NTIRE 2023 Efficient SR Challenge Factsheet
GateFormer is What You Need for SR

Yulong Liu, Jinpeng Shi, Shizhuang Weng
Anhui university
Anhui,China
yl.liu88@outlook.com,jinpeeeng.s@gmail.com

1. Team details

- Team name: **FRL Team 04**
- Team leader name: **Yulong Liu**
- Team leader address, phone number, and email:
  - address: **Anhui University, Hefei, China**
  - phone number: **+86 183 5603 9028**
  - email: **yl.liu88@outlook.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: **ylliu**
  - development entries: **4**
  - validation entries: **1**
- Best scoring entries of the team during development/validation phase:

PSNR	SSIM	Runtime	Params	Extra Data
29.01 (13)	0.83 (12)	0.07 (22)	199922.00 (17)	1.00 (1)

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

2. Method details

2.1. Network Architecture

Our team is committed to proposing a structurally simple and unified transformer to solve the single image super-resolution problems.The overall framework of the proposed GateFormer method is shown in the Figure 1. A 3×3 convolutional layer is firstly used to extract shallow features from low-resolution images. Then multiple gate block modules are stacked to perform deep feature extraction on the shallow features. Finally the SR images are generated by upscaling module. Global skip connection is included to ensure better retention of useful information.

2.2. Gate Block

Inspired by Gated Linear Units [2], we use two alternating gate modules to achieve this unified structure. One Basic Gate (BG) is used to replace the multi-layer perceptron (MLP) [7] in the transformer. We also found that large convolution kernels can better capture global information to some extent. Therefore we upgraded the Basic Gate with a large convolution kernel as the Context Gate(CG) to replace the self-attention (SA) [3, 5] in the transformer.

3. Training strategy

We use DF2K (DIV2K [1], Flickr2K [6]) and LSDIR for datasets, and propose that the channel input is set to 37, 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 β_1 = 0.9, β_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.

4. Experimental results

We test our model on the DIV2K and LSDIR test sets, and the experiments are performed on a V100, using the

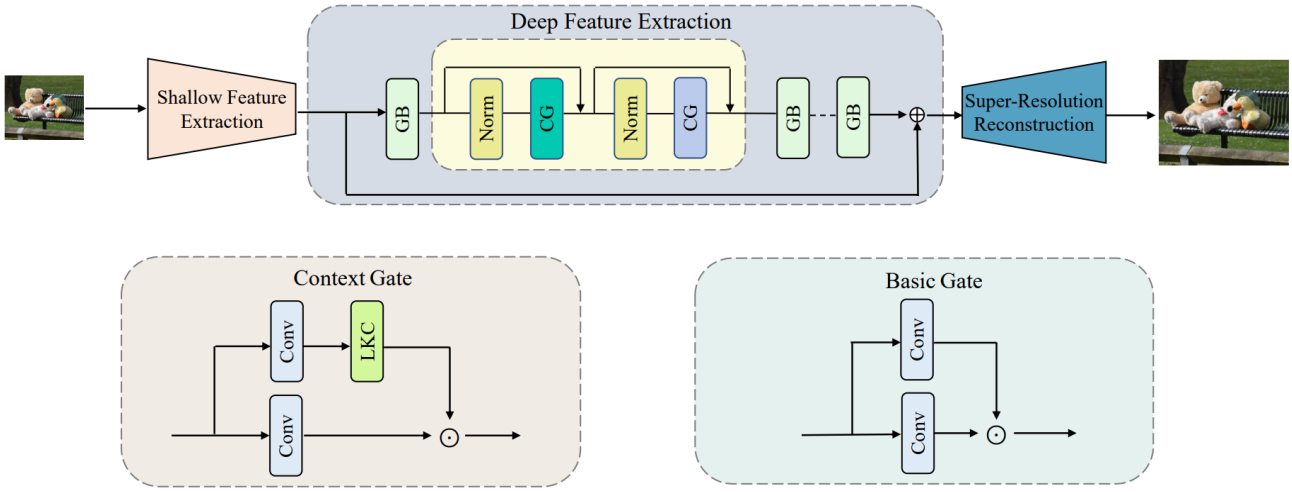

Figure 1. Illustration of the GateFormer.

official code. The results is shown in Table 1.

PSNR	SSIM	Params[K]	FLOPs[G]	Conv	Average Runtime[ms]
27.03 (19)	0.81 (12)	199 (10)	12.75	100	0.05 (16)

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. 1

[2] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 17–33. Springer, 2022. 1

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

[4] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[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. 1

[7] Hugo Touvron, Piotr Bojanowski, Mathilde Caron, Matthieu Cord, Alaaeldin El-Nouby, Edouard Grave, Gautier Izacard, Armand Joulin, Gabriel Synnaeve, Jakob Verbeek, et al. Resmlp: Feedforward networks for image classification with data-efficient training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 1