NTIRE 2025 Efficient SR Challenge Factsheet
-Team Rochester -

Pinxin Liu, Yongsheng Yu, Hang Hua, Yunlong Tang
University of Rochester

1. Introduction

This factsheet mainly contains the contribution and implementaion details of Team Rochester in NTIRE 2025 Efficient Super Resolution Competition.

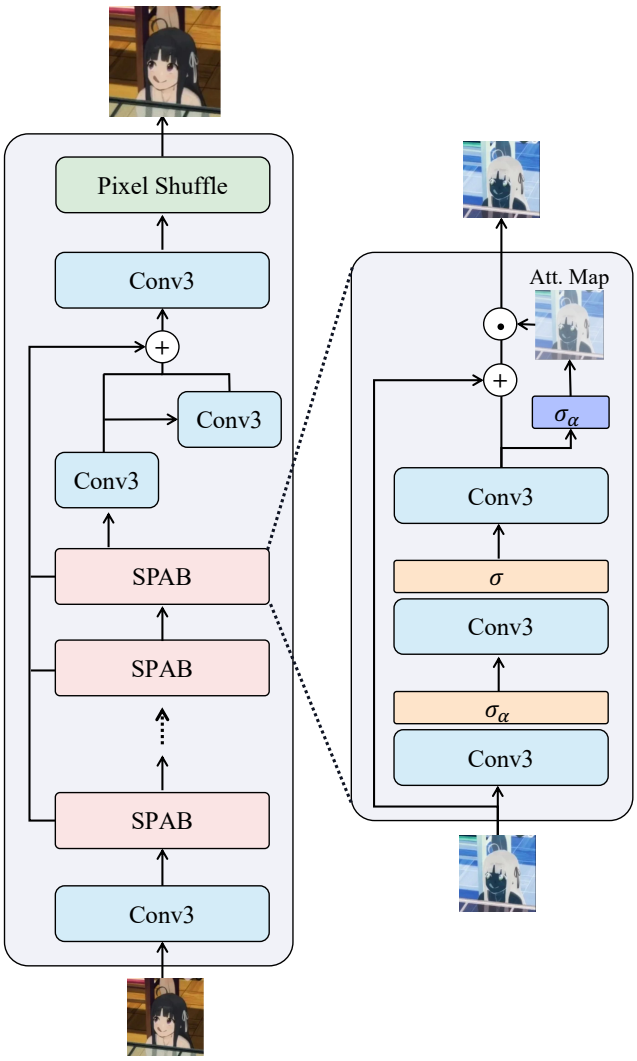

Figure 1. **Network Structure**. We reduce the channel dimension from 48 to 26 from the original design and introduce addition convolution to stablizae the attention feature maps from SPAN blocks.

2. Method details

Our proposed method, **ESRNet**, is an improved and more efficient variant of last year's XiaomiMM SPAN network [1]. The original SPAN network demonstrated strong generation quality but required complex training tricks and model fusion strategies, making it difficult to reproduce and computationally expensive. In contrast, ESRNet achieves similar performance with significantly reduced computational overhead, enhanced training stability, and improved inference speed.

Model Architecture A key aspect of ESRNet's design is its ability to maintain high performance while reducing computational costs. As shown in Fig. 1, our modifications include:

- Retaining the first six SPAN attention blocks as core feature extraction components while introducing a lightweight convolutional layer to refine the extracted feature maps before fusing them with the original features. This modification enhances feature representation while stabilizing the training process.
- Reducing the number of feature channels from 48 to 26, leading to a substantial decrease in both **model parameters and floating-point operations (FLOPs). This reduction not only lowers GPU memory consumption but also improves inference efficiency without degrading performance.
- Improved validation speed, as ESRNet requires fewer computations per forward pass, making it more suitable for real-time applications compared with the baseline method.

Overall, ESRNet has approximately half the number of parameters and FLOPs compared to the baseline EFPN network, yet it maintains a high PSNR score, demonstrating that our modifications achieve an excellent trade-off between efficiency and performance.

Training Methodology We train ESRNet on RGB image patches of size 256×256, applying standard augmentation techniques such as random flipping and rotation to enhance generalization. To ensure stable convergence and optimal performance, we adopt a three-stage training strategy:

1. **Initial Feature Learning:** We train the model with a batch size of **64** using Charbonnier loss, a robust loss function that mitigates the effects of outliers. The Adam optimizer is used with an initial learning rate of 2×10^{-4}, which follows a cosine decay schedule.

2. **Refinement Stage:** We progressively decrease the learning rate linearly from 2×10^{-4} to 2×10^{-5}, allowing the model to refine its learned features while maintaining stable gradients.

3. **Fine-Tuning with L2 Loss:** In the final stage, we adopt L2 loss to fine-tune the model, further enhancing detail restoration. The learning rate is further reduced from 2×10^{-5} to 1×10^{-6} for smooth convergence.

By structuring the training into these stages, we eliminate the need for complex training tricks used in previous approaches while achieving more stable and reliable optimization.

One of the most significant advantages of ESRNet is its improved validation time due to its optimized architecture. Compared to the original SPAN network, ESRNet achieves a similar PSNR score while reducing computational complexity. The model requires significantly fewer FLOPs and parameters, leading to a noticeable reduction in inference time and GPU memory usage**. This makes ESRNet a practical solution for applications requiring both high-quality generation and efficient computation.

References

[1] Cheng Wan, Hongyuan Yu, Zhiqi Li, Yihang Chen, Yajun Zou, Yuqing Liu, Xuanwu Yin, and Kunlong Zuo. Swift parameter-free attention network for efficient super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6246–6256, 2024. 1