NTIRE 2023 Efficient SR Challenge Factsheet
LGTT: Local-Global Term Transformer for SR

Jinchen Zhu, Jinpeng Shi, Hui Li, Shizhuang Weng
Anhui university
Anhui,China
jinchen.Z@outlook.com,jinpeeeng.s@gmail.com

1. Team details

- Team name: **FRL Team 0**
- Team leader name: **Jinchen Zhu**
- Team leader address, phone number, and email:
  - address: **Anhui University, Hefei, China**
  - phone number: **+86 188 5507 6340**
  - email: **jinchen.Z@outlook.com**
- Rest of the team members: **Jinpeng Shi (advisor)**, **Shizhuang Weng (advisor)** and **Hui Li**
- 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: **Jinchen.Z**
  - development entries: **7**
  - validation entries: **8**
- Best scoring entries of the team during development/validation phase:

PSNR	SSIM	Runtime	Params	Extra Data
29.00 (21)	0.83 (20)	0.07 (20)	118243.00 (5)	1.00 (1)

- Link to the codes/executables of the solution(s): github.com/Jinchen2028/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 01**, lead by Jinchen Zhu 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

General method description There is already a lot of work trying to reduce the complexity of SA, including [2, 3, 5]. Our goal is to reduce the complexity of transformer for SR and maintain its performance,so we propose Local-Global Term Transformer (LGTT) for SR , not every group needs self attention. The overall architecture is shown in the Figure 1. In order to establish long-term dependencies efficiently, a pair of SAs is reserved in each group as a Global-term Modeler and using the striped window mechanism [8], avoiding redundant operations. we improve BWSA by adding multi-groups and multi-heads mechanisms to improve computational efficiency,i.e., multi-head group WSA (GMSA), as shown in the Figure 2. In the rest of the blocks we propose efficient pixel mixer (PM) modules to without computational cost and can efficiently model short distance dependencies as a Local-term modeler. which will be mentioned later about the PM module.

Pixel Mixer We propose the pixel mixer module without parameters and computational complexity to create short-distance dependencies in each input Token. The PM module first divides the feature channels into five groups equally and moves the edge feature points to the opposite side in the order of left, right, top, and bottom for the first four groups. By inputting an intermediate feature of H $\times$ W $\times$ C, the PM module enables each input window in SA to obtain different local information according to other channels by linking

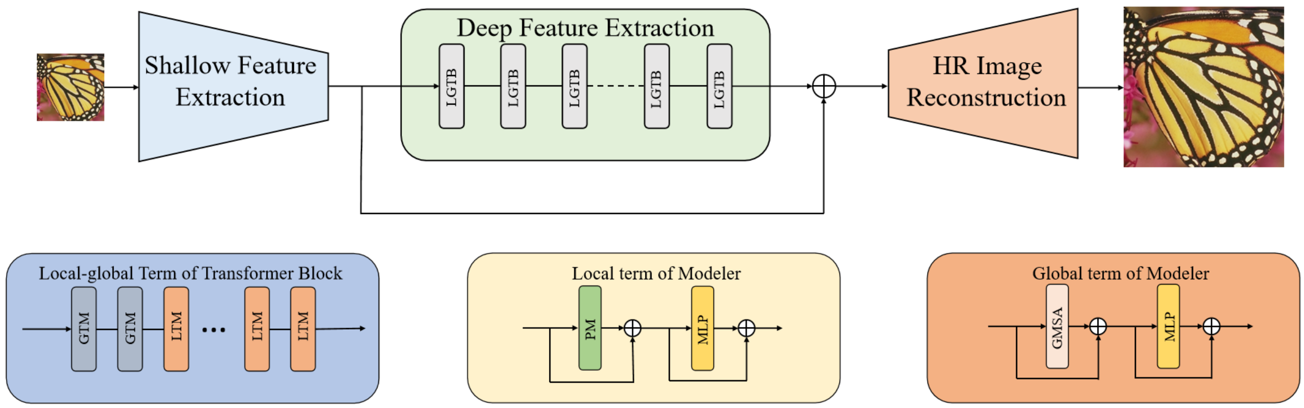

Figure 1. Illustration of the LGTT.

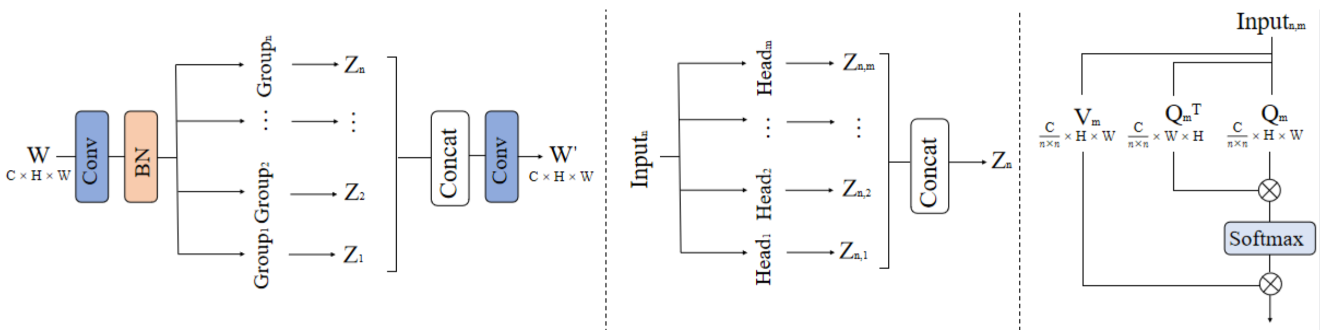

Figure 2. Illustration of our multi-groups multi-heads SA mechanism, updated on the basis of [8]. The left figure shows the multi-groups mechanism, which divides the groups according to the number of different windows of the input, as a way to increase the speed. The middle figure shows the multi-heads mechanism, where each group can have n heads in it, while computing self-attention. The right figure shows the calculation of SA on a head.

the edge feature points with the opposite edge feature points respectively, so that each feature point can be fully utilized and the perceptual field of the later module can be increased.

3. Training strategy

We use DF2K (DIV2K [1], Flickr2K [6])and LSDIR for datasets. and propose that the channel input is set to 30, the data augmentation method with 90°, 180°, 270° random rotation and horizontal flip is used for training, the batchsize is set to 128, and the input patch size of LR is 64×64.Trained using Adam optimizer [4] with $\beta_1 = 0.9$, $\beta_2 = 0.999$. The initialized learning rate is 5×10^{-4} and decays to 1×10^{-6} with the cosine learning rate. The model is optimized using the loss function of L_1 for a total of 1×10^6 iterations. Model training was performed using Pytorch [7] on two NVIDIA V100 32G GPUs.

4. Experimental results

We test our model on the DIV2K and LSDIR test sets, and the experiments are performed on a V100, using the official code. The results is shown in Table 1.

PSNR	SSIM	Params[K]	FLOPs[G]	Conv	Average Runtime[ms]
27.02 (21)	0.81 (17)	115 (3)	7.38	58	0.08 (18)

Table 1. Result of DIV2K and LSDIR test sets

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] Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. Generating long sequences with sparse transformers. *arXiv preprint arXiv:1904.10509*, 2019. 1

[3] Krzysztof Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Davis, Afroz Mohiuddin, Lukasz Kaiser, et al. Rethinking attention with performers. *arXiv preprint arXiv:2009.14794*, 2020. 1

[4] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 2

[5] Nikita Kitaev, Łukasz Kaiser, and Anselm Levskaya. Reformer: The efficient transformer. *arXiv preprint arXiv:2001.04451*, 2020. 1

[6] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 2

[7] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 2

[8] Jinpeng Shi, Hui Li, Tianle Liu, Yulong Liu, Mingjian Zhang, Jinchen Zhu, Ling Zheng, and Shizhuang Weng. Image super-resolution using efficient striped window transformer. *arXiv preprint arXiv:2301.09869*, 2023. 1, 2