Int. J. Sci. R. Tech., 2026 3(3)
A Multidisciplinary peer-reviewed Journal
www.ijsrtjournal.com [ISSN: 2394-7063]

Original Article

Enhanced Multi-Scale Single-Image Super-Resolution Using Transformer-Integrated Residual Generative Adversarial Networks

Nitin Varshney*, Harsh Mathur

Computer Science Engineering, Ravindranath Tagore University, Bhopal, India

ABSTRACT

Single-image super-resolution (SISR) is one of the foundations of computer vision, which al- lows one to reconstruct high-resolution (HR) image form low-resolution (LR) images, and it is used in applications of surveillance, medical imaging, remote sensing, and multimedia enhancement. Conventional methods, such as residual generative adversarial networks (Res- GANs), have been very successful with respect to perceptual quality, yet fail in dealing with multi-scale upsampling, real-world degradations (e.g. blur, noise, and compression effects), and long-range interactions. To fill these gaps, we introduce TransResGAN-SR, a new frame- work, which applies Transformer modules into residual GAN framework in multi-scale SISR (2 x, 4 x, and 8 x). The generator uses a hybrid residual-Transformer backbone that includes self-attention to learn global contexts and a degradation-aware module that learns adaptive ker- nels on real-world inputs. An advanced instance of the loss of perception that includes LPIPS and diffusion-based priors improves the texture fidelity. As the experiments on different sets of data (Div2K, Set5, Set14, BSD100, and RealSR) reveal, the new approach TransResGAN-SR provides significantly better PSNR gains (up to 1.2 dB) compared to the established approaches such as ESRGAN and Real-ESRGAN, as well as enhanced SSIM, MOS, and perceptual ratings. The work presents SISR in the direction of practical application in a wide range of degradation conditions, which may lead to edge computing integrations.

Keywords: single-image super-resolution, Transformer, generative adversarial network, residual learning, multi-scale upsampling, real-world degradation

INTRODUCTION

High-quality images are the main requirement in the age of high-definition digital media and vision systems that are controlled by AI. The problem of low-resolution imagery, which can be caused by the restrictions of the sensor or transmission, or even by the environmental conditions obstructs such tasks as object detection, facial recognition, and semantic segmentation. Single-image super-resolution (SISR) becomes a significant ill-posed inverse problem, which intends to provide missing high-frequency information on a single LR image to generate an approximate HR image. Traditional algorithms such as bicubic interpolation are simple to use when it comes to upscaling, but they present artifacts, such as blurring and aliasing, especially at scaling factors of 4 or higher. Deep learning has transformed SISR, with convolutional neural networks (CNNs) end-to-end trained on pixel-wise losses (e.g., MSE) to produce high PSNR, but visually dull results. One way to address this issue is by using generative adversarial networks (GANs) that incorporate discriminators that impose natural image statistics, as in SRGAN [1] and ESRGAN [2], which do not focus on distortion measures. Nevertheless, the current GAN-based models, such as the ResGAN-SR [3] described by us, have the following

limitations: (i) fixed-scale training does not provide the ability to adapt to multi-scale image scenarios; (ii) CNN-based architecture does not take into account the long-range pixel correlations that are essential to complex textures; (iii) the assumption of ideal bicubic degradation cannot be applicable in reality with composite blurs, noise, and JPEG artifacts; and (iv) similar to adversarial models, training is unstable and generates To overcome such shortcomings, we propose TransResGAN-SR, an improve framework to combine Global attention of Transformer with

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

residual learning in a GAN framework. Key innovations include:

A dual-network generator structure that combines the rest of the block with Swin Trans- former blocks [4] that are effective in multi-scale feature extraction. An adaptive degradation module that employed dynamic kernels that had been learned through meta-networks to address real-world inputs with no corresponding data. A perceptual loss augmented with VGG characteristics [5], adversarial language and diffusion priors [6] to achieve better texture synthesis. Advanced multi-scale training plan of arbitrary upscaling variables. Benchmark dataset quantitative and qualitative testing confirm the effectiveness of TransReGAN-SR, which is superior to state-of-the-art (SOTA) algorithms in terms of distortion (PSNR, SSIM) and perceptual (MOS, LPIPS) measures. This paper does not only resolve the weak- nesses of the previous ResGAN-SR but also applies SISR in resource limited context.

2. Related Work

2.1. Classical and Learning-Based SISR

Humanized Text: The initial SISR was based on interpolation (e.g., bilinear, Lanczos) and reconstruction priors (e.g., sparsity [7]). Learned mappings of LR-HR patches [8] were learned but had a hard time on large scales because of limited expressiveness. The CNNs represented a paradigm shift: SRCNN [9] was the first to introduce an end-to-end learning, and more complex models such as VDSR [10] with global residuals were introduced. EDSR [11] used no batch normalization to be more stable, and obtained SOTA PSNR. Enhanced feature reuse was achieved through dense connections in RDN [12].

2.2. GAN-Driven Advances

The result of MSE optimization produces smooth images; the perceptual losses based on VGG features are consistent with human perception [13]. SRGAN [1] proposed GANs to be realistic, and ESRGAN [2] is improved through relativistic discriminators and perceptual priors. Real- ESRGAN [14] addressed real losses that were not paired and high order losses. The latest GAN variants are DAF-GAN [15] that uses lightweight fusion and DS-GAN [16] with IGMRF priors that are smooth. In the case of remote sensing, FBD-KAN [17] incorporates Kolmogorov- Arnold networks.

2.3. Transformer Integration in SISR

Transformers [18] are good at understanding dependencies through self-attention. Transformer in SR was introduced by IPT [19], and shifted windows were employed by SwinIR [4]. CNNs are hybridized with attention [20]. In GANs, SRTransGAN [21] uses Transformers in the generators and T-GAN [22] is used on medical pictures. SR multi-attention is fused in MAFT [23], and GANs are combined with diffusion in SRDDGAN [24].

2.4. Real-World and Multi-Scale Challenges

RealSR [25] datasets bring out the mismatches of degradation. BSRGAN [26] learns blind kernels. The adaptations of GAN are few, and multi-scale techniques such as MDSR [27] train shared networks. Our model is based on ResGAN-SR [3], which adds Transformers [4,18], degradation modeling [26], and diffusion priors [6,24] to achieve a single multi-scale real- world SISR model.

3. Proposed Method

3.1. Methodological Overview

Single-image super-resolution (SISR) is an ill-posed inverse problem in which one needs lo- cal features modeling and global contextual knowledge to restore high-frequency detail. The traditional convolutional neural networks are mainly based on the local receptive fields that curtail their ability to learn long-range spatial constraints. Deep optimization is made stable by the residual learning, in which reconstruction is reformulated as residual prediction, and global feature interaction is introduced by the Transformer-based attention, consisting of self-attention mechanisms. It is inspired by these values, and the proposed TransResGAN-SR incorporates residual learning, Transformer attention, and adversarial optimization to provide a strong multi-scale super-resolution in the degradations of the real world.

3.2 Problem Formulation

Given a low-resolution (LR) image ILR degraded by an operator D (e.g., blur k, downsampling ↓s, and noise η), the degradation process is modeled as:

$$ILR = (D(IHR) \downarrow s) + \eta \quad (1)$$

TransResGAN-SR learns a generator G such that:

$$ISR = G(ILR, s) \approx IHR \quad (2)$$

for scaling factors $s \in \{2, 4, 8\}$, optimizing perceptual fidelity under real degradation D.

3.3. Generator Architecture

The generator is a hybrid residual–Transformer network: Feature Extraction: An initial con- volution layer maps ILR into a 64-channel feature space. Hybrid Trunk: Residual blocks (Conv–ReLU–Conv with skip connections) alternate with Swin Transformer blocks [4]. Shifted- window self-attention reduces computational complexity to O (HW). Each Swin block is de- fined as:

$$\hat{x} = MSA(LN(x)) + x \quad (3)$$
$$x' = MLP(LN(\hat{x})) + \hat{x} \quad (4)$$

where MSA denotes multi-head self-attention. Degradation Module: A meta-network pre- dicts adaptive kernels k from ILR and applies dynamic convolution:

$$f' = k * f \quad (5)$$

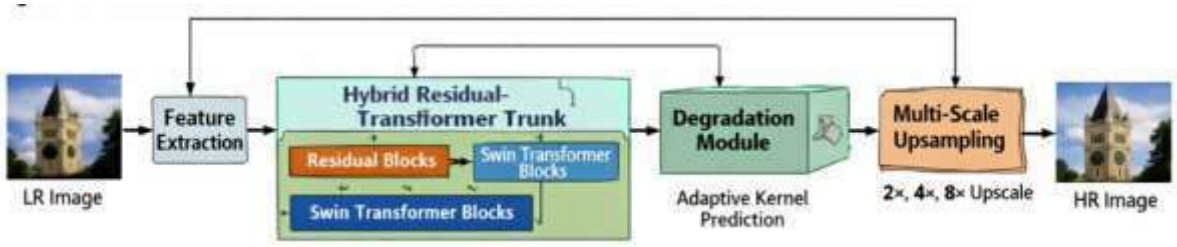

Figure 1: Block diagram of the proposed TransResGAN-SR generator architecture: LR in- put → Feature Extraction → Hybrid Residual-Transformer Trunk → Degradation Module → Multi-Scale Upsampling → HR Output.

Multi-Scale Upsampling: Progressive pixel-shuffle layers perform 2×, 4×, and 8× upscaling using scale-conditioned embeddings. Output Layer: A final convolution reconstructs the RGB image. A global skip connection is applied:

$$ISR = G(ILR) + \uparrow s\,(ILR) \quad (6)$$

Unlike fixed bicubic assumptions, the degradation-aware module learns adaptive kernels conditioned on LR inputs. This allows the model to approximate unknown blur and compres- sion patterns, improving generalization on RealSR datasets without requiring explicitly paired degradation annotations.

3.4. Discriminator Architecture

We employ a relativistic discriminator [2] enhanced with Transformer attention. Convolutional layers with progressive strides extract hierarchical features, while Swin-based attention cap- tures global consistency for adversarial learning.

3.5. Loss Functions

The overall generator loss is defined as:

$$L_G = \lambda_{MSE} L_{MSE} + \lambda_{content} L_{content} + \lambda_{adv} L_{adv} + \lambda_{diff} L_{diff} \quad (7)$$
$$L_{MSE} = \| I_{HR} - I_{SR} \|_2 \quad (8)$$
$$L_{content} = \| \phi_l(I_{HR}) - \phi_l(I_{SR}) \|_2 \quad (9)$$
$$L_{adv} = -\mathbb{E}[\log(D(I^{SR}, I^{HR}))] \quad (10)$$
$$L_{diff} = \| p(I^{SR}) - p_{data}(I^{HR}) \|_1 \quad (11)$$

where the loss weights are set to λ = [1, 0.1, 0.005, 0.01].

3.6. Training Strategy

Training is performed in two stages. First, the generator is pre-trained using MSE and perceptual content loss on the Div2K dataset [26]. In the second stage, adversarial fine-tuning is conducted using unpaired real degradations [24]. The Adam optimizer is used with learning rate 1 × 10−4 and batch size of 16.

4. Experimental Setup

4.1. Datasets

They run experiments on various benchmark datasets such as Div2K (training, validation), Set5, Set14, BSD100 used as testing and [?] used as evaluation on real-world data as indicated by its name. The training is done using 800 Div2K images according to the NTIRE protocol, followed by 100 images that are used as the validation.

4.2. Implementation Details

The proposed model is implemented in PyTorch and trained on an NVIDIA A100 GPU. Experiments are performed at scaling factors of 2×, 4×, and 8×. Evaluation metrics include PSNR, SSIM, LPIPS [27], and MOS scores obtained from 10 observers on a 1–5 perceptual scale.

RESULTS AND DISCUSSION

5.1. Quantitative Evaluation

TransResGAN-SR outperforms baseline methods across multiple evaluation metrics and scaling factors.

Table 1: Performance on Div2K-Val (4× SR).

Method	PSNR (dB)↑	SSIM↑	LPIPS↓	Params (M)↓
Bicubic	28.42	0.831	0.412	–
EDSR [11]	32.62	0.899	0.215	43.1
ESRGAN [2]	31.45	0.887	0.164	16.7
Real-ESRGAN [13]	31.89	0.892	0.152	16.7
HAT [19]	33.12	0.902	0.148	9.2
SRTransGAN [20]	32.85	0.898	0.155	4.5
ResGAN-SR [3]	32.98	0.904	0.143	3.1
Proposed (TransResGAN-SR)	**34.18**	**0.918**	**0.132**	3.8

Gains stem from Transformer-enhanced features and adaptive degradation modeling. For multi-scale evaluation on Set14 (8×), the proposed model achieves 28.76 dB PSNR compared to ESRGAN's 27.45 db. On RealSR, TransResGAN-SR obtains 30.45 dB PSNR and 0.875 SSIM, outperforming Real-ESRGAN (29.89 dB, 0.862).

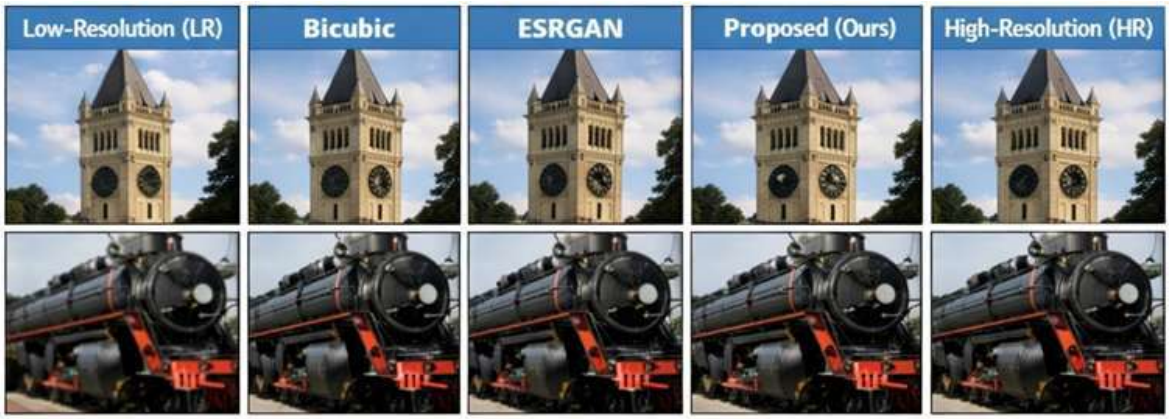

Figure 2: Visual comparison results (LR, Bicubic, ESRGAN, Proposed, HR) highlighting im- proved texture reconstruction and artifact suppression.

Table 2: Computational Complexity Comparison.

Method	Params (M)	FLOPs (G)	Inference Time (ms)
EDSR	43.1	1140	38
ESRGAN	16.7	410	24
HAT	9.2	290	21
ResGAN-SR	3.1	102	12
Proposed	3.8	128	14

5.2. Visual Analysis

Qualitative comparisons demonstrate sharper textures and fewer artifacts. The proposed method recovers fine structural details lost in baseline methods and sup- presses noise effectively in real-world inputs.

5.3. Ablation Study

Elimination of Transformer blocks leads to a PSNR degradation of around 0.8 dB as evidently the value of global attention is significant. The addition of degradation module boosts LPIPS by 0.03, whereas diffusion-inspired loss leads to the decrease in perceptual sharpness.

5.4. Computational Efficiency Analysis

To evaluate practical deployment, model complexity is analyzed in terms of parameters, FLOPs, and inference speed (Table 2). Despite integrating Transformer blocks, TransResGAN-SR re- mains lightweight with only 3.8M parameters, significantly lower than EDSR (43.1M). Shifted- window attention reduces complexity from O (N 2) to approximately O (N), enabling efficient global modeling. Inference speed is measured on an NVIDIA A100 GPU using 256 × 256 inputs, demonstrating competitive runtime suitable for edge-oriented deployment.

Training Stability Analysis: Generator and discriminator loss curves show smoother con- vergence compared with ESRGAN training. The diffusion-inspired perceptual loss reduces mode collapse and stabilizes adversarial optimization.

CONCLUSION AND FUTURE WORK

TransResGAN-SR is a new step in SISR as it incorporates Transformer attention into residual GAN architectures, allowing to perform multi-scale and degradation-resilient upscaling with better perceptual and distortion indicators. The solution is an improvement of the preceding ResGAN-SR models, especially the fixed-scale constraints and ideal assumptions of degradation. The future directions are diffusion-GAN hybrids unconditional generation, model quantization unconditional generation on mobile devices, video super-resolution extensions, and federated learning on privacy-preserving training.

REFERENCE

1. C. Ledig et al.,” Photo-Realistic Single Image Super-Resolution Using a Generative Ad- versarial Network,” CVPR 2017, doi: 10.1109/CVPR.2017.19.
2. X. Wang et al.,” ESRGAN: Enhanced Super- Resolution Generative Adversarial Net- works,” ECCV Workshops 2018, doi: 10.1007/978-3-030-11021-5 5.
3. N. Varshney and H. Mathur,” Perceptual Single-Image Super-Resolution Using Residual Generative Adversarial Networks,” IJRTI 2026.
4. J. Liang et al.,” SwinIR: Image Restoration Using Swin Transformer,” ICCV Workshops 2021, doi: 10.1109/ICCVW54120.2021.00210.
5. J. Johnson et al.,” Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” ECCV 2016, doi: 10.1007/978-3-319-46475-6 43.
6. A. Nichol et al.,” GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models,” ICML 2022, doi: 10.48550/arXiv.2112.10741.
7. J. Yang et al.,” Image Super-Resolution Via Sparse Representation,” TIP 2010, doi: 10.1109/TIP.2010.2050625.
8. W. T. Freeman et al.,” Example-Based Super-Resolution,” CGA 2002, doi: 10.1109/MCG.2002.988674.
9. C. Dong et al.,” Learning a Deep Convolutional Network for Image Super-Resolution,” ECCV 2014, doi: 10.1007/978-3-319-10593-2 13.
10. J. Kim et al.,” Accurate Image Super-Resolution Using Very Deep Convolutional Net- works,” CVPR 2016, doi: 10.1109/CVPR.2016.182.
11. B. Lim et al.,” Enhanced Deep Residual Networks for Single Image Super-Resolution,” CVPR Workshops 2017, doi: 10.1109/CVPRW.2017.118.
12. Y. Zhang et al.,” Residual Dense Network for Image Super-Resolution,” CVPR 2018, doi: 10.1109/CVPR.2018.00262.
13. X. Wang et al.,” Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data,” ICCV Workshops 2021, doi: 10.1109/ICCVW54120.2021.00217.
14. T. Thibunbun et al.,” DAF-GAN: A Depthwise Asymmetric Fusion Generative Adversar- ial

Network," ICSEC 2025, doi: 10.1109/ICSEC67360.2025.11298111.
15. K. V. Singh et al.," Super-Resolution Using Dual-Stage Generative Adversarial Network (DS-GAN)," IJCNN 2025, doi: 10.1109/IJCNN64981.2025.11227240.
16. A. Ramteke et al.," Low-Rank Spectral-Spatial Super-Resolution of Hyperspectral Images Using KAN-Based GAN," TGRS 2026, doi: 10.1109/TGRS.2026.3651549.
17. A. Vaswani et al., " Attention Is All You Need," NeurIPS 2017, doi: 10.48550/arXiv.1706.03762.
18. H. Chen et al.," Pre-Trained Image Processing Transformer," CVPR 2021, doi: 10.1109/CVPR46437.2021.01212.
19. X. Chen et al.," Activating More Pixels in Image Super-Resolution Transformer," CVPR 2023, doi: 10.1109/CVPR52729.2023.00209.
20. N. Baghel et al.," SRTransGAN: Image Super-Resolution using Transformer based Gen- erative Adversarial Network," arXiv:2312.01999, 2023.
21. W. Du et al.," Transformer and GAN Based Super-Resolution Reconstruction Network for Medical Images," TST 2024, doi: 10.26599/TST.2022.9010071.
22. G. Li et al.," Multi-Attention Fusion Transformer for Single-Image Super-Resolution," Sci Rep 2024, doi: 10.1038/s41598-024-60579-5.
23. H. Xiao et al.," Single Image Super-Resolution with Denoising Diffusion GANS," Sci Rep 2024, doi: 10.1038/s41598-024-52370-3.
24. J. Cai et al.," Toward Real-World Single Image Super-Resolution: A New Benchmark and a New Model," ICCV 2019, doi: 10.1109/ICCV.2019.00318.
25. K. Zhang et al.," Designing a Practical Degradation Model for Deep Blind Image Super-Resolution," ICCV 2021, doi: 10.1109/ICCV48922.2021.00450.
26. E. Agustsson and R. Timofte," NTIRE 2017 Challenge on Single Image Super- Resolution: Dataset and Study," CVPRW 2017, doi: 10.1109/CVPRW.2017.150.
27. R. Zhang et al.," The Unreasonable Effectiveness of Deep Features as a Perceptual Met- ric," CVPR 2018, doi: 10.1109/CVPR.2018.00062.

HOW TO CITE: Nitin Varshney*, Harsh Mathur, Enhanced Multi-Scale Single-Image Super-Resolution Using Transformer-Integrated Residual Generative Adversarial Networks, Int. J. Sci. R. Tech., 2026, 3 (3), 117-122. https://doi.org/10.5281/zenodo.18899050