Universe Scientific Publishing

DOI:10.18686/ahe.v7i13.8462

Lightweight Network Based on Interleaved Group Convolution for Image Super-Resolution

Jiexin Zhang, HanWang, Jiaxin Luo, Zheng Zhang

School of Chengdu Technological University, Chengdu 610000, China

Abstract: Deep learning has been successfully applied to single im- age super-resolution problems due to its high data fitting ability. However, the increasing depth and complexity of the network has brought about the disappearance of information, data volume and computational redundancy, and is not suit- able for small devices. To solve these problems, we propose a new lightweight network model based on interleaved group convolution for single image super-resolution reconstruction. The core idea of this algorithm is to broaden the network structure by means of group convolution, enhance the sparsity of the convolution kernel, and achieve the purpose of reduc- ing the amount of calculation and the amount of parameters. After a lot of experimental evaluation, we prove that our al- gorithm can achieve better results with a smaller number of parameters.

Keywords: Super-Resolution; Deep-learning; Lightwe- ight model; Residual Learning; Interleaved group convolution

Fund Project: The work was supported by the University level project of CDTU(2022ZR051), by the College Students' Innovative Entrepreneurial Training Plan Program in Sichuan Province(202211116015).

1. Introduction

Single image Super-Resolution(SISR) reconstruction algo- rithm has gained increasing research attention for decades. SISR is broadly divided into three research directions: in- terpolation based[1], reconstruction based[2], and learning based methods. At present, the learning-based method has become an important research direction of SISR, and the interpolation-based and reconstruction-based methods are usually used as auxiliary processing processes based on learn- ing methods. Among them, the super-resolution reconstruc- tion algorithm SRCNN proposed by Dong et al[3] is based on convolutional neural network for feature extraction, and then achieves super-resolution reconstruction by changing convo- lution kernel and network layer, which has achieved very good results. Based on this, more algorithms try to strengthen the effect of super-resolution algorithm by deepening the neu- ral network. However, this structure has several shortcom- ings: 1) It does not consider the network degradation problem that may occur with the deepening of the network; 2) The data redundancy problem that is inevitable due to the deepening of the structure has not been solved. In order to solve the above problems, we propose a neural network based on interleaved group convolution for image super-resolution.

2. Related works

In order to reduce or even eliminate the redundancy of data, we usually adopt the method of simplifying the network struc- ture or eliminating redundancy in the convolution. This pa- per only discusses the method of eliminating redundancy in the convolution. In order to reduce the amount of data and calculation in this process, it is generally processed for the convolution kernel. Common processing methods areconvert convolution kernels into: 1) low-precision kernels, the most common way is to binarize the convolution kernel. 2) low- rank kernels, the size of the convolution kernel will be re- duced from large to small, FSRCNN[4] is to use this method to reduce the training scale. 3) sparse kernels, that is, increas- ing the number of zeros in the kernels as much as possible to reduce the amount of calculation. 4) the free combination of the above methods.

Based on the above, Dr. Wang proposed a new structure named Interleaved Group Convolutions[5]. It is a convolu- tional neural network convoluted in units of convolutions. The convolutional layer is divided into multiple groups by channel to convolve separately to reduce the amount of pa- rameter data. For the problem that the correlation between different groups may be insufficient for group

convolution, the member exchange between groups is used to solve the problem. This structure reduces the parameters without causing performance degradation.Therefore, we adopt the same mechanism in different ways.

3. Proposed methods

In this section we will describe the proposed model architec- ture. Earlier we mentioned that the residual network and the interleaved group convolution (IGC) can solve the network degradation and data redundancy problem. Therefore, we try To replace the ordinary convolution kernel with the IGC block based on the VDSR structure to optimize the network struc- ture.

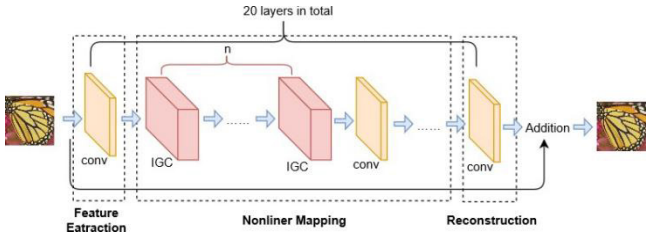

Fig. 1. The overall structure of the model is structure2 when n = 3, structure3 when n = 6, and structure1 when n = 18 and the whole structure has only two regular convolutional layers.

3.1 Network Structure

The overall structure of the proposed algorithm is shown in Figure 1. The model can be easily divided into three parts: feature extraction, nonlinear mapping and reconstruction.

Feature extraction: This part includes a convolution layer to map image channels to various feature maps. Considering that the network structure is widened, the correlation between the convolution layers may be weakened, we expand the con- volution kernel size in the feature extraction step from the usual $3 \times 3 \times 64$ to $3 \times 3 \times 96$.

Nonlinear mapping: The nonlinear mapping consists of regular convolutional layers and an IGC blocks to squeeze the amount of parameters, enhance the sparsity of the convolution kernel, and reduce the amount of computation.The structure of the IGC block will be discussed in 3.2

Reconstruction: Entering the reconstruction part, the re- construction part consists of a convolutional layer with a con- volution kernel size of $3 \times 3 \times 1$. In order to ensure the integrity of the image edge information in the entire network, we use the 0 padding method to keep the image size unchanged.

In order to find a structure with relatively optimal efficiency and results, we experimented with different numbers of convolutional layers and IGC layers. Based on the parameter quantity considerations, we conducted three sets of control experiments with different structures. in which structure1 re- places all convolution kernels except head and tail with IGC; structure2 is based on structure1, inserting a common convo- lution layer after every 3 IGC layers(n = 3); Structure3 is based on structure1, inserting a normal convolution layer af- ter every 6 IGC layers(n = 6). Ensure that the total depth of the above three models is 20 layers. Peak Signal to Noise Ratio (PSNR) and structural similarity index (SSIM) of each model were tested and compared. The results are shown in Table 1.

We have found through comparison experiments that com-pared with VDSR, the structure after adding the IGC layers, where

Table 1. The PSNR and SSIM values of different models on different test sets, where upscale = 2, 3 , the training set is General100 +91images+ BSD100, the test set is Set5, Set14, and para is the total parameter amount of each structure

	structure1	structure2	structure3	VDSR
Set5-x2	37.63 0.9593	37.59 0.9593	37.61 0.9593	37.53 0.9587
Set5-x3	33.57 0.9211	33.67 0.9215	33.65 0.9215	33.66 0.9213
Set14-x2	33.08 0.9133	33.07 0.9133	33.08 0.9131	33.03 0.9124
Set14-x3	29.68 0.8300	29.76 0.8310	29.78 0.8314	29.77 0.8314
para	578KB	1748KB	1163KB	2605KB

upscale is set to 2, the effect is better on each com- mon testset, and the parameter amount is smaller. And under the same training set, the evaluation indicators of structure1 are relatively best, structure3 is the second, and structure2 is the weakest. This shows that expanding the network width can effectively improve the quality of super-resolution recon- structed images at low magnification. However, when we per- formed high magnification amplification, we found that the results were different. From the results shown in Table 1, we found that when the magnification is 3 times, the result of the structure adding regular convolution layers is better than that of the structure using only the IGC blocks. After comprehen- sive consideration, we chose to use a convolution layer for every 6 blocks to enhance the correlation between the convo- lutional layers. After testing, we found that our algorithm can achieve similar or slightly better results than VDSR, and the parameter is about half of the VDSR.

3.2 Broaden and Squeeze

Since the emergence of network structures such as DenseNet[6] and ResNet, we can find that it is particularly important to have long and short branch structures in the neu- ral network. Although the flow of information can be made very good through skip connection, simply deepening the net- work still does not work well, so we try to widen the network. As mentioned in article[5], We divide a convolutional layer into two groups of convolutions by simply dividing the input chan- nel: primary group convolution and secondary group convo- lution, which are respectively on the primary and secondary partitions. Wherein the primary group convolution performs the spatial convolution over each primary partition separately, and the secondary group convolution performs a 1×1 con- volution over each secondary partition, blending the chan- nels across partitions outputted by primary group convolu- tion. The difference is that after each IGC block, we add a shuffle layer to scramble the convolution channel to avoid the problem of weak correlation caused by duplicate grouping in- formation.Its structure is shown in Figure 2.

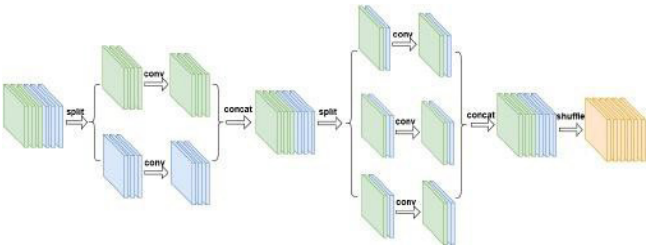

Fig. 2. Illustrating the IGC block. The left part is the primary group convolution, and the input multidimensional matrix is divided into L groups by channel, each group has M channels (L = 2, M = 3 in the figure), and the right part is secondary group convolution.

Suppose we need to convolve an image with n1 channels, and we want to extract n2 feature convolution layers (for ease of calcu- lation, take $n2 = n1 = n = M \times L$ here), the convolution kernel size is f, the total parameter amount $para1 = n \times f \times f \times n$ is required. Using the structure of the IGC proposed above, the required parameter quantity is $para2 = (M \times f \times f \times M) \times L + (L \times 1 \times 1 \times L) \times M$,

according to para1 - para2 > 0 We can see that para2 is less than para1, that is, the use of IGC layer instead of ordi- nary convolutional layer can achieve the purpose of reducing the amount of data. In this paper, L = 24 and M = 4.

ADAM[15] optimization method is adopted. The parameters β1 and β2 are set to 0.9 and 0.999, ϵ = 1- 8e, and the initial learning rate is 0.001, which decreases twice. The learning rate was reduced by 10 times at epoch = 30 and epoch = 45, respectively. According to the training of the L1 training network proposed by Zhao et al.[16], the performance is improved compared with the L2 training network. So this article uses L1 loss as a loss function instead of using the usual L2 loss or mean squared error (MSE). The formula for the L1 loss function is $L1 = \frac{1}{n}\sum_{i=1}^{n} | F_{x_i}(\grave{e}) - y_i |$ where is the reconstracted HR image from the LR image , is the ground truth image, and θ represents all parameters for training. This article sets the batch size to 16. The deep learning framework used in this article is CAFFE.

4. Experiment and results

4.1 Datasets

We use DIV2K[9] to generate training LR and HR patches. DIV2K is a new high quality (2K resolution) image dataset proposed by the SR mission. It consists of 800 training im- ages, 100 verification images and 100 test images. During the test, the data sets Set5[10] and Set14[11] are typically used for SR benchmarks. The B100[12] from the Berkeley split dataset consists of 100 natural images and is used for testing too. In addition, the proposed method was evaluated using the Urban100 dataset[13], which included 100

challenging images. Experiments were performed using a scale factor of 2×, 3×, and 4× between LR and HR images.

4.2 Implementation Details

In order to fully train the data, each image is downsampled by bicubic interpolation and cropped into small nonoverlapping images of size 36 × 36. Because the three-channel image is easier to fall into the local optimal solution when training, this paper transforms the RGB image into the YCbCr color space, and only trains the Y channel, and uses the rectified linear units (ReLU)[14] as the activation function. In this paper, the

(a) (b) (c) (d) (e) (f) (g)

Fig. 3. SR results of "img016"(in Urban100) with scale fac- tor × 3, The picture (a) is a ground truth picture, the pic- ture from (b) to (g) are based on: HR, bicubic, SRCNN, DDSR, VDSR and our algorithm, the parameter values are (PSNR/SSIM): 22.45/0.7992, 24.59/0.8817, 24.77/0.8847, 26.01/0.90, 26.56/0.9150

4.3 Results

In this paper, PSNR and SSIM are chosen as the evaluation criteria for the image super-resolution effect. The test re- sults are shown in Table 2. Based on the quantitative eval- uation of the parameters in the table, our model is superior to other advanced algorithms mentioned in the table. In Figure 4, we can visually compare our algorithm with other algo- rithms. Compared with other algorithms, the algorithm can obtain more clear edge information, and is more sensory than the original image. Therefore, our algorithm can ensure better super-resolution reconstruction while ensuring the weight of the model.

5. Conclusion

In this paper, we propose a lightweight network for super- resolution based on interleaved group convolution. The ex- perimental results of the benchmark datasets show that our model achieves competitive performance compared to current state-of-the-art models with less parameters and faster speed.

Table 2. The average PSNR and SSIM comparison results of different algorithms on the test set (red is the best result, blue is the second)

Dataset	Scale	Bicubic PSNR/SSIM	SRCNN[3] PSNR/SSIM	DDSR[7] PSNR/SSIM	VDSR[8] PSNR/SSIM	OURS PSNR/SSIM
Set5	×2	33.66 / 0.9299	36.66 / 0.9542	37.23 / 0.9574	37.53 / 0.9587	37.73 / 0.9596
	×3	30.39 / 0.8682	32.75 / 0.9090	33.23 / 0.9166	33.66 / 0.9213	33.78 / 0.9233
	×4	28.42 / 0.8104	30.48 / 0.8286	30.82 / 0.8718	31.35 / 0.8838	31.36 / 0.8839
Set14	×2	30.24 / 0.8688	32.42 / 0.9063	32.79 / 0.9102	33.03 / 0.9124	33.11 / 0.9137
	×3	27.55 / 0.7742	29.28 / 0.8209	29.55 / 0.8264	29.77 / 0.8314	29.80 / 0.8320
	×4	26.00 / 0.7027	27.49 / 0.7503	27.69 / 0.7569	28.01 / 0.7674	28.08 / 0.7767
B100	×2	29.56 / 0.8431	31.36 / 0.8879	31.81 / 0.8945	31.90 / 0.8960	31.95 / 0.8970
	×3	27.21 / 0.7385	28.41 / 0.7863	28.73 / 0.7943	28.82 / 0.7976	28.83 / 0.7991
	×4	25.96 / 0.6675	26.90 / 0.7101	27.10 / 0.7183	27.29 / 0.7251	27.31 / 0.7264
Urban100	×2	26.88 / 0.8403	26.50 / 0.8946	/	30.76 / 0.9140	31.07 / 0.9173
	×3	24.26 / 0.7349	26.24 / 0.7989	/	27.14 / 0.8279	27.16 / 0.8284
	×4	23.14 / 0.6577	24.53 / 0.7221	/	25.18 / 0.7524	25.19 / 0.7527

In addition, we have demonstrated the effective combination of IGC blocks and regular convolutional layers, and success- fully applied the idea of broadening the network structure to super-resolution, which provides a new research directions for our future research. Based on its lightweight nature, our next research is to apply this structure to small devices such as mobile phones, but it still needs to be explored.

References:

[1] Francesc Aràndiga, "A nonlinear algorithm for mono- tone piecewise bicubic interpolation," Applied Mathe- matics and Computation, vol. 272, pp. 100–113, 2016.

[2] Xin Yang, Yan Zhang, Dake Zhou, and Ruigang Yang, "An improved iterative back projection algorithm based on ringing artifacts suppression," Neurocomputing, vol. 162, pp. 171–179, 2015.

[3] Chao Dong, Chen Change Loy, Kaiming He, and Xi- aoou Tang, "Learning a deep convolutional network for image super-resolution," in Computer Vision - ECCV 2014 - 13th European Conference, 2014, pp. 184–199.

[4] Chao Dong, Chen Change Loy, and Xiaoou Tang, "Ac- celerating the super-resolution convolutional neural net- work," in Computer Vision - ECCV 2016 - 14th Euro- pean Conference, Amsterdam, 2016, pp. 391–407.

[5] Ting Zhang, Guo-Jun Qi, Bin Xiao, and Jingdong Wang, "Interleaved group convolutions for deep neural net- works," CoRR, vol. abs/1707.02725, 2017.

[6] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao, "Im- age super-resolution using dense skip connections," in IEEE International Conference on Computer Vision, ICCV, 2017, pp. 4809–4817.

[7] Zhang Y Liu SG Guo M. Peng YL, Zhang L, "Deep deconvolution neural network for image super- resolution.(in chinese).," Journal of Software, pp. 29(4): 926–934, 2018.

[8] Jiwon Kim, Jung Kwon Lee, and Kyoung MuLee, "Ac- curate image super-resolution using very deep convolu- tional networks," in 2016 IEEE Conference on Com- puter Vision and Pattern Recognition,CVPR, 2016, pp. 1646–1654.

[9] Radu Timofte, Eirikur Agustsson, and et al, "NTIRE 2017 challenge on single image super-resolution: Meth- ods and results," in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR, 2017, pp. 1110–1121.

[10] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie-Line Alberi-Morel, "Low-complexity single- image super-resolution based on nonnegative neighbor embedding," in British Machine Vision Conference, BMVC, 2012, pp. 1–10.

[11] Roman Zeyde, Michael Elad, and Matan Protter, "On single image scale-up using sparse-representations," in Curves and Surfaces - 7th International Conference, 2010, pp. 711–730.

[12] David R. Martin, Charless C. Fowlkes, Doron Tal, and Jitendra Malik, "A database of human segmented nat- ural images and its application to evaluating segmenta- tion algorithms and measuring ecological statistics," in 2001 IEEE International Conference on Computer Vi- sion, ICCV, 2001, pp. 416–425.

[13] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja, "Single image super-resolution from transformed self-exemplars," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 5197–5206.

[14] Kevin Jarrett, Koray Kavukcuoglu, Marc'Aurelio Ran- zato, and Yann LeCun, "What is the best multi-stage ar- chitecture for object recognition?," in IEEE 12th Inter- national Conference on Computer Vision, ICCV, 2009, pp. 2146–2153.

[15] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," CoRR, vol. abs/1412.6980, 2014.

[16] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in 2017 IEEE Conference on Computer Vision and Pattern Recogni- tion,CVPR, 2017, pp. 5835–5843.