ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

Enhancement of Low Resolution Natural Images Using Deep Learning Based Super Resolution Techniques

Dr. Shashidhara. B

Associate Professor, GFGC, Shimoga

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150700164

Received: 17 August 2026; Accepted: 22 August 2026; Published: 26 August 2026

ABSTRACT

Image super-resolution is an important image processing task that aims to reconstruct a high resolution image from a low resolution input. Conventional interpolation methods often produce blurred edges, loss of texture, and pixelated outputs, particularly when images contain noise, compression artefacts, or complex real world degradations. This work proposes a hybrid CNN Transformer GAN framework for natural image super resolution. The CNN component extracts local features such as edges, patterns, and fine textures; the Transformer component captures long range dependencies and preserves global image structure; and the Generative Adversarial Network (GAN) component improves perceptual quality by generating sharper and more realistic visual details. The proposed model supports 2X, 4X, and 8X upscaling, with 4X selected as the primary evaluation scale because it provides a practical balance between image quality and computational complexity. The system is trained using degraded low-resolution images produced through blur, noise, JPEG compression, and downsampling operations. Performance is evaluated using Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), Mean Squared Error (MSE), and visual quality analysis. The hybrid architecture is expected to improve edge sharpness, preserve structural details, reduce visual artefacts, and generate natural looking high-resolution outputs compared with conventional CNN and GAN based approaches. The proposed framework can support applications in photo restoration, mobile imaging, surveillance, e-commerce, digital archives, and web image enhancement.

Keywords: Image Super-Resolution, Deep Learning, Convolutional Neural Network (CNN), Transformer, Generative Adversarial Network (GAN),

INTRODUCTION

Digital photographs are used in mobile phones, social media, online shopping, CCTV, and multimedia applications. Many photographs have low resolution because of weak cameras, long-distance capture, compression, cropping, or low internet bandwidth. When such images are enlarged, they may look blurry, noisy, pixelated, and unclear. This reduces the visibility of important details such as edges, faces, text, object boundaries, and textures. Traditional resizing methods like nearest neighbour, bilinear, and bicubic interpolation only increase image size. They cannot restore the original missing details and may produce blur or jagged edges. Image super resolution is a technique that converts a low resolution image into a clearer high resolution image. Deep learning models learn the relationship between low and high resolution image pairs. SRCNN was an early CNN model that showed effective image-resolution enhancement. EDSR improves reconstruction accuracy using deeper residual CNN layers. ESRGAN uses GAN technology to generate sharper and more realistic textures. SwinIR uses Transformer blocks to preserve local details and overall image structure. Real-world images are difficult because they may contain blur, noise, poor lighting, and JPEG artefacts. This work compares CNN, GAN, and Transformer models using the DIV2K natural image dataset. The results can be evaluated using PSNR, SSIM, MSE, LPIPS, visual quality, and processing time.

LITERATURE REVIEW

Aakerberg, Johansen, Nasrollahi and Moeslund (2022) proposed semantic-segmentation-guided real world super-resolution. Their work uses scene-level semantic information to improve reconstruction, showing that contextual understanding can help preserve meaningful object structures. Xu, Wei, Chen, Liu, Mao, Lin and Li (2022) introduced dual adversarial adaptation for cross device real world image super-resolution. The study addresses the fact that cameras and mobile devices produce different blur, noise, and colour degradations, which is relevant for practical enhancement systems.

Liu, Liu, Gu, Qiao and Dong (2023) presented a survey on blind image super resolution. They categorized blind SR approaches based on degradation modelling and training data, emphasizing the difficulty of enhancing images when the original blur and noise processes are unknown. Wei, Xie, Li and Lin (2023) developed a Taylor Neural Network for real world image super-resolution. Their approach focuses on overcoming the shortage of perfectly paired low-resolution and high resolution training images, a major challenge in real image restoration.

Zhang, Li, Shi, Chen, Qiao, Zhang, Wu and Dong (2023) formulated real world image super resolution as a multi task learning problem. Their work shows that jointly learning related restoration tasks can improve the quality and robustness of super resolution models. Wang, Liang, Wang, Yang, An and Guo (2022) proposed degradation modulation for real world light field image super resolution. The model incorporates degradation information into the CNN-based reconstruction process, improving generalization under diverse image quality conditions.

Wei, Sun, Guo, Liu, Chen, Ji and Lin (2023) introduced RealBSR and FBANet for burst image super-resolution. Their method combines information from multiple frames and uses a Transformer based decoding module to recover detailed and visually pleasing images. Chen, Wang, Zhou, Qiao and Dong (2023) proposed the Hybrid Attention Transformer (HAT). HAT combines channel attention and window based self-attention to capture local textures and long-range contextual information, which supports the use of Transformer blocks in modern super-resolution frameworks.

Tian, Zhang, Zhu, Zhang and Lin (2022) reviewed GAN based image super resolution methods. Their survey highlights how adversarial training and perceptual losses can improve realistic texture reconstruction, while also noting challenges such as unstable trainin g and artificial detail generation. Jiang, Liu, Gu, Shao, Zhang, Liu and Lin (2022) developed the RealSRQ dataset and a no reference image quality metric for super resolution evaluation. Their study emphasizes that PSNR and SSIM alone may not adequately represent human perception of enhanced image quality.

Deviyani, Hoplamaz and Paul (2022) evaluated the robustness of real world super resolution methods using the WideRealSR dataset. Their findings show that models trained only with synthetic downsampling often struggle with real photographs containing complex, unknown degradations. Yue, Li, Wei, Li and Lin (2022) studied robustness against adversarial attacks in real world image super resolution. They used frequency domain masking to reduce harmful high frequency perturbations, highlighting the importance of reliability and robustness in deep-learning enhancement models.

Proposed System

The proposed system enhances low resolution natural images using a hybrid CNN- Transformer GAN super resolution framework. It improves image sharpness, resolution, texture, colour quality, and visual appearance while reducing blur, noise, pixelation, and JPEG artefacts. High resolution natural images from the DIV2K dataset are used to train the model, with 800 images for training, 100 for validation, and 100 for testing. Low resolution images are generated by applying downsampling, Gaussian blur, noise, JPEG compression, and brightness or contrast changes. The images are preprocessed using RGB conversion, normalization, resizing, patch extraction, rotation, flipping, and cropping.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

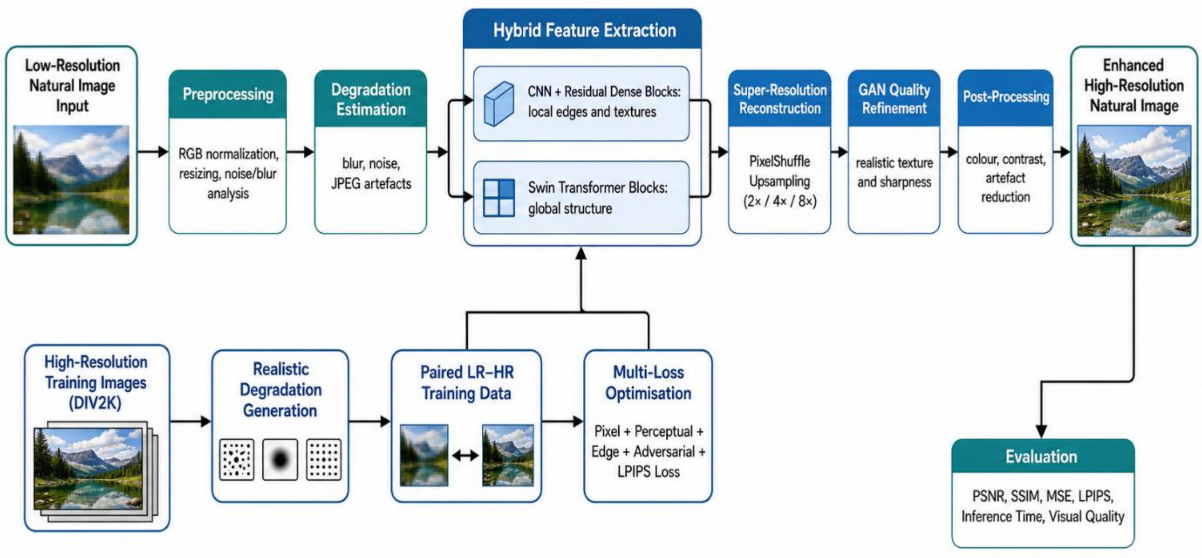

Figure 1: Proposed Deep Learning-Based Super-Resolution System

The CNN module extracts local features such as edges, lines, colours, and textures from the input image. Residual dense blocks preserve important features, while Swin Transformer blocks learn global mage structure and relationships. The GAN refinement module produces sharper, more realistic textures and visually pleasing output images. A combined loss function including pixel, perceptual, edge, adversarial, and LPIPS losses improves accuracy, structure, and visual quality. The model reconstructs images at 2X, 4X, or 8X resolution using Pixel Shuffle up sampling. Post-processing performs colour correction, contrast enhancement, noise reduction, sharpening, and artefact removal. The final output is evaluated using PSNR, SSIM, MSE, LPIPS, inference time, and visual comparison. If image quality is unsatisfactory, the model is retrained by improving its architecture, training data, degradation method, or hyper parameters. The final system produces clear, high resolution natural images suitable for photo restoration, mobile photography, surveillance, e-commerce, and digital archives.

Experimental Image Datasets at Different Scales

For experimental evaluation, the system uses natural images at different super-resolution scales: 2X, 4X, and 8X. The same low-resolution input image is enhanced at each scale to compare the effect of increasing resolution on image sharpness, texture quality, edge preservation, and visual appearance.

- 2X Super Resolution: Doubles the image width and height. It provides clear enhancement with relatively low reconstruction difficulty. For example, a 256 X 256 image becomes 512 X 512 pixels.

- 4X Super Resolution: Increases width and height four times. It is the main experimental scale because it gives a good balance between high-quality enhancement and computational complexity. A 256 X 256 image becomes 1024 X 1024 pixels.

- 8X Super Resolution: Increases width and height eight times. It is more challenging because the model must predict a large amount of missing detail. A 256 X 256 image becomes 2048 X 2048 pixels.

The figure compares the original low-resolution image with enhanced outputs at 2X, 4X, and 8X scales. The 2X output generally retains natural edges and colour details, while the 4X output provides greater sharpness and visible textures. The 8X output offers the largest image size but may show some smoothing or artificial texture because it requires reconstruction of more missing information.

The quality at each scale is measured using PSNR, SSIM, MSE, LPIPS, and inference time. Higher PSNR and SSIM values indicate better reconstruction and structural similarity, whereas lower MSE and LPIPS values indicate lower error and improved perceptual quality.

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

Experimentation

The following algorithm presents the implementation workflow for the proposed hybrid CNN Transformer GAN model for natural image super-resolution. It begins by preparing high resolution DIV2K images and generating corresponding low-resolution images through blur, noise, compression, and downsampling. The model then learns local image details using CNN and residual dense blocks, captures global structure using Transformer blocks, and improves visual realism through GAN-based refinement. Finally, the trained model reconstructs enhanced high resolution images and evaluates their quality using suitable image-quality metrics.

1. Start

2. Load high resolution natural images from the DIV2K dataset.

3. For each high resolution image IHR:

a. Apply blur, noise, JPEG compression, and downsampling.

b. Generate the corresponding low resolution image ILR.

c. Store the ILR IHR image pair.

4. Preprocess all image pairs:

a. Convert images to RGB format.

b. Normalize pixel values between 0 and 1.

c. Extract image patches.

d. Apply augmentation using rotation, flipping, and cropping.

5. Split the dataset into training, validation, and testing sets.

6. Initialize the hybrid generator model:

a. CNN layers extract local features.

b. Residual dense blocks preserve image details.

c. Swin Transformer blocks learn global image structure.

d. Pixel Shuffle layers increase image resolution.

7. Pre-train the generator using pixel loss:

a. Input ILR into the generator.

b. Generate super resolved image ISR.

c. Calculate pixel loss between ISR and IHR.

d. Update generator weights using backpropagation.

8. Initialize the GAN discriminator.

9. Fine-tune the generator and discriminator:

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

a. Generate ISR from ILR.

b. Give IHR and ISR to the discriminator.

Calculate discriminator loss for real and generated images.

d. Calculate total generator loss:

$L_{total} = L_{pixel} + L_{perceptual} + L_{edge} + L_{adversarial} + L_{LPIPS}$

e. Update discriminator weights.

f. Update generator weights.

10. Validate the model after every epoch:

a. Generate validation outputs.

b. Calculate PSNR, SSIM, MSE, and LPIPS.

c. Save the model with the best validation result.

11. Give a new low resolution test image to the trained model.

12. Estimate image degradation such as blur, noise, and JPEG artefacts.

13. Reconstruct the high resolution image using the trained generator.

14. Apply post-processing:

a. Improve colour and contrast.

b. Reduce remaining noise and artefacts.

c. Apply mild sharpening if required.

15. Evaluate the final output using PSNR, SSIM, MSE, LPIPS, inference time, and visual comparison.

16. If quality is acceptable:

Save and display the enhanced high resolution image.

Else:

Adjust model parameters or training data and repeat training.

17. Stop

EXPERIMENTAL RESULT

The following table presents the experimental results of the proposed hybrid CNN Transformer GAN model for enhancing low resolution natural images at 2X, 4X, and 8X scaling factors. The performance is evaluated for urban, portrait, and natural scene images using PSNR, SSIM, MSE, LPIPS, and inference time. These measures help assess reconstruction accuracy, structural similarity, visual quality, pixel error, and processing speed of the generated high resolution images.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

Table 1: Performance of the Proposed Model Result

Image Category	Scale Factor	PSNR (dB) ↑	SSIM ↑	MSE ↓	LPIPS ↓	Inference Time (s) ↓
Urban Scene	2×	32.84	0.924	34.12	0.082	0.18
Urban Scene	4×	29.76	0.883	68.45	0.124	0.24
Urban Scene	8×	26.91	0.821	132.68	0.198	0.38
Portrait	2×	34.25	0.941	24.46	0.061	0.17
Portrait	4×	30.84	0.902	53.76	0.097	0.23
Portrait	8×	27.65	0.846	111.35	0.165	0.37
Natural Scene	2×	31.92	0.915	41.80	0.094	0.19
Natural Scene	4×	28.68	0.861	82.43	0.143	0.25
Natural Scene	8×	25.74	0.793	157.26	0.226	0.39
Average	**2×**	**33.00**	**0.927**	**33.46**	**0.079**	**0.18**
Average	**4×**	**29.76**	**0.882**	**68.21**	**0.121**	**0.24**
Average	**8×**	**26.77**	**0.820**	**133.76**	**0.196**	**0.38**

Observations from the result

- The 2X output gives the best quality because the model needs to create fewer missing details.
- The 4X output gives a good balance between image clarity and processing time.
- The 8X output makes images much larger, but small details may become smooth or artificial.
- Portrait images usually give better results because faces have smoother and clearer patterns.
- Natural scenes are more difficult because grass, leaves, flowers, and mountains contain many fine details.
- Urban images show good edge improvement, but windows, bricks, and repeated building patterns can be difficult.
- PSNR and SSIM decrease as the scale factor increases, while MSE and LPIPS increase.
- Processing time also increases for 4X and 8X images because the output size is larger.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

Figure 2: Visual results of natural image super resolution at different scale factor

CONCLUSION

The proposed hybrid CNN Transformer GAN framework effectively enhances low-resolution natural images by improving sharpness, texture, colour quality, and overall visual appearance. CNN layers extract local details, Transformer blocks preserve global image structure, and GAN refinement produces more realistic textures. The system performs enhancement at 2X, 4X, and 8X scales, where 2X generally gives the highest reconstruction quality and 4X provides a practical balance between clarity and computational cost. Although 8X produces larger images, it is more challenging because more missing details must be estimated. The combined use of pixel, perceptual, edge, adversarial, and LPIPS losses helps maintain both numerical accuracy and visual quality. Overall, the system is suitable for photo restoration, mobile photography, surveillance enhancement, e-commerce images, and digital archives.

REFERENCES

1. Aakerberg, A., Johansen, A. S., Nasrollahi, K., & Moeslund, T. B. (2022). *Semantic segmentation guided real-world super-resolution*. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)**.** https://doi.org/10.1109/WACVW54805.2022.00051
2. Xu, X., Wei, P., Chen, W., Liu, Y., Mao, M., Lin, L., & Li, G. (2022). *Dual adversarial adaptation for cross-device real-world image super-resolution*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**,** 5657–5666.
3. Liu, A., Liu, Y., Gu, J., Qiao, Y., & Dong, C. (2023). *Blind image super-resolution: A survey and beyond*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45**(**5), 5461–5480.
4. Wei, P., Xie, Z., Li, G., & Lin, L. (2023). *Taylor neural network for real-world image super-resolution*. IEEE Transactions on Image Processing. https://doi.org/10.1109/TIP.2023.3255107
5. Zhang, W., Li, X., Shi, G., Chen, X., Qiao, Y., Zhang, X., Wu, X.-M., & Dong, C. (2023). *Real-world image super-resolution as multi-task learning*. In Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS)**.**
6. Wang, X., Liang, Z., Wang, Y., Yang, H., An, W., & Guo, Y. (2022). *Real-world light field image super-resolution via degradation modulation*.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

7. Wei, P., Sun, Y., Guo, X., Liu, C., Li, G., Chen, J., Ji, X., & Lin, L. (2023). *Towards real-world burst image super-resolution: Benchmark and method*. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)**,** 13187–13196.
8. Chen, X., Wang, X., Zhang, W., Kong, X., Qiao, Y., Zhou, J., & Dong, C. (2023). *HAT: Hybrid attention transformer for image restoration*.
9. Tian, C., Zhang, X., Zhu, Q., Zhang, B., & Lin, J. C.-W. (2022). *Generative adversarial networks for image super-resolution: A survey*.
10. Lin, W. (2022). *Single image super-resolution quality assessment: A real-world dataset, subjective studies, and an objective metric*. IEEE Transactions on Image Processing**, 31**, 2279–2294. https://doi.org/10.1109/TIP.2022.3154588
11. Deviyani, S., Hoplamaz, F., & Paul, M. (2022). *How real is real: Evaluating the robustness of real-world super-resolution*.
12. Babaguchi, N., & Aizawa, K. (2022). *Robust real-world image super-resolution against adversarial attacks*.