

COMPARISON OF SUPER RESOLUTION METHODS

Vinnitsia National Technical University

Анотація

Ця стаття має на меті порівняти та розглянути найкращі варіанти використання найновіших методів надвисокої роздільної здатності зображень: EDSR, Real-ESRGAN та SwinIR. Порівняльний аналіз проводиться відповідно до критеріїв метрик PSNR / SSIM, а також перцептивної якості.

Ключові слова: Super Resolution, згортова нейронна мережа, генеративно-змагальна мережа, свін-трансформер.

Abstract

This paper aims to provide the comparison and the best use cases of the most recent image super-resolution methods: EDSR, Real-ESRGAN, and SwinIR. A comparative analysis is carried out according to the criteria of PSNR / SSIM metrics and also perceptual quality.

Keywords: Super Resolution, convolutional neural network, generative adversarial network, swin transformer.

Introduction

The goal of Super Resolution (SR) methods is to recover a high resolution image from one or more low resolution input images. Many SR methods have been proposed in the computer vision community to improve and achieve high-resolution images over the years. In the past few years, there has been a significant advancement in image SR algorithms. This paper aims to provide the comparison and the best use cases of the most recent image SR methods: EDSR, Real-ESRGAN, and SwinIR.

Main methods

Enhanced Deep Residual Network (EDSR) is a convolutional neural network (CNN) with 32 layers and 256 feature maps designed to maximize pixel-wise reconstruction fidelity in single-image SR [1]. It builds upon deep residual learning by removing batch normalization layers from residual blocks, or ResBlocks (Figure 1). Batch normalization layers normalize the features, getting rid of range flexibility from networks. Their removal increases the performance and reduces GPU memory usage. EDSR is optimized using mean squared error (MSE) loss, which directly correlates with traditional full-reference quality metrics such as peak signal-to-noise ratio (PSNR) and Structural Similarity Index Measure (SSIM).

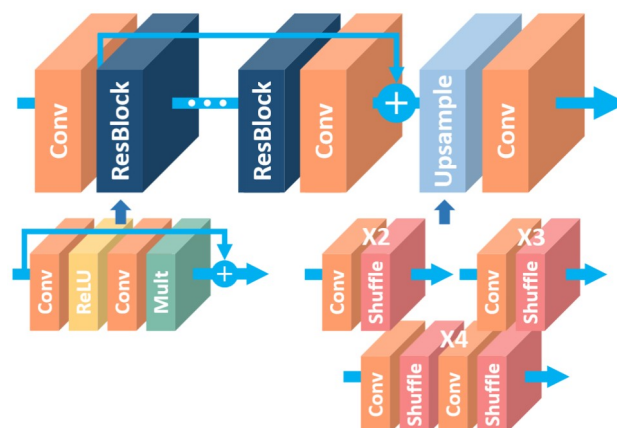


Figure 1 – The architecture of EDSR

Real-ESRGAN is an extension of Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) to address real-world image restoration, where degradations such as noise, blur, compression artifacts, and unknown downsampling kernels are present [2-3]. Instead of training solely on synthetic bicubic downsampling, Real-ESRGAN uses a complex degradation pipeline to simulate realistic image corruption. In generative adversarial networks, two models are trained simultaneously: a generator that generates high-resolution images, and a discriminator that estimates the probability that the training data image comes from something other than the result of the generator's work (Figure 2). Unlike PSNR-oriented methods, Real-ESRGAN prioritizes perceptual quality and robustness over pixel-level accuracy, often producing sharper and more natural-looking images at the cost of numerical fidelity.

SwinIR is a hybrid architecture consisting of three modules: shallow feature extraction (CNN), deep feature extraction (swin transformer), and high quality image reconstruction (CNN) (Figure 3) [4]. It uses shifted window self-attention to model long-range spatial dependencies. The standard multi-head self attention (MSA) module in a classic transformer block is replaced by a module based on shifted windows. For efficient modeling, self-attention is computed within local windows. The windows are arranged to evenly partition the image in a non-overlapping manner. The window grid is being shifted, so patches that were in different windows before end up in the same window after shifting. Over multiple layers information propagates across the entire image without ever doing global attention, that way reducing memory usage [5].

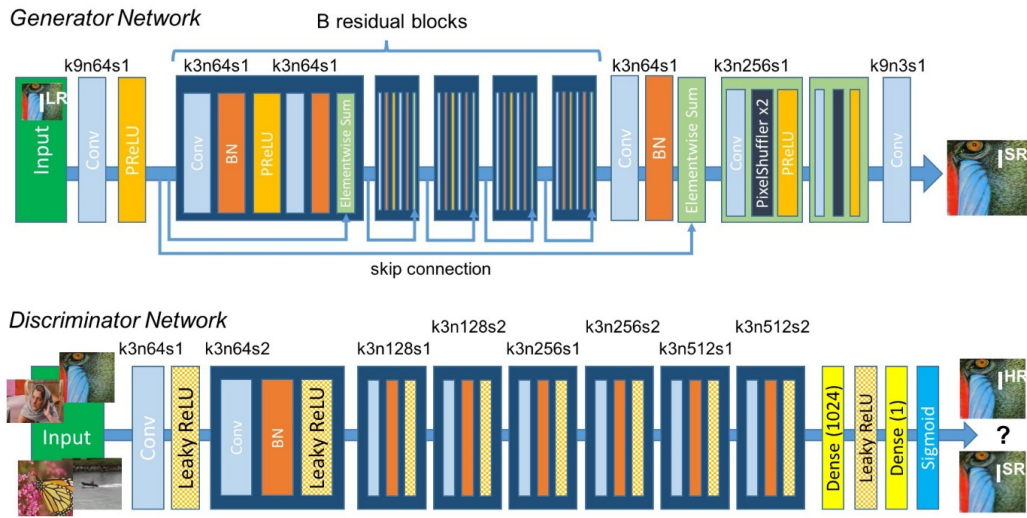


Figure 2 – The architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer

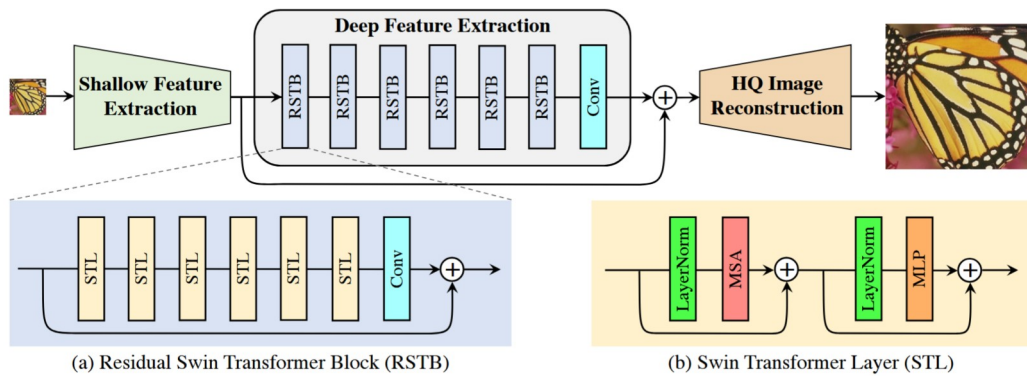


Figure 3 – The architecture of SwinIR

Comparison results

EDSR consistently achieves very high PSNR scores on standard synthetic benchmarks like Set5, Set14, Urban100, making it a widely adopted baseline in SR research (Table 1). However, its reliance on pixel-wise loss often leads to overly smooth textures and limited perceptual realism.

As a result of global context modeling, SwinIR often outperforms CNN-based models such as EDSR while also producing sharper and more coherent textures. It represents a modern compromise between reconstruction fidelity and perceptual quality, albeit at Heavier compute cost.

Unlike PSNR-oriented methods, Real-ESRGAN prioritizes perceptual quality and robustness over pixel-level accuracy, often producing sharper and more natural-looking images at the cost of numerical fidelity. This makes it especially suitable for photo restoration and practical upscaling tasks. Since real degraded images often lack a well-defined ground truth, numerical fidelity metrics become unreliable, and perceptual measures such as qualitative evaluations are preferred.

Table 1 – Quantitative comparison (average PSNR/SSIM) of EDSR and SwinIR on benchmark datasets

Dataset	Scale	EDSR	SwinIR
Set5	×2	38.11 / 0.9601	38.42 / 0.9623
	×3	34.65 / 0.9282	34.97 / 0.9318
	×4	32.46 / 0.8968	32.92 / 0.9044
Set14	×2	33.92 / 0.9195	34.46 / 0.9250
	×3	30.52 / 0.8462	30.93 / 0.8534
	×4	28.80 / 0.7876	29.09 / 0.7950
Urban100	×2	32.93 / 0.9351	33.81 / 0.9427
	×3	28.80 / 0.8653	29.75 / 0.8826
	×4	26.64 / 0.8033	27.45 / 0.8254

Conclusions

CNN models like EDSR are simple in architecture and fast in inference. Zbut an attempt to restore every lost pixel can produce too smooth textures and lack of perceptual sharpness.

Generative adversarial network such al Real-ESRGAN can produce Visually realistic textures and are robust to noise and blur but they can also hallucinate little details. This make them more fitting for the task of old or degraded images upscaling.

Swin Transformers like SwinIR bring a strong balance between PSNR and perceptual quality to the table by capturing global context. But it comes at the cost of heavier compute cost and, consequentially, slower performance. They are fitting for high-quality SR tasks where compute resources are available.

REFERENCES

1. Lim B., Son S., Kim H., Nah S., Lee K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution / 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) – pp. 1132-1140.
2. Wang X., Xie L., Dong C., Shan Y. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data / 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) – pp. 1905-1914.
3. Ledig C., Theis L., Huszar F., Caballero J., Aitken A.P., Tejani A., Totz J., Wang Z., & Shi W. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network / 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) – pp. 105-114.
4. Liang J., Cao J., Sun G., Zhang K., Gool L.V., Timofte R. SwinIR: Image Restoration Using Swin Transformer / 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) – pp. 1833-1844.
5. Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., Lin S., Guo B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows / 2021 IEEE/CVF International Conference on Computer Vision (ICCV) – pp. 9992-10002.

Сичова Марина Сергіївна – студентка групи 1КН-25м, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: marinasychova55707@gmail.com

Мельник Олеся Дмитрівна – к.ф.н., доцент кафедри іноземних мов, Вінницький національний технічний університет, e-mail: prysyazna_lesya@vntu.edu.ua

Sychova Maryna S. – student of group 3KN-24b, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, Ukraine, e-mail: marinasychova55707@gmail.com

Melnyk Olesia D. – Candidate of Philological Sciences, Assistant Professor of the Department of Foreign Languages, Vinnytsia National Technical University, Vinnytsia, e-mail: prysyazna_lesya@vntu.edu.ua