NTIRE 2024 Efficient SR Challenge Factsheet - Intermittent Feature Aggregation with Distillation for Efficient Super-Resolution-

ECNU_MViC
East China Normal University
Shanghai, China
10205501423@stu.ecnu.edu.cn

1. Team details

- Team name
  ECNU_MViC
- Team leader name
  Bohan Jia[1]
- Team leader address, phone number, and email
  [1]East China Normal University, China
  (+86)18862988901
  10205501423@stu.ecnu.edu.cn
- Rest of the team members
  Jincheng Liao[1], Junbo Qiao[1,2], Yunshuai Zhou[1], Yun Zhang[2,3], Wei Li[2], Shaohui Lin[1]
- Affiliations
  [1]East China Normal University, China
  [2]Huawei Noah's Ark Lab, China
  [3]The Hong Kong University of Science and Technology, China
- Affiliation of the team and/or team members with NTIRE 2024 sponsors (check the workshop website)
  The affiliation of the team is not on the list of sponsors on the website.
- User names and entries on the NTIRE 2024 Codalab competitions (development/validation and testing phases)
  All user names and entries of us are shown in Tab 1.
- Best scoring entries of the team during development/validation phase
  The best scoring entries of are shown in Tab 2.
- Link to the codes/executables of the solution(s) following https://github.com/BhJia/NTIRE2024_ESR

Table 1. User names and entries of our team

User name	entries	
	development	testing
BHJia	2	3
liaojc	0	3
super_zys	0	1

Table 2. Best scoring entries of our team

development		testing	
PSNR	SSIM	PSNR	SSIM
26.96	0.80	27.00	0.81

2. Method details

- **General method description**
  As shown in Fig. 1, we propose an intermittent feature aggregation network named IFADNet. The IFADNet chitecture comprises three parts: the shallow feature extraction, the deep feature extraction based on alternating BFEB blocks and RFMB blocks, and the reconstruction stage.
  We employ a single 3× 3 convolution to extract the shallow feature $F_s \in \mathbb{R}^{C\times H\times W}$ in the first stage H_F:

$$F_s = H_F(I_i), \tag{1}$$

  where I_i, C, H, W are the input image, the embedding channel dimension, height and width of the input, respectively.

  During the second stage, six intermittent blocks are used to extract the deep feature $F_d \in \mathbb{R}^{C\times H\times W}$:

$$F_d = H_D(F_s). \tag{2}$$

  Specifically, H_D consists of an alternating blueprint feature extraction block (BFEB) and a reparamaterized

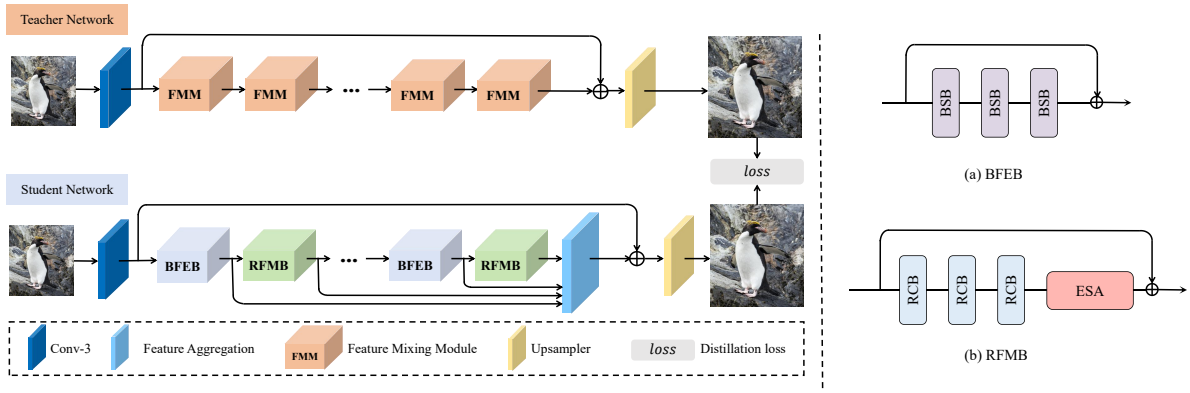

Figure 1. The pipeline of IFADNet. (a) Structure of RCB. (b) Structure of RFMB.

feature modulation block (RFMB). The detail of BFEB and RFMB will be introduced in the next paragraph. By using F_s and F_d as inputs, the high-quality image I_r is generated in the reconstruction stage denoted by H_R as:

$$I_r = H_R(F_s + F_d), \tag{3}$$

where H_R involves a single 3×3 convolution followed by a pixel shuffle operation [6].

Inspired by the blueprint shallow residual block [5], we design blueprint feature extraction block to reduce the computation time, which did not significantly reduce model performance. As shown in Fig. 1 (a), the BFEB contains three blueprint shallow blocks (BSB) which contain a 1×1 point-wise convolution with a 3 × 3 depth-wise convolution followed by GELU activation. The reparamaterized feature modulation block (RFMB) consists of three reparameterized convolution blocks (CB) and an enhanced spatial attention block [3] are employed to extract and modulate the deep feature fully. The detailed structure is illustrated in Fig. 2. It is noticed that BFEB and RFMB are intermittent with the attention mechanism only used in RFMB. We find that the intermittent setting of blocks significantly reduces model complexity while not largely impairing model effectiveness. At the end of the second stage, the extracted features of each block are concatenated and aggregated using two convolutions.

- **Reparameterization**
  Reparameterization has shown a strong ability to improve the feature presentation. Different from the reparameterization module design of high level tasks, We design isotropic edge-oriented convolutional block in our model. As shwon in Fig. 2(a), the Sobel-Dx and Sobel-Dy are empoly isotropic Sobel function to enhance the network's representation capabilities. For inference, the output is computed in a simplified 3 × 3 convolution, which significantly reduces computation cost.

- **Training Strategy**
  We use DIV2K [1] and the first 10K images of LSDIR [4] to train our model. The training dataset is augmented with horizontal flips and 90-degree rotations. Knowledge distillation is applied to improve the model performance. We use the large version of pre-trained SAFMN [7] as our teacher model.

  Our student model IFADNet has 6 blocks(3 BFEB and 3 RFMB). The channel of our network is 36. The training details are as follows:
  1. Training from scratch. The HR patch size is 256. The mini-batch size is set to 64. The model is trained by minimizing L1 loss and distillation loss(also L1 loss) with Adam optimizer [2]. The initial learning rate is set to 2×10^{-3} and halved at $\{100k, 500k, 800k, 900k, 950k\}$-iteration. The total number of iterations is 1000k.

  2. Finetuning with larger patches. The HR patch size is set to 640. The model is finetuned with MSE loss. Other settings are the same as in the previous step.

- **Experimental Results**
  We compare our IFADNet with RLFN on the same machine that is equipped with RTX 3090 GPU. As shown in Tab 3, our method achieves a PSNR result of 26.90 dB on the validation phase data and 27.00 dB on the testing phase data. The parameters, FLOPs, and GPU memory consumption of our model are much lower than the baseline RLFN, which shows promising performance on low-computing devices.

3. Other details

- We may submit a solution(s) description paper at NTIRE 2024 workshop.

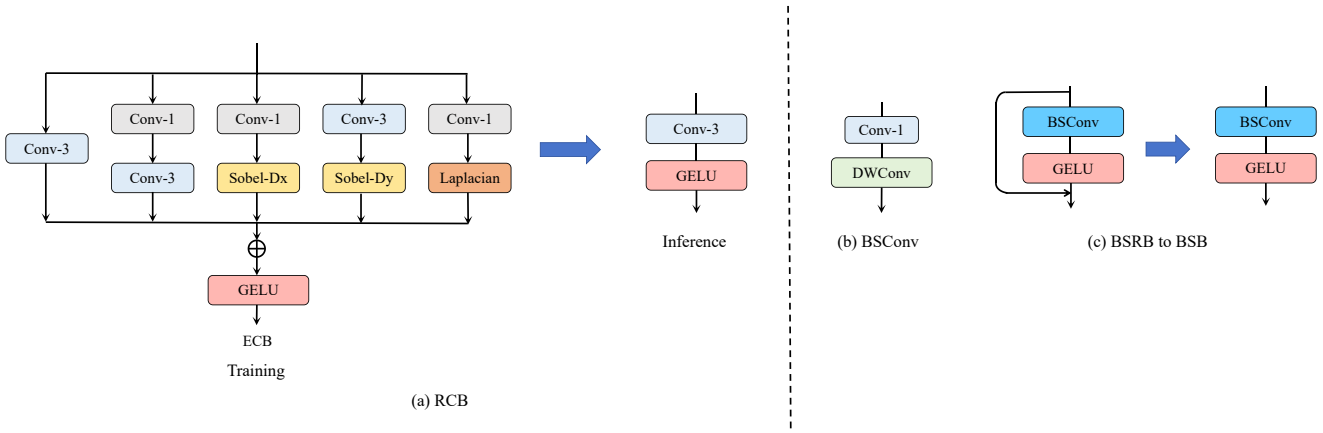

Figure 2. (a) Structure of BFEB. (b) Structure of BSConv. (c) From BSRB to BSB.

Method	Time[ms]			PSNR[dB]		#Params	Flops	GPU Mem
	Ave.	Val.	Test	Val.	Test	[M]	[G]	[M]
RLFN [3]	13.56	13.69	13.44	26.96	27.07	0.317	19.67	774.28
IFADNet	13.36	13.53	13.19	26.90	27.00	0.163	9.78	508.40

Table 3. The comlexity of IFADNet that test on an RTX 3090 GPU

- General comments and impressions of the NTIRE 2024 challenge.
  Challenge and Innovation. NTIRE 2024 promises to push boundaries in image restoration and enhancement, fostering innovation and creativity among participants.
  Practical Applications, With a focus on real-world scenarios, NTIRE 2024 emphasizes the practical utility and impact of image processing advancements.

- What do you expect from a new challenge in image restoration, enhancement and manipulation?
  Promote the development of efficient super-resolution networks

- Other comments: encountered difficulties, fairness of the challenge, proposed subcategories, proposed evaluation method(s), etc. There are difficulties in measuring inference time. It is hard to measure an accurate average inference time.

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] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. 2

[3] Fangyuan Kong, Mingxi Li, Songwei Liu, Ding Liu, Jingwen He, Yang Bai, Fangmin Chen, and Lean Fu. Residual local feature network for efficient super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 766–776, 2022. 2, 3

[4] 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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1775–1787, 2023. 2

[5] Zheyuan Li, Yingqi Liu, Xiangyu Chen, Haoming Cai, Jinjin Gu, Yu Qiao, and Chao Dong. Blueprint separable residual network for efficient image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 833–843, June 2022. 2

[6] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 2

[7] Long Sun, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Spatially-adaptive feature modulation for efficient image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023. 2