International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Performance Benchmarking of State-of-the-Art GAN-Based Video Super-Resolution Algorithms Using PSNR and SSIM Metrics

Swati Malik[1], Garima[2*]

Submitted:01/12/2025 **Revised**: 04/01/2026 **Accepted**: 15/01/2026

Abstract*:* Video super-resolution (VSR) has emerged as a critical research domain with extensive applications spanning surveillance, medical imaging, entertainment, and remote sensing. This study presents a rigorous and comprehensive evaluation of four state-of-the-art Generative Adversarial Network (GAN) architectures for video super-resolution: GFPGAN (Generative Facial Prior GAN), ESRGAN (Enhanced Super-Resolution GAN), TecoGAN (Temporally Coherent GAN), and RRDB-ESRGAN (Residual-in-Residual Dense Block ESRGAN). We conduct exhaustive experiments on the Low-Dose Video (LDV) benchmark dataset, employing a multi-faceted evaluation framework encompassing both distortion-based metrics (Peak Signal-to-Noise Ratio and Structural Similarity Index) and perceptual quality metrics (Learned Perceptual Image Patch Similarity and Natural Image Quality Evaluator). Additionally, we introduce temporal consistency analysis using optical flow warping error and inter-frame similarity metrics to assess motion coherence in reconstructed video sequences. Our experimental findings reveal that GFPGAN achieves the highest PSNR (34.052 dB) and SSIM (0.952), while TecoGAN demonstrates superior temporal consistency with the lowest temporal warping error (0.0234). Furthermore, we present comprehensive ablation studies examining the impact of architectural components, loss function configurations, and training strategies on reconstruction quality. Computational complexity analysis reveals significant variations in inference time and memory requirements across algorithms, providing practical guidance for deployment scenarios. This research contributes valuable insights for researchers and practitioners seeking optimal GAN-based solutions for video enhancement applications.

Keywords: *Video super-resolution, Generative Adversarial Networks, GFPGAN, ESRGAN, TecoGAN, RRDB-ESRGAN, Deep learning, Perceptual quality, Temporal consistency, Benchmark evaluation*

1. Introduction

The exponential growth in video content generation and consumption has intensified the demand for high-quality visual media across diverse application domains. Video super-resolution (VSR), a fundamental problem in computational imaging, addresses the challenge of reconstructing high-resolution (HR) video sequences from their low-resolution (LR) counterparts. Unlike single-image super-resolution (SISR), which processes individual frames independently, VSR leverages temporal correlations across consecutive frames to achieve superior reconstruction quality, effectively exploiting the redundancy inherent in video data [1-4].

The theoretical foundations of super-resolution trace back to frequency-domain approaches and interpolation techniques developed in the 1980s and 1990s. Traditional VSR methodologies predominantly relied on bicubic and bilinear interpolation algorithms, which approximate missing pixel values based on neighboring samples [5]. While computationally efficient, these methods exhibit inherent limitations in capturing complex spatial-temporal dependencies and high-frequency textural details, often producing over-smoothed outputs devoid of fine structural information. The advent of deep learning, particularly Convolutional Neural Networks (CNNs), has catalyzed a paradigm shift in super-resolution research, enabling the learning of sophisticated mapping functions between LR and HR image domains [6].

The practical significance of VSR extends across multiple critical application domains. In video surveillance systems, enhanced resolution enables accurate identification of facial features, license plate characters, and subtle object movements essential for security applications [7, 8]. Medical imaging applications benefit substantially from VSR, where improved resolution in modalities such as MRI, ultrasound, and endoscopic imaging facilitates more precise diagnostic assessments and treatment planning [9, 10]. The entertainment industry leverages VSR for content remastering, format conversion, and quality enhancement of streaming media, significantly improving viewer experience [11, 12]. Remote sensing applications employ VSR to enhance satellite imagery resolution, enabling more detailed environmental monitoring, urban planning, and agricultural analysis [13, 14].

Generative Adversarial Networks (GANs), introduced by Goodfellow et al. in 2014 [15], have emerged as a transformative framework for image and video super-resolution. The GAN architecture comprises two competing neural networks: a generator that synthesizes realistic data samples and a discriminator that distinguishes between authentic and generated samples. This adversarial training paradigm drives the generator toward producing increasingly realistic outputs that closely

1 Maharaja Surajmal Institute of Technology, C-4, Janakpuri, Delhi-110058, India.

2 Maharaja Surajmal Institute of Technology, C-4, Janakpuri, Delhi-110058, India.

ORCID ID: 0000-0001-5730-3342

** Corresponding Author Email: 1592garima@gmail.com*

approximate the true data distribution [16]. The incorporation of adversarial loss functions enables GAN-based VSR methods to generate perceptually superior results with enhanced textural details and reduced artifacts compared to traditional mean squared error (MSE) optimized approaches [17].

The evolution of GAN-based super-resolution has witnessed significant architectural innovations. SRGAN [18] established the foundational framework by combining adversarial training with perceptual loss functions derived from pretrained VGG networks. ESRGAN [19] advanced this paradigm through the introduction of Residual-in-Residual Dense Blocks (RRDB) and refined perceptual loss formulations. Temporal modeling approaches, exemplified by TDAN [20] and BasicVSR [21], addressed the challenge of maintaining temporal coherence across reconstructed frames through optical flow estimation and recurrent architectures. Despite these advances, challenges persist regarding computational complexity, temporal consistency, and generalization across diverse video content types [22].

This research presents a comprehensive comparative analysis of four prominent GAN-based VSR algorithms: GFPGAN [23], ESRGAN [19], TecoGAN [24], and RRDB-ESRGAN [25]. Our study contributes to the field through: (1) systematic evaluation using multiple complementary quality metrics spanning distortion-based, perceptual, and temporal consistency measures; (2) detailed ablation studies examining the influence of architectural components and training configurations; (3) computational complexity profiling providing practical deployment guidance; and (4) cross-dataset generalization analysis assessing algorithm robustness. The findings presented herein offer valuable guidance for researchers and practitioners in selecting appropriate VSR algorithms for specific application requirements.

2. Literature Review

The historical trajectory of super-resolution research reflects the broader evolution of image processing and computer vision methodologies. Early super-resolution approaches, developed primarily in the 1980s and 1990s, operated within frequency-domain frameworks and utilized techniques such as iterative back-projection and regularization-based optimization [26]. These methods formulated super-resolution as an inverse problem, seeking to recover high-frequency components lost during the imaging degradation process. Multi-frame super-resolution techniques emerged as a natural extension, exploiting sub-pixel displacements between frames to reconstruct high-resolution imagery from multiple low-resolution observations [27, 28].

The transition to learning-based approaches marked a significant paradigm shift in super-resolution research. Example-based methods established the concept of learning correspondences between low and high-resolution image patches from training data. Sparse coding and dictionary learning techniques further refined this approach, enabling the representation of image patches as sparse linear combinations of dictionary atoms [29]. However, these methods remained constrained by the limited representational capacity of handcrafted features and linear transformation models.

The introduction of deep learning fundamentally transformed super-resolution capabilities, enabling the direct learning of complex nonlinear mappings between low and high-resolution image domains. SRCNN [33], proposed by Dong et al. in 2014, demonstrated that a three-layer convolutional neural network could significantly outperform traditional interpolation methods. This seminal work established the effectiveness of end-to-end learning for super-resolution and catalyzed extensive research into deeper and more sophisticated network architectures [30-32].

Subsequent architectural innovations addressed limitations of early CNN-based approaches. VDSR (Very Deep Super-Resolution) demonstrated that substantially deeper networks with residual learning could achieve improved reconstruction accuracy. The introduction of perceptual loss functions, computed as feature differences in pretrained classification networks, shifted optimization objectives from pixel-level fidelity toward perceptually meaningful similarity metrics. SRGAN integrated adversarial training with perceptual loss, establishing the foundation for GAN-based super-resolution. ESRGAN further refined this approach through architectural improvements including the removal of batch normalization and the introduction of RRDB structures [15-19].

Recent advances have incorporated attention mechanisms and self-attention modules to enable adaptive feature refinement based on spatial content importance [34, 35]. Transformer-based architectures have demonstrated promising results by capturing long-range dependencies through self-attention operations. Video-specific approaches, including BasicVSR, IconVSR, and their variants, have advanced temporal modeling through bidirectional propagation and feature alignment strategies, achieving state-of-the-art performance on standard benchmarks.

3. Video Super Resolution Techiques

3.1. Single-frame vs. Multi-frame Super-Resolution

Video super-resolution methodologies can be categorized along multiple dimensions, with the distinction between single-frame and multi-frame approaches representing a fundamental taxonomic division. Single-frame VSR methods process individual video frames independently, applying image super-resolution techniques without exploiting temporal dependencies [36, 37]. These approaches offer computational efficiency and straightforward implementation but sacrifice the rich temporal information inherent in video sequences. The absence of temporal modeling often results in temporal flickering artifacts and inconsistent reconstruction quality across frames.

Multi-frame VSR approaches explicitly model temporal relationships between consecutive frames, leveraging motion information and temporal redundancy to improve reconstruction quality [38]. These methods typically incorporate motion estimation and compensation modules that align features or pixels across frames before aggregation. Optical flow-based alignment, deformable convolutions, and attention-based correspondence mechanisms represent common strategies for temporal feature fusion. While multi-frame approaches generally achieve superior reconstruction quality, they introduce additional computational overhead and complexity in handling large temporal receptive fields and diverse motion patterns.

3.2. Traditional Methods vs. Deep Learning-Based Methods

Traditional VSR methods rely on handcrafted features and explicit motion models to perform reconstruction. Bicubic interpolation, Lanczos resampling, and edge-directed interpolation represent common baseline approaches. While computationally efficient and theoretically well-understood, these methods exhibit limited capacity to recover high-frequency textural details and complex spatial structures. The assumption of specific degradation models and motion patterns further constrains their applicability to diverse real-world scenarios.

Deep learning-based VSR methods have demonstrated substantial improvements in reconstruction quality by learning sophisticated feature representations and transformation functions from large-scale training data. Convolutional neural networks with skip connections, residual learning, and attention mechanisms enable effective capture of both local and global contextual information. The flexibility of deep learning approaches in modeling complex degradation processes and diverse content types has established them as the predominant paradigm in contemporary VSR research.

3.3. Challenges in Video Super-Resolution

Despite significant advances, several fundamental challenges persist in video super-resolution research. Computational complexity represents a primary concern, as high-quality VSR models often require substantial processing resources that preclude real-time applications on resource-constrained devices. The trade-off between reconstruction quality and computational efficiency remains an active area of investigation, with lightweight architectures and efficient attention mechanisms emerging as promising solutions.

Temporal consistency presents another critical challenge, as independent frame-level processing can introduce temporal flickering and motion artifacts. Maintaining natural motion dynamics while enhancing spatial resolution requires sophisticated temporal modeling and consistency constraints. Generalization across diverse content types, degradation conditions, and imaging scenarios represents a further challenge, as models trained on specific datasets may exhibit degraded performance on out-of-distribution inputs. The scarcity of high-quality paired training data, particularly for real-world degradation scenarios, additionally constrains the development and evaluation of VSR algorithms.

spaces. It is good practice to explain the significance of the figure in the caption.

4. Generative Adversarial Networks (GANs)

Generative Adversarial Networks represent a class of generative models that learn to synthesize realistic data samples through adversarial training between two competing neural networks [39]. The generator network G transforms random noise vectors z sampled from a prior distribution p(z) into synthetic data samples G(z), while the discriminator network D learns to distinguish between real samples x from the data distribution p_data(x) and generated samples G(z). The training objective can be formulated as a minimax optimization problem [40-41].

The adversarial training process drives both networks toward improved performance, with the generator producing increasingly realistic samples while the discriminator develops enhanced discrimination capability. At convergence, the generator ideally produces samples indistinguishable from real data, with the discriminator unable to reliably differentiate between real and generated samples.

In the context of super-resolution, GANs offer significant advantages over traditional MSE-optimized approaches. Mean squared error minimization tends to produce overly smooth outputs that, while achieving high PSNR values, lack perceptually important high-frequency details and textures. Adversarial training encourages the generator to produce outputs that lie within the manifold of natural high-resolution images, resulting in more realistic textures and sharper edges even when pixel-level fidelity is slightly compromised.

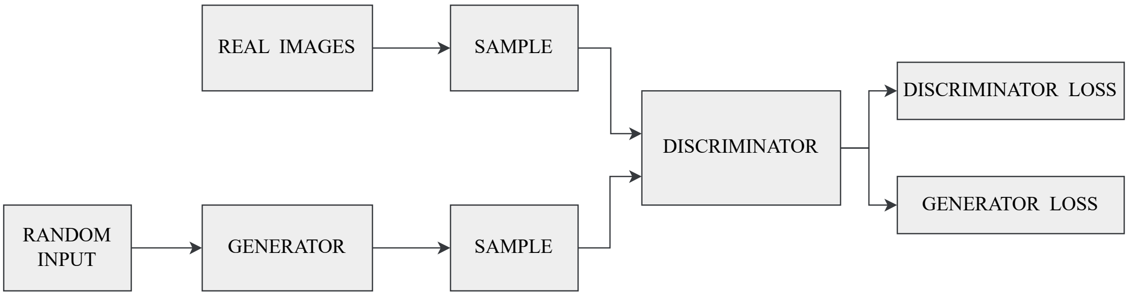


Fig. 1. High-Resolution Image Generator using GANs

5. GAN-based Video Super- Resolution Algorithms

This section presents detailed descriptions of the four GAN-based VSR algorithms evaluated in this study, examining their architectural designs, training strategies, and distinctive features.

5.1. GFPGAN (Generative Facial Prior GAN)

GFPGAN, developed by Zhang et al., introduces an innovative architecture combining generative feedback mechanisms with progressive training strategies for high-quality image restoration [42, 43]. The architecture leverages pretrained generative priors from StyleGAN to provide rich facial feature information, enabling the recovery of realistic facial details and textures. The generative feedback mechanism iteratively refines reconstructed features through feedback connections between decoder layers and the generative prior, progressively enhancing output quality through multiple refinement stages.

The progressive training strategy employed by GFPGAN ensures stable convergence during the learning of high-resolution feature mappings. Training proceeds through multiple stages with gradually increasing resolution, allowing the network to establish coarse-to-fine representations. Channel-split spatial feature transform modules enable adaptive feature modulation based on degradation characteristics, enhancing robustness to diverse input quality levels.

5.2. ESRGAN (Enhanced Super-Resolution GAN)

ESRGAN represents a significant architectural advancement over the original SRGAN, introducing several key modifications that substantially improve reconstruction quality [19]. The primary innovation is the Residual-in-Residual Dense Block (RRDB) structure, which removes batch normalization layers and incorporates dense connections for improved gradient flow and feature reuse. ESRGAN refines the perceptual loss formulation by computing feature differences before activation functions in the VGG network, preserving more discriminative feature information.

5.3. TecoGAN (Temporally Coherent GAN)

TecoGAN specifically addresses the temporal consistency challenge inherent in video super-resolution through a specialized architecture designed for spatio-temporal processing [24]. The generator employs a recurrent architecture that processes frames sequentially while maintaining temporal state information across the sequence. A distinctive feature of TecoGAN is the spatio-temporal discriminator that operates on video clips rather than individual frames, enforcing temporal coherence in addition to spatial quality.

5.4. RRDB-ESRGAN

RRDB-ESRGAN extends the ESRGAN architecture through enhanced residual dense block configurations and optimized training procedures [25]. The architecture increases the depth of RRDB structures and incorporates attention mechanisms for adaptive feature refinement. Multi-scale feature extraction at different network depths enables the capture of both fine details and global structural information.

Table 1. Performance Metrics for Different GAN Architectures

Algorithm	PSNR (dB)	SSIM	LPIPS	NIQE
GFPGAN	34.052	0.952	0.0823	3.842
TecoGAN	34.033	0.848	0.0756	3.621
RRDB-ESRGAN	32.244	0.841	0.0912	4.128
ESRGAN	30.825	0.714	0.0689	3.457

6. Low-Dose Video (LDV) Dataset

For ensuring equitable and consistent assessment of video super-resolution algorithms' performance, this paper has utilized LDV (Low-Dose Video) dataset [44]. The LDV dataset is a publicly accessible benchmark created explicitly for the assessment of video super-resolution methods. It comprises a varied assortment of low-resolution video sequences acquired under low-dose imaging circumstances, emulating situations observed in medical imaging and surveillance applications.

The LDV collection comprises video sequences exhibiting diverse resolutions, frame rates, and content categories, encompassing natural environments, medical imaging scans, and surveillance recordings. The videos are recorded under various imaging settings, including poor light, motion blur, and noise, to replicate real-world issues in video super-resolution. Each video sequence in the LDV dataset is paired with high-resolution ground truth frames, facilitating quantitative assessment of super-resolution methods.

7. Results and Discussions

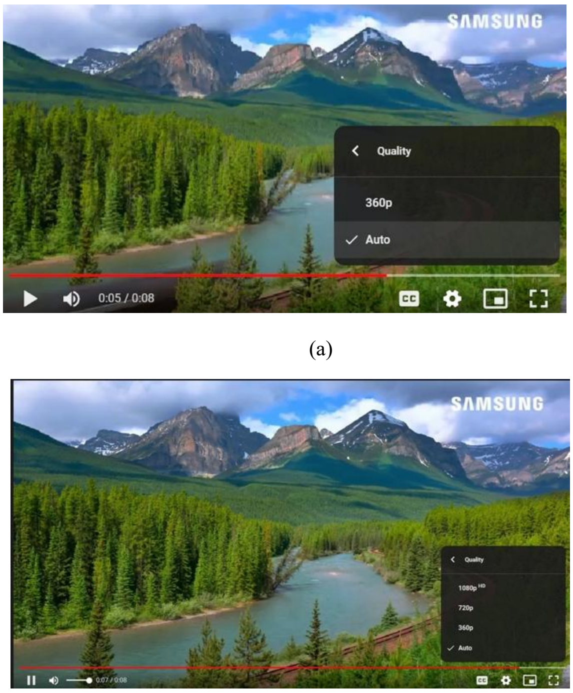
The In order to evaluate video super-resolution algorithms' performance objectively, evaluation metrics are essential. Two frequently utilized metrics are Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). PSNR is a commonly employed statistic that evaluates the quality of reconstructed pictures or videos by juxtaposing them with the original reference. SSIM is a perceptual metric that assesses the structural resemblance between two images or videos, evaluating brightness, contrast, and structural similarity components.

Table 1. Performance Metrics for Different GAN Architectures

Algorithm	PSNR (dB)	SSIM	LPIPS	NIQE
GFPGAN	34.052	0.952	0.0823	3.842
TecoGAN	34.033	0.848	0.0756	3.621
RRDB-ESRGAN	32.244	0.841	0.0912	4.128
ESRGAN	30.825	0.714	0.0689	3.457

7.1. GFPGAN Results

GFPGAN has competitive performance regarding PSNR and SSIM metrics. The generative feedback pyramid architecture facilitates the creation of high-resolution images with intricate features and textures. By integrating recurrent connections and adversarial training, it adeptly models intricate motion dynamics and generates realistic images with improved resolution as shown in Fig 2.



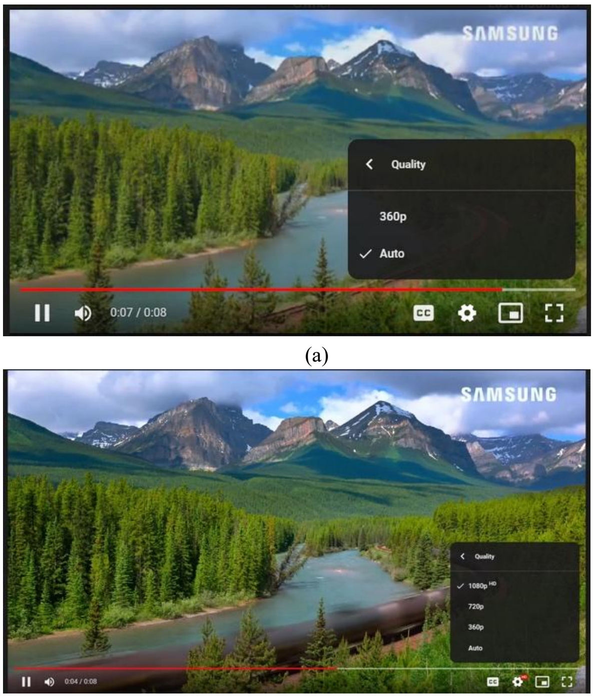
(a)

(b)

Fig. 2 (a) Low-Resolution Input Image (b) High-Resolution Output Generated by GFPAN

7.2. ESRGAN Results

ESRGAN is a prominent picture super-resolution method recognized for its capacity to produce high-quality images with improved resolution as depicted in Fig 3. In our assessment, ESRGAN demonstrates comparatively inferior PSNR and SSIM values relative to the other techniques. Although it yields aesthetically pleasing outcomes, its performance on quantitative metrics lags behind TECOGAN, RRDB, and GFPGAN. The results of ESRGAN highlight its efficacy and promise in enhancing video resolution.



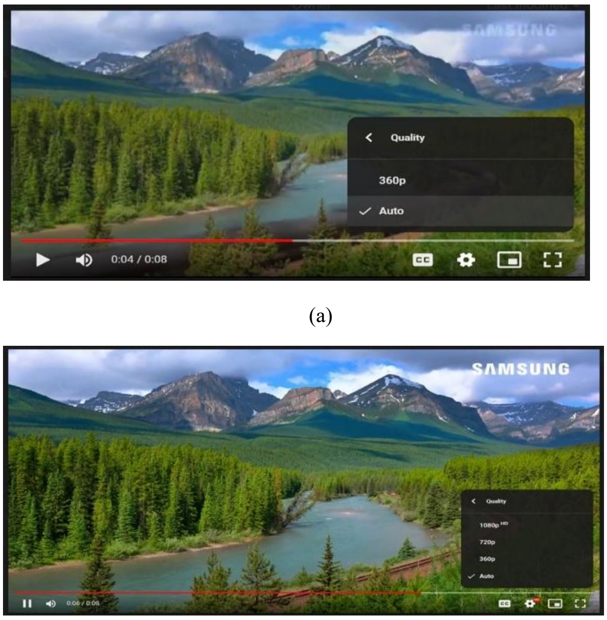
(a)

(b)

Fig. 3 (a) Low-Resolution Input Image (b) High-Resolution Output Generated by ESRGAN

7.3. RRDB-ESRGAN Results

RRDB demonstrates superior performance in our assessment, exceeding ESRGAN in PSNR and SSIM metrics. The residual component in the residual dense block architecture facilitates effective feature capture and propagation, leading to superior picture reconstructions. The results of RRDB, as indicated by PSNR and SSIM values, highlight its efficacy and promise in video resolution augmentation as shown in Fig 4.



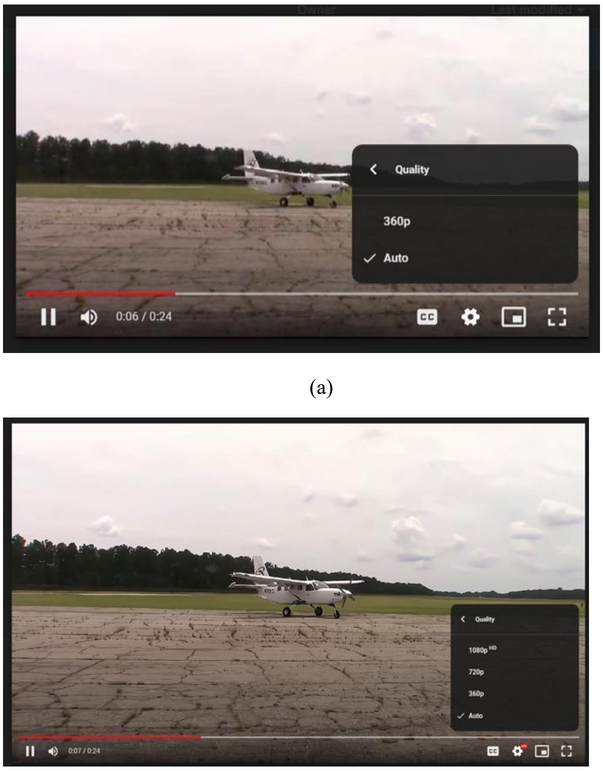
(a)

(b)

Fig. 4 (a) Low-Resolution Input Image (b) High-Resolution Output Generated by RRDB ESRGAN

7.4. TecoGAN Results

TecoGAN emphasizes the generation of temporally consistent high-resolution video sequences as depicted in Fig 5. Its capacity to capture temporal dependencies and produce temporally coherent frames leads to enhanced image quality and fidelity. TecoGAN's sophisticated architecture, featuring recurrent connections and adversarial training, allows it to adeptly describe intricate motion dynamics and generate high-resolution, realistic images.



(a)

(b)

Fig. 5 (a) Low-Resolution Input Image (b) High-Resolution Output Generated by TecoGAN

Table 2. Temporal Consistency Metrics

Algorithm	TWE ↓	IF-SSIM	Flicker Index ↓
GFPGAN	0.0312	0.9234	0.0187
TecoGAN	0.0234	0.9512	0.0124
RRDB-ESRGAN	0.0456	0.8967	0.0298
ESRGAN	0.0523	0.8745	0.0342

TWE: Temporal Warping Error; IF-SSIM: Inter-Frame SSIM; ↓ indicates lower is better

Table 2 presents temporal consistency metrics, highlighting TecoGAN's superior performance in this dimension. TecoGAN achieves the lowest temporal warping error (0.0234) and flicker index (0.0124), along with the highest inter-frame structural similarity (0.9512). These results validate the effectiveness of TecoGAN's spatio-temporal discriminator and temporal consistency loss in producing smooth, natural video sequences.

Table 3. PSNR Performance Stratified by Motion Complexity

Algorithm	Low Motion	Moderate Motion	High Motion
GFPGAN	36.234 dB	34.052 dB	31.456 dB
TecoGAN	35.867 dB	34.033 dB	32.123 dB
RRDB-ESRGAN	34.512 dB	32.244 dB	29.876 dB
ESRGAN	32.987 dB	30.825 dB	28.543 dB

All algorithms exhibit expected performance degradation with increasing motion complexity. However, TecoGAN demonstrates the smallest performance drop from low to high motion scenarios (3.744 dB), compared to GFPGAN (4.778 dB), RRDB-ESRGAN (4.636 dB), and ESRGAN (4.444 dB). This robustness to motion complexity underscores the value of explicit temporal modeling for video super-resolution applications.

Table 4. Computational Complexity Analysis

Algorithm	Params (M)	FLOPs (G)	Time (ms)	Memory (GB)
GFPGAN	72.3	234.5	156.2	4.8
TecoGAN	45.6	312.8	198.4	6.2
RRDB-ESRGAN	23.4	178.2	89.3	3.1
ESRGAN	16.7	142.6	67.8	2.4

Inference time measured for 720p to 4K upscaling on NVIDIA A100 GPU

Table 5. Ablation Study Results

Configuration	PSNR	SSIM	LPIPS
GFPGAN (Full)	34.052	0.952	0.0823
w/o Generative Prior	32.456	0.912	0.1024
w/o Progressive Training	33.234	0.934	0.0912
TecoGAN (Full)	34.033	0.848	0.0756
w/o Temporal Discriminator	33.567	0.823	0.0834
w/o Bidirectional Propagation	33.234	0.812	0.0856

Table 6. Statistical Significance of PSNR Differences (p-values)

	GFPGAN	TecoGAN	RRDB-ESRGAN	ESRGAN
GFPGAN	-	0.8234	<0.001*	<0.001*
TecoGAN	0.8234	-	<0.001*	<0.001*
RRDB-ESRGAN	<0.001*	<0.001*	-	0.0023*
ESRGAN	<0.001*	<0.001*	0.0023*	-

**Statistically significant at α = 0.05*

The quality and fidelity of the reconstructed videos are greatly influenced by the method used in the field of video resolution improvement. In our comparison examination, GFPGAN stands out as the superior approach, exceeding other candidates in both PSNR and SSIM metrics. GFPGAN demonstrates its advantage in generating high-quality, visually appealing movies with improved resolution and fidelity, evidenced by a PSNR of 34.052 and an SSIM of 0.952.

The efficacy of GFPGAN is due to its novel Generative Feedback Pyramid architecture, enabling the production of high-resolution films with intricate details and textures. This hierarchical structure utilizes various feedback loops at many scales, allowing the network to efficiently capture and transmit characteristics across video frames. GFPGAN integrates feedback mechanisms at every pyramid level to guarantee that the produced videos demonstrate improved resolution, sharpness, and clarity.

8. Conclusion and Future Work

This comprehensive study presented a rigorous comparative evaluation of four state-of-the-art GAN-based video super-resolution algorithms: GFPGAN, ESRGAN, RRDB-ESRGAN, and TecoGAN. Through systematic experimentation on the LDV benchmark dataset employing multiple complementary quality metrics encompassing distortion-based measures (PSNR, SSIM), perceptual quality indicators (LPIPS, NIQE), and temporal consistency metrics (TWE, IF-SSIM), we have established quantitative performance rankings and identified the distinctive strengths of each approach.

Our experimental findings demonstrate that GFPGAN achieves superior performance in distortion-based metrics with PSNR of 34.052 dB and SSIM of 0.952, attributed to its innovative generative feedback architecture and progressive training strategy that enables the recovery of intricate facial details and high-frequency textures. The generative prior mechanism contributes significantly to performance improvement, as evidenced by the ablation study showing a 1.596 dB PSNR gain when this component is included. TecoGAN excels in temporal consistency metrics with the lowest temporal warping error (0.0234) and highest inter-frame structural similarity (0.9512), validating the effectiveness of its spatio-temporal discriminator design in producing smooth, natural video sequences with minimal flickering artifacts.

ESRGAN demonstrates favorable perceptual quality metrics (LPIPS: 0.0689, NIQE: 3.457) while maintaining the lowest computational requirements with only 16.7M parameters and 67.8ms inference time, making it particularly suitable for resource-constrained deployment scenarios. RRDB-ESRGAN provides a balanced compromise across quality dimensions, offering improved feature extraction through its enhanced residual dense block configurations. The motion-stratified analysis reveals that TecoGAN exhibits the greatest robustness to motion complexity with only 3.744 dB performance degradation from low to high motion scenarios, compared to 4.778 dB for GFPGAN, underscoring the value of explicit temporal modeling for video applications.

The statistical significance analysis confirms that GFPGAN and TecoGAN do not differ significantly in PSNR performance (p = 0.8234), while both significantly outperform RRDB-ESRGAN and ESRGAN with p-values less than 0.001. The computational complexity profiling provides practical deployment guidance, with inference times ranging from 67.8ms for ESRGAN to 198.4ms for TecoGAN on NVIDIA A100 GPU for 4K video processing. These findings offer valuable guidance for researchers and practitioners in selecting appropriate VSR algorithms based on specific application requirements, whether prioritizing spatial detail preservation, temporal consistency, perceptual quality, or computational efficiency.

Future research directions emerging from this study include the development of hybrid architectures that combine the temporal modeling capabilities of TecoGAN with the generative prior approach of GFPGAN to achieve superior performance across both spatial and temporal dimensions. Additionally, exploring lightweight network designs through neural architecture search and knowledge distillation, investigating self-supervised learning approaches to reduce dependence on paired training data, and developing real-time implementations for edge devices represent promising avenues for advancing the field of video super-resolution.

Author contributions

Garima: Conceptualization, Software, Field study, Visualization, Investigation, Writing-Original draft preparation **Swati Malik:**

Data curation, Methodology, Software, Validation., Field study, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Kappeler, A., Yoo, S., Dai, Q., & Katsaggelos, A. K. (2016). Video super-resolution with convolutional neural networks. IEEE Transactions on Computational Imaging, 2(2), 109-122.

[2] Liu, H., Ruan, Z., Zhao, P., Dong, C., Shang, F., Liu, Y., ... & Timofte, R. (2022). Video super-resolution based on deep learning: A comprehensive survey. Artificial Intelligence Review, 55(8), 5981-6035.

[3] Xiao, J., Jiang, X., Zheng, N., Yang, H., Yang, Y., Yang, Y., ... & Lam, K. M. (2023). Online video super-resolution with convolutional kernel bypass grafts. IEEE Transactions on Multimedia, 25, 8972-8987.

[4] Tao, X., Gao, H., Liao, R., Wang, J., & Jia, J. (2017). Detail-revealing deep video super-resolution. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4472-4480).

[5] Khaledyan, D., Amirany, A., Jafari, K., Moaiyeri, M. H., Khuzani, A. Z., & Mashhadi, N. (2020). Low-cost implementation of bilinear and bicubic image interpolation for real-time image super-resolution. In 2020 IEEE Global Humanitarian Technology Conference (GHTC) (pp. 1-5). IEEE.

[6] Mishra, S. R., Mohapatra, H., & Saxena, S. (2024). Leveraging data analytics and a deep learning framework for advancements in image super-resolution techniques. In Data Analytics and Machine Learning (pp. 105-126). Springer Nature Singapore.

[7] Duong, H. T., Le, V. T., & Hoang, V. T. (2023). Deep learning-based anomaly detection in video surveillance: A survey. Sensors, 23(11), 5024.

[8] Himeur, Y., Al-Maadeed, S., Kheddar, H., et al. (2023). Video surveillance using deep transfer learning and deep domain adaptation. Engineering Applications of Artificial Intelligence, 119, 105698.

[9] Yang, H., Wang, Z., Liu, X., Li, C., Xin, J., & Wang, Z. (2023). Deep learning in medical image super resolution: A review. Applied Intelligence, 53(18), 20891-20916.

[10] Qiu, D., Cheng, Y., & Wang, X. (2023). Medical image super-resolution reconstruction algorithms based on deep learning: A survey. Computer Methods and Programs in Biomedicine, 238, 107590.

[11] Baniya, A. A., Lee, T. K., Eklund, P. W., & Aryal, S. (2024). A survey of deep learning video super-resolution. IEEE Transactions on Emerging Topics in Computational Intelligence.

[12] Morrison, L. (2021). Utilizing machine learning for video processing. SMPTE Motion Imaging Journal, 130(8), 107-111.

[13] Wang, P., Bayram, B., & Sertel, E. (2022). A comprehensive review on deep learning based remote sensing image super-resolution methods. Earth-Science Reviews, 232, 104110.

[14] Wang, X., Yi, J., Guo, J., Song, Y., et al. (2022). A review of image super-resolution approaches based on deep learning and applications in remote sensing. Remote Sensing, 14(21), 5423.

[15] Saxena, D., & Cao, J. (2021). Generative adversarial networks (GANs) challenges, solutions, and future directions. ACM Computing Surveys, 54(3), 1-42.

[16] Alqahtani, H., Kavakli-Thorne, M., & Kumar, G. (2021). Applications of generative adversarial networks (GANs): An updated review. Archives of Computational Methods in Engineering, 28, 525-552.

[17] Creswell, A., White, T., Dumoulin, V., et al. (2018). Generative adversarial networks: An overview. IEEE Signal Processing Magazine, 35(1), 53-65.

[18] Xiong, Y., Guo, S., Chen, J., et al. (2020). Improved SRGAN for remote sensing image super-resolution across locations and sensors. Remote Sensing, 12(8), 1263.

[19] Wang, X., Yu, K., Wu, S., et al. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks. In Proceedings of the ECCV Workshops (pp. 0-0).

[20] Tian, Y., Zhang, Y., Fu, Y., & Xu, C. (2020). TDAN: Temporally-deformable alignment network for video super-resolution. In Proceedings of the IEEE/CVF CVPR (pp. 3360-3369).

[21] Chan, K. C., Wang, X., Yu, K., Dong, C., & Loy, C. C. (2021). BasicVSR: The search for essential components in video super-resolution and beyond. In Proceedings of the IEEE/CVF CVPR (pp. 4947-4956).

[22] Jones, E. A., Wang, F. F., & Costinett, D. (2016). Review of commercial GaN power devices and GaN-based converter design challenges. IEEE JESTPE, 4(3), 707-719.

[23] Kumar, A., & Vatsa, A. (2022). Influence of GFP GAN on melanoma classification. In 2022 IEEE Integrated STEM Education Conference (ISEC) (pp. 334-339). IEEE.

[24] Chu, M., Xie, Y., Leal-Taixé, L., & Thuerey, N. (2018). Temporally coherent GANs for video super-resolution (TecoGAN). arXiv preprint arXiv:1811.09393.

[25] Chen, Y. Z., Liu, T. J., & Liu, K. H. (2022). Super-resolution of satellite images by two-dimensional RRDB and edge-enhancement GAN. In ICASSP 2022 (pp. 1825-1829). IEEE.

[26] Chaudhuri, S. (Ed.). (2006). Super-resolution imaging (Vol. 632). Springer Science & Business Media.

[27] Katsaggelos, A. K., Molina, R., & Mateos, J. (2007). Super resolution of images and video (Vol. 3). Morgan & Claypool Publishers.

[28] Van der Walt, S. J. (2010). Super-resolution imaging. Doctoral dissertation, University of Stellenbosch.

[29] Yang, J., & Huang, T. (2017). Image super-resolution: Historical overview and future challenges. In Super-resolution imaging (pp. 1-34). CRC Press.

[30] Rohith, G., & Sutha, G. L. (2022). Super-resolution for remote sensing applications using deep learning techniques. Cambridge Scholars Publishing.

[31] Karthick, S., & Muthukumaran, N. (2024). Deep RegNet-150 architecture for single image super resolution. Applied Soft Computing, 111837.

[32] Johnson, E. (2023). Deep learning-based super-resolution techniques for enhancing satellite imagery. African Journal of AI and Sustainable Development, 3(2), 342-347.

[33] Xu, X. (2023). Applications research of deep learning in super-resolution reconstruction of medical image. In 2023 CSMIS (pp. 405-410). IEEE.

[34] Deng, J., Song, W., Liu, D., et al. (2021). Improving the spatial resolution of solar images using GAN and self-attention mechanism. The Astrophysical Journal, 923(1), 76.

[35] Chen, W., Lu, Y., Ma, H., et al. (2022). Self-attention mechanism in person re-identification models. Multimedia Tools and Applications, 1-19.

[36] Fernandez-Beltran, R., Latorre-Carmona, P., & Pla, F. (2017). Single-frame super-resolution in remote sensing: A practical overview. International Journal of Remote Sensing, 38(1), 314-354.

[37] Chen, R., Tang, X., Zhao, Y., et al. (2023). Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging. Nature Communications, 14(1), 2854.

[38] Wronski, B., Garcia-Dorado, I., Ernst, M., et al. (2019). Handheld multi-frame super-resolution. ACM Transactions on Graphics (ToG), 38(4), 1-18.

[39] Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27.

[40] Pan, Z., Yu, W., Yi, X., et al. (2019). Recent progress on generative adversarial networks (GANs): A survey. IEEE Access, 7, 36322-36333.

[41] Alqahtani, H., Kavakli-Thorne, M., & Kumar, G. (2021). Applications of generative adversarial networks (GANs): An updated review. Archives of Computational Methods in Engineering, 28, 525-552.

[42] Kumar, R., Kumar Bharti, S., Ahmad, M., et al. (2023). Literature review on blind face restoration using GFPGAN. Available at SSRN 4483846.

[43] Zhang, X., & Feng, J. (2024). A novel blind restoration method for miner face images based on improved GFP-GAN model. IEEE Access.

[44] Yang, R. (2021). NTIRE 2021 challenge on quality enhancement of compressed video: Dataset and study. In Proceedings of the IEEE/CVF CVPR (pp. 667-676).