NTIRE 2023 Efficient SR Challenge Factsheet
Multi-level Dispersion Residual Network for Efficient Image Super-Resolution

1. Factsheet Information

1.1. Team details

- Team name:
  TelunXupt

- Team leader name:
  Yanyu Mao

- Team leader address, phone number, and email
  Address:
  Xian University of Posts and Telecommunications, Xi'an, China
  Phone Number: +86 15137104348
  Email: bolt35982@gmail.com

- Rest of the team members
  Members:
  Yanyu Mao[1] (bolt35982@gmail.com), Nihao Zhang[1], Qian Wang[1,2]

- Affiliation
  [1]Xian University of Posts and Telecommunications, Xi'an, China
  [2]Crime Scene Investigation Unit, Department of Public Security, Shaanxi, China

- Affiliation of the team and/or team members with NTIRE 2023 sponsors (check the workshop website)
  None

- User names and entries on the NTIRE 2023 Codalab competitions (development/validation and testing phases)
  Development phase: (29.00dB)
  User Name: Bolt; **Entries:** 6
  User Name: zwyczhang; **Entries:** 2
  Testing phase: (27.09dB)
  User Name: Bolt; **Entries:** 1
  User Name: zwyczhang; **Entries:** 2

- Best scoring entries of the team during development/validation phase
  As shown in Fig. 1.

EfficientSR								
#	User	Entries	Date of Last Entry	PSNR ▲	SSIM ▲	Runtime per image [s] ▲	Parameters ▲	Extra Data [1] / No Extra Data [0] ▲
2	Bolt	6	03/18/23	29.00 (16)	0.83 (17)	0.09 (27)	95396.00 (4)	1.00 (1)
4	zwyczhang	2	03/18/23	29.00 (16)	0.83 (17)	0.09 (27)	95396.00 (4)	1.00 (1)

Figure 1. Best scoring entries during validation phase.

- Link to the codes/executables of the solution(s) https://github.com/bbbolt/MDRN

1.2. Method details

The TelunXupt team proposed a multi-level dispersion residual network (MDRN). As shown in Fig. 2, MDRN still uses the basic SR framework, inspired by IMDN [4], RFDN [7] and BSRN [6]. The difference is that MDRN introduces the improved attention mechanisms into the proposed basic block enhanced attention distillation block (EADB). Specifically, EADB mainly contributes to improving the original attention mechanism on the basis of the original lightweight distillation framework, including two dimensions of space and channel. EADB replaces original enhanced spatial attention (ESA) [8] and contrast-aware channel attention (CCA) [4] with two proposed more efficient attention mechanisms, multi-level dispersion spatial attention (MDSA) and enhanced contrast-aware channel attention (ECCA), respectively, as shown in Fig. 3. MDSA divides the attention calculation in the original ESA into two parts, namely, the dispersion branch and the refinement branch, and extends the single-level dispersion branch to multiple levels, as shown in Fig. 4. The branch-A is defined as dispersion branch that carries out the single-size spatial compression and attention weight dispersion process (i.e., strided convolution or pooling operation, and interpolation operation), and the branch-B, as refinement branch, uses one 1×1 convolution connection to map high-resolution features to the end. In order to better capture the area with rich structural information, MDSA further introduces the local variance calculation (L-var) into the dispersion branch. To avoid introducing too many operations, local variance is only introduced into the dispersion branch D7. Besides, MDSA reduces the depth of Conv Groups in ESA to balance per-

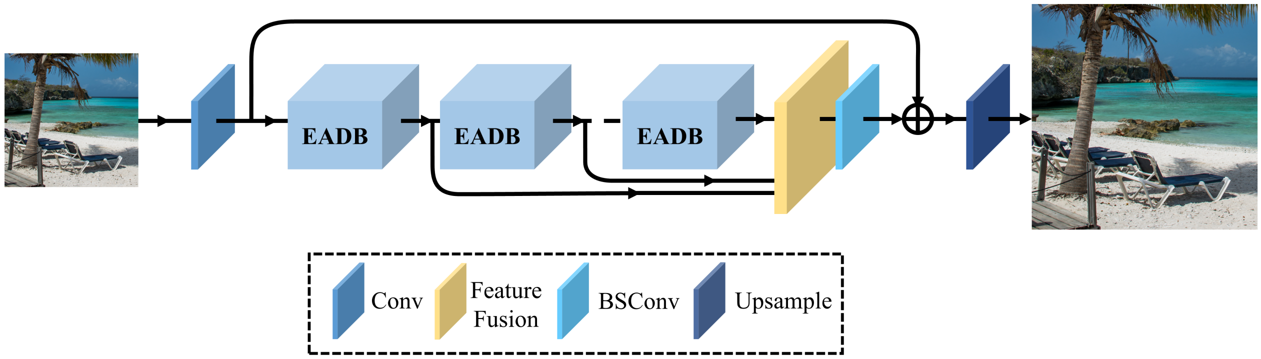
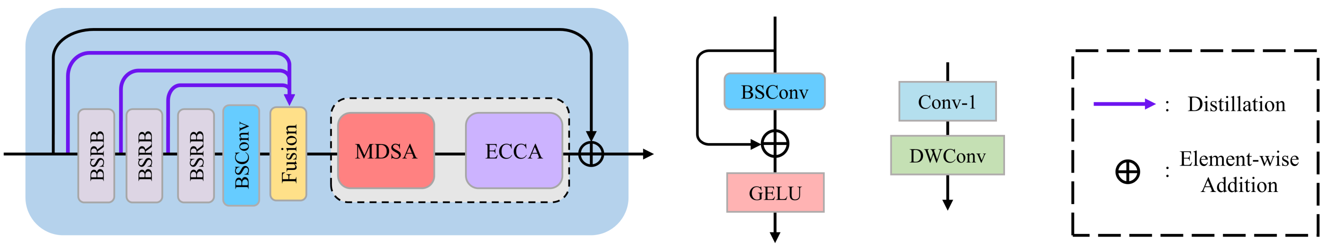
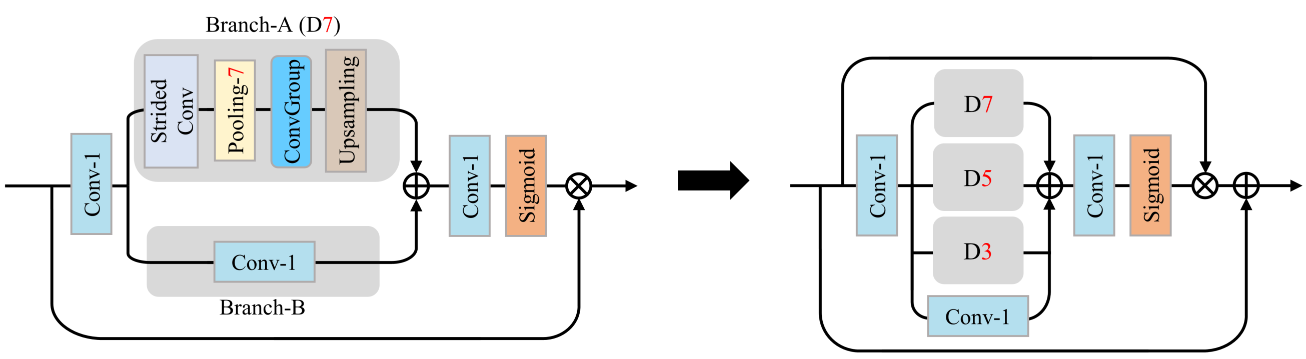


Figure 2. The whole framework of Multi-level Dispersion Residual Network (MDRN)



(a) EADB (b) BSRB (c) BSConv

Figure 3. Enhanced attention distillation block (EADB).



(a) ESA (b) MDSA wo L-var

Figure 4. Enhanced spatial attention (ESA) and multi-level dispersion spatial attention (MDSA) without local variance (L-var).

formance and model complexity. Inspired by RCAN [13], ECCA combines blueprint shallow residual block (BSRB) in BSRN with CCA to form a residual structure as shown in Fig. 5. Then, ECCA remove the residual connection in BSRB, and insert the CCA module between the point-wise and depth-wise convolution layers.

1.3. Training strategy

The proposed MDRN has 8 EADBs, in which the number of feature channels is set to 28. The details of training steps are as follows:

1. Pretraining on DIV2K [1]. HR patches of size 384 × 384 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 × 384 and 64, respectively. The model

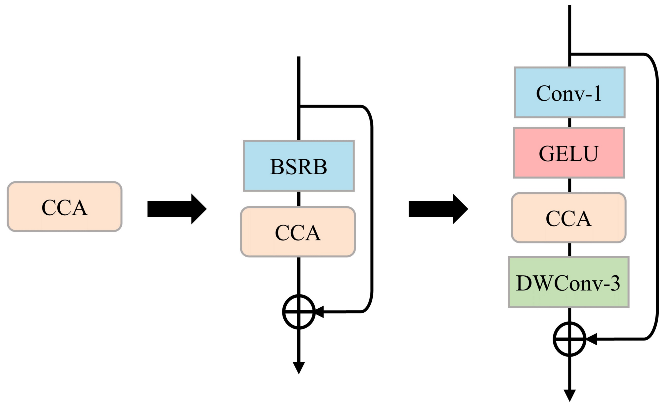


Figure 5. Enhanced contrast-aware channel attention (ECCA).

Table 1. Quantitative comparision with some teams that perform well on the model complexity track of NTIRE 2022 Efficient Super-Resolution [5]. The average peak signal to noise ratio (PSNR) and the structural similarity (SSIM) [12] on the RGB channels are exploited as the evaluation metrics.

#Team	#Params[M]	#FLOPs[G]	Set5 [2]	Set14 [11]	Urban100 [3]	BSDS100 [9]	Manga109 [10]
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
NEESR	0.272	16.86	29.95/0.8596	26.61/0.7378	24.24/0.7524	26.15/0.7090	27.97/0.8694
NJUST_ESR	0.176	8.73	30.14/0.8625	26.78/0.7427	24.50/0.7605	26.23/0.7119	28.50/0.8786
Xpixel	0.156	9.50	30.10/0.8616	26.78/0.7421	24.49/0.7600	26.20/0.7110	28.53/0.8766
TelunXupt (ours)	0.095	5.58	30.15/0.8617	26.78/0.7422	24.56/0.7623	26.22/0.7115	28.58/0.8770

is fine-tuned by minimizing Charbonnier loss function. The initial learning rate is set to 5×10^{-4} and halved at $\{100k, 500k, 800k, 900k, 950k\}$-iteration. The total number of iterations is 1000k.

3. Finetuning on 800 images of DIV2K and the first 10k images of LSDIR again. HR patch size and the mini-batch size are set to 480 × 480 and 64, respectively. The model is fine-tuned by minimizing L2 loss function. The initial learning rate is set to 2×10^{-4} and halved at $\{100k, 300k, 600k\}$-iteration. The total number of iterations is 650k.

1.4. Experimental results

In this section, we compare MDRN on five common benchmarks with the teams that have achieved outstanding results on the model complexity track of NTIRE 2022 Efficient Super-Resolution [5]. As shown in Tab. 1, MDRN achieves competitive performance with few parameters and computations. In addition, MDRN achieves 29.00 dB on the DIV2K validation set.

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] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 3

[3] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 3

[4] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th acm international conference on multimedia*, pages 2024–2032, 2019. 1

[5] Yawei Li, Kai Zhang, Radu Timofte, Luc Van Gool, Fangyuan Kong, Mingxi Li, Songwei Liu, Zongcai Du, Ding Liu, Chenhui Zhou, et al. Ntire 2022 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1062–1102, 2022. 3

[6] 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*, pages 833–843, 2022. 1

[7] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 41–55. Springer, 2020. 1

[8] 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

[9] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001. 3

[10] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76:21811–21838, 2017. 3

[11] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pages 711–730. Springer, 2012. 3

[12] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 3

[13] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 2