PASR: Parallel Attention Image Super Resolution

Aradhana Mishra, Hamza Shafiq, Bumshik Lee*

Chosun University

{aradhana.mishra, hamzashafiq, bslee}@chosun.ac.kr

PASR: 병렬 주의 이미지 초해상도

미쉬라 아라다나, 함자 샤피크, 이범식*
조선대학교

Abstract

Transformer-based super-resolution techniques often face the challenge of effectively capturing local and global features, limiting their ability to produce high-quality super-resolved images. In response, we propose the Parallel Attention Super-Resolution block (PASR). This architecture, where channel attention and spatial attention work in parallel, allows PASR to dynamically recalibrate feature responses, capture long-range dependencies, and emphasize spatially significant regions simultaneously. Through comprehensive experiments, PASR not only showcases substantial improvements over existing methods but also achieves superior quantitative metrics and generates visually compelling super-resolved images with enriched details and coherence. This outstanding performance underscores PASR's potential to effectively address the challenges in image super-resolution and push the boundaries of current methodologies.

Ⅰ. Introduction

In the field of Single Image Super-Resolution (SISR)[1], recent advancements in deep learning, shown by cutting-edge models like SWINIR[2], HAN[8], and SAN[6], have significantly improved the resolution of pictures that feature complex structures like buildings and architectural components. Nevertheless, these technological breakthroughs often prove inadequate when faced with photos that exhibit intricate facial characteristics, which are essential for tasks such as facial recognition.

To overcome this constraint, we provide a system for Single Image Super-Resolution (SISR)[5][7] that utilizes Transformer architecture. This approach incorporates spatial and cross-attention mechanisms, as well as positional encoding. Our architecture is designed to address the difficulties presented by facial content. Its main goal is to improve the reconstruction of facial characteristics by using spatial attention and refining pixel-level features, resulting in improved outcomes. By doing a thorough analysis of many datasets and circumstances, we have shown the efficiency of our method in addressing the challenging issue of picture restoration, specifically in improving face details, resulting in considerable outcomes.

Ⅱ. Method

The architecture comprises three key stages: Shallow Feature Extraction, Deep Feature Extraction, and High-Quality Image Restoration, as shown in Figure 1. In the Shallow Feature Extraction stage, a Swin Transformer[3][4] encoder captures low-level picture information, laying the foundation for subsequent processing. This initial step aims to extract fundamental representations necessary for the restoration process. The Deep Feature Extraction stage is the core of our architecture, consisting of 7 Parallel Attention Blocks (PAB). Each PAB is meticulously crafted to capture intricate and high-level image features essential for restoration. The parallel attention mechanism, comprising both Spatial Attention (SA) and Channel Attention (CA), enhances the capability of these blocks to capture diverse contextual information shown in Figure 1. This design allows for the extraction of detailed and comprehensive feature representations. In this stage, the architecture captures local and global dependencies within the image, ensuring robust feature extraction for subsequent processing. Every PAB is linked using a skip connection, as shown in Figure 1. Initially, the model executes the process of patch unembedding, followed by the application of simultaneous connection attention with patch embedding. This output is directed towards layer normalization and window attention. Finally, in the High-Quality Image Restoration stage, a Swin Transformer decoder is employed to reconstruct the high-quality image.

To evaluate our model's performance, we conducted experiments using traditional testing datasets commonly employed for benchmarking state-of-the-art [5][6][8] (SOTA) methods. These datasets include Set5, Set14 and Urban100. Table 1 shows quantitative results evaluated using PSNR and SSIM metrics.

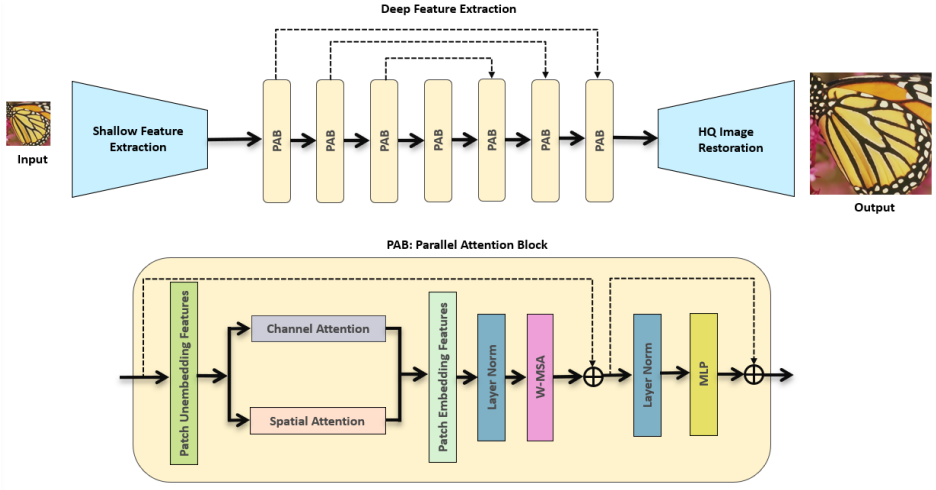

Figure 1: The architecture of the proposed model for image super-resolution network with PAB Blocks.

Table 1: Quantitative comparison (average PSNR/SSIM) with state-of-the-art methods for classical image SR on DIV2K training datasets. Results on ×2 resolution are provided here.

Methods	PSNR (in dB)	SSIM
HAN [8]	38.27	0.9614
RCAN [5]	38.27	0.9614
SAN [6]	38.31	0.9620
PASR(Ours)	38.33	0.9622

Ⅲ. Conclusion

In conclusion, our proposed architecture, consisting of Shallow Feature Extraction, Deep Feature Extraction with Parallel Attention Blocks, and High-Quality Image Restoration using Swin Transformer modules, presents a robust approach to image restoration. By effectively capturing and utilizing shallow and deep image features, our method demonstrates promising results in restoring high-quality images. The parallel attention mechanism enhances the extraction of intricate image features crucial for restoration, ultimately leading to visually pleasing outputs with reduced artifacts.

ACKNOWLEDGMENT

This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(MOE) (2021RIS-002) and also by the National Research Foundation of Korea (NRF) funded by the Korean Government under Grant 2022R1I1A3065473

REFERENCES

[1] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, Shi W. Photo-realistic single image super-resolution using a generative adversarial network. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4681-4690).

[2] Liang J, Cao J, Sun G, Zhang K, Van Gool L, Timofte R. Swinir: Image restoration using swin transformer. InProceedings of the IEEE/CVF international conference on computer vision 2021 (pp. 1833-1844).

[3] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 2020 Oct 22.

[4] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł , Polosukhin I. Attention is all you need. Advances in neural information processing systems. 2017;30.

[5] Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y. Image super-resolution using very deep residual channel attention networks. InProceedings of the European conference on computer vision (ECCV) 2018 (pp. 286-301).

[6] Dai T, Cai J, Zhang Y, Xia ST, Zhang L. Second-order attention network for single image super-resolution. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 11065-11074).

[7] Zhou S, Zhang J, Zuo W, Loy CC. Cross-scale internal graph neural network for image super-resolution. Advances in neural information processing systems. 2020;33:3499-509.

[8] Niu B, Wen W, Ren W, Zhang X, Yang L, Wang S, Zhang K, Cao X, Shen H. Single image super-resolution via a holistic attention network. InComputer Vision– ECCV 2020: 16th European Conference, Glasgow, UK, August 23– 28, 2020, Proceedings, Part XII 16 2020 (pp. 191-207). Springer International Publishing.