NTIRE 2024 Efficient SR Challenge Factsheet -Multi-Scale Attention Network

Zhijian Wu, Jingchao Wang, Dingjiang Huang
School of Data Science and Engineering, East China Normal University
Shanghai, China
zjwu_97@stu.ecnu.edu.cn, jcwang@stu.ecnu.edu.cn, djhuang@dase.ecnu.edu.cn

1. Team details

- Team name

  DASE-IDEALab

- Team leader name

  Zhijian Wu

- Team leader address, phone number, and email

  Address: Shanghai, China

  Phone number: +86 15651768331

  Email: zjwu_97@stu.ecnu.edu.cn

- Rest of the team members Jingchao Wang, Dingjiang Huang

- Team website URL (if any)

  N/A

- Affiliation

  School of Data Science and Engineering, East China Normal University

- Affiliation of the team and/or team members with NTIRE 2024 sponsors (check the workshop website)

  N/A

- User names and entries on the NTIRE 2024 Codalab competitions (development/validation and testing phases)

  User name: zjwu

  testing phases: 1

- Best scoring entries of the team during development/validation phase

PSNR	Runtime	Params	Extra Data
26.89	0.12	87612.00	0

- Link to the codes/executables of the solution(s)

  https://github.com/wjc0602/DASE-IDEALab-NTIRE2024-ESR.git

2. Method details

2.1. The Overall Architecture

The overall architecture of our MSN mainly consists of three parts: shallow feature extraction, deep feature extraction, and upscaling reconstruction.

Specifically, we use a 3×3 convolution for extracting shallow features f_0 from the input LR image I_{LR}:

$$f_0 = H_s(I_{LR}) \tag{1}$$

where H_s represents a 3×3 convolutional layer. Subsequently, the shallow feature is used for the deep feature extraction by a stack of modified distillation attention blocks (MDAB). This process can be formulated as:

$$f_k = H_k(f_{k-1}), \quad k = 1, ..., n \tag{2}$$

where H_k denotes the k-th MDAB. f_{k-1} and f_k indicate input feature and output feature of the k-th MDAB. We employ a 3×3 convolution at the tail of the deep feature extraction, and add the residual f_0 to the output feature:

$$f_r = Conv(f_k) + f_0 \tag{3}$$

Finally, the SR images are constructed by the resulting deep features as follows:

$$I_{SR} = H_r(f_r) \tag{4}$$

where I_{SR} denotes the reconstructed SR image. H_r represents the reconstruction module composed of a pixel shuffle operator and a 3×3 convolution layer.

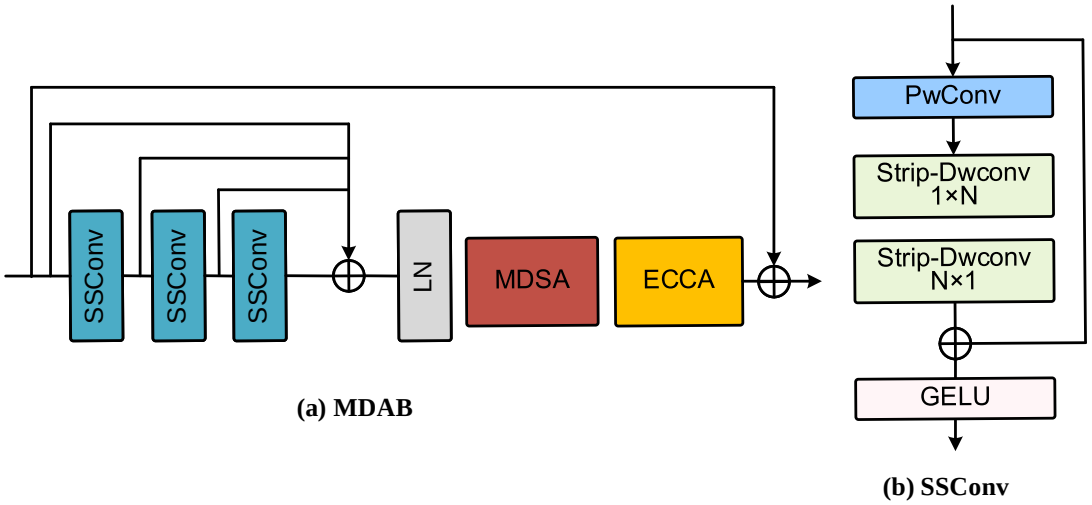

Figure 1. Proposed modified distillation attention block (MDAB).

2.2. Modified Distillation Attention Block

The proposed MDAB is shown in Fig. 1. Our approach is improved based on the Ntire2023ESR Challenge model complexity track champion MDRN. The main changes include three. First, we replace the base BSConv with our SSConv by decomposing the regular depth-wise convolution into two strip depth-wise convolutions. This design helps us to further reduce the model parameters. Second, we introduce layer normalization techniques in MDSA. Since the element-wise multiplication of MDSA leads to training instability when using large learning rates. Our design allows the use of larger learning rates making the model converge faster. Finally, we abandoned the design of using concat and convolution for feature fusion in favor of cheaper addition operations.

2.2.1 Training Details

Pretraining on DIV2K and the first 10k images of LSDIR. HR patch size and mini-batch size are set to 384×384 and 64, respectively. The model is trained by minimizing L1 loss function with Adam optimizer. The initial learning rate is set to 2×10^{-3} and halved at 100k iteration. The total number of iterations is 400k. In finetun stage, HR patch size and the mini-batch size are set to 480×480 and 64, respectively. The model is fine-tuned by minimizing L2 loss function. The initial learning rate is set to 5×10^{-5} and halved at 100k iteration. The total number of iterations is 400k.

2.3. Novelty degree of the solution

Our solution is an improvement on the existing model.

3. Other details

- Planned submission of a solution(s) description paper at NTIRE 2024 workshop.
- General comments and impressions of the NTIRE 2024 challenge.
- What do you expect from a new challenge in image restoration, enhancement and manipulation?
- Other comments: encountered difficulties, fairness of the challenge, proposed subcategories, proposed evaluation method(s), etc.