NTIRE 2023 Efficient SR Challenge Factsheet
-Symmetrical Visual Attention Network for Efficient Image Super-Resolution-

Jing Hu, Chengxu Wu, Qinrui Fan, Chengming Feng, Xi Wu
Chengdu University of Information Technology
No.24, Xuefu Road, Southwest Airport Economic Development Zone, Chengdu, China
woox929@163.com

Ziwei Luo
Uppsala University
P.O. Box 256, SE-751 05 Uppsala, SWEDEN

Xin Wang, Shu Hu, Siwei Lyu
University at Buffalo
12 Capen Hall, Buffalo, New York, USA

1. Introduction

Single-image super-resolution (SISR) is a hot research topic in the field of image processing. In traditional single-image super-resolution research, especially the super-resolution algorithms based on deep learning, good reconstruction results can be obtained. However, these models still have some shortcomings, including a large number of network parameters, a long training time, and high computational overhead. It cannot meet the task of ensuring the reconstruction effect and running fast, efficiently, and lightweight. Therefore, we propose a lightweight Symmetrical Visual Attention Network (SVAN) model in NTIRE 2023 Efficient Super-Resolution Challenge, which is inspired by the Visual Attention Network [2] and VapSR [5], the large kernel convolution is decomposed by deep-wise convolution operation, which ensures a large receptive field while greatly reducing the number of parameters, and achieving an efficient and lightweight attention structure. the model guarantees better X4 refactoring results with lightweight parameter optimization.

2. Factsheet Information

This section covers team details and modeling methods, including Symmetrical Visual Attention Network(SVAN) and training details

2.1. Team details

- Team name: CUIT_SRLab
- Team leader name: Jing Hu
- Chengdu University of Information Technology, Department of Computer Science, Chengdu, China, +8618788913979, woox929@163.com
- Rest of the team members: Chengxu Wu, Chengming Feng, Qinrui Fan, Ziwei Luo, Xin Wang, Shu Hu, Siwei Lyu, Xi Wu
- User names and entries on the NTIRE 2023 Codalab competitions: AweWoo, Efficient Super-Resolution Challenge
- Best scoring on development: PSNR 28.6db validation: PSNR 26.6db
- Link to the codes: https://github.com/IridescentW/SVAN

2.2. Method details

SVAN is divided into three parts, shallow feature extraction module, deep feature extraction module and pixel-shuffle reconstruction module.The input LR is obtained by bicubic downsampling, and the shallow feature map is generated by a 3 x 3 convolution. The deep feature extraction contains 7 Symmetrical Large Kernel Attention Block (SLKAB), each SLKAB is expanded from 32 to 64 channels by 1 x 1 convolution and performs GELU activation to facilitate more information. The deep features are obtained by two symmetric attention blocks, each consisting of a 5 x 5 depth-wise convolutionand a depth-wise dilation convolution with a kernel size of 5 and dilation of 3 and a 1 x 1 convolution. The convolution acceptance domain with kernel 17 can be obtained through the convolution combination,

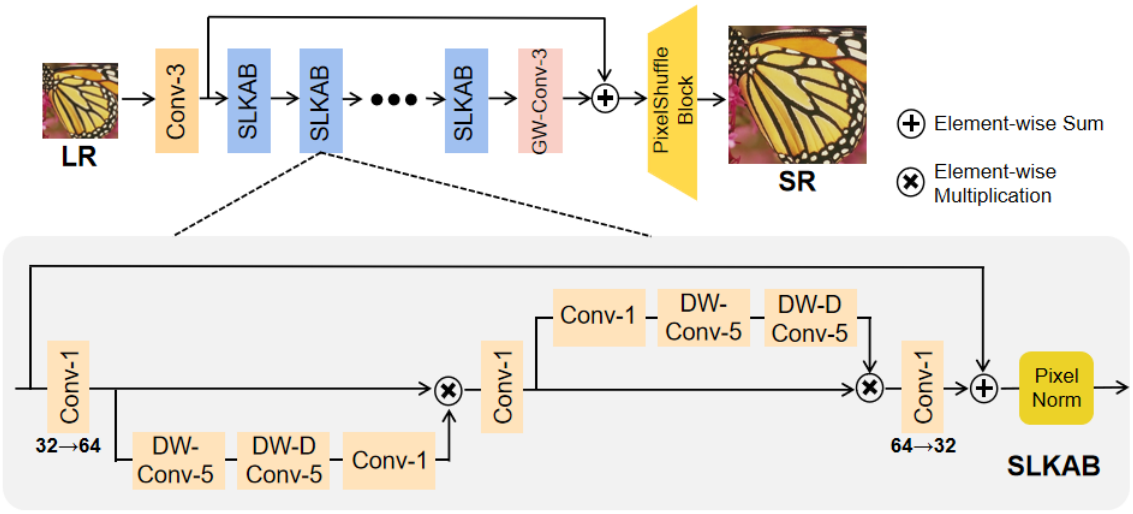

Figure 1. Symmetrical Visual Attention Network (SVAN) and Symmetrical Large Kernel Attention Block (SLKAB)

greatly reduce the number of computational parameters. Attention branch generated features are fused with the original features using element-wise implementation. Another 1×1 convolutional layer will reduce the number of channels to 32. Finally, pixel normalization is used to increase the stability in training. After the deep and shallow features are fused, the features are upsampled using the reconstruction module to reach the HR size. The reconstruction module contains two X2 pixel-shuffle layers to achieve the X4 upsampling size, and both pixel-shuffle layers are preceded by convolutional layers to adjust the number of channels.

2.3. Train details

During the training, HR patches of size 256×256 are randomly cropped from HR images, the minibatch size is set to 196, the number of feature channels is set to 64. The learning rate is set to 1e-4 for 500 epochs of pre-training using the LSDIR [3] dataset and the Adam optimizer minimizing the L1 loss function. DIV2K [1] and Flickr2K [4] datasets are used for continue training the proposed model, Optimizer and loss are unchanged, the initial learning rate is set to 5e-5 and halved at every 200 epochs. After 3000 epochs, L2 loss is used for fine-tuning. And The initial learning rate is set to 1e-4 and halved at every 200 epochs. The total number of epochs is 10000.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 2

[2] Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, and Shi-Min Hu. Visual attention network. *arXiv preprint arXiv:2202.09741*, 2022. 1

[3] Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, et al. Lsdir: A large scale dataset for image restoration. 2

[4] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017. 2

[5] Lin Zhou, Haoming Cai, Jinjin Gu, Zheyuan Li, Yingqi Liu, Xiangyu Chen, Yu Qiao, and Chao Dong. Efficient image super-resolution using vast-receptive-field attention. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, pages 256–272. Springer, 2023. 1