NTIRE 2024 Efficient SR Challenge Factsheet -Attention Guidance Distillation Network for Efficient Image Super-Resolution-

Shuli Cheng, Hongyuan Wang, Ziyan Wei, and Qingting Tang
Xinjiang University
Ürümqi, China
cslxju@xju.edu.cn, {why5200, 107552204088, tqt}@stu.xju.edu.cn

1. Team details

- Team name
  XJU_100th Ann

- Team leader name
  Shuli Cheng

- Team leader address, phone number, and email
  Team leader address: School of Computer Science and Technology, Xinjiang University, Ürümqi, China
  Phone number: +86 18290857626
  E-mail: cslxju@xju.edu.cn

- Rest of the team members
  Hongyuan Wang, Ziyan Wei, and Qingting Tang

- Affiliation
  School of Computer Science and Technology, Xinjiang University, Ürümqi, China

- Affiliation of the team and/or team members with NTIRE 2024 sponsors (check the workshop website)
  No affiliation

- User names and entries on the NTIRE 2024 Codalab competitions (development/validation and testing phases)
  User name: wanghongyuan
  The entry records are as follows.

26	wanghongyuan	1	03/05/24	26.90 (16)	0.79 (12)
14	wanghongyuan	3	03/19/24	27.02 (12)	0.81 (13)

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

	PSNR	Extra Data
val	26.90	1.00

- Link to the codes/executables of the solution(s)
  https://github.com/daydreamer2024/NTIRE2024-ESR-XJU_100th-Ann

2. Method details

General method description.

We propose a more lightweight attention guidance distillation network (AGDN) for efficient image super-resolution, which is influenced by existing studies such as IMDN [2], RFDN [5], BSRN [4], and MDRN [7], and further improved based on these studies. Fig.1 illustrates the overall architecture of our network, including shallow feature extraction, deep feature extraction, multi-layer feature fusion, and reconstruction. The effectiveness of this overall architecture has been extensively validated in previous studies.

It is found that the deep feature extraction phase is still a key limiting factor for model performance, and this phase consists of a series of feature distillation blocks. Each block can be divided into two main parts: distillation in the pre-phase and enhancement in the post-phase. In the distillation part, it has been a developmental process from channel splitting operations to feature distillation connections to blueprint separable convolutions instead of traditional convolutions; in the enhancement part, it has evolved from channel attention to spatial attention to spatial channel attention fusion. Based on the above analysis, we have rethought the feature distillation block and primarily improved the following aspects.

To enhance the features, we propose the new attention guidance distillation block (AGDB) with more efficient spatial attention, channel attention and self-attention as the base block of AGDN. As shown in Fig.2, we use the multi-level variance-aware spatial attention (MVSA) and reallocated contrast-aware channel attention (RCCA) as alternatives to the enhanced spatial attention (ESA) [6] and contrast-aware channel attention (CCA) [2], and introduce sparse global self-attention (SGSA) [3] to achieve further feature enhancement.

In MVSA, we consider the impact of multi-level branching and local variance on performance. Multi-level branches with small windows cannot cover a sufficient

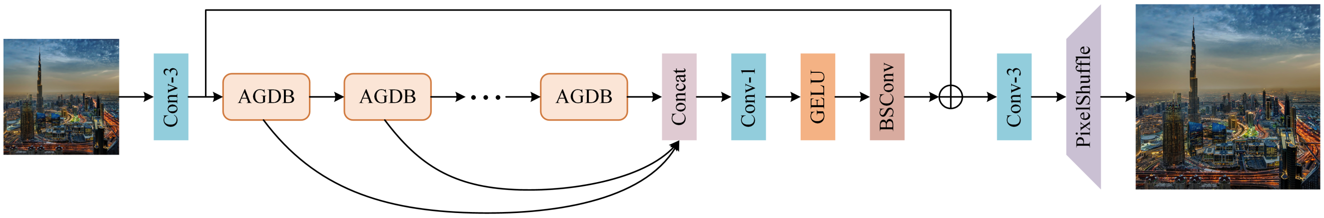
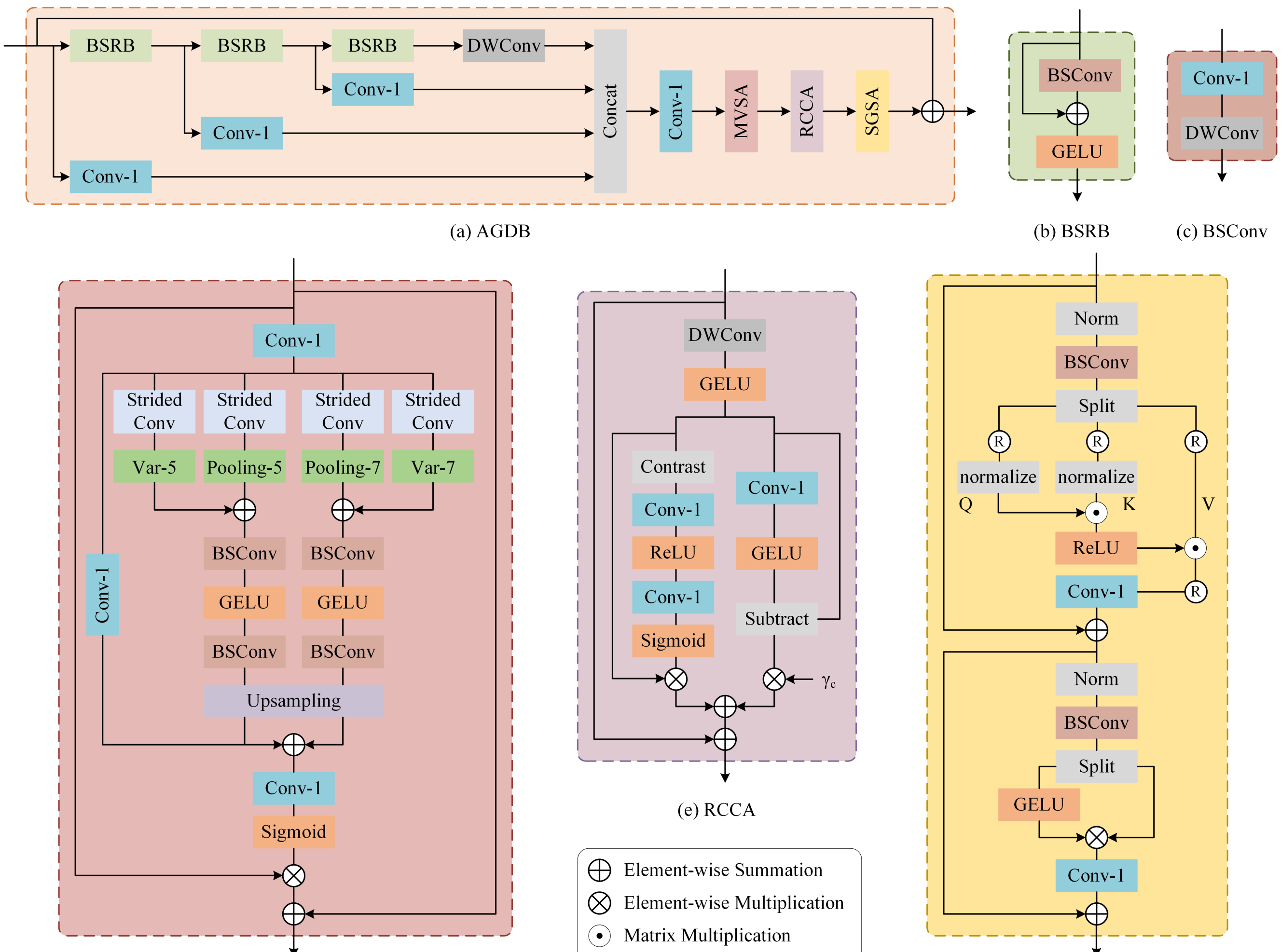


Figure 1. The overall architecture of attention guidance distillation network (AGDN).



(a) AGDB (b) BSRB (c) BSConv

(d) MVSA (e) RCCA (f) SGSA

Figure 2. The details of each component. (a) AGDB: Attention Guidance Distillation Block; (b) BSRB: Blueprint Shallow Residual Block; (c) BSConv: Blueprint Separable Convolution; (d) MVSA: Multi-level Variance-aware Spatial Attention; (e) RCCA: Reallocated Contrast-aware Channel Attention; (f) SGSA: Sparse Global Self-attention.

range of information, while using local variance in a single branch can lead to large differences in weights between branches. Therefore, we designed D5 and D7 branches that contain both local variance to better capture structurally information-rich regions while balancing performance and model complexity. In RCCA, we not only consider the reallocation of weights across channels by traditional channel attention but also enhance the treatment of common information across all channels. We added complementary branches with 1×1 convolution and GELU activation representations to reallocate complementary channel information, promoting the uniqueness of each channel.

Finally, we introduce SGSA to select the most useful similarity values, aiming to better utilize the most useful global features for image reconstruction. Here, we follow the original computational approach of SGSA [3], where global attention in image restoration usually has a gap between the training and testing stages. Therefore, we use the test-time localizer converter (TLC) [1] approach during the testing phase.

Training strategy.

The proposed AGDN has 4 AGDBs, in which the number of feature channels is set to 24. The details of the training steps are as follows:

1. Pretraining on the DIV2K and Flickr2K datasets. HR patches of size 256 $\times$ 256 are randomly cropped from HR images, and the mini-batch size is set to 64. 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, 500k, 800k, 900k, 950k}-iteration. The total number of iterations is 1000k.

2. Finetuning on 800 images of DIV2K and the first 10k images of LSDIR. HR patch size and mini-batch size are set to 384 $\times$ 384 and 32, respectively. The model is fine-tuned by minimizing the L2 loss function. The initial learning rate is set to 5×10^{-4} and halved at 50k iteration. The total number of iterations is 100k.

We conducted all experiments within the PyTorch framework, utilizing an NVIDIA RTX 4090 GPU.

Experimental results.

	PSNR	Params(M)	FLOPs(G)	Acts(M)	Mem(M)
RLFN	26.96	0.317	19.67	80.05	467.70
AGDN	26.90	0.069	4.39	190.92	612.01

We compared the differences between the RLFN and the AGDN on the DIV2K_LSDIR_valid dataset. As can be seen from the table, our model AGDN has a great advantage in the number of Params and FLOPs. During the testing phase, our proposed model scored 27.02.

3. Other details

- Planned submission of a solution(s) description paper at NTIRE 2024 workshop.

  We are planning to submit the solution description paper to NTIRE 2024 workshop.

- General comments and impressions of the NTIRE 2024 challenge.

  The NTIRE 2024 challenge was extremely exciting, and our team gained valuable knowledge that had a positive impact on us. This competition encouraged us to expand our applications beyond our usual research focus, leading to the development of new and innovative ideas. The outcomes of this challenge are expected to contribute significantly to cutting-edge advancements in this field.

- What do you expect from a new challenge in image restoration, enhancement and manipulation?

  Overall, a new challenge in image restoration, enhancement, and manipulation should foster creativity, collaboration, and advancements that benefit both the research community and practical applications of image processing technology.

- Other comments: encountered difficulties, fairness of the challenge, proposed subcategories, proposed evaluation method(s), etc.

  A more detailed explanation of the challenge requirements would be appreciated if possible, as it is essential to ensure that the submitted work meets the challenge's criteria.

References

[1] Xiaojie Chu, Liangyu Chen, Chengpeng Chen, and Xin Lu. Improving image restoration by revisiting global information aggregation. In *Proceedings of European Conference on Computer Vision*, pages 53–71. Springer, 2022. 3

[2] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the Acm International Conference on Multimedia*, pages 2024–2032, 2019. 1

[3] Xiang Li, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Dlgsanet: lightweight dynamic local and global self-attention networks for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12792–12801, 2023. 1, 3

[4] 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 Workshops*, pages 833–843, 2022. 1

[5] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In *Proceedings of the European Conference on Computer Vision Workshops*, pages 41–55, 2020. 1

[6] Jie Liu, Wenjie Zhang, Yuting Tang, Jie Tang, and Gangshan Wu. Residual feature aggregation network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2359–2368, 2020. 1

[7] Yanyu Mao, Nihao Zhang, Qian Wang, Bendu Bai, Wanying Bai, Haonan Fang, Peng Liu, Mingyue Li, and Shengbo Yan. Multi-level dispersion residual network for efficient image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1660–1669, 2023. 1