ISSN: 2710-0588

Image Super-resolution Networks based on Structure Reparameterized Convolution

Tianya Xing, Hua Huo

College of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China

Abstract

The traditional image super resolution method based on convolutional neural network generally brings two problems: First, the model has only a single scale of receptive field, which cannot use the image information in a smaller range; Second, the size of the feature map will become smaller and smaller in the process of continuous convolution, and it is necessary to continuously carry out edge zeroing operation to maintain the original size, which leads to the loss of part of the edge information. To solve the above two problems, we propose a network structure based on structure re-parameterized convolution, which sets a small convolution kernel in parallel beside a large convolution kernel in the same layer, trains the two cores simultaneously, and finally merges the two cores. The experimental results show that in this way, we make the large convolution kernel could capture smaller information, which not only improves the high-frequency details of the reconstructed image but also avoids frequent edge filling to reduce the information density, and effectively speeds up the reasoning speed after the model is re-parameterized. Compared with the advanced methods, our method achieves good results.

Keywords

Super-resolution; Structure Reparameterization; Residual Network; Image Processing.

1. Introduction

Images collected from nature are often degraded due to various problems, thus reducing the original image quality. Image super resolution, a classical computer vision task, aims to reverse reasoning the process of image degradation mapping, to reconstruct high-quality high-resolution images (HR) before degradation from low-resolution images (LR). Such methods have been proven to be effective and inexpensive. SRCNN of Dong et al.[1] designed a three-layer convolutional neural network to perform image super-resolution tasks, namely, feature extraction layer, convolution layer and reconstruction layer. Feature extraction layer is used to extract shallow features from input LR images, convolutional layer is used to process the extracted features in higher order, and reconstruction layer is used to expand image resolution. Although SRCNN can obtain far more data than the traditional interpolation method on the test set, because it is a shallow neural network with small field of perception, it can only process low-order features of low-resolution images, and the processing ability of higher-order features of images is insufficient. Kim et al.[2] found the shortcomings in SRCNN and proposed the importance of receptive field for SR model. The VDSR model proposed by their team introduced residual structure to deepen network depth and increase network receptive field to improve reconstruction quality. In addition, a structure is designed to allow VDSR to process LR images with different magnifications in a single model and obtain better image reconstruction results than SRCNN.

ISSN: 2710-0588

With the continuous development of artificial intelligence technology, more and more latest technologies have been applied to the field of image super resolution, resulting in more and more advanced SR models, such as ESRGAN[3], KernelGAN[4] and Real-ESRGAN[5] based on generative adversarial network; RCAN[6], A2N[7] based on attention mechanism; SwinIR based on Transformer [8]; SRDIFF[9] based on diffusion model, and traditional convolutional neural network SR models such as SRCNN[1], SRResNet[10], EDSR[11] and VDSR[2] no longer have particularly obvious advantages compared with these models.

After analysis, it is found that the performance of traditional image SR neural network is insufficient due to the following reasons:

(1) Traditional neural network models are generally based on stacking various carefully designed convolutional structures[12], and the final reconstruction quality is improved by iterating between network layers, but this will make the original input image smaller and smaller in the process of continuous convolution. Different from computer vision tasks such as image classification[13, 14], which continuously convolve until high-order features are extracted, image super-resolution tasks have certain requirements on image output size. Therefore, to meet such requirements, the traditional convolutional neural network super-division model will maintain image size by continuously filling edges. In this way, the model will lose part of the information of the original image, which will affect the final reconstruction quality. The deeper the layer, the more affected the network model will be.

(2) Generally, only a single-scale convolution kernel is set in the nonlinear mapping layer, and the model has only a single-scale receptive field, which can only capture a specific form of image feature information in a specific region, which is obviously not conducive to the diversity of image reconstruction details.

To solve the above problems, we propose an image super-resolution network RepSRCNN based on structure reparameterized convolution. Specifically, we designed a structural re-parameterized convolutional block to replace the convolutional layer in the traditional image SR network with its stacked multi-layer network, which cleverly deepens the network depth and increases the network receptive field[15], thus providing more abundant feature information for image reconstruction. In a single structure reparameterized convolution block, parallel large convolution kernel and small convolution kernel are set for learning training at the same time, and multiple convolution nuclei are merged into one after training, which greatly reduces the complexity of the model and increases the inference speed. In addition, fusion of image information learned by multiple convolution cores gives large convolution cores the ability to capture small-scale image information, effectively improving the reconstruction result. This method can also reduce the impact of partial information lost in the convolution process on the final output result by setting convolution cores with the size of 1*1. Thus, to some extent, the adverse effects of edge zeroing operation in traditional image SR network are avoided.

The work of this paper is summarized as follows:

(1) A structural re-parameterized convolutional block is designed. Compared with the traditional convolutional module, the newly designed module uses only one edge zero filling operation and sets 1*1 convolution to avoid adverse effects caused by the zero-filling operation. The image size does not change before and after the input module, so a very deep network can be stacked with this module.

(2) The application of BN (Batch Normalization) layer[16] in image super-resolution network model is analyzed and explored. Existing views[2, 11] hold that BN layer is not suitable for application in image super-resolution tasks. However, this paper believes that adding BN layer to a specific structure will greatly improve the results without affecting the training stability.

ISSN: 2710-0588

(3) A new network called RepSRCNN is proposed for single-image super-resolution tasks. The network can not only learn the global distribution of the input image, but also capture the local information in a small range and accelerate the inference speed effectively after structural reparameterization. The performance of the network is better than that of the traditional image super-resolution network in training and testing.

2. Proposed Method

This paper proposes an image super-resolution network based on structural reparameterized convolution. The overall network structure draws on the traditional SRCNN[1]. The advantage of this structure is that it is simple in structure and can be trained easily. The method proposed in this paper adds a transpose layer on top of the original three-layer structure to adjust the size of the image. The original nonlinear mapping layer is a single layer structure based on the general convolution kernel, and only one convolution operation is performed on the image, which is replaced by the structural re-parameterized convolution block proposed in this paper. The overall structural design of the network model is shown in Fig. 1:

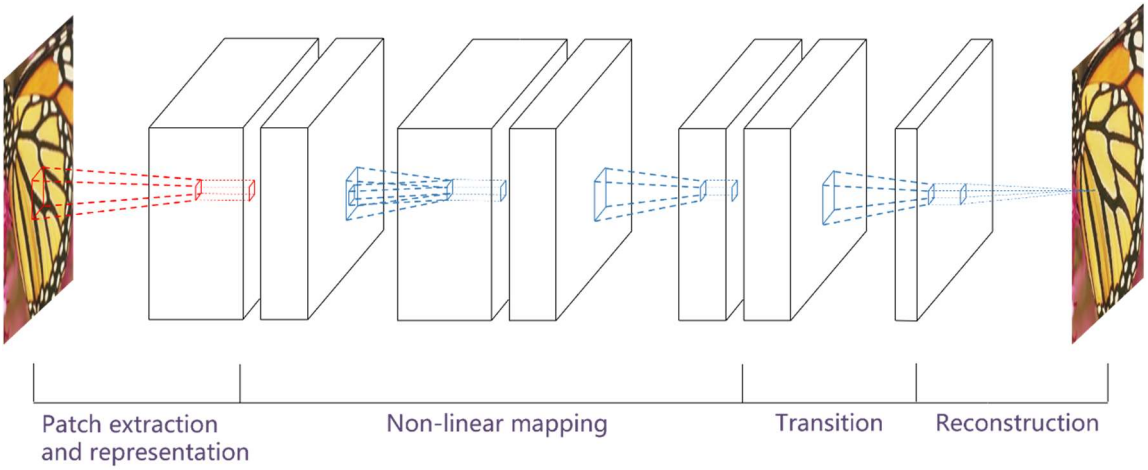

Fig. 1 Overall structural framework of RepSRCNN

The patch extraction and representation layer of the network is only realized by a simple convolution layer, which performs shallow feature extraction and nonlinear activation of input low-resolution images. The convolution kernel size is 3*3, the number of channels is 128, and the nonlinear activation function adopts Relu function. The mathematical description of this layer is shown in equation (1):

$$F_1(Y) = \max(0, W_1 * Y + B_1) \tag{1}$$

where, Y represents the input image, W_1 is the weight of the convolution kernel, B_1 is the convolution kernel bias, and $*$ represents the convolution operation.

The nonlinear mapping layer is formed by stacking structure re-parameterized convolutional blocks to form a deep network. Due to the reparameterization of this module, various internal complex structures can finally be fused into a convolutional kernel with various learned information and a skip connection structure is set, so it can be abstracted mathematically into a convolutional kernel with residual learning structure, as shown in equation (2):

ISSN: 2710-0588

$$F_2(Y) = \max(0, W_2 * Y + B_2) + Y \tag{2}$$

where W_2 and B_2 represent the weight and bias of the convolution after the fusion of the layer. The function of the transition layer is to transpose the image in terms of channel, size and adjust the shape of the image to facilitate the final reconstruction. Its interior is composed of a simple 2D convolution stack, the convolution kernel size is set to 3*3, the step size is 1, and the number of channels for the final output image is 32. The mathematical expression is (3):

$$F_3(Y) = \max(0, W_3 * Y + B_3) \tag{3}$$

The final reconstruction layer is a single convolution block with a size of 3*3 and a stride of 1, without activation function. Its function is to reconstruct the feature graph processed by the network through up-sampling. The mathematical expression is as follows (4):

$$F_4(Y) = W_4 * Y + B_4 \tag{4}$$

the image is reconstructed after passing through this layer and becomes the final output HR image.

2.1. Structure Reparameterized Convolution Blocks

The section headings are in boldface capital and lowercase letters. Second level headings are typed as part of the succeeding paragraph (like the subsection heading of this paragraph). All manuscripts must be in English, also the table and figure texts, otherwise $_s w$ we cannot publish your paper. Please keep a second copy of your manuscript in your office.

In the nonlinear mapping layer, the network adopts the structure re-parameterized convolution block designed in this paper to replace the traditional 2D convolution. In the algorithm proposed in this paper, only two convolution kernels are set. The specific structure can be expressed as a 1*1 convolution kernels set in parallel beside a normal 3*3 convolution kernels. The two kernels are called large kernels and small kernels respectively. The large and small nuclei have different regional information capturing abilities, that is, they have different scale receptive fields. During training, both the large and small cores are learned at the same time, and the whole structure can be described by formula (5):

$$F(Y) = \max[0, BN(W_l * Y + B_l) + BN(W_s * Y + B_s)] + Y \tag{5}$$

where W_l and B_l represent the weight and bias parameters of large convolution kernel, W_s and B_s represent the weight and bias parameters of small convolution kernel, and $BN(\cdot)$ represents the mapping of Batch Normalization layer. The mathematical formula is as follows:

$$BN(x) = \frac{\lambda x}{\sqrt{Var(x) + \varepsilon}} + (\beta - \frac{\lambda E(x)}{\sqrt{Var(x) + \varepsilon}}) \tag{6}$$

where λ and β are the learnable parameters of the BN layer. Fig. 2 shows the structural design of the entire module:

ISSN: 2710-0588

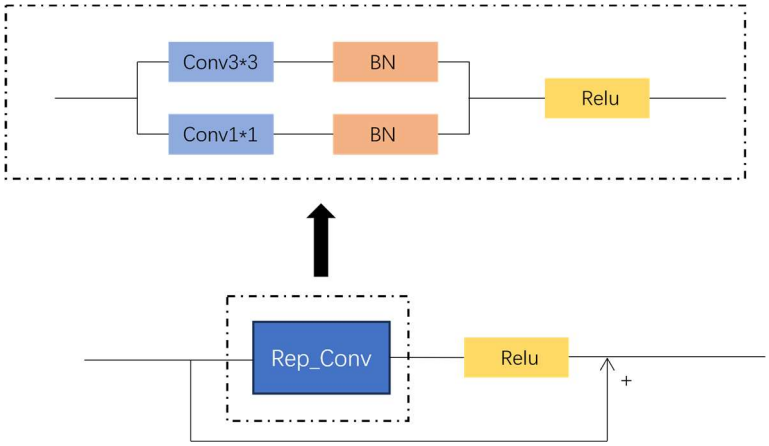

Fig. 2 Structural reparameterized convolutional block

Image super-resolution tasks have strict requirements on the output size of images. In order not to affect the output size, an ordinary convolution kernel will carry out edge filling operations on the convolution feature graph after a convolution operation, and the resulting image will then be transferred to the next layer for processing. In this process, part of the edge information of the image will be lost, and such information loss will be superimposed with the deepening of the network. Such superposition of a deep network further affects the final reconstruction quality of the image. If convolution kernels of multiple scales are set in parallel, the weighted addition of the result can enrich image information more than the convolution layer of convolution kernels of single size. Convolution kernels of 1*1 size will not change the image size. Therefore, after merging the results of convolution kernels of general convolution and convolution kernels of 1*1 size, Part of the edge information lost due to the edge filling operation can be largely supplemented, as shown in Fig. 3:

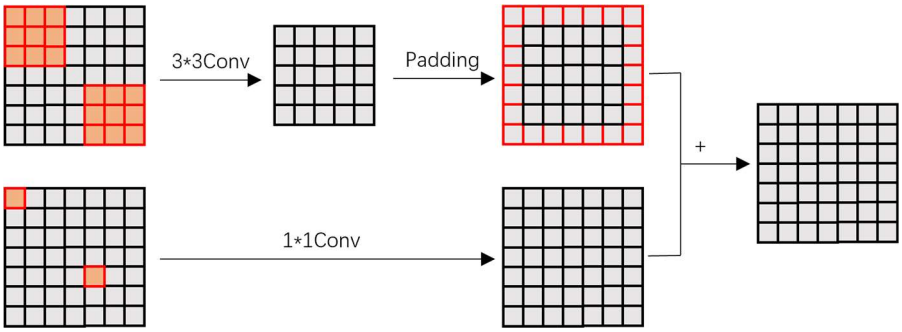

Fig. 3 Fusion process of convolutional results

The residual network can also achieve the above purpose to some extent because of its internal skip connection structure, which can send the original input image directly to the end of the module for addition. However, if the residual structure is simply used without any improvement, The problem of skip join is that the input image containing rich low frequency information is directly transferred to the end of the input image, and the higher-order information contained in the fused feature map is obviously inferior to the result of multi-scale convolutional high-order extraction and fusion of structural reparameterized convolutional blocks. For example, for the same low-resolution input image, the difference between the reconstruction results of the VDSR network that only uses depth stacking residual blocks and the RepSRCNN network that uses structural reparameterized convolutional blocks proposed in this paper is shown in Fig. 4:

ISSN: 2710-0588

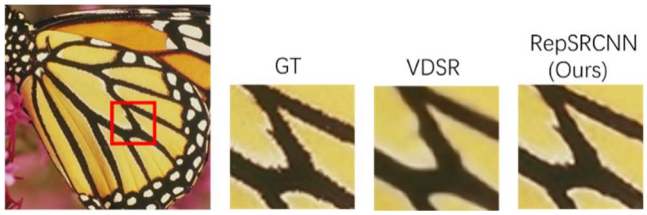

Fig. 4 Effect comparison

It can be clearly seen from the figure that the overall image reconstructed by the network of stacked residual blocks is somewhat smooth. Although good reconstruction effect is achieved in terms of shape and color, many details are missing in the texture edges. In contrast, the image reconstructed by using structural re-parameterized convolutional blocks has better subjective perception.

In addition, two convolution cores will be merged into one convolution kernel after structural reparameterization fusion, and the parameters learned by a small kernel can be embedded into a large kernel, and the two cores can learn at the same time during training. In this way, while the large kernel ensures that the model can learn basic features, the small kernel can also capture small-scale patterns, thus improving the final reconstruction effect of the model. Using branch structure in training can effectively improve the performance of the network and using structure reparameterization to change it into single structure in reasoning, which can reduce the memory and speed up the reasoning speed. Table 1 lists the parameters:

Table 1. The time of reasoning

status	Inference time
before	26.603ms
after	19.594ms
differ	7.009ms

After the model is reparameterized, the inference time is shortened by 26.35%, which proves the validity and necessity of the structural reparameterization algorithm.

2.2. Structure Reparameterization

Ding et al.[17-19] found that multiple convolutions check of the same image with size compatibility, and the resulting sum can be equivalent to the result of a fusion convolution check of the same input. The fusion convolution kernel can be used to replace the original multiple convolution cores, to reduce the complexity of the model. To make the network model easier to train, this process becomes structural reparameterization.

In the structural reparameterized convolution block we used, only a large convolution kernel of size 3*3 and a small convolution kernel of size 1*1 are set. The large and small convolution kernels can be merged into a new convolution kernel with their respective subsequent connected BN layers. This process is mathematically described as Equation (7) (8):

$$\begin{cases} W_{fl} = \dfrac{\lambda W_1}{\sqrt{Var(W_l + \varepsilon)}} \\ B_{fl} = \beta - \dfrac{\lambda W_1}{\sqrt{Var(W_l + \varepsilon)}} \end{cases} \tag{7}$$

ISSN: 2710-0588

$$\begin{cases} W_{fs} = \dfrac{\lambda W_s}{\sqrt{Var(W_s + \varepsilon)}} \\ B_{fs} = \beta - \dfrac{\lambda W_s}{\sqrt{Var(W_s + \varepsilon)}} \end{cases} \tag{8}$$

suppose the two new convolution kernels after merging are K_l and K_s, where W_1 and W_s respectively represent the weights of the large and small convolution kernels, λ and β are the parameters learned at the BN layer, W_{fl} and W_{fs} represent the weights of the merged new convolution kernels K_l and K_s, and B_{fl} and B_{fs} represent the respective biases of K_l and K_s. Then K_l and K_s can be combined into one kernel again, as described in equation (9):

$$\begin{cases} W_f = W_{fl} + W_{fs} \\ B_f = B_{fl} + B_{fs} \end{cases} \tag{9}$$

let the new convolution kernel after merging be K_f, where W_f and B_f are the weight and bias of K_f respectively. Fig. 5 shows the fusion process.

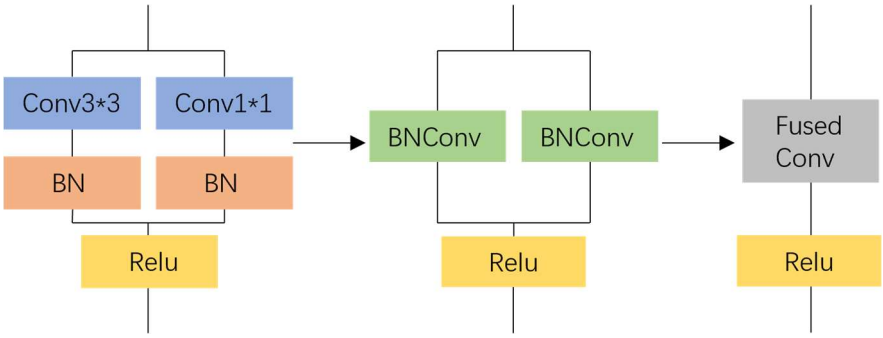

Fig. 5 Integration process

After structure reparameterization, the original convolution block structure is finally merged into a convolution kernel, which reduces the complexity of the model and can effectively accelerate the model inference speed.

2.3. About the BN layer

The addition of BN layer can make the data distribution of the neural network into a distribution of mean 0 and variance 1, so that it can achieve the effect of accelerating model convergence, controlling gradient explosion, preventing overfitting, etc., which has become one of the most used algorithms in deep learning, and has been widely used in most computer vision tasks such as image classification and target detection. However, the existing studies[2, 11] pointed out that the BN layer could not play a good role in the image super-resolution task, because the image super-resolution task had strict requirements on the information difference between the input and output images of the network, and the network only allowed to change the resolution and part of the details of the image, while the BN layer would normalize the data distribution. It will destroy the contrast and other information of the original image and have a great impact on the reconstruction result, so the BN layer is rarely used in the image super-resolution network.

ISSN: 2710-0588

Residual structure can avoid some of the problems mentioned above to a certain extent, BN layer will destroy the image contrast and other information, but the residual structure can pass the image's original contrast and other information directly to the next layer by jumping connection, so the original information of the image is retained, while maintaining the advantages of BN layer to eliminate its adverse effects.

RepSRCNN proposed in this paper only has a residual structure in the structural reparameterized convolutional block, so inserting BN layer into the structural reparameterized convolutional block can effectively improve the network performance. The experimental results show that:

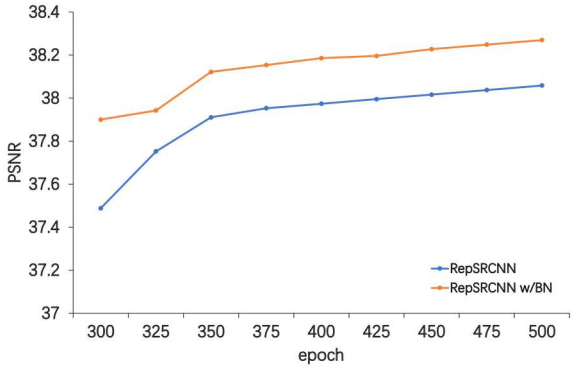

Fig. 6 The difference between models with and without BN layers

The lines in the figure represent RepSRCNN with BN layer and the original RepSRCNN respectively. The model has a better reconstruction effect on HR image after adding BN layer to the module.

3. Experiments

The experiment adopts 32GB running memory, IntelCore i7-12700KF CPU, NVDIA GeForce RTX 3060Ti GPU, 8GB video memory, The software environment is Windows10, pytorch1.13.1, CUDA12.2, and python3.9.13.

Set5, Set14, BSD100 and Urban100, commonly used in the field of image super-resolution reconstruction, are used in the test data set, with a total of 219 images, including various types of images with high and low frequency features ranging from simple to complex, and structures ranging from discrete to compact. The training data was selected from DIV2K data set, and the types covered people, buildings, natural scenery, animals, and other types, and it was cut into 33×33 image blocks, and the image blocks were degraded by double cubic interpolation method. The optimizer adopts SGD algorithm, the learning rate is set to 0.001, the momentum is set to 0.9, the Batch size of the whole training is 128, and the epoch is 500.

The mean square error loss MSE is used as the loss function for training, and the formula is (10):

$$Loss = \frac{1}{n}\sum_{i=1}^{n} \| F(Y_i;\theta) - X_i \|^2 \tag{10}$$

where n is the set value of Batch size, θ is the parameter set of the model, including the weight and bias of each convolutional kernel of the network, X_i represents the real image HR, and Y_i represents the low-resolution image LR corresponding to the label. Using MSE as a loss function

ISSN: 2710-0588

can help the network model get a higher PSNR value, forcing the network to optimize towards the label of the real image.

3.1. Comparison Experiment of the Number of Blocks

As mentioned above, the number of structural re-parameterized convolution blocks has a great impact on the quality of reconstruction results, so this experiment is to explore the number of blocks stacked to achieve the best reconstruction effect. In the experiment, Set5 test set with magnification of 2 was used to compare the test results under the number of re-parameterized convolution blocks with different structures, and PSNR and SSIM were used as indicators to measure the reconstruction effect. The specific experimental results are shown in Table 2:

Table 2. Comparison of reconstruction results with different module numbers

Blocks	Scale	PSNR	SSIM
1-Block	×2	38.02	0.9585
2-Block	×2	38.07	0.9597
4-Block	×2	38.23	0.9623
6-Block	×2	38.21	0.9618

As can be seen from Table 2, with epoch set to 500, PSNR is optimal when the number of blocks is 4. If the number of blocks is too small, the model features will not be fully obtained. Compared with the case of the number of blocks 1 and 2, the reconstruction effect of the model using 4 blocks is increased by 0.55% and 0.42%, respectively. However, when the number of blocks is set too large, the increase in the number of model parameters will lead to training difficulties, and it will also lead to performance degradation under 500 training cycles. Compared with the case of 6 blocks, the reconstruction effect of the model using 4 blocks is improved by 0.05%.

To sum up, in the nonlinear mapping layer of the model, four structural re-parameterized convolution blocks are used as the optimal construction, and the following experiments are conducted on the premise that the number of blocks is set to 4.

3.2. Ablation Experiment

As mentioned above, the number of structural re-parameterized convolution blocks has a great impact on the quality of reconstruction results, so.

To explore the influence of each structure in the image super-resolution model based on structure reparameterized convolution on the final reconstruction performance, this section designed an ablation experiment with the control of a single variable, that is, only one structure in the model was changed in each experiment, while the other structures remained unchanged. In the experiment, Set5 was used as the test set, the magnification was 2, and PSNR and SSIM were used as indicators to measure the reconstruction effect. The experimental results are shown in Table 3:

Table 3. Ablation experiment

Model	Rep-Conv	BN	PSNR	SSIM
1	√	√	38.23	0.9623
2	√	×	38.06	0.9607
3	×	√	36.90	0.9412
4	×	×	37.02	0.9466

ISSN: 2710-0588

In the table, Model 1 has two structures: structural re-parameterized convolutional block and BN layer. Model 2 has only structural re-parameterized convolutional block structure, and its comprehensive performance decreases by 0.45% compared with model 1; Model 3 only adds BN layer after the general convolutional layer, and its comprehensive performance decreases by 3.60% compared with model 1. The experiment shows that the addition of two kinds of structures can improve the reconstruction effect.

At the same time, it is noted that model 4 has neither structural re-parameterized convolutional blocks nor BN layer, and its comprehensive performance is improved by 0.33% compared with model 3. This data proves that BN layer is not suitable for general convolutional structures in image super-resolution models, and the addition of this layer will lead to the reduction of reconstruction effect. By comparing model 1 and model 4, it can be obtained that, Adding BN layer to the structural reparameterized convolution block will improve the overall performance, which confirms the previous view.

3.3. Contrast with Advanced Methods

To further analyze the performance of the model, this experiment compared the model proposed in this paper with other existing models and tested the average PSNR and SSIM of the reconstructed images of each model on the four test sets of Set5, Set14, BSD100 and Urban100 respectively, and selected three scales of magnification of 2, 3 and 4 times. Among them, Bicubic is the traditional bicubic interpolation algorithm, SRCNN[1] is the shallow convolutional neural network model, EDSR[11] and VDSR[2] are the deep residual network model, and the experimental results are shown in Table 4:

Table 4. Assessment results

Method	Scale	Set5 (PSNR/SSIM)	Set14 (PSNR/SSIM)	BSD100 (PSNR/SSIM)	Urban100 (PSNR/SSIM)
Bicubic	2	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403
SRCNN	2	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946
FSRCNN	2	37.05/0.9560	32.66/0.9090	31.53/0.8920	29.88/0.9020
VDSR	2	37.53/0.9590	33.05/0.9130	31.90/0.8960	30.77/0.9140
EDSR	2	38.11/0.9602	33.92/0.9159	32.32/0.9013	32.93/0.9351
RepSRCNN	2	**38.23/0.9623**	**34.07/0.9212**	**32.36/0.9021**	**33.01/0.9378**
Bicubic	3	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349
SRCNN	3	32.75/0.9090	29.30/0.8215	28.41/0.7863	26.24/0.7989
FSRCNN	3	33.18/0.9140	29.37/0.8240	28.53/0.7910	26.43/0.8080
VDSR	3	33.67/0.9210	29.78/0.8320	28.83/0.7990	27.14/0.8290
EDSR	3	34.65/0.9280	30.52/0.8462	29.25/0.8093	28.80/0.8653
RepSRCNN	3	**34.71/0.9305**	**30.58/0.8481**	**29.28/0.8097**	**28.91/0.8705**
Bicubic	4	28.42/0.8104	26.00/0.7027	25396/0.6675	23.14/0.6577
SRCNN	4	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221
FSRCNN	4	30.72/0.8660	27.61/0.7550	26.98/0.7150	24.62/0.7280
VDSR	4	31.35/0.8830	28.02/0.7680	27.29/0.0726	25.18/0.7540
EDSR	4	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/0.8033
RepSRCNN	4	**32.56/0.8991**	**28.84/0.7892**	**27.72/0.7436**	**26.66/0.8048**

The data in the table show that the RepSRCNN model proposed in this paper is superior to other algorithm models under the two objective evaluation criteria of PSNR and SSIM. Under 2x, 3x and 4x magnification, the PSNR values of the model proposed in this paper are respectively

ISSN: 2710-0588

higher than SRCNN method by 4.57dB, 4.32dB and 4.14dB, and higher than EDSR method by 0.12dB, 0.06dB and 0.10dB, respectively. Experiments show that the RepSRCNN model proposed in this paper has certain advantages in the task of image super-resolution reconstruction.

To feel the reconstruction effect of high-resolution images, this experiment selected four images more intuitively from the test set and used RepSRCNN and other models to carry out super-resolution reconstruction respectively to obtain a more intuitive comparison of subjective evaluation effects, as shown in Fig. 7:

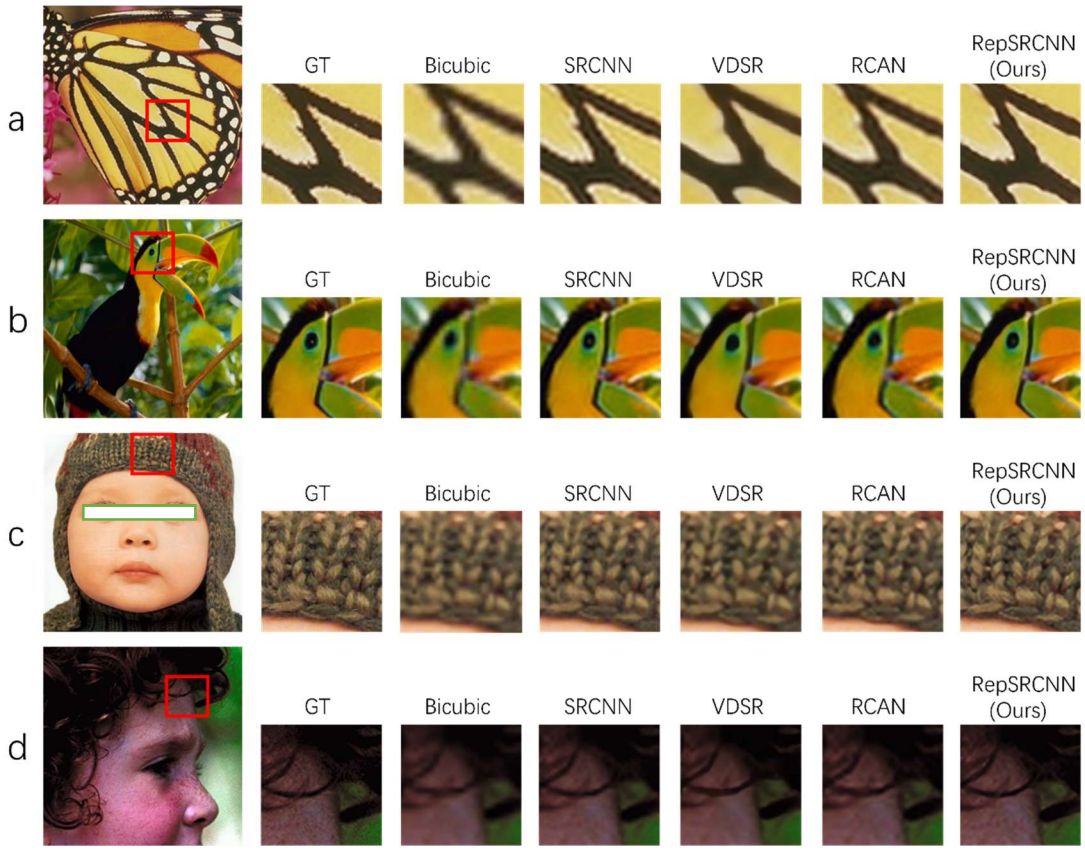

Fig. 7 Image super-resolution reconstruction results

In the figure, GT represents the real image without degradation, Bicubic is the image amplified by bicubic interpolation, and SRCNN, VDSR and RCAN are three different types of image super-resolution models. As can be seen from Fig. 3-11, the high-resolution image interpolation after bicubic processing has obvious traces and a high degree of blur, which is most inconsistent with human subjective feelings. After SRCNN reconstruction, the high-resolution image is blurred a lot, but there are obvious artifacts and jagged edges, and the perception is poor. Compared with the previous methods, the resolution of high-resolution images reconstructed by VDSR and RCAN is significantly improved. However, compared with the real image, the texture edge of high-resolution images reconstructed by VDSR is too smooth, and the reconstruction of hair will appear unreal due to smoothness, while the HR reconstructed by RCAN method will have some noise. In summary, the high-resolution image reconstructed by the RepSRCNN model proposed in this paper is superior to other methods in terms of sharpness, edge jagging and texture effect, which is closer to the real image and more in line with people's subjective feelings.

ISSN: 2710-0588

4. Summary

In this paper, a new image super-resolution network model RepSRCNN based on structural re-parameterized convolutional blocks is proposed. The convolutional block structure in this model can learn and merge multiple convolution cores of different sizes simultaneously by parallel setting, which can enable the nonlinear mapping layer to obtain more specifications and richer image information. And to some extent, the image information lost due to edge filling operation can be avoided. At the same time, the model verifies that adding BN layer to the structural reparameterized convolution block is beneficial to improve the final reconstruction effect. The PSNR and SSIM values obtained by experiments on different data sets prove the effectiveness of the proposed RepSRCNN. However, the experiment also revealed some defects and deficiencies of the RepSRCNN model, such as training difficulties, model convergence difficulties, and poor reconstruction effect of point graphics under high magnification. It is expected that these problems can be further solved in the future work.

References

[1] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13. Springer International Publishing, 2014: 184-199.

[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] Wang X, Yu K, Wu S, et al. Esrgan: Enhanced super-resolution generative adversarial networks[C]. Proceedings of the European conference on computer vision (ECCV) workshops. 2018: 0-0.

[4] 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.

[5] Wang X, Xie L, Dong C, et al. Real-esrgan: Training real-world blind super-resolution with pure synthetic data[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 1905-1914.

[6] 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.

[7] Chen H, Gu J, Zhang Z. Attention in attention network for image super-resolution[J]. arxiv preprint arxiv:2104.09497, 2021.

[8] Liang J, Cao J, Sun G, et al. Swinir: Image restoration using swin transformer[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 1833-1844.

[9] Li H, Yang Y, Chang M, et al. Srdiff: Single image super-resolution with diffusion probabilistic models[J]. Neurocomputing, 2022, 479: 47-59.

[10] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681-4690.

[11] 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.

[12] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1637-1645.

[13] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[14] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[15] Luo W, Li Y, Urtasun R, et al. Understanding the effective receptive field in deep convolutional neural networks[J]. Advances in neural information processing systems, 2016, 29.

ISSN: 2710-0588

[16] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. International conference on machine learning. pmlr, 2015: 448-456.

[17] Ding X, Zhang X, Ma N, et al. Repvgg: Making vgg-style convnets great again[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.

[18] Ding X, Guo Y, Ding G, et al. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1911-1920.

[19] Ding X, Hao T, Tan J, et al. Resrep: Lossless cnn pruning via decoupling remembering and forgetting[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 4510-4520.