PRED: A PARALLEL NETWORK FOR HANDLING MULTIPLE DEGRADATIONS VIA SINGLE MODEL IN SINGLE IMAGE SUPER-RESOLUTION

Guangyang Wu, Lili Zhao, Wenyi Wang, Liaoyuan Zeng, Jianwen Chen

School of Information and Communication Engineering
University of Electronic Science and Technology of China
Chengdu 611731, China

ABSTRACT

Existing SISR (single image super-resolution) methods mostly assume that a low-resolution (LR) image is bicubicly downsampled from its high-resolution (HR) counterpart, which inevitably give rise to poor performance when the degradation is out of assumption. To address this issue, we propose a framework PRED (parallel residual and encoder-decoder network) with an innovative training strategy to enhance the robustness to multiple degradations. Consequently, the network can handle spatially variant degradations, which significantly improves the practicability of the proposed method. Extensive experimental results on real LR images show that the proposed method can not only produce favorable results on multiple degradations, but also reconstruct visually plausible HR images.

***Index Terms*—** SISR, multiple degradation, CNN, PRED

1. INTRODUCTION

Single Image Super Resolution (SISR) aims at reconstructing high-resolution (HR) image from its low-resolution (LR) counterpart [1–3]. Generally, the relation between LR-HR images can vary depending on the situation. Many studies assumed that LR image is a bicubicly downsampled version of HR image, but other degradations such as blur, decimation, or noise can also exist in practical applications.

As a classic problem, SISR is still an active but challenging research topic due to its ill-poseness nature and high practical values [4], and can serve as a built-in module for other image restoration or recognition tasks. Recently, deep convolutional neural networks (CNNs) have been successfully applied to SISR [5–7] and significantly improve performance in terms of peak signal-to-noise ratio (PSNR).

However, these works suffer from a common defect, that is, they can only deal with one specific situation, such as bicubic downsampling, but fail to solve multiple degradations with single model. Because the practical degradations of SISR is much more complex [8, 9], and the performance of learned CNN models may deteriorate seriously when the assumed degradation deviates from the true one, making them less effective in practical scenarios.

In this case, this paper aims at answering the following questions: (i): Can CNN robustly handle multiple degradations with single model? (ii): Is it possible to use synthetic data to train a model with high practicability?

In the view of above, the main contributions of this paper are summarized as follows:

- We propose an effective CNN framework, indicated as PRED (parallel residual and encoder-decoder network), with an innovative training strategy. The proposed method goes beyond the widely-used bicubic degradation assumption and works for multiple and even spatially variant degradations.
- We propose IBF (Integrated Block Fusion) to blockwise reconstruct the full image instead of convolving directly on the whole image, which shows prominent improvements.
- We show that the proposed network learned from synthetic training data can produce competitive results against the state-of-the-art SISR method on both synthetic and real LR images.

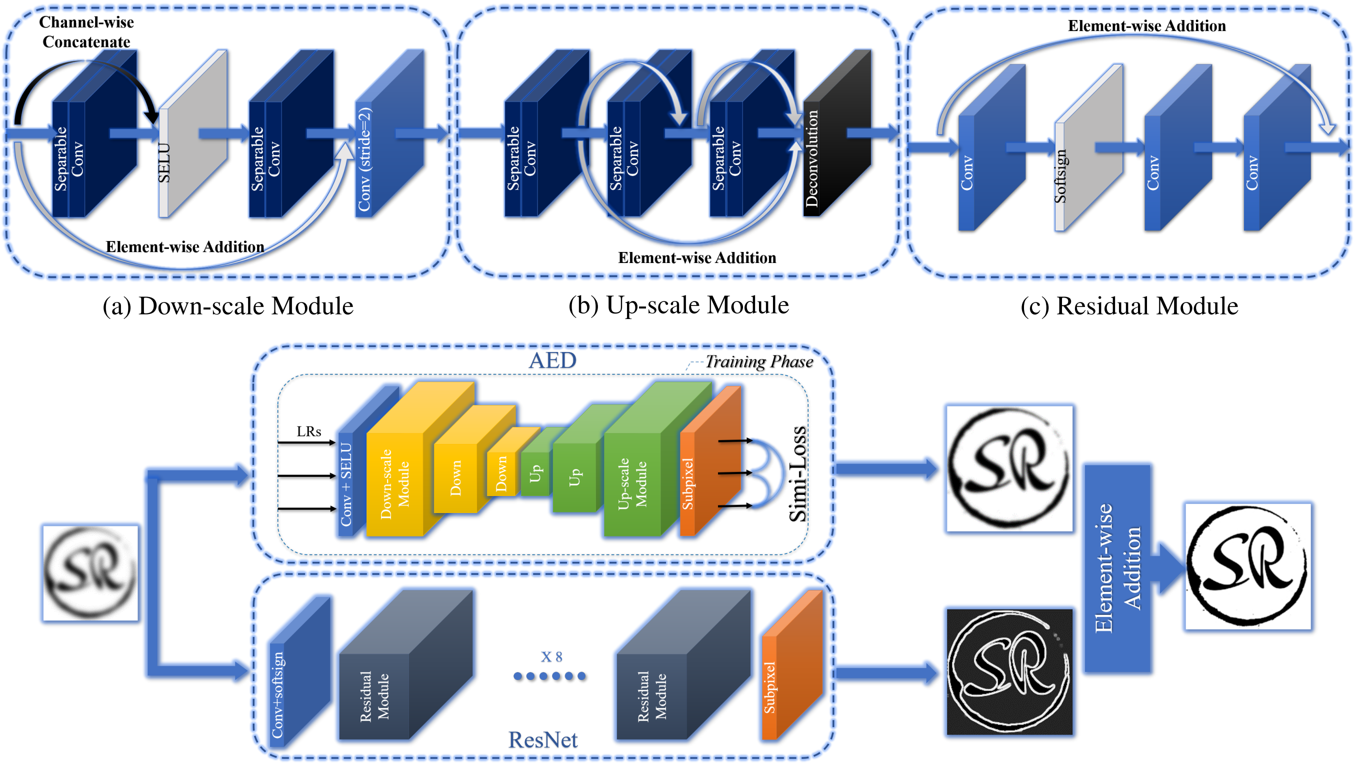
The rest of the paper is organized as follows: Section 2 introduces the proposed method mentioned above. Section 3 and 4 focus on the experiments and evaluation both on synthetic data and real LR images. Section 5 draws the conclusions according to the experimental results.

2. PROPOSED METHOD

The proposed PRED framework has two parallel networks, denoted as AED (auto encoder-decoder) and ResNet (residual network), which will be introduced in the following parts.

2.1. AED network

The AED is applied as an image purifier to preserve the primary components of the contents in the image and meanwhile eliminate the randomly generated corruptions [10].



(a) Down-scale Module (b) Up-scale Module (c) Residual Module

(d) PRED Framework

Fig. 1. Demonstration of the proposed PRED structure and Simi-Loss (d). The AED network consists 3 Down / Up-scale Modules (a / b), while the ResNet connected in parallel consists 8 Residual Modules (c). The AED takes multiple randomly degraded LR images as inputs in the training phase and minimizes the Simi-Loss which will be introduced in Section 2.1.

The encoder primarily consists of three Down-scale modules (Fig. 1 (a)), reducing the spatial scale to one-eighth of the input. The decoder is mainly composed of three Up-scale modules (Fig. 1 (b)), and a pixel shuffling layer [11,12] is cascaded to obtain the image magnified by S times, where S is the super-resolution upscaling factor. Considering parameter-efficient, most layers use the separable convolution [13, 14]. We use SELUs [15] (scaled exponential linear units) as activation layers to induce self-normalizing properties.

To deal with multiple degradations, the most conventional way is to train the AED network with LR-HR data pairs generated from multiple degradations, and directly minimize the MSE (Mean Square Error) loss to achieve acceptable performance in terms of peak signal-to-noise ratio (PSNR). We refer to this training strategy as the "regular" method.

In order to enhance the robustness of the network, we introduce an innovative training strategy by **controlling the consistency of AED output to multiple LR images generated from one HR image with different degradations.**

In the training phase, we generate n LR images from one HR image by applying n different degradations. And we further assume that the n LR images share similar SR estimation from the proposed AED. Accordingly, the Simi-loss can be formulated as:

$$L_{simi} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} mse(SR_i, SR_j) \tag{1}$$

where mse is the mean square error [4].

In the meanwhile, the Mean-MSE loss is needed for reconstruction quality, which can be formulated as:

$$L_{mmse} = \frac{1}{n} \sum_{i=1}^{n} mse(SR_i, HR) \tag{2}$$

in which SR_i denotes the ith output image of the AED network and HR denotes the corresponding ground truth HR image.

The final cost function applied in training phase is the linear combination of L_{simi} and L_{mmse}:

$$L_{multin} = \alpha L_{mmse} + \beta L_{simi} \tag{3}$$

where α and β are hyperparameters. We refer to this proposed training strategy as "multin" since it has multiple inputs.

Table 1. Average PSNR results for general degradations on dataset Set5 [16]. We use nearest-neighbor as downsampler with scale factor 2. The best two results are highlighted in red and blue colors, respectively.

Degradation Methods		EDSR	PRED-regular	PRED-multin-a	PRED-multin-b	PRED-multin-pro
blur σ_b	noise σ_n	PSNR (dB)				
0.5	1	25.28	33.61	33.29	*33.74*	**34.24**
0.5	2	21.83	33.42	32.95	*33.55*	**34.00**
0.5	3	19.93	*32.94*	32.52	32.35	**33.71**
1.5	1	26.04	31.39	31.48	*32.37*	**33.54**
1.5	2	22.29	31.54	30.97	*31.80*	**32.73**
1.5	3	20.21	30.94	30.14	*31.13*	**32.03**

2.2. ResNet in Parallel

Although the proposed AED is robust to multiple degradations, its robustness is somehow obtained by sacrificing the reconstruction quality. In order to generate high quality SR images, we parallelly combined a ResNet [17] with 8 Residual modules (Fig. 1 (c)) to the AED to compensate the details of the SR image. The non-linearities applied here are softsigns in consideration of the compensation which could be both in positive and negative value. In the end, we cascade a subpixel layer and add the outputs of two parallel parts to obtain the final SR image.

2.3. IBF

According to our literature review, most CNN based SISR method [7,10,17] directly reconstruct the whole HR image in the testing phase no matter the network is trained block-wise or image-wise. In our experiments, we found that block-wise reconstruction shows prominent improvements compared with directly reconstructing the whole image. Therefore, we proposed to use an integrated block fusion (IBF) method to block-wise reconstruct a full image in the testing phase. Specifically, we decompose the image into blocks with the size of 24. The blocks are horizontally and vertically overlapped with 12 pixels ($st = 12$), and the reconstruction of overlapped area will take the average value from each block. We refer to this reconstruction method as "PRED-multin-pro" in this paper.

3. IMPLEMENTATION

3.1. Data

The training data are sub-images in shape of $(48 \times 48 \times 1)$, which are extracted from the DIV2K dataset (only Y channel in YCbCr color space is considered). Sub-images are degraded randomly with σ_b and σ_n (standard deviation of Gaussian blur kernel and additive Gaussian noise), and then downsampled directly (using nearest-neighbor downsampler) with scale factor S. We set the range of σ_b to $(0, 2)$ with $stride = 0.2$ and σ_n to $(0, 5)$ with $stride = 1$. There are 66 degradations covered by synthetic training data.

Table 2. Hyperparameters settings of different models. Note that the PRED-multin-pro will reconstruct full image block-wise (IBF applied), while other models directly reconstruct the whole image (IBF not applied) in testing phase.

Models	AED Training				IBF
	training strategy			input number	st
	regular	multin			
		α	β		
PRED-regular	√	-	-	1	-
PRED-multin-a	-	9	1	3	-
PRED-multin-b	-	9	10	4	-
PRED-multin-pro	-	9	10	4	12

3.2. Training details

We trained four models with different settings as shown in Table 2. Adam [20] is used as the optimizer with learning rate $\eta = 0.0001$. Considering the gradient dispersion caused by increasing depth, we use weight-normalization [21] to drive higher learning rate (i.e., 0.001) while fine-tuning ResNets.

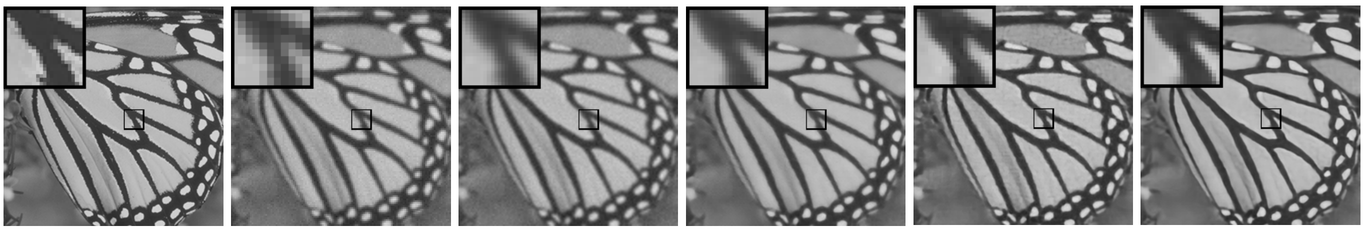
4. EXPERIMENT RESULTS

4.1. Experiment on Synthetic LR Images

In this subsection, we evaluate the performance on synthetic LR images. We compare the proposed method with EDSR [7] on different degradations in Table 1. The performance of EDSR deteriorates seriously when the assumed bicubic degradation deviates. Meanwhile, the PRED-multin-pro benefiting from the "multin" training strategy and the IBF method outperforms PRED-regular both inside (Table 1) and outside (Table 3) the training space, which further proves the scalability of the proposed method on spatially variant degradations. The visual comparison is given in Fig. 2. One can see that PRED-regular produce more visually pleasant results than EDSR. However, they can not recover edges as sharper as PRED-multin-pro.

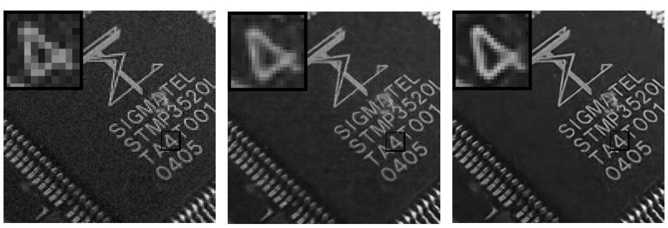
Table 3. Average PSNR results on datasets for bicubic / nearest downsampler (scale factor = 2) with $\sigma_b = 1.5$ and $\sigma_n = 0$ applied here. Note that the bicubic degradation is out of the training space here. Best results are highlighted in red color.

Dataset	Downsampler	PRED-regular	PRED-multin-a	PRED-multin-b	PRED-multin-pro
		PSNR (dB)			
Set5 [16]	bicubic	29.15	29.44	29.68	**29.73**
	nearest	33.11	33.41	32.56	**33.80**
Set14 [18]	bicubic	27.76	27.69	27.73	**27.97**
	nearest	29.95	30.45	29.73	**30.83**
BSD100 [19]	bicubic	27.21	**27.50**	27.31	27.35
	nearest	28.63	29.15	28.68	**29.54**



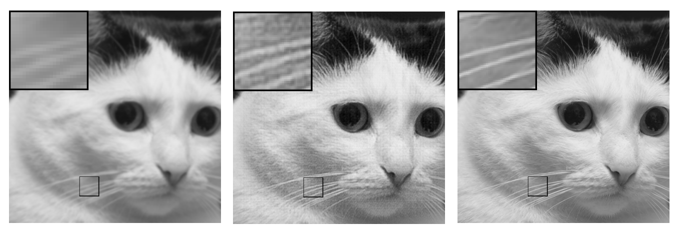
(a) Ground Truth (b) LR (zoomed) (c) Bicubic (d) EDSR (e) PRED-regular (f) PRED-multin-pro

Fig. 2. SISR performance comparison on image "Butterfly" from Set5 [16]. The degradation involves Gaussian blur kernel with $\sigma_b = 1.2$, additive Gaussian noise with $\sigma_n = 3$ and nearest-neighbor downsampler with scale factor 2. Results in terms of PSNR are (c) Bicubic / 22.05dB, (d) EDSR / 24.526dB, (e) PRED-regular / 26.30dB, (f) PRED-multin-pro / 27.49dB.



(b) LR image (c) EDSR (c) PRED-multin-pro

Fig. 3. SISR results on real image "Chip" with scale factor 2.



(b) LR image (c) EDSR (c) PRED-multin-pro

Fig. 4. SISR results on real image "Cat" with scale factor 2.

4.2. Experiments on Real Images

Besides the above experiments on LR images synthetically downsampled from HR images with known degradations, we also do experiments on real LR images to demonstrate the practicability of the proposed method. Since there are no ground-truth HR images, we only provide the visual comparisons. Fig. 3 and Fig. 4 illustrate the SISR results on two real LR images Chip and Cat, respectively. The EDSR [7] is used as one of the representative CNN-based methods for comparison. It can be observed from the visual results that the PRED-multin-pro can produce much more visually plausible HR images than the competing methods.

5. CONCLUSION

In this paper, we propose an effective framework handling multiple degradations via a single model. PRED network and Simi-loss are proved to be valid on enhancing the robustness of the framework. IBF shows significant improvements on reconstruction quality. The results on synthetic LR images demonstrate that the proposed method can perform favorably not only on traditional but also spatially variant degradations. Moreover, the results on real LR images show that the proposed method can reconstruct visually plausible HR images.

6. ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under grant No.61701094.

7. REFERENCES

[1] Shuhang Gu, Wangmeng Zuo, Qi Xie, Deyu Meng, Xiangchu Feng, and Lei Zhang, "Convolutional sparse coding for image super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1823–1831.

[2] Radu Timofte, Vincent De Smet, and Luc Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Asian Conference on Computer Vision*. Springer, 2014, pp. 111–126.

[3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[4] Simon Baker and Takeo Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[5] 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*, 2016, pp. 1646–1654.

[6] Ding Liu, Zhaowen Wang, Bihan Wen, Jianchao Yang, Wei Han, and Thomas S Huang, "Robust single image super-resolution via deep networks with sparse prior," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3194–3207, 2016.

[7] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *The IEEE conference on computer vision and pattern recognition (CVPR) workshops*, 2017, vol. 1, p. 4.

[8] Yaniv Romano, John Isidoro, and Peyman Milanfar, "Raisr: rapid and accurate image super resolution," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 110–125, 2017.

[9] Chih-Yuan Yang, Chao Ma, and Ming-Hsuan Yang, "Single-image super-resolution: A benchmark," in *European Conference on Computer Vision*. Springer, 2014, pp. 372–386.

[10] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," *arXiv preprint arXiv:1606.08921*, 2016.

[11] 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*, 2016, pp. 1874–1883.

[12] Augustus Odena, Vincent Dumoulin, and Chris Olah, "Deconvolution and checkerboard artifacts," *Distill*, vol. 1, no. 10, pp. e3, 2016.

[13] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[14] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *arXiv preprint arXiv:1801.04381*, 2018.

[15] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter, "Self-normalizing neural networks," in *Advances in Neural Information Processing Systems*, 2017, pp. 971–980.

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

[17] Jiahui Yu, Yuchen Fan, Jianchao Yang, Ning Xu, Zhaowen Wang, Xinchao Wang, and Thomas Huang, "Wide activation for efficient and accurate image super-resolution," *arXiv preprint arXiv:1808.08718*, 2018.

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

[19] 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 statistics," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, vol. 2, pp. 416–423.

[20] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[21] Tim Salimans and Durk P Kingma, "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 901–909.