NTIRE 2024 Efficient SR Challenge - Cao Group - ID 29

Bohao Liao, Peizhe Xia, Long Peng, Zhibo Du, Xin Di, Yang Wang, Wei Zhai, Yang Cao, Zhengjun Zha
University of Science and Technology of China
liaobh@mail.ustc.edu.cn

Renjing Pei, Jiaming Guo, Songcen Xu
Huawei Noah's Ark Lab
{peirenjing, guojiaming5, xusongcen}@huawei.com

1. Factsheet Information

1.1. Team details

- Team name: Cao Group
- Team leader name: Bohao Liao
- Team leader address: Hefei, China phone number: +86 15014931419 email: liaobh@mail.ustc.edu.cn
- Rest of the team members: Peizhe Xia, Long Peng, Zhibo Du, Xin Di, Yang Wang, Wei Zhai, Renjing Pei, Jiaming Guo, Songcen Xu, Yang Cao, Zhengjun Zha
- Team website URL: No
- Affiliation: University of Science and Technology of China
- Affiliation of the team and/or team members with NTIRE 2024 sponsors (check the workshop website): not related
- User names and entries on the NTIRE 2024 Codalab competitions (development/validation and testing phases): zero_zoe
- Best scoring entries of the team during development/validation phase: Entry 3, shown in Table 1
- Link to the codes/executables of the solution(s): R2Net

1.2. Method details

1.2.1 Network Design

Our proposed network architecture, as illustrated in Figure 1, is termed the Reparameterization 2 Super-Resolution Network (R2Net). Inspired by previous works DIPNet [1] and SRN [2], we observed that DIPNet employs reparameterization techniques to enhance network performance, whereas SRN, for the sake of reducing training time, does not utilize reparameterization. During inference, SRN has fewer convolutional modules but a higher number of channels. Therefore, we aim to boost the network's expressive capability through reparameterization while simultaneously reducing the number of channels to fully exploit the potential of the SRN network.

Specifically, we build upon the SRN framework by combining Residual Blocks (RB) and Enhanced Spatial Attention (ESA) to form a new feature extraction module with reparameterization, named reNRB. We remove the residual connections within RB and the 1x1 convolutions on the residual connections in ESA, retaining only the global residual connections. Additionally, we replace all 3x3 convolutions in the network with RRRB [3]. Inspired by the Feed Forward Network (FFN) module structure in transformers, we independently propose a reparameterization module to enhance the performance of 1x1 convolutions by expanding the intermediate channel count, thereby harnessing the representation capability of complex structures during optimization. During inference, we utilize structural reparameterization to degrade it into a single 1x1 convolution as shown in Figure 2, thereby improving the network's representational capacity without incurring additional computational costs. Furthermore, drawing from DIPNet, we find that preserving the last convolution before the residual connection significantly boosts network performance. For the sake of lightweight design, we set its kernel size to 1x1. Notably, we eliminate biases in all convolutional layers, as this not only accelerates inference speed but also slightly enhances network performance. Finally, the feature learning part of our overall network consists of four reNRBs connected in series, employing 3x3 convolutions for feature mapping and reconstruction.

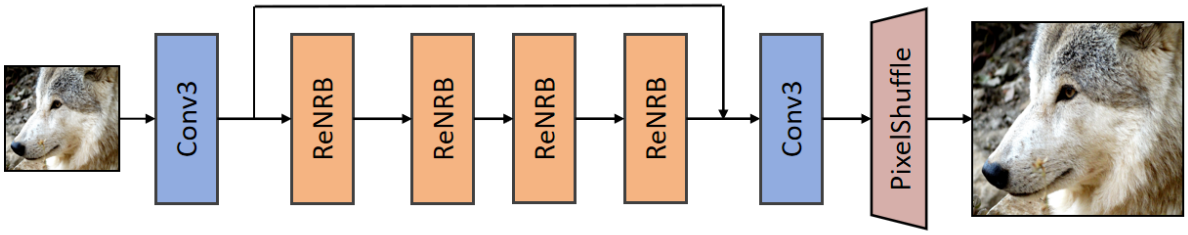

Figure 1. The arhchitecture of our proposed network

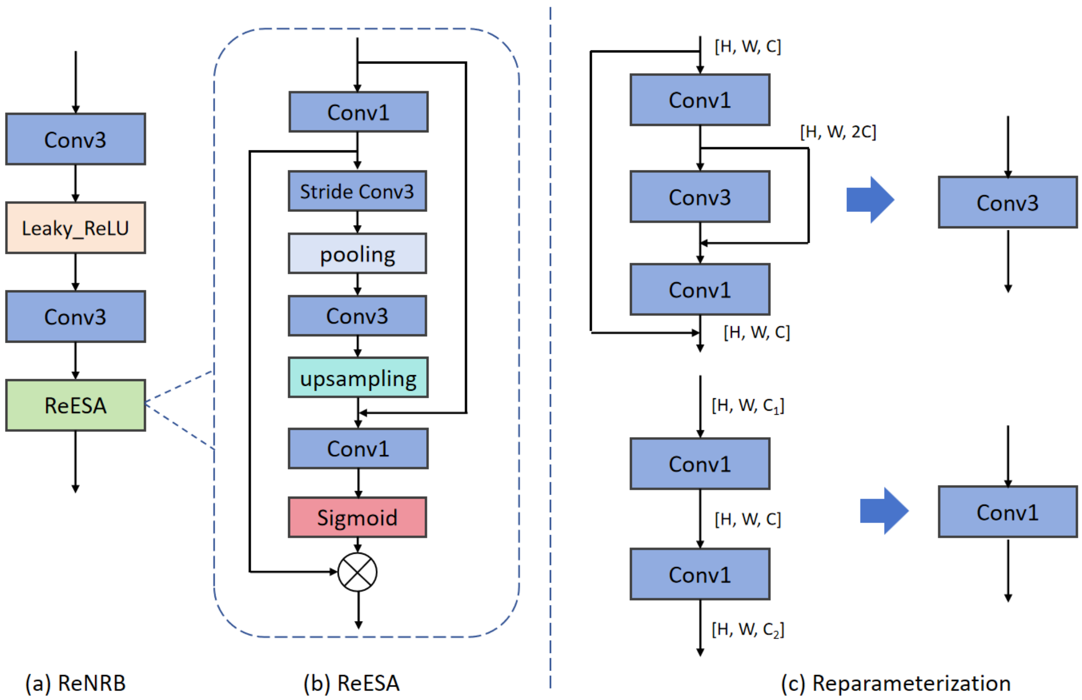

Figure 2. The arhchitecture of our network block and reparameterization

1.2.2 Experiments

Datasets. We only adopt DIV2k as training datasets.

Training Details. We train the network on RGB channels and enhance the training data through random flipping and rotation. The number of ESA channels is set to 16, while the number of feature channels for other convolutional layers is set to 48. The training process is divided into three stages.

1. In the first stage, we randomly crop 256x256 HR image patches from ground truth images, with a batch size of 64. We use the Adam optimizer, setting $\beta 1 = 0.9$ and $\beta 2 = 0.999$, and minimize the Charbonnier loss function. The initial learning rate is set to 2e-4, with a cosine learning rate decay strategy. The number of iterations in the first stage is set to 2e6.

2. In the second stage, we increase the size of the HR image patches to 512x512, with other settings remaining the same as in the first stage.

3. In the third stage, the batch size is set to 32, and the L2 loss is minimized over 2e6 iterations. The initial learning rate is set to 2e-5. Throughout the entire training process, we employ an Exponential Moving Average (EMA) strategy to enhance the robustness of training.

References

[1] Lei Yu, Xinpeng Li, Youwei Li, Ting Jiang, Qi Wu, Haoqiang Fan, and Shuaicheng Liu. Dipnet: Efficiency distillation and iterative pruning for image super-resolution. In *Proceedings*

Table 1. Performance comparison on DIV2K validation set for x4 upscaling , the average cost time is tested with an LR image input of size 300×400 while the others are tested with an LR image input of size 256×256 on an NVIDIA 3090 GPU

Net	flops	param	avg time	val PSNR	test PSNR
RLFN [4]	19.67	0.3172	8.004	26.96	27.07
R2Net	13.05	0.2148	4.223	26.9042	27.0023

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1692–1701, 2023. 1

[2] Yucong Wang and Minjie Cai. A single residual network with esa modules and distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1970–1980, 2023. 1

[3] Zongcai Du, Ding Liu, Jie Liu, Jie Tang, Gangshan Wu, and Lean Fu. Fast and memory-efficient network towards efficient image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 853–862, 2022. 1

[4] 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. 3