NTIRE 2023 Efficient SR Challenge Factsheet
-Efficient Feature Distillation Network-

Mingjian Zhang , Jingpeng Shi
Anhui University, Hefei, China
hukangyy@gmail.com,jinpeeeng.s@gmail.com

1. Team Details

- Team name: **FRL Team 02**
- Team leader name: **Mingjian Zhang**
- Team leader address, phone number, and email:
  - address: **Anhui University, Hefei, China**
  - phone number: **+86 136 6742 1529**
  - email: **zhang9317112@gmail.com**
- Rest of the team members: **Jinpeng Shi (advisor)**
- Team website URL (if any): github.com/Fried-Rice-Lab/FriedRiceLab
- Affiliation: **Anhui University**
- Affiliation of the team and/or team members with NTIRE 2023 sponsors (check the workshop website): **N/A**
- User names and entries on the NTIRE 2023 Codalab competitions (development/validation and testing phases):
  - user name: **zhang9317112**
  - development entries: **2**
  - validation entries: **17**
- Best scoring entries of the team during development/validation phase:

PSNR	SSIM	Runtime	Params	Extra Data
29.02 (10)	0.83 (9)	0.09 (28)	244558.00 (21)	1.00 (1)

- Link to the codes/executables of the solution(s): github.com/zhang9317112/NTIRE2023_ESR

Fried Rice Lab (FRL) is organized by students from Anhui University who are interested in image restoration. FRL is dedicated to proposing clean and efficient image restoration solutions and contributing to the image restoration open source community. **FRL Team 02**, led by Mingjian Zhang and advised by Jinpeng Shi, was among the teams that FRL sent to compete in the NTIRE 2023 ESR competition, with *Someone (replace if any)* completing the roster.

2. Method Details

The structure is inspired by classical SR structure——CNN(Convolutional Neural Network) and Transformer structure which is a common-used in SR area. Nowadays, transformer is widely used in lightweight sr tasks. However, we know from ConvneXt [1] that CNN is not inferior to trans in performance in many cases. from this idea we analyze the advantages of each of Transformer(Likes Swin transformer [2]) and CNN, and propose RB(Residual Block) based on the most classic SR network EDSR [3] with changes, adding a DSB(Dimensional Separable Block) which is similar to the transformer' self-attention mechanism to obtain long-range connections.We also add a nonlinear activation function to stabilize the training.

As shown in Fig1, our structure have three modules:(1)3*3 kernel convolution(shallow feature extraction).(2)Basic Layer(deep feature extraction).(3)3*3 kernel convolution and sub-Pixel(HR-image reconstruction).Our main contribution is to creatively present a useful layer called Basic Layer,which can process information efficiently.

2.1. Block Layer

Block Layer is the main layer in this structure.This Layer includes the following two modules——Dimensional separable Block and Residual Block.

2.2. Residual Block

Residual Block is improved by the classical convolutional stacking module of EDSR, which removes the BatchNorm and considers it redundant. But we found that LayerNorm [4] is suitable with SR tasks, so we added LayerNorm to that module to improve the performance.

2.3. Dimensional Separable Block

Poolformer [5] confirmed that th advantage of transformer is its backbone structure and its self-attention can be replaced by convolution. Inspiring by Depthwise Separable Convolution [6],We propose Dimensional Separable Block to divide the information features into H(Height), W(Weight), and C(Channel). From Mobilenets, [7] the feature information is divided into PW(PointWise) and DW(DepthWise), and we process the H and W information together for PW (intra-channel feature information) and then use c for DW (inter-channel feature information). From Simple baselines for image restoration [8], we know that the information of H and W is multiplied in front to achieve his nonlinearity, then finally outputs features. From Inception-ResNet [9], we know that the convolution kernel is decomposed so as to maintain performance while greatly reducing the parameters,So we use grouped convolution for feature extraction of H and W.

2.4. Training Details

Training was performed on DIV2K [10] and Flickr2K [11] images. HR patches of size 256 × 256 were randomly cropped from the HR images and the mini-batch size was set to 128. The model was trained with the ADAM optimizer [?], where $\beta_1 = 0.9$ and $\beta_2 = 0.9999$. The initial learning rate was set to 5×10^{-4} with cosine learning rate decay. The L2 loss was used for ab initio training and the number of iterations The model was implemented using Pytorch 1.10.1 and trained on 2 GeForce RTX 3090 GPUs.

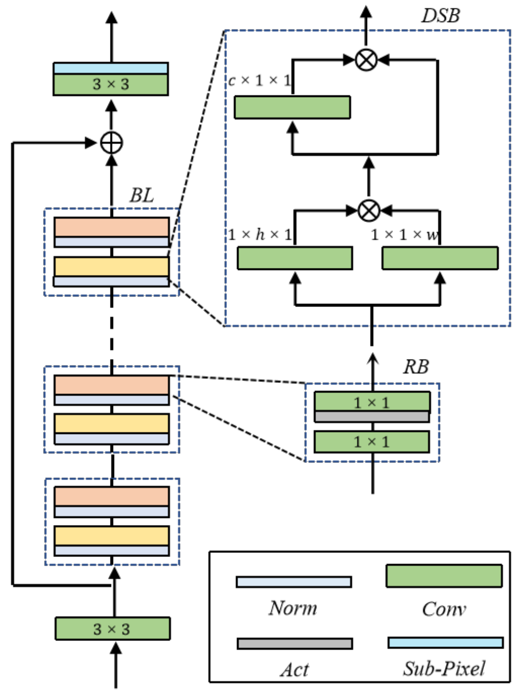
2.5. Experimental Results

The experimental result is shown in Tbale 1. FLOPs and Activation are tested on an LR image of size 256 × 256. PNSR[val] is tested on DIV2k validation dataset, while PSNR[test] is calculated on a combination of DIV2K and LSDIR test data. The runtime is is evaluated on DIV2K and LSDIR test datasets using a single GeForce RTX 3090 GPU.

3. Other details

- Planned submission of a solution(s) description paper at NTIRE 2023 workshop.

  We are not planning to submit the solution description paper to NTIRE2023 workshop, since it has been submitted to other conference.



For the purpose of easier understanding,the residual connections are omitted.

Figure 1. The architecture of Super Resolution Net-X (SRneXt)

PSNR[val]	PSNR[test]	Params[M]	FLOPs[G]
29.02	27.02	0.2445	15.376
GPU Mem.[M]	Activation[M]	Average Runtime[ms]	Conv2d
448.6826	422.903	91.16	158

Table 1. Result for NTIRE2023 ESR Challenge. FLOPs and Activation are tested on an LR image of size 256 × 256. The runtime is is averaged on DIV2K and LSDIR test datasets using a single NVIDIA RTX 3090 GPU.

- General comments and impressions of the NTIRE 2023 challenge.

  The organizers provided detailed processes and instructions and we appreciate the effort they spent!

References

[1] Liu Z, Mao H, Wu C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 11976-11986. 1

[2] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022. 1

[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. 1

[4] Ba J L, Kiros J R, Hinton G E. Layer normalization[J]. arXiv preprint arXiv:1607.06450, 2016. 2

[5] Yu W, Luo M, Zhou P, et al. Metaformer is actually what you need for vision[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 10819-10829. 2

[6] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258. 2

[7] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017. 2

[8] Chen L, Chu X, Zhang X, et al. Simple baselines for image restoration[C]//Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII. Cham: Springer Nature Switzerland, 2022: 17-33. 2

[9] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1). 2

[10] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In CVPR workshops, pages 126–135, 2017. 2

[11] 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 CVPR workshops, pages 114–125, 2017. 2