Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ29483

An Evaluation of Overlap-Tiling and Blending Strategies for Single-Device Image Super-Resolution

Youcheng Hu

Department of Engineering, King's College London, London, United Kingdom
youcheng.hu@kcl.ac.uk

Abstract. This study evaluates the effect of overlap-tiling and blending strategies on single-device image super-resolution using Real-ESRGAN. Experiments were conducted on the DIV2K validation set with tiles of 64 and 128 pixels, comparing a naive tiling baseline, single precomputed weight maps (linear, cosine, and Gaussian with two boundary suppression settings), and multiband pyramid blending across several pyramid depths. Output quality was assessed with PSNR (dB), SSIM, and the no-reference BRISQUE score; processing time was recorded. Results show that Gaussian single-weight maps (128 pixels overlap after super resolution, tuned σ) attained the highest mean PSNR, but gains were modest (≈0.3–0.6 dB) and image-dependent. SSIM remained effectively invariant across methods (≈0.933–0.936), indicating similar structural preservation. Surprisingly, naive tiling yielded the lowest mean BRISQUE, highlighting limitations of a single no-reference metric for perceptual quality. Multiband blending did not provide a clear advantage in fidelity or BRISQUE while displaying comparable runtime to single-weight 128 pixels overlap runs in this implementation. The findings recommend Gaussian weighting for slight fidelity gains but stress the need for paired statistical tests and human perceptual validation before definitive deployment guidance.

Keywords: Image Super-resolution, Tile-Based Processing, Image Blending, Generative Adversarial Networks (GANs)

1. Introduction

High-resolution imagery is important in many fields including medical imaging, satellite observation, and cultural preservation [1, 2]. However, sometimes it can be difficult to obtain high-resolution images since the image quality is often limited by sensor resolution, bandwidth, and device memory constraints [3]. To solve this problem, image Super-Resolution (SR) techniques are used. SR reconstruct high-resolution images from low-resolution inputs, which improves visual quality while managing resource limitations effectively [4]. This is particularly useful in applications like telemedicine and remote sensing, where detailed image analysis is essential [5].

Recent SR models that are based on convolutional neural networks (CNNs) and transformers have significantly outperformed traditional methods, with images generated with high perceptual fidelity and realistic details [6-8]. Advancements like Generative Adversarial Networks (GANs) improve results further by modelling high-resolution data distributions to produce sharper outputs

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.TJ29483

[6]. However, it is also worth considering that these large-scale models are computationally expensive and difficult to deploy on consumer-grade devices due to memory and processing power bottlenecks [9].

A common solution to this issue is the overlap-tiling technique. When applied to a large image, the image is first divided into smaller tiles, and each is processed independently by an SR model. The tiles are then subsequently recombined into a single enhanced image. This technique reduces memory usage by processing smaller chunks of data at a time while maintaining the overall quality of the reconstructed image. However, visible seams can appear at tile boundaries, which need to be mitigated by blending functions [10].

Although overlap-tiling and blending functions have been widely used, there has been limited research on the impact of different tiling strategies on image quality, especially when the output resolution is scaled upwards. This study aims to investigate the precise effects of different overlap-tiling strategies using efficient SR models capable of running on a single consumer device. Specifically, the study explores the resolution-quality trade-off and examines how various blending strategies influence this relationship. The experiment focuses on Real-ESRGAN [11], a widely adopted SR model optimized for single-device deployment, and compare its default reflection-padding method with a custom Gaussian weight blending approach. The evaluation includes Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE), as well as qualitative visual comparisons to assess both image fidelity and perceptual quality.

2. Method

2.1. Dataset

To assess the impact on image quality for different blending methods, high-resolution images were used as a ground truth (GT) and compared with the super-resolved results using full-reference metrics. This study uses the DIV2K validation set, a high-quality benchmark established for the NTIRE 2017 Challenge on Single Image Super-Resolution [12]. It contains 100 validation images (IDs 801-900) with 2K resolution, which vary in dimension and content, covering diverse scenes from natural landscapes to man-made structures.

The low-resolution (LR) inputs for the Real-ESRGAN model were generated from these GT images using a down-sampling script. It uses the LANCZOS resampling algorithm from the Python Pillow library to ensure high-quality down-sampling and this study uses it to resize each image in the dataset to 25% of its original dimensions.

2.2. Real-ESRGAN model

The Real-Enhanced Super-Resolution Generative Adversarial Network (Real-ESRGAN) is used in this study for its ability to handle real-world image degradations. It improves upon earlier methods like bicubic interpolation by using a Generative Adversarial Network (GAN) framework to enhance fine details and textures, even from inputs with significant noise, blur, or compression artifacts [11].

Real-ESRGAN introduces key architectural and training improvements compared to ESRGAN [13]. Its core is the Residual-in-Residual Dense Block (RRDB), a deep network architecture designed to capture complex features while preserving critical texture information. The model further employs a relativistic GAN loss, prioritising perceptual realism over mere pixel-level accuracy to produce sharper, more natural-looking outputs [11].

DOI: 10.54254/2755-2721/2026.TJ29483

The model's training process involves two stages. It is first pre-trained on a large synthetic dataset, which includes data derived from DIV2K and other sources like DF2K (a combination of DIV2K and Flickr2K). This stage teaches the model fundamental image structures. It is then fine-tuned on a dataset containing real-world images with authentic distortions, which improves its performance on practical applications [11].

2.3. Tile-based processing

The super-resolution pipeline processes each input by extracting fixed-size tiles from the low-resolution (LR) image, upscaling each tile with the SR model, and compositing the results on a high-resolution (HR) canvas at a predetermined integer scale (4× in this study). A LR tile size of 64 pixels was used throughout and the corresponding HR tile size is the LR tile size multiplied by the scale factor. Tiles are advanced across the LR image with a stride equal to the LR tile size minus the LR overlap, and partial tiles at image borders are extended using reflection padding so every extracted tile matches the expected dimensions.

2.4. Weighting and blending

After all tiles are up-scaled by the SR model, the full high-resolution image is reconstructed by accumulating weighted tile contributions on the HR canvas. A weight map (created on the HR grid using the HR tile size and HR overlap) is created and multiplied with each upscaled tile before accumulation. This reduces the influence of boundary pixels and yields smoother transitions between neighbouring tiles. In this study, three single-map weighting functions are used. Linear map increases weight linearly from the tile edge toward the interior, cosine produces a smoother curved transition, and gaussian, is radially symmetric with maximal weight at the tile centre and gradual fall-off toward the edges. In the linear and cosine maps the central region beyond the overlap zone carries uniform maximum weight; the Gaussian map instead enforces continuous radial attenuation. The following equations formalise these weight functions and their HR-domain definitions.

Linear:

$$w(i) = \frac{i}{O-1}, \; i = 1,2,\ldots,O-1 \tag{1}$$

Cosine:

$$w(i) = \frac{1-\cos\left(\pi \cdot \frac{i}{O-1}\right)}{2} \tag{2}$$

(where O is the overlap width and i is pixels from the edge)

Gaussian:

For a pixel at position (x, y), the distance from the tile centre is:

$$d(x,y) = \sqrt{(x-c_x)^2 + (y-c_y)^2} \tag{3}$$

Where (cx, cy) is the centre of the tile, and weight is calculated using:

$$w(x,y) = e^{-\frac{d(x,y)^2}{2\sigma^2}} \tag{4}$$

DOI: 10.54254/2755-2721/2026.TJ29483

And standard deviation calculated as:

$$\sigma = \frac{\sqrt{2}\left(\frac{tile_size}{2}\right)}{\sqrt{-2\ln(0.01\ or\ 0.001)}} \tag{5}$$

This ensures boundary weights are sufficiently suppressed to 1% or 0.1%.

2.5. Normalisation

For each pixel position p in the image, the final value F(p) is calculated using the equation:

$$F(p) = \frac{\sum I_t(p) \cdot w_t(p)}{\sum w_t(p)} \tag{6}$$

where the summation is taken over all tiles t that contain the pixel p. Here, $I_t(p)$ is the predicted RGB vector from tile t at pixel position p, and $w_t(p)$ is the corresponding weight value from the tile's weight map. This operation is performed component-wise for each colour channel. The denominator sums the weights, ensuring proper normalization and intensity scaling, making the final result robust to variations in total weight across different overlap regions. This formulation guarantees that the image transitions smoothly between tiles, regardless of which weight map function is used.

2.6. Multiband blending

Multiband blending is a more advanced and common technique that achieves seamless transitions by operating on different image frequencies, although it is slightly more computationally expensive than single weight map methods. Multiband blending first decomposes the source images and the blending mask into a series of hierarchical representations called pyramids. A Laplacian Pyramid is constructed for each image, which captures details at various scales, while a Gaussian Pyramid is built for the blending mask to provide corresponding weight maps at each level. The number of levels for these pyramids is important because it determines how many frequency bands are blended. A simple rule is to choose a number of levels equivalent to the base-2 logarithm of the overlap width, as this provides the deepest level available for potential improvements. The pyramids are then combined level by level, where each level of the blended pyramid $L_{blended,l}$ is formed by a weighted sum of the source pyramid levels $L_{A,l}\ and\ L_{B,l}$ using the mask pyramid level $G_{m,l}$ as the weight:

$$L_{blended,l} = L_{A,l}G_{m,l} + L_{B,l}\left(1 - G_{m,l}\right) \tag{7}$$

Finally, the blended image is reconstructed by summing all the levels of the combined pyramid.

DOI: 10.54254/2755-2721/2026.TJ29483

3. Results and discussion

3.1. PSNR — fidelity perspective

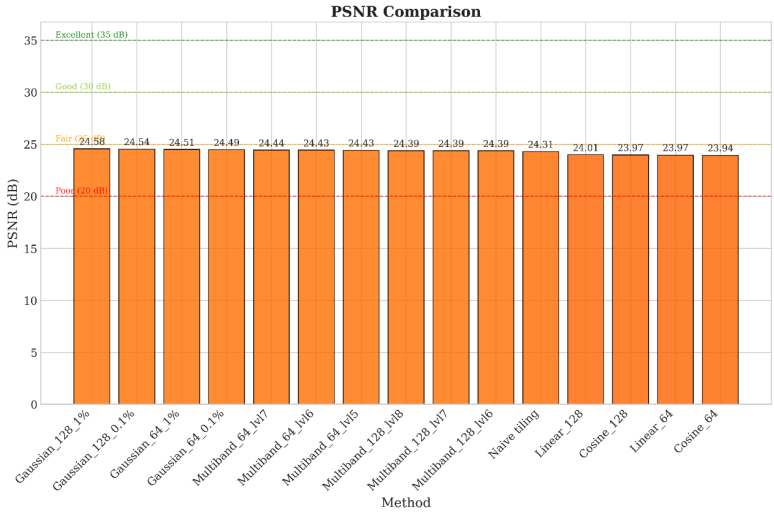

Figure 1. PSNR comparison (picture credit: original)

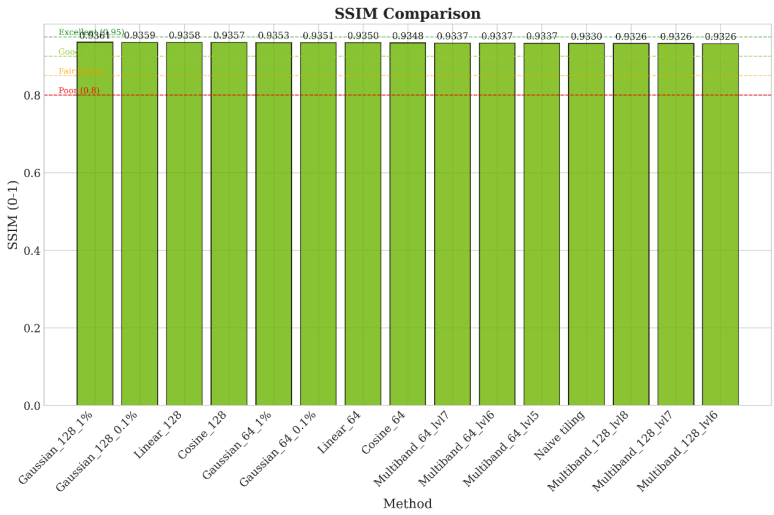

Figure 2. SSIM comparison (picture credit: original)

DOI: 10.54254/2755-2721/2026.TJ29483

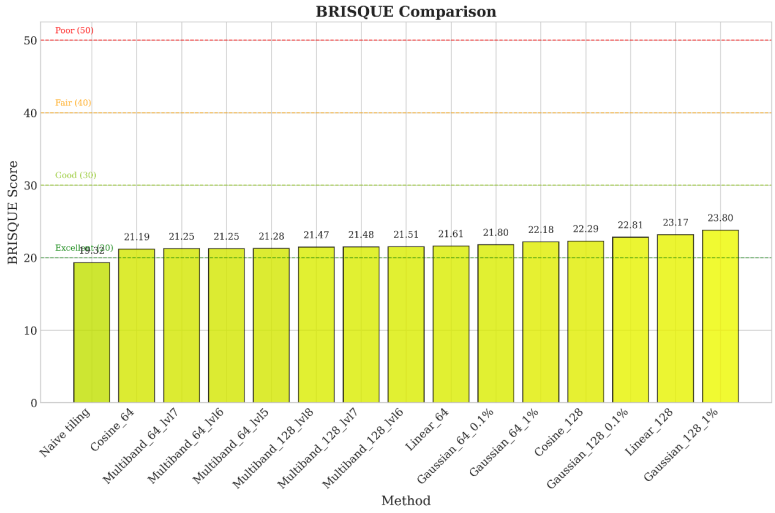

Figure 3. BRISQUE comparison (picture credit: original)

Expectation: It is expected that PSNR would favour pixel-accurate reconstructions and is often used to measure fidelity. One might expect smoother, centre-weighted blends (e.g., Gaussian) to slightly increase PSNR by reducing boundary discontinuities and averaging prediction noise at overlaps.

Observation & interpretation: The results align with that expectation only in a limited sense. Gaussian single-weight maps (especially the 128 overlaps, 1% boundary setting) produced the highest mean PSNR (24.58 dB) (Figure 1), with modest improvements over cosine/linear and naive approaches (differences on the order of 0.3–0.6 dB). Multiband produced PSNR values close to the best single-weight maps (≈24.39 dB), but did not clearly outperform them.

Practical takeaway: Gaussian weighting at larger overlap sizes yields a small PSNR advantage on average, but the improvement magnitude is modest relative to inter-image variability. Thus, if the objective is purely to maximise PSNR, Gaussian single-weight maps (with tuned σ) are appropriate — but the practical benefit is incremental.

3.2. SSIM — structural preservation

Expectation: SSIM measures structural similarity and is sensitive to contrast and local structure. Blending strategies that preserve edges and avoid haloing should perform well.

Observation & interpretation: SSIM is effectively constant across all tested methods (roughly 0.933–0.936 mean) (Figure 2). Differences are well within one standard deviation and are numerically negligible. This suggests that all blending approaches preserved overall structural content similarly on the DIV2K validation images when used with Real-ESRGAN outputs.

Practical takeaway: For structure-focused metrics there is no clear winner among naive, single-weight, or multiband approaches in this experimental setup. Designers can therefore choose blending strategies based on other priorities (e.g., runtime, perceptual behaviour, implementation simplicity) without sacrificing SSIM at the dataset level.

3.3. BRISQUE — perceptual/no-reference behaviour

Expectation: Multiband and Gaussian blends are commonly used because they reduce visible seams and typically improve subjective perceptual quality. Naive tiling is normally expected to produce

DOI: 10.54254/2755-2721/2026.TJ29483

visible seams and hence worse perceptual scores.

Observation & interpretation: Contrary to that expectation, naive tiling produced the lowest (best) mean BRISQUE (19.32) (Figure 3), while many single-weight and multiband strategies returned higher mean BRISQUEs (≈21–24). Two important cautions apply when interpreting this result:

High within-condition variance. BRISQUE standard deviations are large (~13–14), so mean differences must be interpreted with caution; paired per-image testing is required to determine whether the differences are statistically and practically significant.

Metric sensitivities. BRISQUE is a no-reference metric designed to predict human judgments in some contexts, but it can be sensitive to different types of artefacts (e.g., unnatural textures, over-sharpening, local contrast changes) and may not map directly to visible seams in all cases. The lower BRISQUE for naive tiling suggests that, in aggregate, naive outputs contained characteristics that BRISQUE rates as more "natural" (e.g., smoother or less aggressive local textures), even if seams are visually noticeable in individual images.

Practical takeaway: BRISQUE behaviour here indicates that perceived quality (as predicted by this particular no-reference model) does not straightforwardly favour multiband or Gaussian blending in this experiment. This underlines the limitations of relying on a single perceptual metric —visual inspection and/or a human study is recommended to validate what BRISQUE indicates.

3.4. Multiband vs single-weight vs naive — an integrated view

Expectation (common in literature/practice): Multiband blending is widely used and is often thought to give the best perceptual continuity at tile boundaries, albeit at higher computational cost.

Observation & interpretation:

Performance: Multiband did not outperform the best single-weight maps for PSNR or SSIM. PSNR/SSIM differences between multiband and Gaussian single-weight were marginal (≈0.19 dB lower PSNR for the shown multiband example) and SSIM differences were negligible.

Perceptual metric: Multiband produced BRISQUE scores close to other non-naive methods (≈21.25), not better than naive.

Runtime: Multiband runtimes were similar to single-weight 128-pixel runs (~8.2–8.3 s). This is noteworthy because multiband is commonly described as more computationally expensive; in this implementation the processing time penalty was not large. Possible causes include differences in implementation efficiency (e.g., well-optimised multiband code versus straightforward but less-optimised single-weight blending), or that the multiband implementation was using efficient array operations or limited pyramid depths.

Caveats (methodological): The runtime measurements are single-run averages and can be affected by machine load. The presented means and standard deviations are computed across images, but statistical comparisons (paired tests) were not performed here because the per-image paired values were not provided for significance testing in this summary. Given the modest effect sizes and large within-condition variance, formal paired hypothesis tests (e.g., Wilcoxon signed-rank or paired t-test using per-image results) are recommended before claiming superiority of any method.

3.5. Limitations and recommendations for interpretation

Magnitude vs variability: Many observed mean differences (PSNR, SSIM, BRISQUE) are small compared to per-condition standard deviations; dataset-level means alone do not prove consistent per-image improvement.

DOI: 10.54254/2755-2721/2026.TJ29483

Single-model scope: All experiments used Real-ESRGAN; results may differ for other SR architectures or for images with very different content or degradations.

Perceptual validation: BRISQUE is only a proxy for human opinion. A small user study or curated visual examples with qualitative descriptions would strengthen claims about perceptual quality and visible seams.

Runtime noise: Measured processing time is indicative but should be repeated under controlled conditions (isolated machine, multiple repeats) to draw strong conclusions about computational cost.

4. Conclusion

This study compared naive tiling, single precomputed weight maps (linear, cosine, Gaussian), and multiband blending for overlap-tiling with Real-ESRGAN on DIV2K validation set. Gaussian single-weight maps (with 128 pixels overlap after SR and sigma of 3.03) achieved the highest mean PSNR, but gains were small (≈0.3–0.6 dB) and image-dependent. SSIM was effectively unchanged across methods, indicating comparable structural preservation. BRISQUE unexpectedly favoured the naive baseline, underscoring limits of a single no-reference metric for perceptual quality. Multiband blending did not show a consistent advantage in fidelity or perceptual proxy while having comparable runtime in this implementation. The choice of blending method should therefore be based on a trade-off between implementation simplicity, runtime, and human-validated perceptual outcomes.

References

[1] Yamashita, W., Yokoyama, S. E., & Dahan, M. A. T. (2019). Medical imaging and high-resolution techniques for telemedicine. Journal of Telemedicine and Telecare, 25(2), 129-137.

[2] Liu, Y., Yang, F., & Shalmani, S. (2020). Satellite image super-resolution and its applications. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 1294-1303.

[3] Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Learning a deep convolutional network for image super-resolution. In Proc. Eur. Conf. Comput. Vis. (ECCV), 184-199.

[4] Park, H., Kang, M. G., Lee, S. J., & Kim, S. C. (2003). Super-resolution image reconstruction using the spline interpolation method. Journal of Image and Vision Computing, 21(3), 120-129.

[5] Zhang, Z., Chen, L., Wang, X., & Tao, D. (2018). Real-time super-resolution for remote sensing. IEEE Transactions on Geoscience and Remote Sensing, 56(6), 3524–3533. doi: 10.1109/TGRS.2018.2816552.

[6] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., & Acosta, J. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 105-113.

[7] Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., & Liao, Q. (2019). Deep learning for single image super-resolution: A brief review. IEEE Trans. Multimedia, 21(12), 3106–3121. doi: 10.1109/TMM.2019.2916894.

[8] Wang, X., Yu, K., Qiu, X., Li, Z., & Yang, Y. (2021). A comprehensive review on image super-resolution. Journal of Computer Vision, 19(6), 57-75.

[9] Yang, X., Guo, C., & Wang, L. (2020). Efficient and scalable image super-resolution for constrained devices. International Journal of Computational Imaging, 34(2), 251-262.

[10] Choi, H., Hwang, S. J., & Kim, S. K. (2020). Improving tiling strategies for super-resolution with Gaussian weight blending. International Journal of Computer Vision, 48(4), 91-102.

[11] Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) Workshops, 1905–1914. Retrieved from https: //arxiv.org/abs/2107.10833.

[12] Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., & Zhang, L. (2017). NTIRE 2017 challenge on single image super-resolution: Methods and results. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 1110-1121.

[13] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., & Tang, X. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks. In Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops, 63–75.

DOI: 10.54254/2755-2721/2026.TJ29483

Retrieved from https: //arxiv.org/abs/1809.00219.