WaveMixSR-V2: Enhancing Super-resolution with Higher Efficiency (Student Abstract)

Pranav Jeevan*, Neeraj Nixon*, Amit Sethi
Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai 400076, India
pjeevan@iitb.ac.in, 20d070056@iitb.ac.in, asethi@iitb.ac.in

Abstract

Recent advancements in single image super-resolution have been predominantly driven by token-mixers and transformer architectures. WaveMixSR utilized the WaveMix architecture, employing a two-dimensional discrete wavelet transform for spatial token mixing, achieving superior performance in super-resolution tasks with remarkable resource efficiency. In this work, we present an enhanced version of the WaveMixSR architecture by (1) replacing the traditional transpose convolution layer with a PixelShuffle operation and (2) implementing a multi-stage design for higher resolution tasks ($4\times$). Our experiments demonstrate that our enhanced model – WaveMixSR-V2 – outperforms other architectures in multiple super-resolution tasks, achieving state-of-the-art for the BSD100 dataset, while also consuming fewer resources and exhibiting higher parameter efficiency and throughput.

Introduction

Single-image super-resolution (SISR) is a key task in image reconstruction, aiming to transform low-resolution (LR) images into high-resolution (HR) by predicting and restoring missing details. This process requires capturing both local information and global context. Recent advancements in super-resolution, particularly with attention-based transformers like SwinFIR (Zhang et al. 2023) and hybrid attention transformer (Chen et al. 2023), have surpassed traditional CNN approaches due to their ability to capture long-range dependencies. However, transformers face challenges with quadratic complexity in self-attention, leading to high resource demands and requiring large datasets. To overcome this, token-mixer models such as WaveMixSR (Jeevan et al. 2024), which uses a two-dimensional discrete wavelet transform, have shown potential for improved efficiency and even superior performance. Building on the strengths of WaveMixSR, we propose enhancements to the model by rethinking its upsampling strategy inside the WaveMix blocks and changing the single stage design.

*These authors contributed equally.

Architectural Improvements

Multi-stage Design

We have made significant improvements to the WaveMixSR model, focusing on two key aspects. First, we addressed how the model handles SR tasks higher than $2\times$. In the original WaveMixSR (Jeevan et al. 2024), all SR tasks were performed by directly resizing the LR image to HR using a single upsampling layer. This layer relied on non-parametric upsampling techniques, such as bilinear or bicubic interpolation, which limited the model's ability to fine-tune and optimize the SR process across different scales. Our approach involved transitioning from this single-stage design to a more robust multi-stage design. In our new architecture, we introduced a series of resolution-doubling $2\times$ SR blocks, which progressively doubles the resolution step by step. This multi-stage approach allows for better SR performance at higher scales while reducing resource consumption.

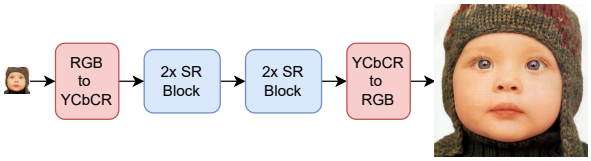

Figure 1: Architecture of WaveMixSR-V2 showing $4\times$ SR with two $2\times$ SR blocks in series.

For instance, in a $4\times$ super-resolution task, instead of directly upsampling the LR image to HR using a single interpolation layer, the model now proceeds through a series of two $2\times$ SR blocks as shown in Fig. 1 By incrementally increasing the resolution (doubling in each stage), the model is better able to refine the details at each step, leading to superior super-resolution performance compared to the single upsampling operation used in the original WaveMixSR as shown in Tab. 1 and Fig. 4.

PixelShuffle

We introduce a key modification to the WaveMixSR model by replacing the transposed convolution operation in the WaveMix blocks with a PixelShuffle (Shi et al. 2016) operation followed by a convolution layer (WaveMixSR-V2 block) as shown in Fig. 2. While the original WaveMixSR

used transposed convolutions, which involved numerous parameters and high computational cost, PixelShuffle upsamples the image more efficiently by rearranging pixels from feature maps. This significantly reduces the number of parameters, enhancing the model's efficiency. The subsequent convolution layer after PixelShuffle allows the model to continue learning and refining features effectively. Moreover, PixelShuffle avoids the checkerboard artifacts commonly introduced by transposed convolutions, producing smoother and more natural-looking images while maintaining high-quality super-resolution outputs as shown in Fig. 3.

Results

Incorporating these improvements in the architecture has enabled WaveMixSR-V2 to achieve new state-of-the-art (SOTA) performance on the BSD100 dataset (Martin et al. 2001). Notably, it accomplishes this with less than half the number of parameters, lesser computations and lower latency compared to WaveMixSR (previous SOTA) as shown in Tab. 1 and Tab. 2. Similar to WaveMixSR, WaveMixSR-V2 requires less training data than transformer-based models and outperforms even those pre-trained on the ImageNet-1k dataset.

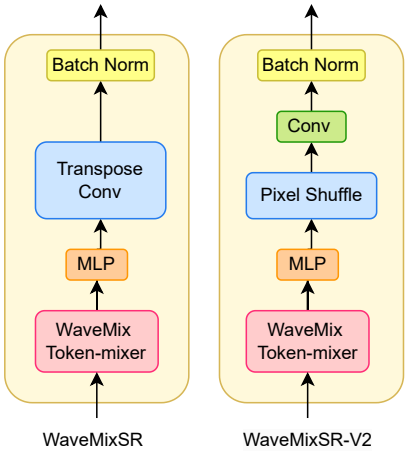

Figure 2: Simplified block diagram of WaveMix block in WaveMixSR (on the left) and WaveMixSR-V2 block (on the right).

Model	#Params.	#Multi-Adds.
SwinIR (Liang et al. 2021)	11.8 M	49.6 G
HAT (Chen et al. 2023)	20.8 M	103.7 G
WaveMixSR (Jeevan et al. 2024)	1.7 M	25.8 G
WaveMixSR-V2	**0.7 M**	**25.6 G**

Table 1: Model complexity comparison of WaveMixSR-V2 with other state-of-the-art methods such as WaveMixSR, SwinIR and HAT on 4× SR of 64×64 input patch.

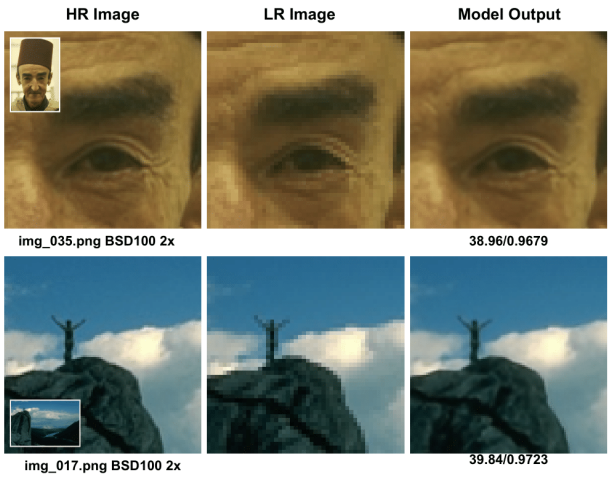

Figure 3: Visual results of 2× SR on BSD100 dataset. Each column from the left shows a patch from the HR image (shown as a small image near the corner), the same patch extracted from the LR image, and a patch taken from the model output respectively. The filename of the image is given below the HR image and the PSNR/SSIM of the model output is reported at below the model output. The values displayed are computed for the whole image and not just the patch.

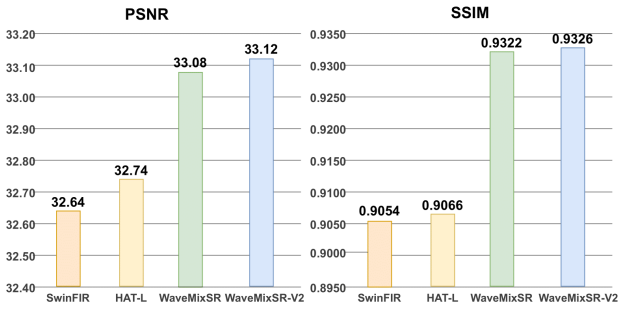

Figure 4: Comparison of PSNR and SSIM for 2× SR on BSD100 dataset shows WaveMixSR-V2 surpasses the previous state-of-the-art WaveMixSR and other methods such as HAT and SwinFIR.

Model	Training Latency ↓ (ms)	Training Throughput↑ (fps)	Inference Latency ↓ (ms)	Inference Throughput ↑ (fps)
WaveMixSR	22.8	43.8	18.6	53.7
WaveMixSR-V2	19.6	50.8	12.1	82.6

Table 2: Comparison of latency and throughput of WaveMixSR-V2 and WaveMixSR shows that WaveMixSR-V2 is significantly faster than WaveMixSR

References

Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; and Dong, C. 2023. Activating More Pixels in Image Super-Resolution Transformer. arXiv:2205.04437.

Jeevan, P.; Srinidhi, A.; Prathiba, P.; and Sethi, A. 2024.

WaveMixSR: Resource-Efficient Neural Network for Image Super-Resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 5884–5892.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Gool, L. V.; and Timofte, R. 2021. SwinIR: Image Restoration Using Swin Transformer. arXiv:2108.10257.

Martin, D.; Fowlkes, C.; Tal, D.; and Malik, J. 2001. 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, 416–423 vol.2.

Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; and Wang, Z. 2016. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. arXiv:1609.05158.

Zhang, D.; Huang, F.; Liu, S.; Wang, X.; and Jin, Z. 2023. SwinFIR: Revisiting the SwinIR with Fast Fourier Convolution and Improved Training for Image Super-Resolution. arXiv:2208.11247.