NTIRE 2024 Efficient SR Challenge Factsheet
- SlimRLFN: Making RLFN Smaller and Faster Again -

Shenglong Zhou
USTC
Hefei, China
slzhou96@mail.ustc.edu.cn

Binbin Chen
HUST
Wuhan, China
tinyshine@yeah.net

1. Team Details

- HiSR Group
- Team leader name: Shenglong Zhou[1]
- Addr: University of Science and Technology of China, Hefei, Anhui China,

  Tel: (+86) 13635695560,

  Email: slzhou96@mail.ustc.edu.cn
- Rest of the team members: Binbin Chen[2]
- [1]University of Science and Technology of China

  [2]Huazhong University of Science and Technology
- User name: SCong (#13 on the test phase)
- Best scoring: 26.914 (valid) / 27.026 (test).
- Code link: https://github.com/Blcony/SlimRLFN

2. Method details

Network Architecture. We propose the SlimRLFN for the efficient super-resolution task. The network architecture is inspired by the design of RLFN [3], while fully exploring the capacity of reparameterizable convolution, light distillation, and iterative model pruning. The whole architecture is shown in Fig.1, which mainly consists of six SRLFB modules and a pixel shuffle module. Reparameterizable convolutions are utilized in the SRLFB module, aiming to improve the super-resolution capability without introducing any additional parameter overhead during the inference stage. Meanwhile, the network is optimized by the pixel-wise loss such as charbonnier loss or L2 loss, along with the distillation loss provided by a light but efficient teacher model. Last but not least, we use iterative pruning to shrink the model size while maintaining the promising performance at the last training stage.

Reparameterizable Convolution Reparameterization skill [2] is widely used in computer vision tasks, and has been proven to be effective in improving the lightweight model's performance. As shown in Fig.1, each SRLFB has three 3×3 convolutions with 30 channels and an ESA [4] module. We first light the ESA architecture by removing the 1×1 residual convolution following SRN [5]. Then We replace the regular 3×3 convolutions including the one in the ESA module with reparameterizable convolutions. During the inference stage, the reparameterizable convolution can be converted to the regular convolution easily.

Light Distillation. Relying on the prominent performance of the teacher model, knowledge distillation [5, 6] can improve the super-resolution ability of the lightweight student model. Therefore, we train an RLFN model consisting of six blocks with 196 channels as our teacher model, and regard our SlimRLFN as the student model. It is worth mentioning that, different from using a huge model such as HAT [1], we choose the relatively lighter RLFN model as the teacher. Compared to the huge model, our teacher model not only offers faster training speed but also maintains reliable knowledge distillation capability. During the training stage, in addition to using the groundtruth HR image for charbonnier loss, we also constrain the difference between the output image of the teacher model and our SlimRLFN.

Iterative Pruning. Model pruning is effective in shrinking the model size by the pre-defined loss function. Inspired by previous pruning work [7], we use L2 pruning to shrink our SlimRLFN and fine-tune it to make promising results. The pruning method is utilized in an iterative manner, and we stop it until the model can not make the right super-resolution images.

Implementation Details. We choose DIV2K, Flickr2K and LSDIR datasets as our training datasets, and we augment them with horizontal/vertical flips and rotations during the training stage. We set the batch size as 64 for all training stages and other hyperparameters such as patch size or learning rate are determined by the specific training stage. We will summarize our whole training process as follows.

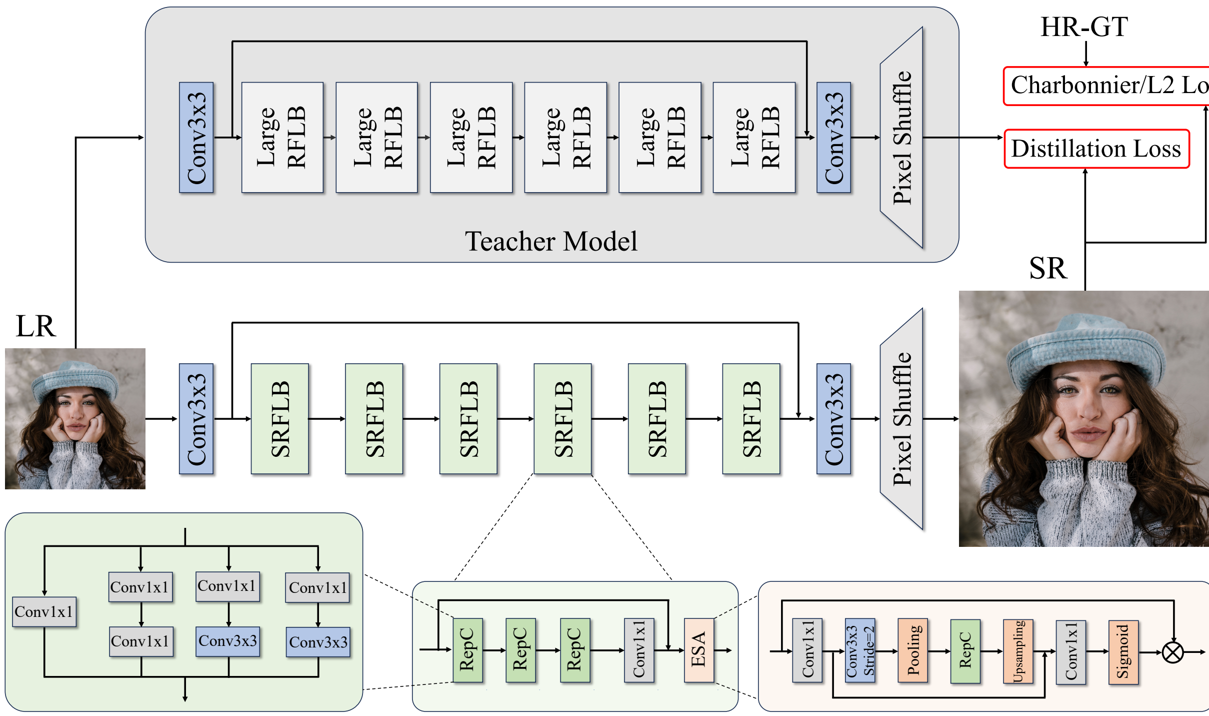

Figure 1. Network architecture of our SlimRLFN.

1) Training the teacher model. We choose the large RLFN as our teacher model. We set the patch size as 256×256, and use charbonnier loss and Adam optimizer for optimization. We train the teacher model for 100 epochs, and set the initial learning rate as 1e-3. The learning rate decay follows cosine annealing with T_{max} as 100 and eta_{min} as 1e-5.

2) Training the SlimRLFN with light distillation. We set the patch size as 256×256, and use two loss functions for training. The first one is the regular charbonnier loss with groundtruth HR image, and the second one is the charbonnier loss between SlimRLFN's output and teacher's output. We also choose the Adam optimizer for optimization. We train the SlimRLFN model for 100 epochs, and set the initial learning rate as 1e-3. The learning rate decay follows cosine annealing with T_{max} as 100 and eta_{min} as 1e-5. Then we repeat this stage two more times without distillation loss, and the pretrained model is adopted from the last stage.

3) Training the SlimRLFN with a larger patch size progressively . We set the patch size as {384×384, 512×512}, and set the initial learning rate as {5e-4, 2.5e-4} respectively. Each stage's pretrained model is adopted from the last stage, and we train the model under the same patch size three times in total. The other details are the same as before.

4) Training the SlimRLFN with the patch size of 640×640. We use the L2 loss in this stage, and set the initial learning rate as 1e-4. We also adopt cosine annealing with T_{max} as 100 and eta_{min} as 1e-6 for learning rate decay.

5) [Optional] SlimRLFN pruning stage. After training from the above several epochs, we adopt the iterative pruning for the SlimRLFN which has obtained promising performance.

3. Other details

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

  Reply: Yes, we plan to submit our SlimRLFN to the NTIRE 2024 workshop with more details.

- General comments and impressions of the NTIRE 2024 challenge.

  Reply: NTIRE 2024 challenge gives an excellent starting point for freshmen in the image restoration field just like us, and we appreciate organizers for holding such wonderful challenges. We have learnt a lot from this challenge, and we hope we can obtain a good ranking result and look forward to this challenge next year!

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

  Reply: Efficient super-resolution challenge is always a necessary and competitive challenge, we would like to continue participating next year.

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

  Reply: The method of calculating inference time may not be the optimal solution.

References

[1] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023. 1

[2] Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13733–13742, 2021. 1

[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. 1

[4] 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

[5] 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

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

[7] Michael Zhu and Suyog Gupta. To prune, or not to prune: exploring the efficacy of pruning for model compression. *arXiv preprint arXiv:1710.01878*, 2017. 1