude faster than deconvolution layer, which allows ESPCN to implement 1080P video real-time SR on a single K2 GPU.

Recently, many state-of-the-art super-resolution based on deep CNNs have been proposed, such as LapSRN [21], DBPN [22], DRCN [23], DRRN [24]. Although these networks significantly improve the previous CNNs for image super-resolution, their network complexity are often several orders higher than the aforementioned real-time SR methods [11-13]. Hence, there is an essential demand to develop a new deep CNN for real-time applications which outperforms the existing real-time SR networks for lower complexity.

In this paper, we propose two network models to tackle the weakness of existing state-of-the-art real-time image super-resolution networks based on deep learning. Specifically, the proposed networks called as SSNet-M and SSNet utilize depthwise separable convolutions to replace the standard convolutions, in order to significantly reduce the overall complexity. To maintain the same level of model parameters, we further introduce the recursive depthwise separable convolutions to reuse the shared model parameters for achieving a higher level of image quality through recursive iterations. Finally, we propose a novel technique for super-resolution called as super-sampling to over-sample the output image followed by adaptive down-sampling, in order to improve the generation of final image. The proposed models are extremely lightweight models with merely 7k and 22k model parameters, which outperform existing real-time super-resolution methods based on deep learning, such as SRCNN [11], FSRCNN [12], ESPCN [13], VDSR [14], in terms of subjective and objective quality evaluations, and model complexity.

In summary, our main contributions are as follows:

- We introduce the first recursive depthwise separable convolution network for image super-resolution, in order to formulate the real-time models with extremely low complexity, which improves the image quality while maintaining the same level of model parameters.
- We propose a novel technique called as super-sampling for generating the higher quality HR image for super-resolution. The key step is to over-sample the output image followed by adaptive down-sampling, in order to extracts more abundant HR information for the final HR generation.
- In comparisons with existing state-of-the-art real-time super-resolution models based on deep learning, the proposed models give obviously better subjective and objective quality but require lower model complexity in terms of model parameters and multiplication operations.

The rest of organization of this paper is as follows. Section II describes the related works of various techniques used in our models. Section III gives the details of our recursive depthwise separable convolution networks. Section IV shows the experimental results and ablation study of the proposed models. Section V concludes this paper and gives a future direction of our proposed works.

II. RELATED WORKS

A. Depthwise Separable Convolution

Andrew et al. proposed a depthwise separable convolution to greatly accelerate the computations of convolution neural network [15], in which depthwise convolution and pointwise convolution can be separated from standard convolutions. It shows that depthwise separable convolutions require much lower computations with slight performance loss.

Due to the high performance-computation ratio, many methods in different research fields have begun to apply depthwise separable convolution to implement the CNN. Chollet et al. proposed Xception Net, which introduces depthwise separable convolution into the task of image classification [16]. Xception is based on the inception net v3 proposed by Szegedy et al., which achieves better performance than Inception Net v3 with essentially the same parameters as Inception Net V3 [17]. Hence, it illustrates the practicality of depthwise separable convolution.

In the field of semantic segmentation, Liang-Chieh et al. deeply studied Xception and applied depthwise separable convolution to Atrous Spatial Pyramid Pooling and decoder modules, and proposed faster and stronger encoders-decoder network, DeepLabv3 [18]. Inspired by Xception and ByteNet [19], Lukasz et al. proposed to apply depthwise separable convolution to machine translation [20]. The method named as Slice Net achieved better performance while reducing the amount of model parameters compared to ByteNet.

The aforementioned researches has led us to discover the practicality of applying depthwise separable convolution to real-time image SR to propose SSNet-M and SSNet models.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2929223, IEEE Access

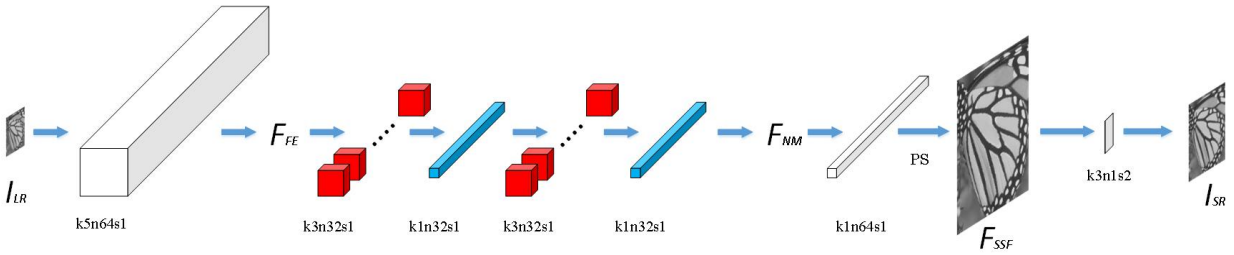

FIGURE 1. Proposed SSNet-M model

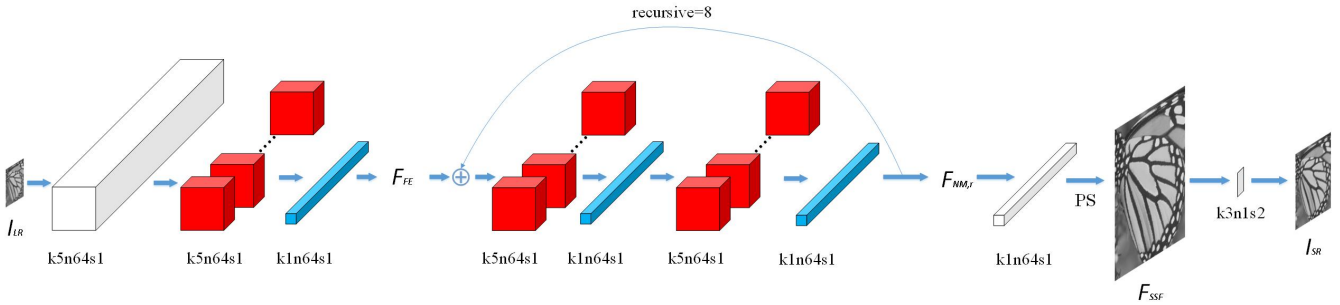

FIGURE 2. Proposed SSNet model

A. Up-down-sampling convolution for image SR

Lai et al. proposed the Laplacian Pyramid Super-Resolution Network (LapSRN) to progressively reconstruct the sub-band residuals of high resolution images [21]. At each pyramid level, LapSRN takes coarse-resolution feature maps as input, predicts the high-frequency residuals, and uses transposed convolutions for up-sampling to the finer level. Compared to gradually up-sampling input image in LapSRN, Muhammad et al. proposed DBPN, which uses a network structure in which up-sampling and down-sampling are performed alternately to improve the performance of the model by concatenating the information generated by these convolution kernels [22]. DBPN has achieved good results on tasks with large magnifications such as 8× magnification.

The SSNet-M and SSNet we proposed in this paper are inherently different from aforementioned methods [21-22]. The proposed super-sampling methods estimate the output image with size larger than the target HR image, and then the over-sampled output image is adaptively down-sampled by a convolution kernel to obtain the final output. Such design can estimate more abundant HR information to be adaptively selected (by the convolution kernel) for generating the final image. On the contrary, LapSRN [21] and DBPN [22] do not over-sample the output image larger than the target image.

C. Recursive Convolution Network

Kim et al. proposed the deeply-recursive convolutional network (DRCN) [23]. DRCN improves the performance of very deep super-resolution (VDSR) [14] by using a very deep recursive layer. Experiments show that when the CNN was trained to convergence, the parameters of each convolution layer are very similar. For this reason, the shared convolution kernels can greatly reduce the model parameters while preserving the model performance without significant deterioration. Tai et al. further improved DRCN and proposed the deep recursive residual network (DRRN) [24] which allows the model to learn local and global features for enhancing the expressive ability of image signals of the model. DRRN is a recursive CNN with a depth of 52 layers, but the amount of model parameters of the DRRN is not very high due to extensive recursive learning.

Through the study of DRCN and DRRN, the proposed SSNet model formulates the recursive depthwise separable convolution network using two shared depthwise separable convolution layers to greatly improve the performance of the model while maintaining the minimal model parameters.

III. PROPOSED METHOD

In this section, we will systematically describe the proposed lightweight deep models for real-time image super-resolution and analyze the techniques used in various parts of the models. In terms of techniques, we will introduce depthwise separable convolution, recursive depthwise separable convolution network, and super-sampling block. In terms of models, we proposed SSNet-M and SSNet based on aforementioned skills, as shown in Figure 1 and Figure 2. The proposed SSNet-M is a lightweight version, and SSNet can be considered an enhanced version. Let us describe the formulations of SSNet-M and SSNet in details in the following subsections.

A. SSNet-M

Our proposed SSNet-M is based on ESPCN [13] for significant modifications to develop the final lightweight model which provides better image quality than ESPCN but requires significantly lower computations. Different from

ESPCN that utilizes three standard convolutions and a pixel shuffling layer, the proposed SSNet-M uses one standard convolution, two depthwise separable convolutions, and a super-sampling block (including a pixel shuffling layer and a down-sampling layer) to perform super-resolution.

Let us describe the SSNet-M in details as follows. As shown in Table I and Figure 1, our proposed SSNet-M initially uses a 5×5 convolution kernel to generate a 64-channel feature map from the input LR image I_{LR},

$$F_{FE} = F_{Conv}(I_{LR}) \quad (1)$$

where F_{Conv} represents the standard convolution and F_{FE} represents the feature map extracted from the LR image. Then, two depthwise separable convolutions are adopted to perform non-linear mapping of extracted features,

$$F_{NM} = F_{PwConv}(F_{DwConv}(F_{PwConv}(F_{DwConv}(F_{FE})))) \quad (2)$$

where F_{DwConv} represents the depthwise convolution, F_{PwConv} represents the pointwise convolution, and F_{NM} represents the non-linearly mapped features generated from the first and second depthwise separable convolutions (separated into depthwise and pointwise convolutions) with kernel size 3×3. After that, the super-sampling block (SSBlock) is used to generate the super-sampling feature F_{SSF},

$$F_{SSF} = PS(F_{PwConv}(F_{NM})) \quad (3)$$

where the pointwise convolution F_{PwConv} expands the number of channels of the feature map F_{NM} from ($scale^2$) to ($scale^2$*SSS), where $scale$ refers to super-resolution factor and SSS refers to super-sampling-scale which controls the over-sampling rate of the target HR image. After that, the pixel shuffling (PS) [13] is applied to obtain the super-sampling feature F_{SSF} as shown in Figure 1.

Finally, SSBlock applies a 3×3 convolution F_{Conv} to downscale the super-sampled feature F_{SSF} by setting stride to 2 to generate the final output image I_{SR}, as follows,

$$I_{SR} = F_{Conv}(F_{SSF}) \quad (4)$$

As illustrated in Figure 1, k5n64s1 means that the size of the convolution kernel (k) is 5×5, the output channel (n) is 64, and the convolution stride (s) is 1. k3n1s2 means that the size of the convolution kernel (k) is 3×3, the output channel (n) is 1, and the convolution stride (s) is 2. Moreover, the red block indicates depthwise convolution, and the blue block indicates pointwise convolution.

Overall, the complexity of SSNet-M in terms of multiplication operations and model parameters is about 30% of ESPCN [13], but performance degradation caused by deep separable convolution is compensated by super-sampling block, such that the performance of SSNet-M is higher than ESPCN, as shown in the experimental section.

B. SSNet

Built upon the lightweight SSNet-M, we propose a enhanced version using a larger convolution kernel with multi-recursive learning, which is named as SSNet, as illustrated in Figure 2. Specifically, the SSNet inputs the extracted features F_{FE} from the input image to depthwise separable convolutions for non-linear mapping,

$$F_{NM} = F_{PwConv}(F_{DwConv}(F_{FE})) \quad (5)$$

where F_{NM} represents non-linearly mapped features generated from depthwise separable convolutions using 5×5 convolution kernel. After that, two shared depthwise separable convolutions are recursively implemented to iteratively enhance the non-linear mapping, as follows,

$$F_{NM,r} = F_{PwConv2}(F_{DwConv2}(F_{PwConv1}(F_{DwConv1}(F_{NM,r-1} + F_{NM})))) \quad (6)$$

where r means the number of recursion. In SSNet, the size of depthwise convolution kernel (red) and pointwise convolution kernel (blue) are 5×5 and 1×1 respectively, as shown in Figure 2. Moreover, the channels of convolution increases from 32 to 64 from SSNet-M to SSNet. Finally, the super-sampling block is applied to reconstruct the final HR image from the recursively estimated feature $F_{NM,r}$,

$$F_{SSF} = PS(F_{PwConv}(F_{NM,r})) \quad (7)$$

$$I_{SR} = F_{Conv}(F_{SSF}) \quad (8)$$

The PSNR performance of SSNet for 4× super-resolution on Set5 is 0.75dB higher than ESPCN [13], while the model parameters of SSNet is 22k, which is 2k lower than ESPCN. The reason for such significant improvement is the recursive depthwise separable convolutions as described in details in the following subsection. The details of network architecture of SSNet-M and SSNet are shown in Table I.

TABLE I
NETWORK ARCHITECTURES OF ESPCN AND PROPOSED SSNET-M AND SSNET FOR 4× SUPER-RESOLUTION

	ESPCN [13]	Proposed SSNet-M (Non-recursive)	Proposed SSNet (Recursive)
Feature extraction	Conv(5,64,1)	Conv(5,64,1)	Conv(5,64,1)
Nonlinear mapping	Conv(3,64,32)	Depthwise-Conv(3,64,32) Pointwise-Conv(1,32,32) Depthwise-Conv(3,32,32) Pointwise-Conv(1,32,32)	Depthwise-Conv(5,64,64) Pointwise-Conv(1,64,64) Depthwise-Conv(5,64,64) - Shared Pointwise-Conv(1,64,64) - Shared Depthwise-Conv(5,64,64) - Shared Pointwise-Conv(1,64,64) - Shared
Feature reconstruction	Conv(3,32,16) -Pixel shuffling (16→1 channel) -4× Super-resolution	Pointwise-Conv(1,32,64) -pixel shuffling (64→1 channel) -8× Super-resolution	Pointwise-Conv(1,64,64) -pixel shuffling (64→1 channel) -8× Super-resolution
Feature down-sampling	-	Conv(3,1,1)-strike = 2 -2× down-sampling	Conv(3,1,1)-strike = 2 -2× down-sampling
SR up-down ratio	1×→4×	1×→8×→4×	1×→8×→4×

C. MSE loss function

As mean squares error is the most widely used metric for image super-resolution, we define the loss function $L_{MSE}(.)$ as follows,

$$L_{MSE}(I_{LR}, I_{HR}) = \left\| F(I_{LR}) - I_{HR} \right\|^2 = \left\| I_{SR} - I_{HR} \right\|^2 \quad (9)$$

where I_{HR} represents the original HR image, I_{SR} represents the estimated HR image from proposed models, $F(.)$ is overall interference function of proposed SSNet or SSNet-M models for estimating the HR image from the LR image.

D. Cost of depthwise separable convolution for image SR

To explain the cost of standard convolutions and the depthwise separable convolutions, we consider one layer of convolutions to justify the substantial complexity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2929223, IEEE Access

reductions achieved by our proposed SSNet-M and SSNet for real-time image super-resolution.

Let us define the size of the input feature map to be $H\times W\times M$, and the size of the feature map generated by the convolution to be $H\times W\times N$, where H, W, M, N represents the height, width, input and output channels of the feature map. Hence, the cost of multi-add operations of the standard convolution can be calculated as [15]

$$Cost_{Conv} = K^2 \times H \times W \times M \times N \quad (10)$$

where K represents the height and width of the convolution kernel. In other words, the standard convolution multiples each M input features with the N convolution kernels with size $K\times K$ to generate the output features.

Depthwise separable convolution divides the standard convolution into depthwise convolution and pointwise convolution, where depthwise convolution applies element-wise product of a convolution kernel and each channel of input features independently, hence the cost of multi-add operations of depthwise convolution is,

$$Cost_{DwConv} = K^2 \times H \times W \times M \quad (11)$$

where depthwise convolution can greatly increase the speed of convolution by reducing multi-add operations by N times. Different from standard convolution, this step does not combine feature maps to generate new features, so we must add a 1×1 pointwise convolution followed by depthwise convolution to fuse the features together. Since K=1, the cost of multi-add operations of pointwise convolution is,

$$Cost_{PwConv} = H \times W \times M \times N \quad (12)$$

As a result, the ratio of the cost of depthwise separable convolution compared with the standard convolution is

$$\frac{Cost_{DwConv} + Cost_{PwConv}}{Cost_{Conv}} = \frac{1}{N} + \frac{1}{K^2} \quad (13)$$

For example, let us refer to k3n32s1, where K is 3 and N is 32, depthwise separable convolution requires about seven times lower computation than standard convolutions for only a small reduction in accuracy as adopted in our SSNet-M and SSNet model, as explained in preceding sub-sections.

E. Recursive depthwise separable convolution network

In the proposed SSNet, the recursive depthwise separable convolution network is proposed for achieving higher image quality without increasing the model parameters. Let us explain the methodology of recursive depthwise separable convolution network in this section.

For single-recursive block as shown in Figure 3, let us refer the input feature after the feature extraction layer as F_{FE}. Let us apply two independent non-sharing depthwise separable convolutions for initial non-linear mapping,

$$F_{NM} = F_{PwConv}(F_{DwConv}(F_{PwConv}(F_{DwConv}(F_{FE})))) \quad (14)$$

where F_{NM} represents the output of two depthwise separable convolutions, F_{DwConv} and F_{PwConv} represent two non-sharing depthwise convolutions and pointwise convolutions. Let us utilize a single depthwise separable convolution (depthwise convolution and pointwise convolution) as the structure of single-recursive block (SRB), which can be represented as,

$$F_{NM,r} = F_{PwConv}(F_{DwConv}(F_{NM,r-1} + F_{FE})) \quad (15)$$

where r represents the number of recursions and the depthwise and pointwise convolutions are shared. Since the convolution kernel is used recursively, the convolution kernel parameters W_r and bias values b_r are equal for each convolution operation, which can be illustrated as,

$$W_1, b_1 = W_2, b_2 = W_3, b_3 = \cdots = W_r, b_r \quad (16)$$

where W_r and b_r represent the weight and bias of the shared depthwise separable convolution. Although the recursive use of the convolution kernel will not increase the model parameters, the network performance usually improves due to deeper convolutions to increase the receptive field.

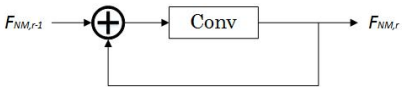

FIGURE 3. Single-recursive block (SRB).

In our paper, we use two shared recursive depthwise separable convolutions to formulate the multi-recursive block (MRB), in order to further improve the model performance. The structure of the multi-recursive block is shown in Figure 4, which can be formulated as follows,

$$F_{NM} = F_{PwConv}(F_{DwConv}(F_{FE})) \quad (17)$$

$$F_{NM,r} = F_{PwConv2}(F_{DwConv2}(F_{PwConv1}(F_{DwConv1}(F_{NM,r-1} + F_{FE})))) \quad (18)$$

$$W_{1,1}, W_{1,2} = W_{2,1}, W_{2,2} = \cdots = W_{r,1}, W_{r,2} \quad (19)$$

$$b_{1,1}, b_{1,2} = b_{2,1}, b_{2,2} = \cdots = b_{r,1}, b_{r,2} \quad (20)$$

where $F_{DwConv1}$ and $F_{DwConv2}$ represent the first and the second shared depthwise convolutions, the first subscript of W and b represents the number of recursion, and the second subscript represents the first or the second depthwise separable convolution and bias. Although the multi-recursive block requires the same model parameters of the single recursive block, our experimental results show that MRB provides much better performance than SRB.

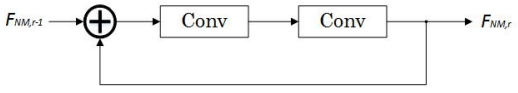

FIGURE 4. Multi-recursive block (MRB).

F. Super-sampling block

Let us explain the methodology of the proposed super-sampling block (SSBlock) in details in this section given in Figure 5. The objective of SSBlock is to estimate more abundant HR information by over-sampling the features for final HR image reconstruction. In other words, SSBlock reconstructs the output image beyond the size of the target image, where the over-sampled image is adaptively down-sampled by a convolution layer with strike=2 to the desired image size. Such design can significantly improve the final HR reconstruction with slight computation increment.

As illustrated in Figure 5, the proposed super-sampling block for SSNet-M and SSNet inputs the features F_{NM} from the output of the preceding nonlinear mapping layer. Then, the pointwise convolution k1n64s1 and pixel shuffling (*PS*)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2929223, IEEE Access

are used together to generate the feature maps to a size larger than the size of the target HR image, called as the super-sampling features F_{SSF}. After that, SSBlock applies a 3×3 convolution k3n1s2 to downscale the feature map by setting stride (s) to 2 to generate the final output I_{SR}.

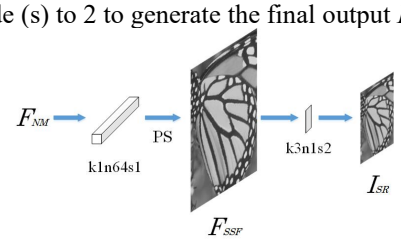

FIGURE 5. Super-sampling block (SSBlock).

IV. EXPERIMENTAL RESULTS

In this section, we will show extensive experimental results of the proposed SSNet-M and SSNet, including the training and testing details, as well as ablation study of different components of our models. Specifically, we analyze the contribution of each sub-network structure to the performance and complexity of the model. Finally, we compare SSNet-M and SSNet to other real-time image super-resolution methods on standard benchmark datasets.

A. Training and testing details

We used DIV2K dataset [25] as the training database for training the models, where DIV2K contains 800 training images, 100 validation images and 100 testing images. The 800 training images were down-sampled using bicubic interpolation to generate the LR training images. The patch size of LR samples is 32×32 and the batch size is 32. We set the initial learning rate to 0.001 and decreased the learning rate by a factor of 10 at each 30 epochs. All the weights in the model were initialized with Kaiming distribution. For optimization, we used Adam with $\beta1 = 0.9$, $\beta2 = 0.999$, and epsilon=10^{-8}. For the platform, we used Tensorflow 1.12.0 on Ubuntu 16.04.4, which is installed with CUDA 9 and CUDNN 7. The GPU for training and testing the models is Nvidia 980Ti which runs for few hours to finish the training process. We trained the models for 100 epochs to ensure convergence, as shown in Figure 6.

For the testing settings, the Y channels in YCbCr space of the output of models are evaluated using PSNR and SSIM measurements. For test datasets, we used Set5 [26], Set14 [27], Urban100, BSDS100 [28] and Manga109 [29].

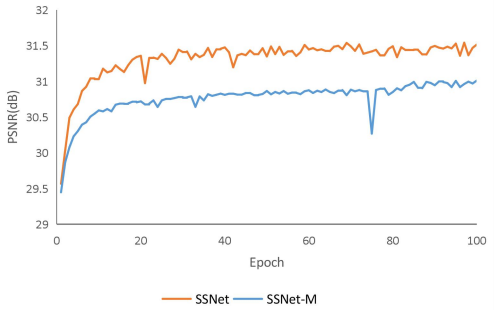

FIGURE 6. PSNR(dB) of each epoch on Set5 during training

B. Ablation study

(1) SSNet-M

In this section, we study the contribution of each component of the SSNet-M in Table II. Since the ESPCN is the baseline of SSNet-M, we study the contribution of each change from ESPCN that formulates SSNet-M. PSNR evaluations on Set5 of trained models using 100 epochs are shown in different configurations of Table II. Initially, we replace the standard convolutions in ESPCN by depthwise separable convolutions, which results into 0.2 dB drop in PSNR with reduction of 19k model parameters, as shown in configuration 2 of Table II.

To compensate for the performance loss due to depthwise separable convolution, the super-sampling block is applied. To evaluate the contributions of the super-sampling mechanism with minimal increment of model parameters, we increase the channel output of the last pointwise convolution of non-linear mapping layers before the SSBlock from Pointwise-Conv(1,32,32) to Pointwise-Conv(1,32,64). Hence, we remove the first pointwise convolution F_{PwConv} in the SSBlock, i.e., Conv(1,32,64) in Table I. As shown in configuration 3 of Table II, this incomplete SSBlcok increases the PSNR by 0.18 dB with the increment of 1k model parameters.

After that, the 1×1 pointwise convolution F_{PwConv} is added as the complete SSBlock which further increases the PSNR by 0.24 dB for another 1k increment of model parameters, as shown in configuration 4 of Table II. We believe that adding a 1×1 pointwise convolution at the front of SSBlock can fuse the output of preceding non-linear mapping layers before super-sampling, which significantly improves the results.

For the SSNet-M experiments, we also changed the super-sampling-scale (*SSS*) to 2×, 3× and 4× respectively. Throughout experiments, we found that it is best to extraordinarily over-sample the output image by 2 times and then down-sample it by 2 times. The higher over-sampling rate does not help to improve the performance of the model. The experimental results can be seen in configurations 4, 5, 6 in Table II. Based on these results, we used 2× super-sampling scale in the following experiments.

(2) SSNet

In the SSNet experiments, we tested the model with single-recursive block and set the number of recursive times to 4, 8, 12, and 16 times. We can see from the results in Table II (configurations 7 to 10) that when the number of recursions exceeds 12 times, the increase in recursions will have the opposite effect, as illustrated in Figure 7 (left). The best results can be achieved when the number of recursions is 12, and the corresponding PSNR is 31.36 dB.

Then, we use multi-recursive convolution to further improve performance. First, we use a 3×3 depthwise convolution kernel for 8 and 12 recursions, as shown in Table II (configurations 11 and 12). It can be seen that multi-recursive block gives a significant improvement in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2929223, IEEE Access

model performance compared to single-recursive block, but we believe there is still room for improvement.

In order to increase the receptive field of the model, we replaced the 3×3 depthwise convolution kernel with a 5×5 depthwise convolution kernel. As shown in Table II (configurations 13-16), the experimental results show that the larger convolution kernel significantly improves the performance of the model. Moreover, the results of images generated by the different recursive times of the final model are shown in Figure 7 (right) and Figure 8, where 8 recursions are better than those of 12 and 16 recursions. Experimental results show that when the number of recursions is too excessive, the performance of the model will be saturated because the shared model parameters of the convolution kernel are limited. The results of the 8 recursive models are superior to other models using a multi-recursive convolution kernel. The number of parameters of the SSNet model is 22k, which is still 2k less than ESPCN. However, the final SSNet model achieves 0.75 dB gain in PSNR value compared with ESPCN.

Finally, let us evaluate the visual quality of the SSNet with and without the SSBlock as shown in Figure 9. The effectiveness of SSBlock is affirmative as illustrated in Figure 9 which shows that the edges and textures are much better reconstructed using the SSBlock to over-sample the output image followed by down-sampling. Specifically, the edges are sharper and smoother by using the SSBlock to generate the image with less halo artifacts. Moreover, the output of SSBlock images contain some checkerboards which will be eliminated by the strike-2 convolution after down-sampling, as shown in middle columns of Figure 9.

TABLE II
ABLATION STUDY OF PROPOSED SSNET-M AND SSNET FOR 4× SR ON SET5

Config.		Depthwise separable convolution	SSBlock	PwConv 1×1	Super-Sampling Scale	Single-Recursive Block	Multi-Recursive Block	Recursion times	Kernel size of convolutions	PSNR(dB) (Set5)	Model parameters
1	ESPCN [13]	-	-	-	-	-	-	-	3×3	30.79	24k
2	-	√	-	-	-	-	-	-	3×3	30.59	5k
3	-	√	√	-	2×	-	-	-	3×3	30.77	6k
4	**SSNet-M**	√	√	√	**2×**	**-**	**-**	**-**	**3×3**	**31.01**	**7k**
5	-	√	√	√	3×	-	-	-	3×3	30.91	10k
6	-	√	√	√	4×	-	-	-	3×3	30.93	13k
7	-	√	√	√	2×	√	-	4	3×3	31.16	19k
8	-	√	√	√	2×	√	-	8	3×3	31.26	19k
9	-	√	√	√	2×	√	-	12	3×3	31.36	19k
10	-	√	√	√	2×	√	-	16	3×3	31.24	19k
11	-	√	√	√	2×	-	√	8	3×3	31.45	19k
12	-	√	√	√	2×	-	√	12	3×3	31.40	19k
13	-	√	√	√	2×	-	√	4	5×5	31.36	22k
14	**SSNet**	√	√	√	2×	**-**	√	**8**	**5×5**	**31.54**	**22k**
15	-	√	√	√	2×	-	√	12	5×5	31.48	22k
16	-	√	√	√	2×	-	√	16	5×5	31.47	22k

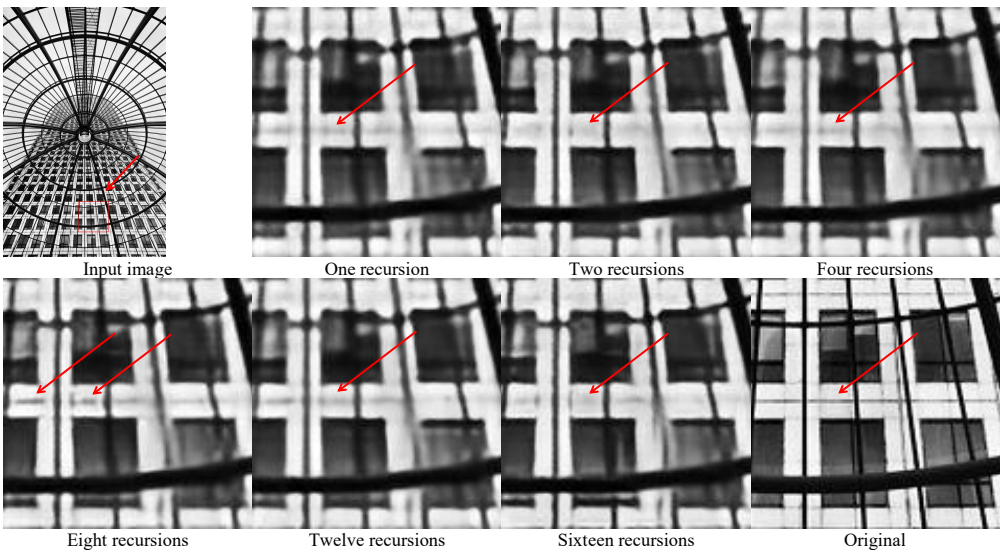

FIGURE 8. SR results of different recursion times in our SSNet model on img_72 in Urban100 data set

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2929223, IEEE Access

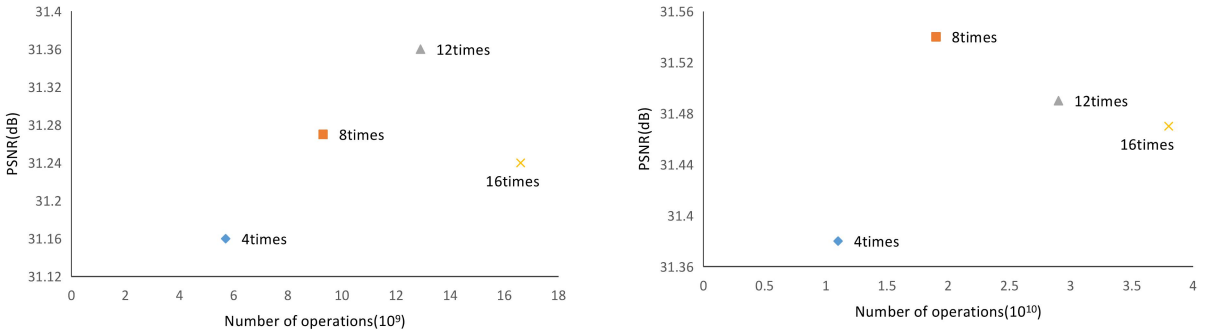

FIGURE 7. Appropriate number of recursions for single-recursive block (left) and multi-recursive block (right)

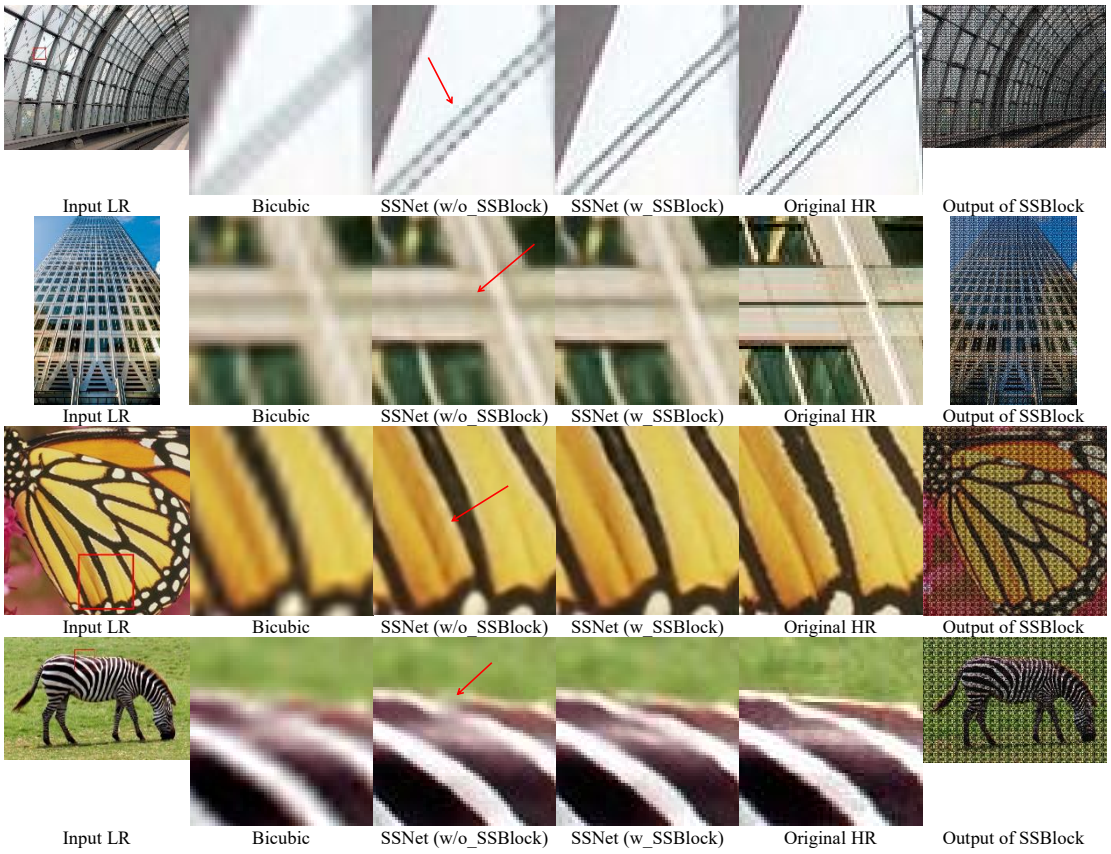

FIGURE 9. SR results of effectiveness of SSBlock in our SSNet model on different images in data sets

C. Comparison with state-of-the-art real-time image SR methods

(1) Complexity evaluations and execution time

In Table III, we compare the complexity in terms of model parameters, multiplication operations and runtime of our models with state-of-the-art real-time or recursive image SR methods, including SRCNN [11], FSRCNN [12], ESPCN [13], VDSR [14], DRCN [23] and DRRN [24]. For calculating the operations, we consider the multiplications operations for test images with size 540×360. For the run-time, we measure the 4× SR for baby image in Set5. Specifically, the proposed SSNet-M requires significantly lower model parameters, operations and runtime than ESPCN, FSRCNN, and SRCNN. Hence, SSNet-M is an extremely lightweight network model for a very broad range of applications in the real world, such as real-time video SR and other scenarios where the computing speed is critical, as shown in Figure 10 and Figure 11.

SSNet is an enhanced version of SSNet-M with a parameter count of only 22k, which is 2k lower than ESPCN. As illustrated in Table III, Figure 10 and Figure 11, SSNet requires much lower complexity in terms of parameters and multiple operations than the comparable models, such as VDSR, DRCN, DRRN. Hence, SSNet is suitable for scenarios which requires high super-resolution performance and computational efficiency. It has the same very wide application prospects as SSNet-M.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2929223, IEEE Access

(2) Subjective quality and PSNR/SSIM evaluations

Table IV shows the PSNR/SSIM comparisons of the proposed models with state-of-the-art image SR methods, including A+ [27], SCN [30], RFL [31], SRCNN [11], FSRCNN [12], ESPCN [13], VDSR [14], DRCN [23] and DRRN [24]. The PSNR/SSIM performance of proposed SSNet-M outperforms A+, SRCNN, SCN, RFL, ESPCN and FSRCNN but requires lower network complexity. The performance of SSNet is very competitive compared to VDSR, DRCN and DRRN, but the model parameters and operations of SSNet is at least an order lower than these methods, as shown in Table III. Figure 10 and Figure 11 give the PSNR comparisons with respect to model parameters and operations of various methods. In summary, our SSNet-M and SSNet obviously outperform existing state-of-the-arts SR methods in terms of efficiency for the same or better PSRN/SSIM performance.

Figure 12 to Figure 16 give the subjective evaluations of various SR methods for 4× SR on the test datasets. Specifically, the edges and texture regions reconstructed by SSNet-M are better than those reconstructed by A+, SRCNN, FSRCNN, and ESPCN. Although the SSNet-M requires much lower complexity than the aforementioned methods, the edge sharpness and clearness of SSNet-M is obviously better, as pointed by blue arrows. Moreover, the halo artifacts of ESPCN are more obvious than SSNet-M.

For SSNet, the subjective quality of images generated by SSNet is affirmatively better than VDSR and the aforementioned SR methods, as pointed by red arrows. Specifically, VDSR and other SR methods produce more blurry and aliased edges and texture regions, which are better reconstructed by the proposed SSNet.

Overall, the subjective quality evaluations in Figure 12 to Figure 16 generally agree with the objective PSNR/SSIM evaluations in Table IV. The objective and subjective quality of SSNet-M and SSNet is on a par with or better than existing state-of-the-art SR methods but the proposed SSNet-M and SSNet require much lower complexity.

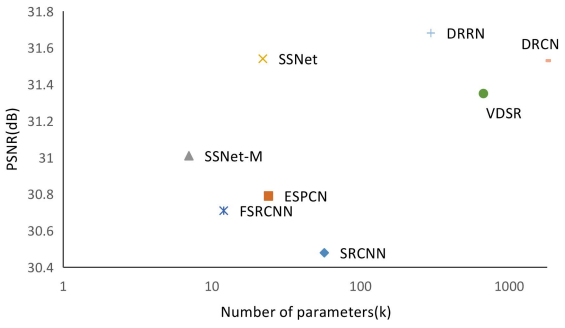

FIGURE 10. PSNR(dB) vs parameters on Set5

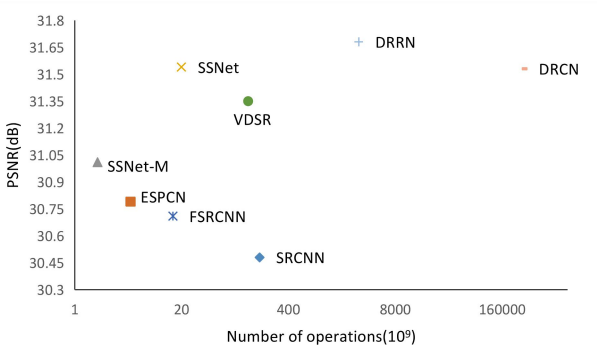

FIGURE 11. PSNR(dB) vs operations on Set5

TABLE III
COMPLEXITY OF IMAGE SR METHODS FOR 4× SR

	Layers	Parameters	Operations	Runtime
SSNet-M	4	7k	$1.9\text{x}10^{9}$	7.17ms
ESPCN [13]	3	24k	$4.79\text{x}10^{9}$	12.53ms
FSRCNN [12]	8	12k	$1.57\text{x}10^{10}$	24.22ms
SSNet	19	22k	$1.99\text{x}10^{10}$	135.33ms
VDSR [14]	20	670k	$12.9\text{x}10^{10}$	245.91ms
SRCNN [11]	3	57k	$1.78\text{x}10^{11}$	616.84ms
DRRN [23]	52	297K	$2.87\text{x}10^{12}$	1426.20ms
DRCN [24]	20	1774K	$2.81\text{x}10^{14}$	41910.80ms

TABLE IV
PSNR/SSIM RESULTS OF 2×, 3×, 4× IMAGE SR

Algorithm	Scale	Set5	Set14	Urban100	B100	Manga109
Bicubic	2	33.65/0.930	30.34/0.870	26.88/0.841	29.56/0.844	30.84/0.935
A+[27]		36.54/0.954	32.40/0.906	29.23/0.894	31.22/0.887	35.33/0.967
SRCNN[11]		36.66/0.954	32.45/0.906	29.50/0.894	31.53/0.892	35.72/0.968
FSRCNN[12]		37.00/0.956	32.63/0.908	29.88/0.902	31.58/0.890	36.62/0.971
SCN[30]		36.52/0.953	32.42/0.904	29.50/0.896	31.24/0.884	35.47/0.966
RFL[31]		36.55/0.954	32.36/0.905	29.13/0.891	31.16/0.885	35.08/0.966
ESPCN[13]		36.91/0.954	32.61/0.907	29.73/0.898	31.29/0.888	36.06/0.968
SSNet-M		37.08/0.956	32.78/0.908	29.92/0.900	31.46/0.900	36.47/0.970
SSNet		37.51/0.958	33.19/0.912	30.88/0.914	31.79/0.894	37.55/0.974
VDSR[14]		37.53/0.958	33.03/0.912	30.76/0.914	31.90/0.896	37.16/0.974
DRCN[23]		37.63/0.957	33.04/0.912	30.75/0.913	31.85/0.894	-
DRRN[24]		37.74/0.959	33.23/0.914	31.23/0.919	32.05/0.897	-
Bicubic	3	30.39/0.868	27.64/0.776	24.46/0.736	27.21/0.740	26.98/0.858
A+[27]		32.60/0.908	29.24/0.821	26.05/0.798	28.30/0.784	29.91/0.911
SRCNN[11]		32.75/0.909	29.28/0.821	26.24/0.799	28.41/0.786	30.58/0.913
FSRCNN[12]		33.16/0.914	29.43/0.824	26.43/0.808	28.53/0.791	31.09/0.920
SCN[30]		32.60/0.907	29.24/0.819	26.21/0.801	28.32/0.782	30.21/0.912
RFL[31]		32.45/0.905	29.15/0.819	26.21/0.801	28.32/0.782	30.21/0.912
ESPCN[13]		33.06/0.912	29.36/0.821	26.27/0.800	28.31/0.785	30.70/0.912
SSNet-M		33.24/0.915	29.51/0.824	26.44/0.805	28.42/0.788	31.01/0.918
SSNet		33.89/0.922	29.83/0.830	27.09/0.826	28.69/0.795	32.14/0.932
VDSR[14]		33.66/0.921	29.77/0.831	27.14/0.827	28.82/0.798	31.99/0.933
DRCN[23]		33.82/0.923	29.76/0.831	27.15/0.828	28.80/0.796	-
DRRN[24]		34.03/0.924	29.96/0.835	27.53/0.838	28.95/0.800	-
Bicubic	4	28.42/0.810	26.00/0.702	23.14/0.658	25.96/0.668	24.89/0.780
A+[27]		30.28/0.860	27.32/0.749	24.32/0.718	26.82/0.709	27.02/0.850
SRCNN[11]		30.48/0.863	27.49/0.750	24.52/0.727	26.90/0.710	27.58/0.85
FSRCNN[12]		30.71/0.865	27.59/0.756	24.61/0.727	26.98/0.715	27.89/0.859
SCN[30]		30.39/0.862	27.48/0.751	24.52/0.725	26.87/0.710	27.39/0.856
RFL[31]		30.15/0.853	27.33/0.748	24.20/0.740	26.75/0.707	26.80/0.840
ESPCN[13]		30.79/0.866	27.66/0.750	24.49/0.718	26.75/0.707	27.65/0.800
SSNet-M		31.01/0.873	27.61/0.755	24.62/0.725	26.84/0.710	27.87/0.858
SSNet		31.54/0.884	28.01/0.764	25.08/0.746	27.08/0.738	28.90/0.879
VDSR[14]		31.35/0.882	28.01/0.770	25.18/0.753	27.24/0.726	28.82/0.886
DRCN[23]		31.53/0.885	28.02/0.767	25.14/0.751	27.23/0.723	-
DRRN[24]		31.68/0.889	28.21/0.772	25.44/0.764	27.38/0.728	-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2929223, IEEE Access

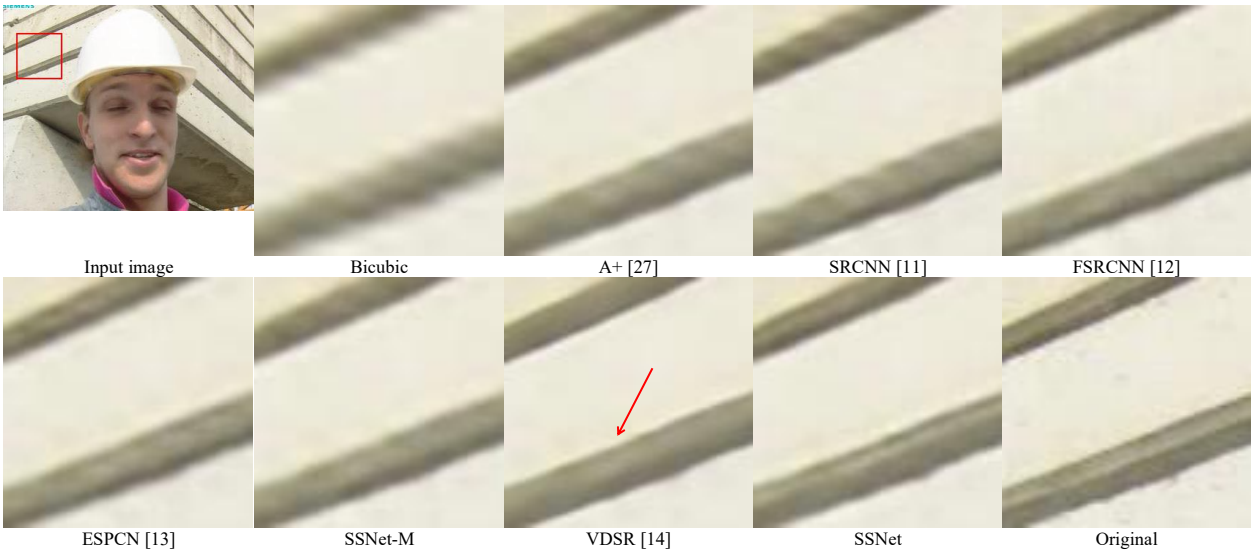

FIGURE 12. 4× Image SR results of foreman in Set14 data set

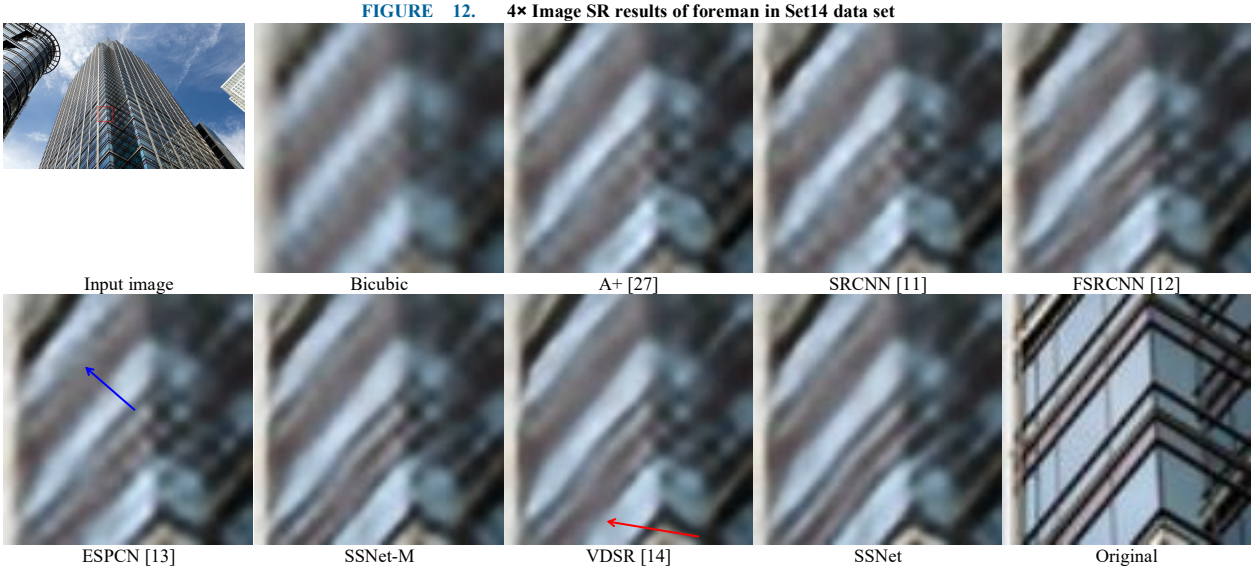

FIGURE 13. 4× Image SR results of img_047 in Urban100 data set

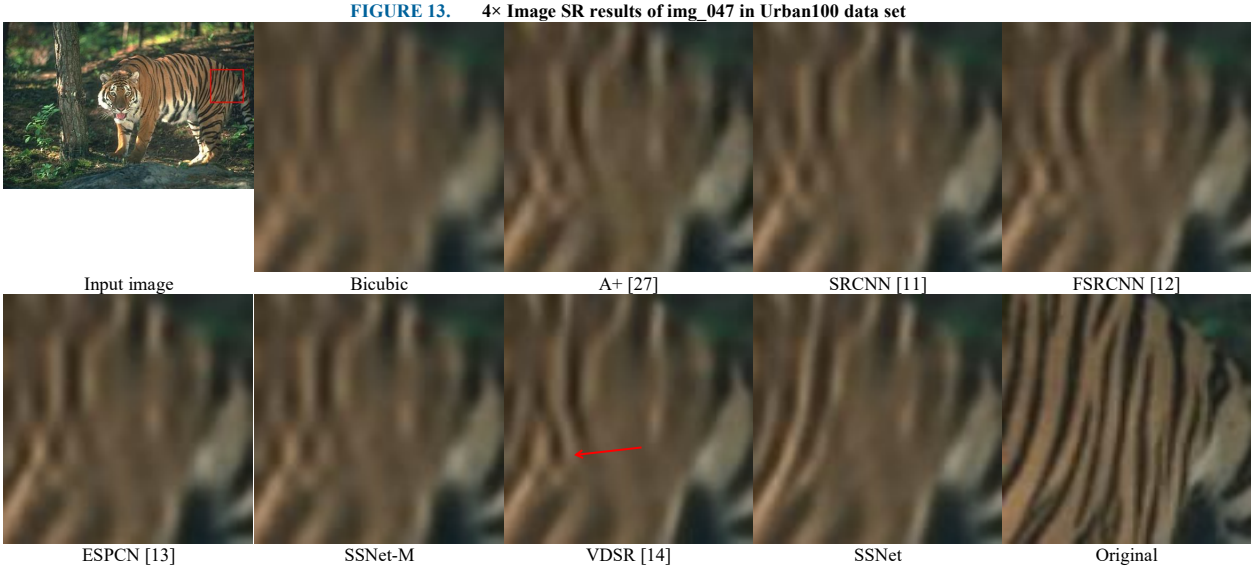

FIGURE 14. 4× Image SR results of img_007 in B100 data

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2929223, IEEE Access

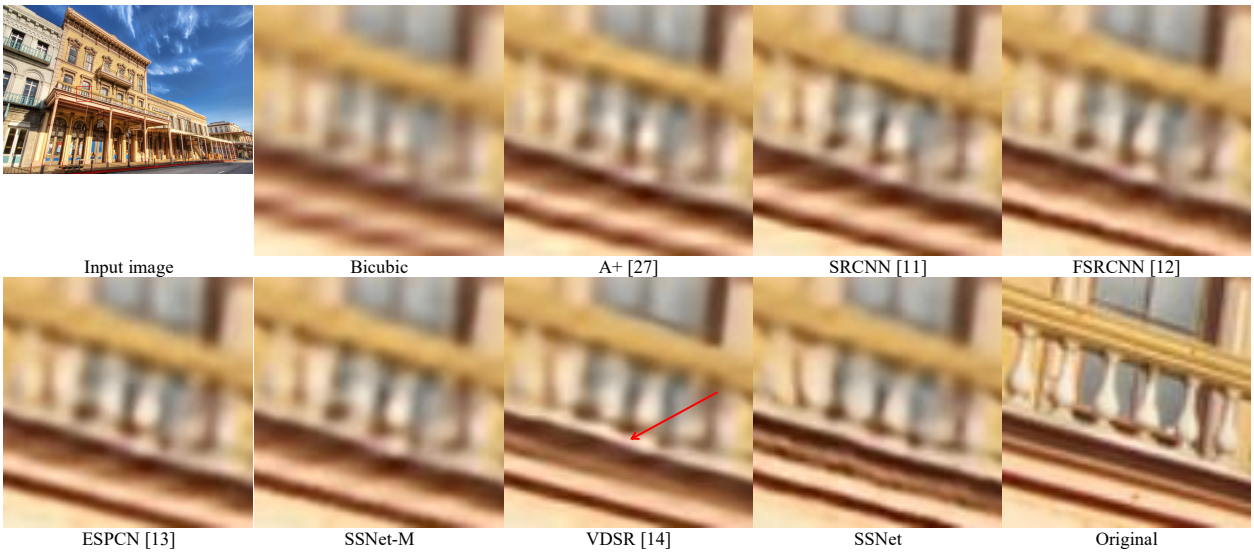

FIGURE 15. 4× Image SR results of img_017 in Urban100 data set

FIGURE 16. 4× Image SR results of Dual Justice in Manga109 data set

V. CONCLUSION

Due to the wide applications of real-time image super-resolution in various research areas, it is highly desirable to propose new algorithms for improving the existing real-time image super-resolution networks based on deep learning. In this paper, we propose two novel real-time image super-resolution models using recursive depthwise separable convolutions and super-sampling technique. Specifically, the complexity of our models are extremely low by replacing the standard convolutions with depthwise convolution and pointwise convolutions. Moreover, the recursive learning is incorporated to recursively refine the features without increasing model parameters. Eventually, a novel technique called as super-sampling is used to over-sample the output image for estimating more abundant high-resolution information followed by adaptive down-sampling to generate the final output HR image.

The proposed SSNet-M and SSNet models are extremely lightweight networks which require merely 7k and 22k model parameters with less multiplication operations. Hence, proposed models are suitable for implementations on memory-limited devices such as mobile phones and embedded systems, etc. Our experiments show that the proposed SSNet-M model can perform real-time video SR for converting 128×128 to 512×512 for over 139 fps using 980 Ti GPU without optimization. Compared with existing real-time SR methods, our models achieve better subjective and objective image quality with lower complexity. We demonstrate the effectiveness of our approaches through a series of experiments and validate our models on Set5, Set14, Urban100, BSDS100 and Manga109 datasets.

The future direction of this work is to investigate real-time image SR using perceptual loss. Our initial results using VGG19 network for formulating the loss function show some promising results with low model complexity.

REFERENCES

[1] Wan-Chi Siu, Zhi-Song Liu, Jun-Jie Huang, and Kwok-Wai Hung, "Learning Approaches for Super-Resolution Imaging," in *Learning Approaches in Signal Processing*, CRC Press, Wan-Chi Siu, Lap-Pui Chau, Liang Wang and Tieniu Tang (Editors), 2018.

[2] Kwok-Wai Hung and Wan-Chi Siu; , "Robust soft-decision interpolation using weighted least squares," *IEEE Trans. on Image Processing*, vol.21, no.3, pp.1061-1069, March 2012

[3] Kwok-Wai Hung and Wan-Chi Siu, "Fast image interpolation using bilateral filter", *IET Image Processing*, vol. 6, no. 7, pp. 877-890, October 2012.

[4] Kwok-Wai Hung, Wan-Chi Siu, "Learning-based Image Interpolation via Robust k-NN Searching for Coherent AR Parameters Estimation", *Journal of Visual Communication and Image Representation*, Vol 31, pp. 305-311, August 2015.

[5] Kwok-Wai Hung, Wan-Chi Siu, "Novel DCT-based Image Up-sampling Scheme using Learning-based Adaptive k-Nearest Neighbor MMSE Estimation", *IEEE Trans. on Circuit and Systems for Video technology*, vol.24, no.12, pp.2018,2033, Dec. 2014.

[6] Kwok-Wai Hung, Kun Wang, Jianmin Jiang, "Image Up-sampling using Deep Cascaded Neural Networks in Dual Domains for Images Down-sampled in DCT domain", *Journal of Visual Communication and Image Representation*, Volume 56, October 2018, Pages 144-149.

[7] Kwok-Wai Hung, Wan-Chi Siu, "Single-Image Super-Resolution Using Iterative Wiener Filter based on Nonlocal Means", *Signal Processing: Image Communication*, Vol. 39, Part A, pp 26-45, November 2015.

[8] Qinglong Chang, Kwok-Wai Hung, Jianmin Jiang, "Deep Learning Based Image Super-resolution for Nonlinear Lens Distortions", *Neurocomputing*, Volume 275, January 2018, Pages 969-982.

[9] Jianmin Jiang, Hossam M Kasem, Kwok-Wai Hung, "A Very Deep Spatial Transformer Towards Robust Single Image Super-Resolution", *IEEE Access*, Volume 7, 45618-45631, 2019.

[10] Kwok-Wai Hung, Kun Wang, Jianmin Jiang, "Image interpolation using convolutional neural networks with deep recursive residual learning", *Multimedia Tools and Applications*, 2019. DOI: 10.1007/s11042-019-7633-1

[11] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Learning a Deep Convolutional Network for Image Super-Resolution", *European Conference on Computer Vision* (ECCV), Zurich, September 2014.

[12] Chao Dong, Chen Change Loy, and Xiaoou Tang, "Accelerating the Super-Resolution Convolutional Neural Network", *European Conference on Computer Vision* (ECCV), Amsterdam, October 2016., pp. 1646-1654.

[13] Wenzhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, Zehan Wang, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Las Vegas, NV, 2016, pp. 1874-1883.

[14] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks", *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Las Vegas, NV, 2016, pp. 1646-1654.

[15] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., An-dreetto, M., Adam, H.: "Mobilenets: Efficient convolutional neural networks for mobile vision applications", *arXiv preprint* arXiv:1704.04861.

[16] Francois Chollet. "Xception: Deep learning with depthwise separable convolutions" *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Honolulu, Hawaii, USA, 2017

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. "Rethinking the inception architecture for computer vision". *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Las Vegas, NV, 2016

[18] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: "Encoder-decoder with atrous separable convolution for semantic image segmentation". *arXiv preprint* arXiv:1802.02611, 2018

[19] Nal Kalchbrenner, Lasse Espeholt, Karen Simonyan, Aaron van den Oord, Alex Graves, and Koray Kavukcuoglu. "Neural machine translation in linear time". *arXiv preprint* arXiv:1610.10099v2, 2017.

[20] Łukasz Kaiser, Aidan N. Gomez, Francois Chollet, "Depthwise separable convolutions for neural machine translation". *arXiv preprint* arXiv:1706.03059, 2017.

[21] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan, "Yang Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution", *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Honolulu, Hawaii, USA, 2017, pp. 5835-5843.

[22] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita,"Deep Back-Projection Networks For Super-Resolution", *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Salt Lake City, 2018.

[23] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee Department of ECE, ASRI, Seoul National University, Korea, "Deeply-Recursive Convolutional Network for Image Super-Resolution", *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Las Vegas, NV, 2016, pp. 1637-1645.

[24] Ying Tai, Jian Yang, and Xiaoming Liu, "Image Super-Resolution via Deep Recursive Residual Network", *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Honolulu, Hawaii, USA, 2017, pp. 2790-2798.

[25] Ignatov, Andrey and Timofte, Radu, "PIRM challenge on perceptual image enhancement on smartphones: report", *European Conference on Computer Vision Workshops* (ECCVW), January 2019

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

[27] Radu Timofte, Vincent De Smet, and Luc Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution", *Asian Conference on Computer Vision* (ACCV), Singapore, Singapore, 2014, 111-126.

[28] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistcs", *IEEE International Conference on Computer Vision* (CVPR), Vancouver, B.C., Canada, 2001, 416-423.

[29] Fujimoto, Azuma and Ogawa, Toru and Yamamoto, Kazuyoshi and Matsui, Yusuke and Yamasaki, Toshihiko and Aizawa, Kiyoharu, "Manga109 Dataset and Creation of Metadata", *International Workshop on coMics ANalysis, Processing and Understanding* (MANPU), 2016, Cancun, Mexico, pages 2:1-2:5, no. 2

[30] Wang, Z., Liu, D., Yang, J., Han, W., Huang, T. "Deep networks for image superresolution with sparse prior" International Conference on Computer Vision (ICCV), 2015

[31] Samuel Schulter, Christian Leistner, Horst Bischof, "Fast and Accurate Image Upscaling with Super-Resolution Forests", *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2015

[32] Q. Yuan, Q. Zhang, J. Li, H. Shen and L. Zhang, "Hyperspectral Image Denoising Employing a Spatial–Spectral Deep Residual Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1205-1218, Feb. 2019.

[33] Q. Zhang, Q. Yuan, C. Zeng, X. Li and Y. Wei, "Missing Data Reconstruction in Remote Sensing Image With a Unified Spatial–Temporal–Spectral Deep Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4274-4288, Aug. 2018.

KWOK-WAI HUNG received the BEng and Ph.D. degrees from Hong Kong Polytechnic University in 2009 and 2014 respectively. From 2014 to 2016, he was a research engineer with ASTRI and Huawei. From 2016 to date, he is an assistant professor with the Research Institute for Future Media Computing, Shenzhen University, China. His research interests include deep learning and signal processing applications in digital multimedia.

ZHIKAI ZHANG received the BSng degree from Beijing Normal University, Zhuhai in 2018. From 2018 to date, he is a master by research student with the Research Institute for Future Media Computing, Shenzhen University, China. His research interests include image super-resolution applications in digital multimedia processing.

JIANMIN JIANG received the Ph.D. degree from the University of Nottingham, Nottingham, U.K., in 1994. From 1997 to 2001, he was a Full Professor of Computing with the University of Glamorgan, Pontypridd, U.K. In 2002, he joined the University of Bradford, Bradford, U.K., as a Chair Professor of Digital Media, and the Director of the Digital Media and Systems Research Institute. He was a Full Professor with the University of Surrey, Guildford, U.K., from 2010 to 2014, and a Distinguished Chair Professor (1000-Plan) with Tianjin University, Tianjin, China, from 2010 to 2013. He is currently a Distinguished Chair Professor and the Director of the Research Institute for Future Media Computing, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. He has published around 400 refereed research papers. His current research interests include, image/video processing in compressed domain, digital video coding, medical imaging, computer graphics, machine learning and AI applications in digital media processing, and retrieval and analysis. Dr. Jiang was a Chartered Engineer, a fellow of IEE and RSA, a member of EPSRC College in the U.K., and an EU FP-6/7 Evaluator.