A Survey on Image Quality Enhancement using Real-ESRGAN: A Comprehensive Review of Datasets and Approaches

Sarthak Bharambe, Saili Sable, Yash Chavan, Ayush Rai, Atharv Mahajan
Computer Science and Engineering Nutan College of Engineering and Research Pune, India

Abstract— **Rapid developments in image restoration techniques have helped to reduce noise, compression artifacts, and low-resolution images challenges. Real-ESRGAN extends ESRGAN on practical applications in real-world practice by incorporating a high-order degradation model, synthetic training pipelines, and advanced architectures of discriminator. This article reviews its methodologies, provides comparisons of its performance using state-of-the-art techniques, and surveys recent related literature. Its applications, comparative strengths, and future directions are highlighted to bridge existing gaps in the field.**

Keywords— Image Enhancement, Super-Resolution, Real- ESRGAN, ESRGAN, JPEG Artifacts, REAL-ESRGAN, PSNR, SSIM.

I. INTRODUCTION

Advanced imaging technologies and the ubiquitous proliferation of multimedia applications have created an enormous demand for high-quality image restoration and enhancement techniques. Traditional methods that have been in use since decades include interpolation and histogram equalization, which, however, have limitations to deal with complex degradations like noise, blur, and compression artifacts. Deep learning in recent years has given the field a revolutionary leap that not only provides solutions with efficiency but also produces results that are visually appealing and realistic.

The first early models that have been used to set the stage for deep learning in super-resolution are SRCNN (Super- Resolution Convolutional Neural Network) and SRGAN (Super-Resolution GAN), which introduce convolutional networks and adversarial losses, respectively. These models are efficient for enhancing synthetic datasets but do not work well when texture consistency is concerned or in adapting to real-world degradations of images. ESRGAN, an enhanced version of SRGAN, introduced perceptual losses and improved GAN architectures, providing sharper and more natural results. However, its applicability to real-world images was limited due to the inherent challenges of unpredictable noise and compression artifacts in real scenarios. To bridge this gap, real-world ESRGAN was developed for real-world applications by incorporating high- order degradation models and synthetic data generation during training.

Real-ESRGAN is a new state-of-the-art evolution in the domain of image enhancement, simulating complex real- world degradation patterns and addressing the limitations of prior models. Its architecture relies on a combination of residual-in-residual dense blocks, U-Net discriminators, and synthetic pipelines to achieve unprecedented performance in artifact removal and detail restoration. By generating various training pairs that mimic real scenarios, Real-ESRGAN not only bridges the gap between synthetic and real datasets but also proves robust generalization across all domains, including digital photography, medical imaging, and content restoration. This paper digs into the key methodologies involved, comparative strengths, and future possibilities of Real-ESRGAN, highlighting its transformative potential for real-world applications.

II. LITERATURE SURVEY

These papers collectively strengthen and enhance image enhancement models, especially for practical applicability in real-world examples, by focusing on issues such as noise, artifacts due to compression, or resolution limitations that synthetic training data and adaptive models can effectively help address.

[1] Real-ESRGAN: Training Real-World Blind Super- Resolution with Pure Synthetic Data

Focus: This paper introduces Real-ESRGAN: improving image restoration by simulation of real-world degradations with synthetic data, such as noise, blur, and compression. Real-ESRGAN relies on a high-order degradation model, which enables the model to handle a wide variety of real- world image distortions. By training on synthetic data that closely mimics real-world conditions, it generalizes better to unknown degradation scenarios.

[2] StarSRGAN: Adaptive Models for Better Quality Images

Focus: StarSRGAN investigates adaptive super-resolution of images with a scalable architecture, bringing better generalization and robustness to real-world applications. An adaptive degradation model is integrated for varied degradation levels across the images, thus improving the model's performance in different environments. A multi-level residual network is used to further enhance perceptual quality.

[3] Applications in Object Detection

Focus: This experiment aims to combine Real-ESRGAN with various object detection frameworks to get improved performance on low resolution and degraded surveillance footage. Enhance the quality of the image using Real- ESRGAN for boosting up accuracy on object detection algorithms and proving dual advantages of increasing resolution and object recognition.

[4] GANs in Medical Imaging

Focus: This paper demonstrates the application of Real- ESRGAN for improving the quality of medical images, such as MRI scans and X-rays, leading to clearer diagnostic details. GAN is used for noise removal of medical images for clear diagnostic purposes of disease identification, having higher resolution than previously enhanced images. As such real-ESRGAN takes place that addresses and captures real noise for real data degradation for improvement.

[5] Font Reconstruction

Focus: Real-ESRGAN is applied to high-resolution font creation and reconstruction, particularly for digital typography. Because of the capacity of Real-ESRGAN to make low-resolution images more text sharp and clear, this method has great application for the generation and restoration of fonts that are essential for digital design and printing.

[6] A-ESRGAN model achieves better perceptual quality

Focus: A-ESRGAN is an enhanced version of the ESRGAN that introduces dropout degradation and adaptive noise modeling. This approach aims to enhance the visual quality and sharpness of restored images by simulating more realistic real-world noise patterns using dropout techniques, thereby enhancing the perceptual quality of the output.

[7] Cross-Modal Learning for Super-Resolution

Focus: This paper really looks into how Real-ESRGAN is integrated into cross-modal learning, integrating super- resolution techniques with semantic tasks such as segmentation or recognition. Improving the resolution of images and combining this with semantic understanding, this approach can dramatically improve recognition of objects and other related tasks that rely heavily on high-quality visual data.

[8] High-Order Modeling in Image Restoration

Focus: This paper provides insight into the high-order degradation models that Real-ESRGAN employs, showcasing how such models are indeed effective for simulating patterns of degradation found in reality. Real- ESRGAN improves over degradation models by including more complex, multi-step processes, like random resizing, several blur operations, and Poisson noise, thus it is able to cover a much wider range of distortions that occur in the real world and improve results for image restoration.

III. RELATED WORK

These models include SRCNN and SRGAN which were the pioneer of deep learning based super-resolution methods based on the convolutional methodology and generative adversarial loss respectively. These models, however, were challenged by real world artifacts such as noise and compression. Following these lines, ESRGAN improved perceptual quality with deeper residual networks, but this approach also led to difficulty in generalization in the real world. Real-ESRGAN addresses these limitations through the introduction of a high-order degradation model that better approximates real-world image distortions, leading to more effective restoration in real-world scenarios.

SRGAN (Super-Resolution Generative Adversarial Network):

Proposed the use of GAN for the first time for super- resolution, providing significantly improved perceptual quality in image restoration. But it pulled some hiccups in texture details, and real-world image artifacts. Real- ESRGAN is an improvement over SRGAN that adopts the high-order degradation models to engage much more complex real-world degradation rules. This makes Real- ESRGAN better suited for practical purposes as it can better handle noise, compression artifacts, and blurry areas. Real- ESRGAN achieves a simulated degradation of various processes such as noise, blur, and compression noise, which boosts the model's capabilities in restoring high-fidelity details from low-resolution images.

ESRGAN (Enhanced Super-Resolution GAN):

Modeling on top of SRGAN by residual-in-residual dense blocks and perceptual loss functions, providing great texture recovery and sharp images. However, ESRGAN was trained primarily on synthetic datasets such that its generalization to realistic conditions could not be such a good performer. It is well established regarding Real-ESRGAN's endogenous synthetic data modeling of a degradation process to overcome such limitations for just such cases.

EDSR (Enhanced Deep Super-Resolution Network):

Concentrated on increasing the depth and efficiency of neural networks for super-resolution without adversarial training. In spite of EDSR performing well on high-resolution images, it cannot quite handle real-world noise and compression artifacts, something that Real-ESRGAN manages to overcome using its degradation models.

CycleGAN and Pix2Pix:

These models laid the groundwork for image-to-image translation tasks, including style transfer and image restoration, but they largely sidestepped super-resolution. Real-ESRGAN, while based on the GAN framework, is uniquely bespoke for the tasks of image restoration and enhancement within real-world degradation scenarios, such as JPEG artifacts and noise.

Different from CycleGAN and Pix2Pix, which are primarily developed to perform image synthesis and modification, Real-ESRGAN, on the contrary, would serve super- resolution. It is thus capable of restoring fine details to improve image quality without destroying their structural integrity. It can therefore handle more complicated real- world distortions by adding high-order degradation models, hence providing more robustness for practical image enhancement tasks. This renders Real-ESRGAN a much more versatile tool in digital media applications, medical imaging, and surveillance, where high resolution and clarity are critical factors.

IV. COMPARATIVE ANALYSIS

Model	Study	Approach	Dataset	Key Metric	Limitations
SRCNN	Early deep learning-based super-resolution	Three-layer CNN	ImageNet	PSNR 26.98, NIQE 7.12	Limited texture recovery
SRGAN	First GAN-based perceptual improvement	Adversarial loss	DIV2K	PSNR 28.12, NIQE 6.77	Struggled with texture consistency
ESRGAN	Enhanced SRGAN	Residual-in-residual dense blocks	DIV2K, Flickr	PSNR 29.45, NIQE 5.74	Limited to synthetic datasets
RealSR	Domain-specific training for real datasets	Specialized training	RealSR Canon/Nikon	PSNR 30.12, NIQE 6.52	Dataset-dependen t
Real-ESRGAN	High-order	Synthetic	Synthetic datasets	PSNR 31.98,	Requires high

	degradation modeling	data pipeline		NIQE 4.59	computational power
Real-ESRGAN+	Enhanced model with sharpened ground truths	Advanced degradation pipeline	Mixed synthetic datasets	PSNR 32.45, NIQE 4.53	Computationally intensive
EDSR	Enhanced Deep Super-Resolution Network	Improved residual block architecture	DIV2K	PSNR 31.02, NIQE 5.87	Higher memory requirements
LapSRN	Laplacian Pyramid Super-Resolution Network	Progressive upscaling	General-purpose datasets	PSNR 30.41, NIQE 6.23	Limited performance on highly degraded images
TSRN	Text Super-Resolution Network	Attention-based architecture	Synthetic text datasets	Improved OCR accuracy	Optimized only for text super-resolution
DUF	Deep Video Upsampling Super-Resolution Framework	Temporal dependency modeling	Video datasets	Improved temporal consistency	High computational complexity for videos
VDSR	Very Deep Super-Resolution Network	Residual learning	General-purpose datasets	PSNR 30.12, NIQE 6.30	Limited to single image super-resolution
TTSR	Textural Transfer Super-Resolution	Texture transfer mechanism	Specific texture datasets	Improved texture recovery	Limited to texture-focused tasks

Discussion:

Real-ESRGAN is already a radical improvement on the earlier models like SRCNN, SRGAN, and ESRGAN, which dealt with a few shortcomings. While earlier models laid the foundation for deep learning-based super-resolution, they could hardly tackle real-world hindrances such as artifacts caused by compression, noise, and blur. SRGAN, using adversarial loss to aid perceptual improvements, was ineffective in restoring fine detail consistently in real images. Following this, ESRGAN improved on the original with more advanced architecture but was still held back by its training on synthetic data, which limited its ability to generalize to real-world cases.

Reasonable image quality can at least be obtained under the given circumstances when the real-world degradation model is introduced and high-order degradation modeling with synthetic training data is practiced. Conditioned upon the degradation processes like noise, blurring, and compression artifacts, a model handles diverse distortions better in practical applications. Generalization to real-world domains allows the applicability of industry standards ranging from digital media, medical imaging to surveillance. Advantages include tile inference, 16-bit image support, and alpha channel processing. Real-ESRGAN provides a versatile, high-quality solution for image restoration that surpasses its predecessors in both performance and applications.

V. FUTURE DIRECTION

As the adoption of cloud computing grows, realistic advancements in anomaly detection are vital to ensure security while keeping solutions practical, scalable, and accessible. Below are more grounded future directions that aim to address real-world challenges in cloud security without over-reliance on cutting-edge, experimental technologies.

1. Real-Time Optimization

Put the Real-ESRGAN going for online web applications subject to the quick inference requirements for video processing in mobile devices or other embedded devices. It will work along the lines of reducing the model complexity with high levels of quality. Pruning of models, quantization, and knowledge distillation are all techniques that can be used to achieve this trade-off. Real-time optimization would enable the model to handle large datasets or video streams and could be tailored toward applications in gaming, TV broadcasting, or autonomous vehicles.

2. Cross-Modal Integration

Integrating super-resolution with object detection and semantic segmentation could provide utility in better performance forecasting in the domains of surveillance and autonomous systems. Taking into account that super- resolution can furnish images with higher resolution, thus assisting downstream tasks in extracting more relevant information from images, this can help object recognition in low-resolution surveillance videos or help identify little but important features in an autonomous system. The possibilities of integration to be extended to multimodal systems may assist visual and sensor data in robust decision- making processes in the real world.

3. Dataset Expansion

Create additional diverse training datasets with realistic degradations to enhance the generalization capabilities of the model to encounter rare and unseen artifacts in practical scenarios. This means that a whole spectrum of real-world conditions must be simulated, from more complex noise patterns to variable lighting conditions and miscellaneous types of compression artifacts. It would be even more effective to combine this with specific domain datasets, like medical imaging scans or satellite photos, thus improving the model's adaptability to certain industrial concerns. But besides, alongside real-life examples, such synthetic and augmented data can be added to provide a well-rounded training set. There are also dynamic methods that generate datasets that can further facilitate injecting the training set with new degradation patterns, which helps keep the model robust over time.

4. Enhancing Visual Fidelity

Enhancing the value of perceptual quality emphasizes high- precision digital media applications in cinematographic content and professional design. The improvements could include things like the tile generation and edge preservation, which really do an incredible job at providing a visually appealing and believable output. Advanced loss functions such as perceptual and contextual losses can be used to align the model's output to human visual preferences, guaranteeing quality results in critical applications such as film production, photography, and digital artwork.

VI. CONCLUSION

Real-ESRGAN takes image enhancement to the next level by fixing some of the shortcomings in earlier models like SRCNN, SRGAN, and ESRGAN. While these approaches improved perceptual quality, they struggled when faced with real-world types of degradation such as noise and compression artifacts. Real-ESRGAN has also incorporated high-order degradation modelling to simulate intricate kinds of distortions such as blur, noise, and JPEG artifacts. This helps the model generalize better in real-world scenarios to produce more accurate restorations. Along with features like tile inference and support for 16-bit images, Real-ESRGAN is diverse enough for applications across digital media, healthcare, and surveillance. It is powerful for practical image enhancement by effectively dealing with large images in real-life imperfections. Research in the future will be focused on optimization for efficiency and expansion of datasets for broader generalization of techniques.

VII. REFERENCES

[1] Wang, X., Yu, K., & Yang, M., "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data," arXiv, 2021. This paper introduces Real-ESRGAN, which incorporates high-order degradation models for handling real-world distortions and enhances image restoration through synthetic data training.

[2] Zhang, Y., Liu, Z., & Xu, Z., "StarSRGAN: Adaptive Models for Enhanced Image Quality," arXiv, 2023. Introduces StarSRGAN, an architecture with adaptive degradation models and scalable design to handle varying distortion levels in image restoration tasks.

[3] Ledig, C., Theis, L., & Huszár, F., "Photo-Realistic Single Image Super-Resolution Using GANs," CVPR, 2017. SRGAN, the first GAN-based model for super- resolution, introduces adversarial and perceptual losses for realistic texture restoration in low-resolution images.

[4] Wang, X., Zhang, J., & Song, Z., "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," ECCV, 2018. ESRGAN improves SRGAN with residual-in-residual dense blocks and perceptual losses, producing sharper images with better texture

fidelity.
[5] Wei, A., Liu, S., & Ma, F., "A-ESRGAN: Adaptive ESRGAN for Enhanced Perceptual Quality," Springer, 2023. A-ESRGAN incorporates dropout degradation and adaptive noise modeling to improve perceptual quality in real-world image super-resolution tasks.
[6] Lv, Z., Zhang, L., & Sun, Y., "Enhancing Object Detection with Real-ESRGAN," IEEE, 2023.
This paper demonstrates how Real-ESRGAN can improve object detection performance by enhancing the resolution of low-quality surveillance footage.
[7] Zhao, W., Liu, J., & He, M., "Deep Learning for Blind Image Super-Resolution," IEEE Access, 2021. Reviews methods for blind super-resolution, highlighting Real-ESRGAN's ability to restore fine details in images affected by unpredictable noise and distortions.
[8] Shi, W., Caballero, J., & Yang, H., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel CNN," CVPR, 2016. Introduces an efficient sub-pixel CNN for real-time image and video super-resolution, offering ideas for improving the computational efficiency of Real- ESRGAN.
[9] Ronneberger, O., Fischer, P., & Brox, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," MICCAI, 2015. U-Net's encoder-decoder architecture, used in Real-ESRGAN's discriminator network, plays a critical role in stabilizing adversarial training and improving output quality.
[10] Odena, A., Olah, C., & Shlens, J., "Conditional Image Synthesis with Auxiliary Classifier GANs," ICML, 2017. Introduces AC-GAN, which uses auxiliary classifiers to improve image synthesis quality, influencing feature extraction in super-resolution tasks like Real-ESRGAN.
[11] Luo, J., Li, Y., & Jiang, Z., "Synthetic Data for Super- Resolution Models," Springer,2022.Discusses how synthetic datasets help improve the generalization of super-resolution models, particularly in tasks where real-world data is limited.
[12] Tian, C., Wang, X., & Liu, B., "Deep Learning Models for Image Denoising and Enhancement," Neurocomputing, 2020. Explores various deep learning methods for denoising and enhancing image quality, providing foundational techniques used in Real-ESRGAN for noise reduction and artifact removal.
[13] Ronneberger, O. et al., "U-Net: Convolutional Networks for Biomedical Image Segmentation," MICCAI, 2015. U-Net architecture serves as the basis for Real-ESRGAN's discriminator, stabilizing adversarial training and improving output quality.