Comparative Analysis of Super-Resolution Techniques

Jacob Lapin
Carnegie Mellon University
Pittsburgh, PA 15213
jelapin@andrew.cmu.edu

Yichen Zhang
Carnegie Mellon University
Pittsburgh, PA 15213
yichenzh@andrew.cmu.edu

Hyunwoo Park
Carnegie Mellon University
Pittsburgh, PA 15213
hp2@andrew.cmu.edu

1 Introduction

The escalating demand for high-resolution visual content in various domains such as medical imaging, satellite imagery, and consumer electronics necessitates advanced super-resolution (SR) techniques that can enhance image quality beyond the limitations of sensor hardware. Traditional SR methods often struggle to balance detail enhancement with artifact suppression, prompting a shift towards more sophisticated machine learning models. This paper conducts a comparative analysis of contemporary SR techniques, primarily focusing on their performance quantified through the Peak Signal-to-Noise Ratio (PSNR).

Our study evaluates multiple leading SR algorithms across diverse datasets ranging from portraits to urban landscapes. By systematically comparing these methods, we aim to identify their strengths and weaknesses in terms of detail reproduction, and computational efficiency. The expected results of our investigation will provide a comprehensive bench marking of existing SR methods, thus facilitating a deeper understanding of their operational dynamics in real-world scenarios.

In addition to utilizing standard evaluation metrics such as PSNR, our research introduces a novel metric based on image saliency called Visual Attention-based Quality Metric (VAQM). This new metric aims to assess the effectiveness of SR methods in preserving and enhancing the perceptual relevance of salient features within images, which are crucial for tasks requiring high levels of visual attention. By integrating saliency into our evaluation framework, we seek to provide a more holistic measure of image quality that aligns closely with human visual perception.

2 Dataset, Task, Evaluations

Datasets:

- **Set5**: Introduced by Bevilacqua et al. in 2012, this dataset is primarily used for initial testing of super-resolution methods due to its small size, which allows for quick performance evaluations.
- **Set14**: Compiled by Zeyde et al. in 2010, this dataset contains a variety of scenes, offering a broader scene diversity. It is utilized to evaluate the versatility of super-resolution algorithms across different types of images.
- **BSDS100**: Part of the larger Berkeley Segmentation Dataset created by Martin et al. in 2001. This subset is used for a comprehensive assessment of an algorithm's generalization capabilities across a broad array of natural scenes.
- **Urban100**: Developed by Huang et al. in 2015, this collection features high-resolution urban images and is ideal for testing super-resolution performance on complex urban scenes, focusing on man-made structures.
- **Sun-Hays 80**: This dataset includes a variety of images used to test and evaluate the effectiveness of super-resolution algorithms on both natural and urban environments.

Preprint. Under review.

Tasks:

- Image Quality Evaluation: Assess the visual quality of the super-resolved images using PSNR and VAQM.
- Processing Speed Analysis: Measure the time taken by each algorithm to process images of varying sizes and complexities.
- Robustness to Noise: Evaluate the performance of super-resolution techniques in the presence of different levels of noise in the input images.
- Generalization Assessment: Test the algorithms on diverse image types across the datasets to evaluate their generalization capabilities.

Evaluation:

The Peak Signal-to-Noise Ratio (PSNR) is a widely used metric in image and video quality assessment, particularly effective in the domains of image compression and super-resolution. It quantitatively measures the quality of a reconstructed image compared to its original version by assessing the noise level introduced due to the reconstruction process.

PSNR is defined based on the Mean Squared Error (MSE) between the original image I and the distorted image K. For images with dimensions $M \times N$, MSE is calculated as:

$$MSE = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (I(i,j) - K(i,j))^2$$

where $I(i,j)$ and $K(i,j)$ are the pixel values at position (i,j) in the original and distorted images, respectively. PSNR is then computed using:

$$PSNR = 20 \times \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right)$$

Here, MAX_I is the maximum possible pixel value of the image, which is typically 255 for 8-bit images.

A higher PSNR indicates lower noise and thus better image quality, while a lower PSNR indicates poorer quality with more distortion. Although a fundamental measure, PSNR sometimes does not correlate well with human visual perception, especially when distortions are not typical noise but are due to processing artifacts or specific types of image content.

PSNR is crucial in evaluating the efficiency of image processing algorithms, especially for applications involving image compression and super-resolution. However, its limitation lies in its assumption of uniform noise perception across various contents, which does not align with the non-linear nature of human visual sensitivity to noise.

3 Related Work

1. [Don+14] introduce SRCNN, the first deep learning-based super-resolution model, which uses a three-layer convolutional neural network to upscale low-resolution images. The SRCNN model set a new benchmark for image super-resolution tasks, demonstrating significant improvements over traditional methods.
2. [Led+17] present SRGAN, a generative adversarial network (GAN) for image super-resolution. The model uses a deep residual network as the generator and a convolutional network as the discriminator. SRGAN is known for its ability to produce high-quality, photo-realistic images, addressing the issue of overly smooth textures in previous methods.
3. [Zha+18] introduce RCAN, a very deep residual channel attention network for image super-resolution. The model incorporates channel attention mechanisms to adaptively re-scale channel-wise features, leading to significant performance improvements. RCAN sets new state-of-the-art results on several benchmark datasets.
4. [Wan+18] improve upon SRGAN by introducing ESRGAN, which features a novel residual-in-residual dense block (RRDB) without batch normalization, enhancing the network's

capacity and stability. The paper also presents a new perceptual loss function that combines content loss and adversarial loss to generate more realistic and detailed textures in super-resolved images.

5. [Che23] introduces a novel distribution-based metric for super-resolution that accounts for the one-to-many mapping problem. This metric correlates highly with human perception and improves training outcomes for super-resolution networks.
6. [Ma+16] introduce BiAtten-Net, a deep learning-based Full-Reference Image Quality Assessment (FR-IQA) method specifically tailored for super-resolution images. This method incorporates a bi-directional attention mechanism that enhances the detection of distortions by focusing on the dynamics between high-resolution references and super-resolution images.
7. [Cha+23] provides a detailed overview of single-image super-resolution techniques enhanced by deep learning. It reviews the progression from traditional methods to advanced deep learning models, including those using generative adversarial networks (GANs). The review also discusses the performance, challenges, and future directions of these models in improving image resolution.
8. [Zho+23] explores image super-resolution using transformer methods. The authors proposed a method using large-window permuted self-attention that achieves less computation comparing to previous transformer methods. They achieved 33.86dB PSNR score on the Urban100 dataset.

4 Approach

We combined the following 5 datasets with a total of 299 images: BSD100, Set5, Set14, SunHays80, Urban100. By simply averaging all channels of 4 nearby pixels in the input images to produce a single output pixel, we produced a down-sampled version of all 299 images with width and heights halved. We will feed in these half-resolution images to our algorithms to attempt to reconstruct the original images by producing an output image with doubled width and height.

We first attempted 4 baseline methods for image reshaping: Nearest Neighbor, Bilinear Interpolation, Bicubic Interpolation, and Lanczos Resampling. In this order, they have increasing effectiveness and decreasing speed. These methods are commonly used in real-time image/video resizing due to their trivial computation cost, but they are not expected to perform better in reconstruction quality than any more sophisticated methods.

5 Experiments

To evaluate each super-resolution algorithm, we commence by inputting half-resolution images from our datasets. Each algorithm is then tasked with the objective of doubling both the width and height of these images, aiming to restore them to their original full-size dimensions. Following the upscaling process, the output images are rigorously compared to their original, high-resolution counterparts.

During the comparison phase, researchers can select from a diverse array of metrics tailored to suit the specific evaluation needs of the study. These metrics, employed to assess the similarity between the upscaled image and the original, may include options such as the Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), among others. This methodological flexibility enables a comprehensive and customized assessment of each algorithm's ability to replicate the true details and quality of the original images.

The image features a performance comparison between various image upscaling techniques using the PSNR metric. The bar chart (Figure 1) illustrates that the Lanczos method outperforms Nearest Neighbor, Bilinear, and Bicubic interpolation, achieving the highest average PSNR value of 28.76 dB. Accompanying the chart, Table 1 presents PSNR values for these methods across several datasets, with Lanczos resampling showing superior results in individual datasets and maintaining the highest average PSNR of 28.7557 dB. This visual data indicates that Lanczos resampling is the most effective technique among those tested for image upscaling tasks.

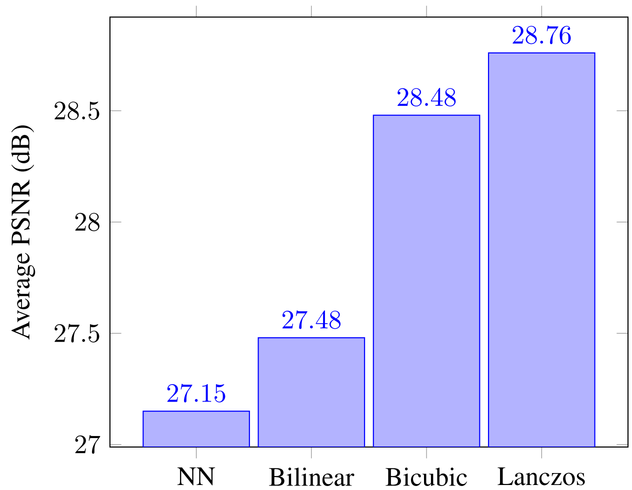

Figure 1: Performance comparison of different image upscaling techniques

Table 1: PSNR comparison on benchmark datasets.

Method / Dataset	BSD100	Set14	Set5	SunHays80	Urban100	Average
Nearest Neighbor	27.3202	26.8949	29.2199	30.2915	24.4016	27.1509
Bilinear Interpolation	27.5094	27.4302	30.4698	30.7725	24.6613	27.4758
Bicubic Interpolation	28.4203	28.4755	31.9631	31.9210	25.6082	28.4783
Lanczos Resampling	**28.6442**	**28.7943**	**32.5046**	**32.2723**	**25.8610**	**28.7557**

6 Plan

6.1 Final Phase (Post-Midway to Conclusion)

Model Evaluation and Comparison:

Individual team members will conduct runs of their respective models and focus on bench-marking performance across the selected datasets using predetermined metrics (PSNR). Team will also look into a novel metric form to better match human perception as mentioned above.

The team will then synthesize and analyze the results. We will then identify strengths and weaknesses of each SR technique in relation to image detail complexity and dataset variety.

Synthesis and Presentation:

The collective findings and insights will be synthesized into a comprehensive poster presentation. We hope to highlight key outcomes and potential implications for future super-resolution research.

References

[Don+14] Chao Dong et al. “Learning a Deep Convolutional Network for Image Super-Resolution”. In: (2014).

[Ma+16] Chao Ma et al. *Learning a No-Reference Quality Metric for Single-Image Super-Resolution*. 2016. arXiv: `1612.05890 [cs.CV]`.

[Led+17] Christian Ledig et al. “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.

[Wan+18] Xintao Wang et al. “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks”. In: *Proceedings of the European Conference on Computer Vision Workshops (ECCVW)*. 2018.

[Zha+18] Yulun Zhang et al. “Image Super-Resolution Using Very Deep Residual Channel Attention Networks”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

[Cha+23] Karansingh Chauhan et al. “Deep Learning-Based Single-Image Super-Resolution: A Comprehensive Review”. In: *IEEE Access* 11 (2023), pp. 21811–21830. DOI: `10.1109/ACCESS.2023.3251396`.

[Che23] Sheng Cheng. *A New Super-Resolution Measurement of Perceptual Quality and Fidelity*. 2023. arXiv: `2303.06207 [cs.CV]`.

[Zho+23] Yupeng Zhou et al. *SRFormer: Permuted Self-Attention for Single Image Super-Resolution*. 2023. arXiv: `2303.09735 [cs.CV]`.