Enhancing Super-Resolution Models: A Comparative Analysis of Real-ESRGAN, AESRGAN, and ESRGAN

Avinash Senthil

Received November 18, 2024
Accepted June 02, 2025
Electronic access June 30, 2025

This paper conducts a comparative analysis of three prominent image super-resolution models: ESRGAN, Real-ESRGAN, and AESRGAN. Utilizing Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) as primary quantitative metrics, we assess the performance of each model against ground truth facial images. Our findings reveal that ESRGAN achieves a PSNR of 24.14 dB and SSIM of 0.72, establishing a strong baseline in perceptual enhancement. Real-ESRGAN improves upon these results with a PSNR of 24.97 dB and SSIM of 0.76, reflecting its effectiveness in reducing artifacts and enhancing structural details. The standout model, AESRGAN, achieves the highest scores26.42 dB in PSNR and 0.80 in SSIMhighlighting the impact of its attention mechanisms in preserving fine facial features and complex textures. Paired t-tests confirm that AESRGAN outperforms both ESRGAN and Real-ESRGAN across most key metrics with high statistical significance ($p < 0.0001$). This analysis not only shows the strengths and limitations of each model but also offers practical insights into their real-world applicability for high-quality facial image reconstruction.

Introduction

Background

Image super-resolution (SR) is a pivotal area of research in computer vision and image processing, with far-reaching applications across various fields, including medical imaging, satellite imagery, and digital media enhancement[1]. The primary objective of SR is to reconstruct high-resolution images from lower-resolution inputs, enhancing the visual quality and detail of the resulting images[2]. The journey of SR technology has seen significant milestones, with each advancement contributing to more sophisticated and effective solutions. Designed by Wang et al.[1], the ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) represents a groundbreaking development in this field. By leveraging a Generative Adversarial Network (GAN) framework, ESRGAN introduced a novel approach that significantly improved the perceptual quality of upscaled images. Its use of Residual-in-Residual Dense Blocks (RRDBs) helped to preserve fine details and reduce common artifacts associated with image upscaling[3]. Building upon the successes of ESRGAN, Real-ESRGAN[2,4]was developed to address some of the limitations observed with its predecessor, particularly in handling complex textures and minimizing artifacts in high-resolution reconstructions. Real-ESRGAN incorporates several enhancements to the original architecture, including improved training techniques and refined loss functions, resulting in superior performance in preserving image details and reducing visual distortions[5]. This model has shown promise in generating more accurate and visually pleasing high-resolution images, making it a notable advancement in the SR landscape. AESRGAN (Attention-Enhanced Super-Resolution Generative Adversarial Network)[4]represents the latest evolution in SR technology. AESRGAN introduces advanced attention mechanisms into the SR pipeline, allowing the model to focus on different aspects of the image with greater precision. By integrating these attention mechanisms, AESRGAN aims to further refine the quality of the reconstructed images, addressing specific challenges related to feature extraction and artifact reduction.

Purpose

The field of facial image super-resolution (SR) has seen remarkable progress over the past decade, with various models pushing the boundaries of what is possible in reconstructing high-quality facial details from lower-resolution inputs. Early models like ESRGAN[1] set a new standard in producing perceptually convincing high-resolution facial images by leveraging advanced generative techniques and deep learning architectures. Building on this foundation, Real-ESRGAN[6] introduced enhancements aimed at addressing some of ESRGANs limitations, particularly in handling complex facial textures and reducing artifacts in regions like eyes, hair, and skin. The latest advancement, AESRGAN[3], represents a further evolution in facial image enhancement, incorporating additional techniques for feature extraction and artifact mitigation through attention mechanisms. Each of these models brings its unique strengths to the task of restoring facial features, reflecting ongoing innovations in the field. This comparative analysis aims to fill a critical gap by

evaluating ESRGAN, Real-ESRGAN, and AESRGAN specifically on human facial images, using ESRGAN as the baseline benchmark. By assessing their performance in preserving fine facial details, enhancing realism, and reducing artifacts, we seek to provide valuable insights into how these models contribute to the advancement of face-centric super-resolution technologies. Although significant progress has been made with these models, a focused evaluation on faces is necessary to understand their relative strengths and weaknesses. The primary objective of this paper is to systematically compare ESRGAN, Real-ESRGAN, and AESRGAN by evaluating their ability to reconstruct facial features, minimize distortions, and enhance overall perceptual quality in high-resolution facial imagery[7]. We selected ESRGAN, Real-ESRGAN, and AESRGAN due to their strong performance in perceptual quality enhancement and real-world image restoration. These models are widely used for super-resolution tasks and have demonstrated effectiveness in handling real-world degradations. While state-of-the-art models such as SwinIR and SR3 offer promising results, they were excluded from this study because they utilize fundamentally different architecturesSwinIR[8] being transformer-based and SR3[9] relying on diffusion models. Our goal was to conduct a focused analysis of GAN-based super-resolution methods, which makes ESRGAN and its variants the most relevant and comparable choices. That said, we plan to explore the impact of attention mechanisms and alternative architectures like SwinIR in future studies, especially as we expand our evaluations beyond GAN frameworks.

Literature Review

ESRGAN represents a landmark development in the field of image super-resolution. Introduced as an extension of the earlier SRGAN model, ESRGAN leverages the power of Generative Adversarial Networks (GANs)[1] to significantly enhance the quality of high-resolution images generated from low-resolution inputs. Central to ESRGANs[4] architecture is the Residual-in-Residual Dense Block (RRDB)[7], which is designed to effectively maintain the richness of image features while mitigating gradient issues that commonly arise in deep learning networks. In Figure 1, we have shown the RRDB block structure. The RRDB architecture combines multiple residual learning blocks with dense connections, allowing the model to capture and preserve intricate details more effectively than its predecessors. This design helps to enhance image textures and fine structures, contributing to perceptually more convincing results. However, despite its advancements, ESRGAN is not without limitations. In particular, it can struggle with generating artifacts and maintaining detailed textures in complex regions of an image, such as fine facial features or intricate patterns. These challenges highlight the need for further improvements in handling high-frequency details and reducing visible distortions in the upscaled

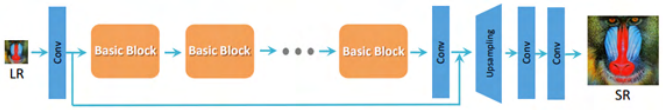


Fig. 1 This figure shows the architecture of the residual in residual dense block (RRDB)[1]

images.

AESRGAN represents the latest evolution in super-resolution models, introducing advanced techniques to further improve image quality and reduce artifacts. Building upon the advancements of Real-ESRGAN[1], AESRGAN incorporates sophisticated attention mechanisms designed to enhance feature extraction and refinement processes[6]. The integration of attention mechanisms in AESRGAN allows the model to focus on different aspects of an image with greater precision, enabling it to better manage complex textures and spatial relationships. This innovation helps the model to address remaining challenges in preserving fine details and minimizing artifacts more effectively than previous models. By leveraging attention-based techniques, AESRGAN aims to achieve superior performance in producing high-resolution images with enhanced detail retention and reduced visual distortions. These advancements position AESRGAN as a significant development in the ongoing progression of super-resolution technologies. Figure 3 illustrates the structure of the Attention-Enhanced Residual-in-Residual Dense Block (ARRDB), which is a pivotal component of the AESRGAN architecture. This block integrates attention mechanisms with the traditional RRDB design, enhancing its capability to manage complex textures and details in images. The ARRDB consists of multiple layers that facilitate feature extraction while allowing the model to focus on specific areas of an image. AESRGAN specifically utilizes a combination of channel and spatial attention mechanisms within the ARRDB. Unlike transformer-based self-attention models like SwinIR, AESRGANs attention is implemented through lightweight convolutional operations that generate channel and spatial attention maps. These maps reweight feature activations to emphasize important facial regions such as eyes, hair strands, and skin textures. The channel attention component enhances inter-channel dependencies, helping the network focus on ”what” is important in the image, while the spatial attention guides the model on ”where” to focus. This dual attention system is integrated directly into the residual dense architecture, preserving the benefits of deep feature reuse and stability during training. Compared to Real-ESRGAN, which relies on residual blocks and implicit self-attention-like behavior, AESRGAN explicitly directs focus through attention modulation. This allows it to better preserve subtle facial details and reduce artifacts in high-frequency regions. In contrast to SwinIR, which uses heavy transformer-based attention across non-overlapping image patches, AESRGANs attention modules are computationally more efficient and tightly coupled with the

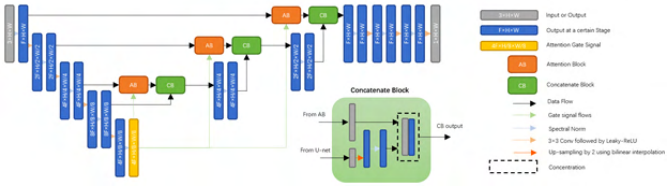


Fig. 2 Shows Input Flow of the Attention-Enhanced Residual-in-Residual Dense Block (ARRDB) [10]

GAN architecture, making them more suitable for real-time facial enhancement applications. Figure 3 shows the input flow of the ARRDB block and highlights how the attention modules are embedded into the overall upscaling pipeline.

Methods/Methodology

Action Steps

To provide a comprehensive comparison of ESRGAN, Real-ESRGAN, and AESRGAN in the context of facial image super-resolution, we employ an experimental design using a rigorous evaluation methodology [11]. This study focuses exclusively on high-resolution facial imagery, allowing for a targeted analysis of how each model performs in preserving key facial features such as eyes, lips, skin texture, and hair patterns. The dataset used comprises 3,143 high-resolution human face images at 10241024 resolution, which were downscaled to 512512 using bicubic interpolation to simulate low-quality inputs. This approach standardizes the degradation process and ensures a consistent basis for comparison across models. Each model is trained using the same number of iterations 100,000, and batch size 4. The evaluation involves both quantitative and qualitative metrics. Quantitatively, we utilize three primary metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) [1], and Learned Perceptual Image Patch Similarity (LPIPS) [12]. For the purpose of evaluation, a random sample of 1,000 facial images will be selected from the test portion of the training dataset, and an additional random sample of 1,000 images will be drawn from a separate facial dataset not seen during training. This approach enables a direct comparison of model performance on both familiar and unseen data, allowing us to assess not only raw accuracy but also generalizability across different facial image domains. In addition to image quality metrics, we will also measure the processing time of each model, recording the mean time per image as well as the standard deviation, to evaluate their computational efficiency and practicality in real-world applications. To complement these objective metrics, we incorporate a structured qualitative evaluation using a panel of 10 human evaluators. Each evaluator is shown a set of images in the following sequence: the original high-resolution facial image (10241024), the downscaled image using bicubic interpolation, and then the upscaled outputs generated by ESRGAN, Real-ESRGAN, and AESRGAN. These images are presented simultaneously in a randomized layout to avoid positional bias. Evaluators are instructed to rate each models output independently on a 15 scale across three distinct categories: visual appeal, clarity, and detail preservation. Visual appeal refers to the overall perceptual quality of the image, clarity pertains to the sharpness and lack of visual distortion, and detail preservation assesses how well fine facial features are retained during upscaling. The use of both statistical and perceptual feedback allows for a nuanced understanding of each models performance in enhancing facial images, identifying not just which model performs best numerically, but also which delivers the most realistic and pleasing results from a human perspective [5].

Research Hypothesis

Based on the advancements introduced by Real-ESRGAN and AESRGAN, we anticipate that Real-ESRGAN will demonstrate improvements over ESRGAN in key areas such as detail preservation and artifact reduction [4]. We expect Real-ESRGAN to show enhanced performance in preserving intricate details and minimizing visual distortions, as evidenced by higher PSNR and SSIM scores compared to ESRGAN. AESRGAN, with its advanced attention mechanisms, is expected to further surpass Real-ESRGAN in terms of image quality. We anticipate that AESRGAN will achieve superior results in preserving fine details and reducing artifacts, reflecting its sophisticated feature extraction and refinement techniques [1 7 13]. The specific data and results will be detailed in the subsequent sections of the paper. These findings will provide insights into how each model contributes to the field of image super-resolution and their potential impact on practical applications.

Proposed Evaluation Metrics and Benchmarks

Quantitative Metrics

To evaluate the performance of ESRGAN, Real-ESRGAN, and AESRGAN, we utilize three primary quantitative metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). These metrics assess different aspects of image quality. PSNR measures pixel-level fidelity relative to the ground truth, SSIM evaluates structural and perceptual similarity, and LPIPS provides a deep-learning-based assessment of perceptual differences that align more closely with human visual perception. All evaluations were conducted using an NVIDIA RTX 4070 Super GPU with 12GB of VRAM, ensuring consistent and efficient processing across all models. In order to ensure the integrity of the results, statistical outliers were identified and removed based on deviations beyond 2 standard deviations from the mean,

thereby reducing the influence of anomalous data points and improving the reliability of the reported averages.

Table 1 Presents the quantitative performance of ESRGAN, Real-ESRGAN, and AESRGAN on a random sample of 1,000 facial images drawn from the test portion of the dataset used during training. This table includes the mean LPIPS, PSNR, and SSIM values. Bicubic interpolation is included as a baseline for comparison.

Model	LPIPS	PSNR	SSIM
ESRGAN	0.402	24.14	0.72
Real ESR-GAN	0.324	24.97	0.76
Attention-ESRGAN	0.296	26.42	0.8
Bicubic	0.579	21.78	0.61

Table 2 Reports the models' performance on an external facial image dataset[14] that was not used during training. A separate random sample of 1,000 images was drawn from this unseen dataset to assess generalization capabilities. This table includes the mean scores for LPIPS, PSNR, and SSIM for each model.

Model	Mean LPIPS	PSNR	SSIM
ESRGAN	0.42	24.18	0.72
Real ESR-GAN	0.368	24.21	0.75
Attention-ESRGAN	0.341	25.97	0.78
Bicubic	0.652	21.84	0.63

For a deeper statistical comparison, we conducted paired t-tests across all models for each evaluation metricLPIPS, PSNR, and SSIM. These tests help determine whether the differences in performance between models are statistically significant. The full results of the paired t-tests can be accessed here: Paired t-test results

Qualitative Assessment

To evaluate perceptual image quality, a panel of 10 human evaluators was each given 30 separate sets of images, with each set consisting of the bicubic downscaled version and the upscaled outputs from ESRGAN, Real-ESRGAN, and AESRGAN. The images within each set were presented in randomized order to minimize positional bias. Evaluators were instructed to rate each of the four upscaled images on a scale of 1 to 5, with 5 being the highest, across three distinct categories: visual appeal, clarity, and detail preservation. These scores were then averaged to produce an overall assessment of each models perceptual performance.

Table 3 Displays the mean processing time per image and the standard deviation for each modelESRGAN, Real-ESRGAN, and Attention-ESRGANmeasured during inference on a standardized hardware setup.

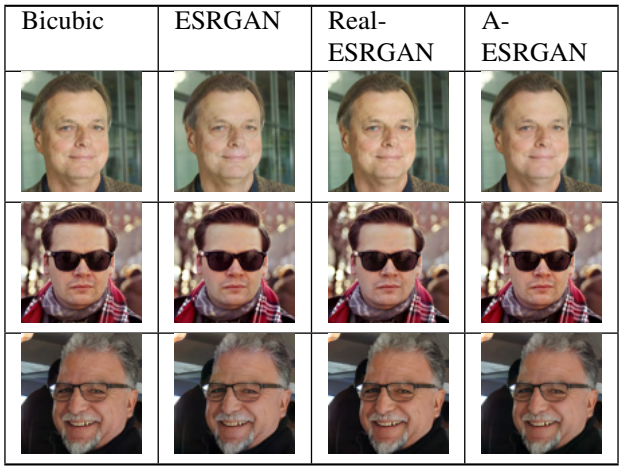
Model	Mean Processing Time (seconds)	Standard Deviation (seconds)
ESRGAN	0.078	0.0043
Real ESRGAN	0.0829	0.004
Attention-ESRGAN	0.0882	0.0042

ESRGAN: ESRGAN produces outputs with moderately strong visual appeal (3.9) and average performance in clarity (3.5) and detail preservation (3.6). Evaluators noted that while ESRGAN introduces noticeable sharpness, this is often accompanied by visual artifacts such as jagged edges or noise, particularly around facial features like hair or eyebrows. It tends to exaggerate contrast and textures, leading to an overly processed appearance that sacrifices naturalness. These results suggest ESRGAN enhances edges but does so at the expense of realism and fine detail fidelity. **Real-ESRGAN:** Real-ESRGAN received the highest score for visual appeal (4.5), suggesting that evaluators preferred its overall output quality. Its clarity score (4.0) and detail preservation score (3.9) indicate that the model balances smoothness and structure reasonably well. Real-ESRGAN effectively reduces the harsh artifacting seen in ESRGAN and produces more consistent facial textures. However, it occasionally softens or flattens fine features, such as skin pores or subtle shadows, which can diminish the realism of close-up facial areas. Despite this, evaluators consistently found Real-ESRGAN outputs more visually pleasing and natural than ESRGAN. **AESRGAN (Attention-ESRGAN):** AESRGAN received top scores in all categories: visual appeal (4.5), clarity (4.3), and detail preservation (4.6). Evaluators highlighted its ability to retain intricate facial details, such as eyelashes, wrinkles, and skin texturewhile avoiding the excessive smoothing seen in Real-ESRGAN. AESRGANs use of attention mechanisms appears to help it adaptively enhance complex textures without introducing harsh artifacts. While some artifacting was still observed around challenging areas like the eyes, the overall balance between realism, clarity, and detail was rated highest by the panel. **Bicubic Interpolation:** Included as a baseline, bicubic interpolation scored the lowest across all categories, visual appeal (2.5), clarity (2.8), and detail preservation (2.2). Evaluators consistently found the images blurry, lacking in sharpness, and unable to restore any meaningful detail. While it maintained the overall structure of faces, it failed to recover textures, making it unsuitable for perceptually convincing super-resolution.

Image Set 1: Presents side-by-side comparisons of facial images processed using four different upscaling methods: Bicubic

Table 4 Presents the average subjective evaluation scores assigned by a panel of 10 human evaluators across 30 sets of facial images. Each modelESRGAN, Real-ESRGAN, AESRGAN, and Bicubic interpolationwas rated on a scale of 1 to 5 in three categories: visual appeal, clarity, and detail preservation. The scores reflect the mean values across all evaluations for each category.

Model	Visual Appeal	Clarity	Detail Preservation
ESRGAN	3.9	3.5	3.6
Real ESRGAN	4.5	4	4.5
Attention-ESRGAN	4.5	4.3	4.6
Bicubic	2.5	2.8	2.2



interpolation, ESRGAN, Real-ESRGAN, and AESRGAN. Each row in the set corresponds to a different facial image, allowing for a direct visual assessment of sharpness, artifact suppression, and detail preservation across models. This comparison highlights how each method performs under the same low-resolution input condition.

Challenges and Considerations

Potential Issues

The advanced features introduced by Real-ESRGAN and AESRGAN, such as attention mechanisms and enhanced feature extraction, significantly improve image quality but also lead to increased computational complexity. These enhancements demand higher processing power and longer training times, which can be a drawback, especially in resource-constrained environments or real-time applications. Balancing the perceptual quality improvements with the added computational overhead is a critical challenge. Furthermore, the increased computational burden may hinder the scalability of these models for large-scale deployments, such as in cloud-based or edge-computing scenarios, where efficiency is crucial. This paper evaluates whether the quality gains justify the increased resource requirements, especially in practical, real-world applications. Overfitting is another concern, particularly for models like AESRGAN that employ sophisticated attention mechanisms and deeper architectures. Overfitting can limit the model's generalizability to new data, reducing its effectiveness in diverse real-world scenarios. To mitigate this, careful model validation and the use of regularization techniques, such as dropout and data augmentation, are essential to ensure optimal performance across a wide range of image types and complexities. Additionally, employing transfer learning and fine-tuning with diverse datasets could help improve adaptability and minimize overfitting. Artifact handling remains an ongoing challenge across all three models, particularly in detailed or high-frequency image regions. While Real-ESRGAN and AESRGAN reduce artifacts more effectively than ESRGAN, they still struggle with specific complex features like eyes or fine textures, demonstrating that further innovations are needed to achieve artifact-free, high-fidelity reconstructions. Exploring advanced loss functions, such as perceptual and adversarial loss variations, and incorporating domain-specific priors could provide potential pathways to address these limitations.

Analysis and Conclusion

The comparative analysis of ESRGAN, Real-ESRGAN, and AESRGAN revealed that AESRGAN outperformed its counterparts across most evaluation metrics. AESRGANs explicit attention mechanisms enabled it to preserve intricate facial details and deliver more perceptually faithful results. This was confirmed by paired t-tests conducted on the trained dataset: AESRGAN significantly outperformed ESRGAN in LPIPS (t = 24.77, $p < 0.0001$), and also showed a statistically significant improvement over Real-ESRGAN (t = 6.29, $p < 0.0001$), indicating superior perceptual similarity to the ground truth. In terms of SSIM, all pairwise comparisons were statistically significant, including Real-ESRGAN vs. AESRGAN (t = -13.82, $p < 0.0001$), highlighting AESRGANs structural fidelity advantage. Although AESRGAN consistently achieved higher average PSNR scores, comparisons such as ESRGAN vs. Real-ESRGAN (t = -6.20, $p < 0.0001$) and Real-ESRGAN vs. AESRGAN (t = -11.92, $p < 0.0001$) showed statistical significance-contrary to earlier assumptions that pixel-level fidelity was not meaningfully different. These results underscore AESRGANs overall superiority across perceptual, structural, and pixel-wise metrics. Furthermore, all GAN-based models significantly outperformed bicubic interpolation across LPIPS, PSNR, and SSIM ($p < 0.0001$ for all comparisons), reinforcing the effectiveness of deep learning approaches for facial super-resolution. While AESRGAN leads in detail preservation, its higher computational

demands may not always justify the performance gains, depending on the application. In addition to the quantitative metrics, runtime efficiency and perceptual quality also played a significant role in evaluating each models practicality and user experience. Inference time measurements revealed that ESRGAN was the fastest model, with a mean processing time of 0.0780 seconds per image, followed by Real-ESRGAN at 0.0829 seconds and AESRGAN at 0.0882 seconds. Although AESRGAN had the longest runtime, the difference was marginaljust over 10 milliseconds slower than ESRGANsuggesting that its improved image quality does not come at the cost of significant latency. These findings affirm that all three GAN-based models are viable for real-time or near-real-time applications, with AESRGAN striking a strong balance between visual performance and computational cost. Qualitative assessments further supported these conclusions. Based on ratings from a panel of ten evaluators across 30 image sets, AESRGAN achieved the highest average scores in clarity (4.3) and detail preservation (4.6), while tying with Real-ESRGAN in visual appeal (4.5). Evaluators favored AESRGAN for its ability to maintain fine facial texturessuch as wrinkles, eyelashes, and skin gradientswithout introducing harsh artifacts. Real-ESRGAN followed closely behind, particularly excelling in smooth, natural textures but occasionally softening finer features. ESRGAN, though faster and sharper, was penalized for exaggerated edges and residual noise, while bicubic interpolation consistently ranked lowest across all categories. Taken together, these results confirm that AESRGAN not only leads in quantitative benchmarks but also delivers the most realistic and visually pleasing outputs in human-centric image enhancement tasks. Future work will focus on developing hybrid models that combine the strengths of existing super-resolution techniques while addressing their limitations, particularly in artifact reduction and computational efficiency. Additionally, we plan to evaluate the robustness of these models under adversarial attacks and perturbations, such as synthetic noise, blur, and compression artifacts. Conducting ablation studies will help isolate the contribution of individual componentslike attention modulesto the overall performance. It will also be important to test the models under a wider variety of real-world degradation types, including low-light conditions and sensor noise. Cross-dataset evaluations using facial and non-facial image domains will further inform how well these models generalize beyond their training data. Finally, we aim to explore the integration of advanced attention mechanisms such as transformer-based or lightweight spatial attention, and to investigate hybrid architectures that combine GANs, transformers, and diffusion models for enhanced performance and versatility.

References

1 X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao and C. C. Loy, Proc. European Conference on Computer Vision (ECCV) Workshops, 2018, pp. 0–0.

2 C. Dong, C. C. Loy, K. He and X. Tang, *Image super-resolution using deep convolutional networks*, https://arxiv.org/abs/1501.00092, 2015, arXiv:1501.00092.

3 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, *Attention is all you need*, https://arxiv.org/abs/1706.03762, 2017, arXiv:1706.03762.

4 Xinntao, *Xinntao/Real-ESRGAN: Real-ESRGAN aims at developing practical algorithms for general image/video restoration*, https://github.com/xinntao/Real-ESRGAN, 2024, GitHub repository.

5 J. Hu, L. Shen and G. Sun, *Squeeze-and-excitation networks*, https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html, 2018, CVPR Open Access.

6 Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong and Y. Fu, Proc. European Conference on Computer Vision (ECCV), 2018, pp. 286–301.

7 Mindspore-Lab, *Mindediting/docs/rrdb.md at master*, https://github.com/mindspore-lab/mindediting/blob/master/docs/rrdb.md, 2023, GitHub documentation.

8 J. Liang *et al.*, *SwinIR: Image Restoration Using Swin Transformer*, https://arxiv.org/abs/2108.10257, 2021, arXiv:2108.10257.

9 C. Saharia, *SR3: Image Super-Resolution via Iterative Refinement*, https://iterative-refinement.github.io/, Accessed: 2025-03-30.

10 Z. Wei, S. Gu and Z. Zhang, *A-ESRGAN: Training real-world blind super-resolution with attention U-Net discriminators*, https://arxiv.org/abs/2112.10046, 2021, arXiv:2112.10046.

11 J. Liu, X. Li, Y. Zhou and M. Wang, *IEEE Xplore*, 2024.

12 X. Wang, K. Yu, C. Dong and C. C. Loy, *Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data*, https://arxiv.org/abs/2107.10833, 2021, arXiv:2107.10833.

13 G. Soldatov, *Flickr-Faces-HQ (FFHQ) small*, https://www.kaggle.com/datasets/tommykamaz/faces-dataset-small?select=faces_dataset_small, 2022, Kaggle dataset.

14 NiveditJain, *Human Faces Dataset*, https://www.kaggle.com/datasets/niveditjain/human-faces-dataset, 2021, Kaggle dataset.