

Degradation-Oriented Adaptive Network for Blind Super-Resolution

Shuo WANG¹ and Lifang CHEN

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China

Abstract. Most of the super-resolution methods based on deep neural networks use a fixed image degradation model. However, when the real image degradation is inconsistent with the model assumptions, the performance of the network will be severely reduced. To address this issue, we propose a degradation-oriented adaptive network (DOANet), which can be used to solve multiple degradation super-resolution reconstruction. The network includes a degradation estimation network and an adaptive reconstruction network. The degradation estimation network extracts the abstract degradation of LR images using continuous residual blocks to provide necessary degradation information for reconstruction network. Then, the adaptive reconstruction network dynamically adjusts the parameters of the network layers based on the estimated degradation information using dynamic convolution and channel modulation to deal with various input images. Experimental results show that the proposed DOANet can cope well with multiple degradation and is superior to other recent blind methods in qualitative and quantitative comparison.

Keywords. Blind super-resolution, multiple degradations, attention mechanism, dynamic convolution

1. Introduction

Super-resolution (SR) as a basic low-level problem in computer vision has received widespread attention, which aims at recovering a high-resolution (HR) image from a low-resolution (LR). It has been widely used in many aspects such as image compression, medical imaging, remote sensing imaging and object recognition. Since SR is an ill-posed problem with multiple plausible solutions, it still faces great challenges.

Recently, convolutional neural networks (CNN) have been widely used in SR tasks. Compared with traditional methods, CNN-based methods have significant improvements in performance and end-to-end applications. Among them, Dong et al. first applied CNN to SR problems. The proposed SRCNN [1] surpassed the traditional method with a simple three-layer convolution network. After that, SR methods based on deep convolutional networks have been widely studied. VDSR [2] proposed by Kim et al. uses residual network to deepen the network depth to 20 layers and achieves better performance. EDSR [3] proposed by Lim et al. further deepens the network by removing the unnecessary BN layer in the SR network. RCAN [4] proposed by Zhang et al. uses residual in residual (RIR) structure to construct a very deep trainable network. However,

¹ Corresponding Author: Shuo WANG, School of Artificial Intelligence and Computer Science, Jiangnan University, China; E-mail: 1919099246@qq.com.

the above-mentioned methods are based on fixed degradation assumptions (such as bicubic interpolation [5]), when the actual image degradation does not match the assumptions, the performance of these networks will decrease rapidly.

So, some studies have proposed for blind SR. KernelGAN [6] uses the internal cross-scale recurrence to learn the internal distribution of image patches to estimate the blur kernel of LR image, but the accuracy of the estimated kernel needs to be improved. IKC [7] uses an iterative network to correct the estimated blur kernel, which improves the accuracy of prediction, but the running time increases significantly. DASR [8] obtains the implicit degradation information of LR images through comparative learning, which further improves the performance of blind SR.

In this paper, we propose a novel and efficient blind SR network to deal with images with multiple degradations. The main contributions of our work are summarized in the following:

- Image degradation is considered and estimated by designed sub-networks.
- The reconstruction network adaptively adjusts the network parameters using degradation information to deal with different degraded images
- Experiments show that our method has achieved advanced results compared with other methods

2. Methodology

2.1. Problem Formulation

In this paper, the degradation model of an image can be formulated as:

$$I^{LR} = (I^{HR} \otimes k) \downarrow_s \quad (1)$$

Where I^{HR} represents high-resolution (HR) image, I^{LR} represents low-resolution (LR) image, k represents the blur kernel, \otimes represents the convolution operation, \downarrow_s represents the down-sampling operation with the scale factor s .

Compared with the real image pairs that are difficult to obtain, although the synthetic data has a certain gap with the real images, it plays an important role in the research of blind SR methods. Following [7][8], isotropic Gaussian kernels and anisotropic Gaussian kernels are used to generate blur images.

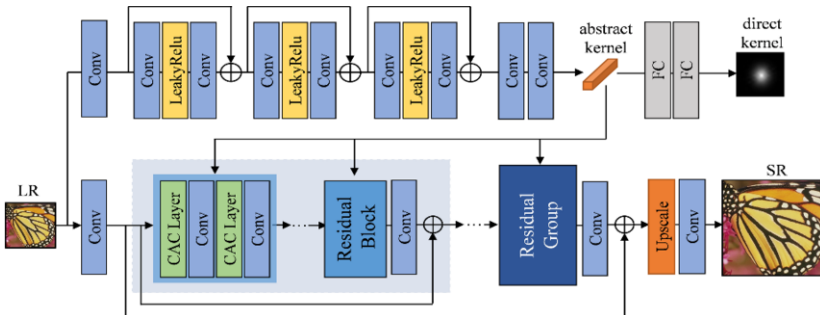


Figure 1. Overall structure of our DOANet

2.2. Structure of DOANet

As shown in Fig 1, proposed DOANet consists of two sub-networks: a degradation estimation network and an adaptive reconstruction network. The LR image is fed to the degradation estimation network to predict the abstract degradation information, and the output is expressed as $v \in \mathbb{R}^{1 \times 256}$. Then, the adaptive reconstruction network takes the LR image with its estimated degradation information to produce the SR result.

• Degradation estimation network

We first need to estimate the degradation of the input image to provide useful information for reconstructing the network. The degradation estimation network takes a degraded LR image I^{LR} as input, and aims to estimate its underlying degradation kernel. The input I^{LR} first undergoes a 3×3 convolutional layer to extract the shallow features, then deep features are extracted by successive three sets of residual blocks and each residual block consist of two 3×3 convolutional layer and one activation layer of Leaky ReLU. Another two 3×3 convolutional layer is placed later for feature reconstruction to generate degenerate information v .

Due to SR tasks are sensitive to kernel estimation, the abstract kernel represents a better generalization ability for estimation errors. So underlying degradation information of the blurred image instead of the ground-truth kernel is predicted by degradation estimation network. After that, the underlying degradation information is fed to two fully connected layers and reshape a blur kernel. The reshaped blur kernel and the ground-truth kernel has the same size for loss calculation.

• Adaptive reconstruction network

Adaptive reconstruction network takes the LR image I^{LR} and its estimated degenerate information which is used to adjust network parameters to deal with the specific input image. Refer to RCAN [4], residual in residual (RIR) structure is used as the backbone of the network due to it can be built deeper network which is easy to train. The network contains several residual groups, and each residual group consists of several residual blocks, the number of groups and blocks is uniformly set to 5. The residual block is composed of two adaptive feature extraction blocks and two 3×3 convolution layers. Input images first undergo a 3×3 convolution for preliminary feature extraction, then the features are fed to RIR for deep feature refinement. The extracted features are up-sampled by the sub-pixel convolutional layer [9] and high-resolution images are reconstructed after a final convolutional layer.

• Adaptive Feature Extraction Block

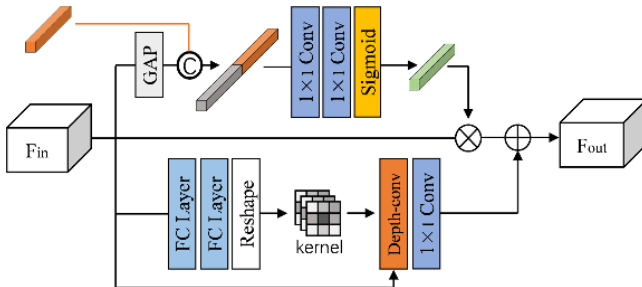


Figure 2. The structure diagram of CAClayer

As shown in Fig. 2, the key part of the adaptive reconstruction network is an adaptive feature extraction block, which consists of two branches: dynamic convolution branch and channel modulation branch.

Actually, dynamic convolution can be thought of as the attention of convolution kernel. In dynamic convolution branch, degradation information is fed to two full-connected layers and reshape a dynamic convolution kernel $k \in \mathbb{R}^{C \times 3 \times 3}$, C represents the number of channels, 3×3 represents the size of the dynamic convolution kernel. Then k is used to perform depth-wise convolution on the input feature maps.

Refer to the channel attention mechanism in SENet [10], the global information of the input feature map is compressed by global average pooling (GAP). The difference with SENet is that later the global information of the feature maps is concatenated with degradation information. After that, the concatenated vector goes through two 1×1 convolutional layers and sigmoid layer to generate channel modulation coefficients. Then the channel modulation coefficients are multiplied by the input feature maps.

The two branches use different attention mechanisms with degradation information to help the network adaptively adjust the network parameters to achieve better generalization capabilities.

2.3. Loss Function

The DOANet is optimized with loss function, which is the sum of the degradation estimation network loss and the adaptive reconstruction loss. The former is used to minimize the difference between the prediction kernel and the ground-truth kernel, and the latter is used to minimize the difference between the predicted SR image and the ground-truth HR image. As done in most previous work, L1 loss function is used to measure the difference for the both parts of the network. The loss function is formulated as follows:

$$loss = \frac{1}{N} \sum_{i=1}^N (\|k_i^p - k_i^{gt}\|_1 + \|I_i^{SR} - I_i^{HR}\|_1) \quad (2)$$

Where, k_i^p and k_i^{gt} represent the predicted kernel and the real kernel, respectively. I_i^{SR} and I_i^{HR} represent the SR image and the HR image.

3. Experiment

3.1. Settings

- **Datasets:** A total of 3450 high quality images from DIV2K and Flickr2K are used for network training. The datasets of Set5, Set14, B100 and Urban100 are used for testing. We degrade images in two ways: isotropic Gaussian blur and anisotropic Gaussian blur. Specifically, for the scale factors of $\times 2$ and $\times 4$, the kernel width ranges of isotropic Gaussian kernels are set to $[0.2, 2.0]$ and $[0.2, 4.0]$, respectively. The shapes of anisotropic Gaussian kernels are determined by parameters λ_1 , λ_2 , and θ . The ranges of λ_1 and λ_2 are set to $[0.2, 4.0]$, the ranges of θ are set to $[0, \pi]$.

- Evaluating indicator: In this paper, the peak signal to noise ratio (PSNR) is used for an objective evaluation index, the formula is expressed as:

$$\text{PSNR} = 10 \times \lg \frac{255^2 \times H \times W}{\sum_{i=1}^H \sum_{j=1}^W (I^{SR}(i, j) - I^{HR}(i, j))^2} \quad (3)$$

Where, H and W represent the height and width of the image, respectively. PSNR is measured in decibel (dB)

- Training details: During training, we first crop the input high-resolution (HR) images into 96×96 patches, and then apply random data enhancement to them, including random horizontals, vertical flips and 90° rotations. After that, the HR patches are blurred using random Gaussian kernels and down-sampling using bicubic interpolation to obtain the corresponding 48×48 low-resolution (LR) patches. The network is trained by the ADAM optimizer, the optimizer parameters $\beta_1=0.9$, $\beta_2=0.999$, $\varepsilon=10^{-8}$, and the batch size is set to 32. The initial learning rate is set to 10^{-3} , and the learning rate is reduced by half every 125 epochs. We use PyTorch to implement our model, the training time on NVIDIA GTX2080 and Intel i7-8700 CPU platform is about 4 days.

3.2. Analysis of performance

- degradation estimation: To demonstrate that the degradation estimation network can effectively extract image blur kernels, we performed a visualization of the estimated blur kernels. Experiments were carried out on the image "baby" from the Set5 dataset, which is blurred with anisotropic Gaussian kernels of different shapes. The estimated result is shown in Fig.3. It can be seen that the degradation estimation encoder can effectively extract the blurred information of the image.

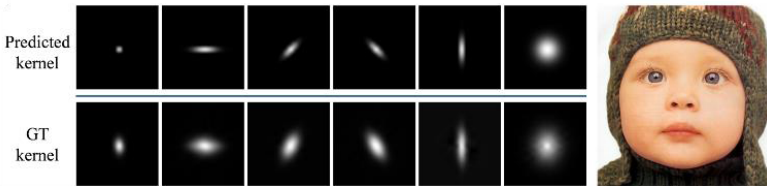


Figure 3. Visualization of the predicted blur kernel on "baby" image

- The plots of loss and PSNR in training: The training process of proposed DOANet is compared with DASR, as shown in Fig. 4. It can be seen from the figures that the loss of DOANet converges to a lower value and the PSNR of DOANet converges to a higher value compared with DASR, which illustrates that the network performance of proposed DOANet is better than DASR.

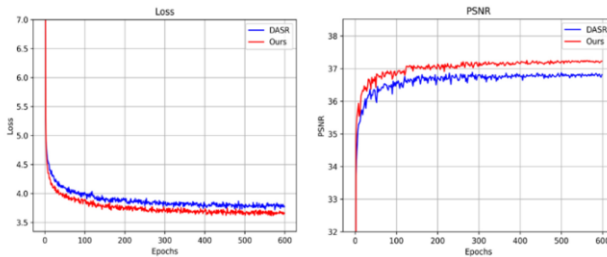


Figure 4. The plots of loss and PSNR in training

- Ablation of adaptive feature extraction block: To demonstrate the effectiveness of adaptive feature extraction block, we conducted ablation experiments on two of its branches. Specifically, we use vanilla convolution instead of dynamic convolution and use residual connection instead of channel modulation. The experiment is tested on B100 dataset with $\times 2$ scale factor, the kernel width is taken as 0 and 1.8 respectively. The experimental results are shown in Table 1:

Table 1. Ablation results of adaptive feature extraction block

model	dynamic convolution branch	channel modulation branch	Kernel width	
			0	1.8
1	✗	✗	32.03	30.54
2	✓	✗	32.16	30.66
3	✗	✓	32.13	30.75
4	✓	✓	32.19	30.87

Table 1 shows that the network without dynamic convolution and channel modulation has the worst performance. There will be improvement when only using dynamic convolution or channel modulation. The best results are obtained by using both, because the adaptive structural design can effectively improve the expression and generalization ability of the network.

3.3. Qualitative results

First, quantitative experiments were carried out on the model trained with isotropic Gaussian kernels, and a few representative methods were selected for comparison. RCAN is a SR method based for bicubic interpolation degradation model. We use the pre-trained model provided by the author for testing. IKC is a blind SR method through iterative blur kernel correction. DASR is a blind SR network by contrast learning to predict degradation information. For testing of scale factors $\times 2$ and $\times 4$, the kernel widths are set to [0.6, 1.2, 1.8] and [1.2, 2.4, 3.6], respectively. Quantitative (PSNR/dB) results are shown in Table 2:








Table 2. Qualitative results on isotropic Gaussian kernel of different methods

Method	Scale	Set14			B100			Urban100		
Kernel width		0.6	1.2	1.8	0.6	1.2	1.8	0.6	1.2	1.8
Bicubic	$\times 2$	29.21	27.13	25.47	28.76	26.93	25.51	26.13	24.46	23.06
RCAN		32.31	28.48	26.33	31.16	28.04	26.26	29.8	25.38	23.44
DAN		32.82	32.78	31.81	31.73	31.72	30.60	30.73	30.38	28.96
DASR		32.96	32.78	31.60	31.78	31.71	30.54	30.71	30.36	28.95
DOANet (Ours)		33.42	33.35	32.16	32.05	32.02	30.87	31.52	31.26	29.84
Bicubic	$\times 4$	25.24	23.83	22.57	25.42	24.2	23.15	22.68	21.62	20.65
RCAN		27.48	24.93	23.41	26.89	25.09	23.93	24.71	22.25	20.99
IKC		28.27	28.07	26.46	27.36	27.19	26.05	25.6	24.85	23.62
DAN		28.49	28.36	27.51	27.51	27.44	26.81	25.73	25.47	24.64
DASR		28.45	28.28	27.45	27.52	27.43	26.83	25.69	25.44	24.66
DOANet (Ours)		28.56	28.45	27.60	27.56	27.50	26.88	25.86	25.74	24.92

It can be observed from Table 1 that the performance of RCAN degrades severely as the kernel width increases, because it is based on bicubic interpolation model. IKC and DASR get a limited improvement due to the use of degradation information. Compared with the above methods, our DOANet achieves a significant improvement because it makes full use of the degradation information to adaptively adjust network parameters through the attention mechanism and dynamic convolution.

In addition, anisotropic Gaussian kernel were used for training data generation. We selected seven typical anisotropic Gaussian kernels for quantitative testing at the scale of $\times 4$ on the Set14 dataset. The experimental results are shown in Table 3, it can be seen that our method has achieved the best qualitative results compared to IKC and DASR.

Table 3. Qualitative results on anisotropic Gaussian kernel of different methods.

Method	Anisotropic Gaussian Kernel						
							
RCAN	26.55	24.61	24.34	24.4	24.31	24.01	23.49
IKC	28.27	27.53	27.31	27.13	27.22	27.01	26.87
DAN	28.28	27.66	27.75	27.82	27.77	27.63	27.21
DASR	28.25	27.81	27.73	27.70	27.63	27.55	27.15
DOANet (Ours)	28.46	28.03	28.03	27.95	27.81	27.81	27.42

3.4. Qualitative results

We also make an intuitive comparison of the visual effects for each method. The benchmark parameters are used for testing on the Urban100 dataset with a kernel width of 2.4 and a scale of 4. IKC takes the results of the last iteration.

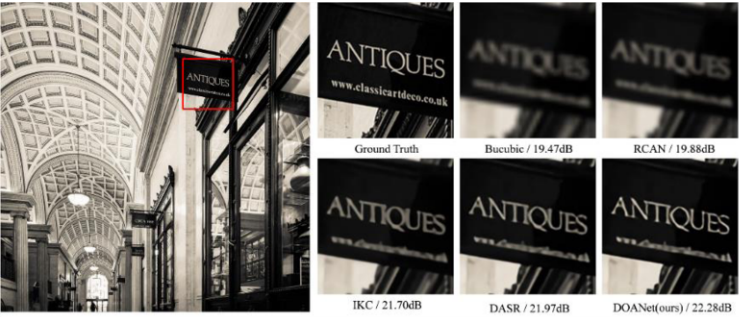


Figure 5. The visual results on the image"083" of the urban100 dataset

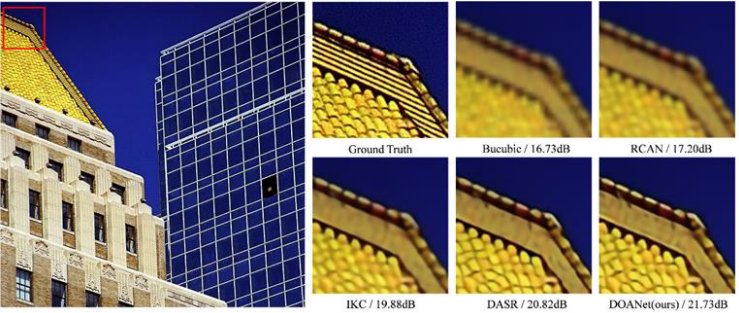


Figure 6. The visual results on the image"063" of the urban100 dataset

As shown in Fig.5 and Fig.6, it can be seen that the performance of RCAN has dropped sharply due to the change of the degradation model, resulting in a very blurred vision effect. IKC uses iterative blur kernel corrects to reduce the influence of different degradations but the results of each iteration of IKC are unstable. DASR achieves better performance than IKC. Compared to the above methods, our DOANet produces results with much clearer details, fewer artifacts and higher perceptual quality, because

conditional adaptive convolutional block can effectively use degradation information to adaptively adjust network parameters to cope with different types of degradation.

4. Conclusion

In this paper, we introduce a degradation-oriented adaptive network (DOANet) for blind SR. Specifically, we propose a degradation estimation network to predict the degradation from LR image and design an adaptive reconstruction network to complete high quality super-resolution reconstruction. It is demonstrated that our method can extract discriminative degradation information and use this information to improve network performance by dynamic convolution and channel modulation, which can effectively deal with different types of degraded images. Experimental results show that our method achieves advanced performance in qualitative and quantitative compared to recent blind super-resolution methods.

References

- [1] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]. European conference on computer vision. Springer, Cham, 2014: 184-199J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [2] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.
- [3] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]. Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 136-144.
- [4] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]. Proceedings of the European conference on computer vision (ECCV). 2018: 286-301.
- [5] Keys R. Cubic convolution interpolation for digital image processing[J]. IEEE transactions on acoustics, speech, and signal processing, 1981, 29(6): 1153-1160.
- [6] Bell-Kligler S, Shocher A, Irani M. Blind super-resolution kernel estimation using an internal-gan[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [7] Gu J, Lu H, Zuo W, et al. Blind super-resolution with iterative kernel correction[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1604-1613.
- [8] Wang L, Wang Y, Dong X, et al. Unsupervised degradation representation learning for blind super-resolution[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 10581-10590.
- [9] Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1874-1883.
- [10] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.