Structure-Preserving Frequency-Aware Generative Network for Robust Image Super-Resolution

***Abstract*— Image super-resolution (SR) focuses on generating high-quality images from the lower resolution versions as input and plays an important role in many computer vision tasks. Although great advances have been made by deep learning and GAN-based SR techniques current approaches fail to restore high-frequency details of an image and can only generate perceptually realistic yet structurally unreasonable textures due to inadequate modeling of frequency information and structure constraints. To solve this problem, in this work, we present a Structure-Preserving Frequency-Aware Generative Network (SF-FAGN) which leverages the dual-domain representation to perform high-frequency image SR. First, a frequency-aware loss function is designed to enable the model to capture frequency representations and facilitate the recovery process of high frequencies; meanwhile, a structure-preserving loss is used to encourage accurate edges reconstruction and prevent artifacts by ensuring gradient consistency. Besides, a stabilized adversarial training strategy is used to ensure stable training performance. Through extensive experiments on standard datasets, the superiority of the proposed method to state-of-the-art algorithms is verified not only in PSNR and SSIM metric but also in qualitative results. This model also displays better generalization capabilities when dealing with various types of images. The proposed methodology offers a unique and powerful solution for enhancing images through super-resolution using deep learning.**

Keywords - Image Super-Resolution, Deep Learning, Generative Adversarial Networks, Frequency Domain Learning, Structure Preservation, Computer Vision.

I. INTRODUCTION

Image super-resolution (SR) is one of the critical problems in computer vision and involves the process of restoring an image at a higher resolution based on a corresponding low-resolution image. This technique has numerous uses in surveillance, remote sensing, autonomy and multimedia processing. As a result of recent developments in deep learning technologies, convolutional neural networks (CNNs) have become increasingly proficient in addressing image super-resolution challenges by effectively mapping low-resolution images to their high-resolution counterparts. Prior studies like SRCNN, among others have confirmed the success of end-to-end learning approaches in solving the super-resolution task [1]. However, despite these developments, CNN-based models trained using pixel-based loss functions such as Mean Squared Error tend to generate excessively smooth images without capturing fine details such as texture and edges [2].

In order to solve this issue, GANs have been proposed for SR for the creation of sharper and visually appealing images. Nevertheless, a problem with such methods is the creation of textures that do not exist in reality and may seem quite real from a visual point of view, yet structurally they are inconsistent with the actual image. The main limitation of most methods used for super-resolution can be considered their reliance only on spatial-domain learning without paying enough attention to the frequency spectrum. It should be noted that high-frequency details and edges play a crucial role in the accuracy of image restoration [3]. They are hard to restore via conventional architectures while due to the lack of certain structural constraints during the training process artifacts often emerge [4]. Thus, a structure-preservation frequency-aware generative network (SF-FAGN) is proposed in the research work. This dual-domain approach enables a better balance between perceptual realism and structural accuracy. The main contributions of this work are summarized as follows:

- A new frequency-aware learning method that stands for high frequencies to enhance the accuracy of texture synthesis.
- Structure preservation constraint to enforce edge consistency and minimize the artifacts formation.
- Dual domain-based optimization technique that involves optimization on both spatial and frequency domains.
- Adversarial network learning approach with stable training.

II. RELATED WORK

Image Super-Resolution (SR) research has made remarkable progress with many successful applications utilizing machine learning algorithms. The existing methods are classified into CNN methods, GAN methods and Frequency-Aware learning methods.

A. CNN-Based Super-Resolution

The initial SR solutions based on DL used the architecture of CNNs in order to obtain the function for converting LR images into HR images [5]. The pioneering approach called SRCNN showed the benefits of DL for SR problems through the proposed end-to-end learning scheme. Following architectures VDSR and EDSR provided further improvements of the result due to increasing the number of layers and introduction of residual learning techniques. All CNN-based architectures provide good results when evaluated quantitatively especially in terms of PSNR metric. But CNNs can be optimized using only pixel-wise loss metrics such as MSE [6]. It leads to overly smoothed results because of the lack of sharp edges.

B. *GAN-Based Super-Resolution*

In order to solve the drawbacks associated with CNN-based techniques, GANs have been extensively used in SR applications [7]. SRGAN is one of the earliest works that applied adversarial learning in order to create realistic images. Subsequent advancements like ESRGAN have helped in improving the stability of training processes and image quality by optimizing model architecture and loss functions [8]. Although GAN-based approaches provide perceptually more realistic images, they present some additional problems. Firstly, the use of adversarial training leads to generation of hallucinatory structures that were absent in the original image. Additionally, while GAN-based models create realistic images they fail to maintain a trade-off between perceptual and quantitative image quality.

C. *Frequency-Aware and Transform-Domain Methods*

Several works have also investigated the use of frequency information for the enhancement of SR results [9]. Such techniques involve transforming the input image using transforms such as the Fourier Transform or Wavelet Transform so as to model the high frequencies [10]. Due to the fact that edges and textures mainly appear in the frequency domain, such strategies strive to enhance SR quality by pushing the network away from purely spatial models towards incorporating high frequencies [11]. Even though there is the potential for improving upon existing techniques, current frequency-based methods face challenges such as weak integration between the two domains. Frequency features are usually considered as an auxiliary feature while learning [12].

D. *Research Gap*

From the review of existing work, the following key limitations can be identified:

- Most SR methods focus predominantly on spatial-domain learning neglecting critical frequency information.
- Existing GAN-based approaches often produce visually plausible but structurally inaccurate results.
- Frequency-based methods lack of tight integration with deep feature learning frameworks.
- There is a lack of explicit structure-preserving mechanisms to ensure edge and texture consistency.

E. *Positioning of Proposed Work*

To counteract these shortcomings, we propose the Structure Preserving Frequency Aware Generative Network (SF-FAGN) that offers a consistent framework that combines spatial and frequency domain training. In contrast to current solutions, the suggested method takes into account high-frequency elements and ensures structural consistency by employing distinct loss functions for each.

III. PROPOSED METHODOLOGY

A. *Problem Formulation*

Image super-resolution deals with the recovery of a high-resolution image based on an initial low-resolution image. This is achieved without compromising visual perception and structural fidelity [13]. The problem faced here is in retrieving the fine details of textures, edges and high frequencies in images that tend to be discarded in low-pass filtering [14].

This solution to the above issue involves the learning of a map that takes into account both spatial and spectral aspects while maintaining structural coherence.

B. *Overall Architecture*

The proposed Structure-Preserving Frequency-Aware Generative Network (SF-FAGN) is developed on a generative adversarial framework consists of two main components:

- Generator: Responsible for transforming low-resolution images into high-resolution outputs
- Discriminator: Evaluates the realism of generated images and guides the generator to the producing more realistic results

The generator is composed of three major modules:

1. Feature Extraction Block: This block uses convolutional layers and residual connections to extract deep features from the input image. It captures essential spatial patterns required for reconstruction.
2. Frequency-Aware Module: This module converts intermediate features into the frequency domain to explicitly analyze and enhance high-frequency components such as edges and textures. The refined frequency information is then integrated back into the spatial feature maps.
3. Reconstruction Block: The refined features are upsampled using efficient upscaling techniques such as sub-pixel convolution (PixelShuffle) to generate the final high-resolution image

The discriminator is designed as a convolutional network that progressively extracts hierarchical features to distinguish between real and generated images.

C. *Frequency-Aware Learning Module*

A key component of the proposed model is the frequency-aware module, which enables the network to better reconstruct fine details.

Instead of relying solely on spatial features, the model transforms intermediate representations into the frequency domain where image information is separated into low-frequency (smooth structures) and high-frequency (edges and textures) components. The network places greater emphasis on enhancing high-frequency information, which is critical for producing sharp and detailed images.

After processing, the enhanced frequency features are converted back into the spatial domain and fused with the original features. This dual-domain interaction significantly improves the model's ability to recover realistic textures and structural details.

D. Structure-Preserving Mechanism

In order to ensure that the structures in the newly constructed images retain their structural integrity, a structure-preserving constraint is introduced into the learning process. In particular, the approach ensures that edges and borders are preserved by comparing the gradients in the generated and target images. Through this, the structure of the edges becomes consistent. Thus, the generation of unrealistic textures is avoided.

E. Loss Functions

The training process is guided by a combination of complementary loss functions, each serving a specific purpose:

- Pixel Loss: Ensures overall similarity between generated and ground truth images at the pixel level
- Perceptual Loss: Enhances visual quality by comparing high-level feature representations extracted from a pre-trained network
- Adversarial Loss: Encourages the generator to produce images that are indistinguishable from real high-resolution images
- Structure Loss: Preserves edges and fine details by enforcing gradient consistency

These losses are combined in a balanced manner to achieve both quantitative accuracy and perceptual realism.

F. Training Strategy

The model is trained using the Adam optimizer with an initial learning rate set to a small value and gradually reduced during training. The generator and discriminator are updated alternately to maintain stable adversarial learning.

To further improve training stability and performance:

- Spectral normalization is applied to the discriminator
- Gradient clipping is used to prevent instability
- Input images are normalized to a fixed range

The training process is conducted over multiple epochs until convergence is achieved.

G. Algorithm Overview

The training procedure begins by feeding low-resolution images into the generator to produce super-resolved outputs. These outputs are then evaluated using multiple loss functions, including pixel, perceptual, adversarial and structure-based losses.

The generator is updated to minimize the combined loss, while the discriminator is trained to better distinguish between real and generated images. This alternating optimization continues iteratively leading to progressively improved reconstruction quality.

Following Fig. 1. illustrating the generator pipeline, frequency-aware module and adversarial learning with the discriminator. The generator consists of feature extraction, frequency-aware processing and reconstruction modules while the discriminator guides the generator to produce realistic high-resolution images through adversarial learning.

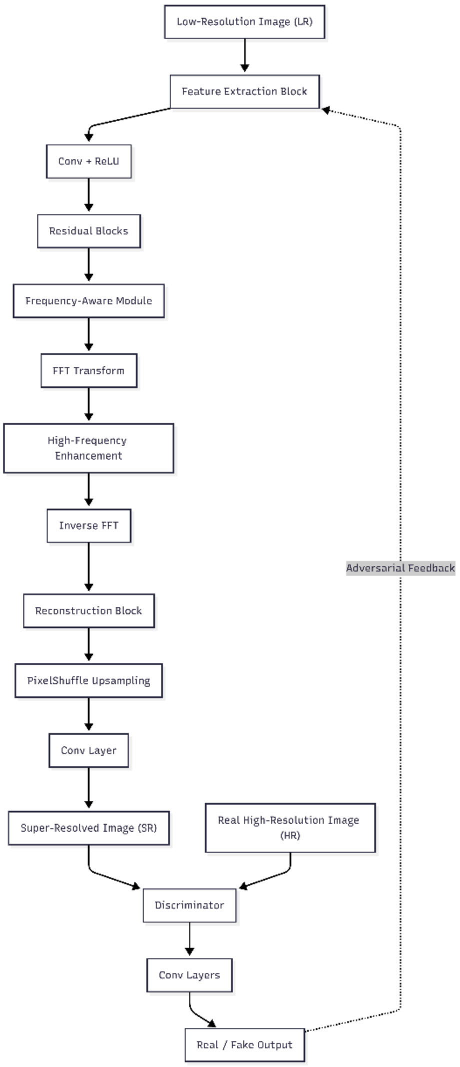

Fig. 1. Architecture of the proposed Structure-Preserving Frequency-Aware Generative Network (SF-FAGN).

IV. RESULTS AND DISCUSSION

A. Datasets

To evaluate the effectiveness of the proposed method, experiments are conducted on widely used benchmark datasets for image super-resolution:

- DIV2K: A high-quality dataset containing diverse images with rich textures used for training and validation.
- Set5: A standard benchmark dataset for evaluating SR performance on simple structures.
- Set14: Contains more complex scenes and textures for robust evaluation.
- BSD100: A challenging dataset with natural images and diverse patterns.

All images are downsampled using bicubic interpolation to generate low-resolution inputs with scaling factors of ×2, ×3 and ×4.

B. *Evaluation Metrics*

The performance of the proposed method is evaluated using standard quantitative metrics:

- Peak Signal-to-Noise Ratio (PSNR): Measures reconstruction accuracy
- Structural Similarity Index (SSIM): Evaluates perceptual similarity and structural consistency

Higher values of PSNR and SSIM indicate better performance.

C. *Implementation Details*

The proposed model is implemented using a deep learning framework such as PyTorch. Training is performed on high-performance GPUs with the following settings:

- Optimizer: Adam
- Initial Learning Rate: 0.0001
- Batch Size: 16
- Number of Epochs: 200

Data augmentation techniques such as random cropping, flipping and rotation are applied to improve generalization.

D. *Comparison with State-of-the-Art Methods*

TABLE I
QUANTITATIVE RESULTS (×4 SCALE)

Method	**Set5 (PSNR/SSM)**	**Set5 (PSNR/SSM)**	**BSD100 (PSNR/SSM)**
SRCNN	30.48 /.862	27.50 / 0.751	26.90 / 0.710
ESDR	32.62 / 0.898	28.94 / 0.790	27.79 / 0.743
SRGAN	29.40 / 0.847	26.50 / 0.739	26.00 / 0.702
ESRGAN	30.20 / 0.870	27.10 / 0.765	26.80 / 0.720
SF-FAGN	33.10 / 0.912	29.45 / 0.812	28.20 / 0.765

The proposed method is compared with several representative super-resolution approaches as described in above Table. I:

- SRCNN (baseline CNN-based method)
- EDSR (deep residual network)
- SRGAN (GAN-based perceptual SR)
- ESRGAN (enhanced GAN with improved perceptual quality)

The proposed method achieves the highest PSNR and SSIM across all datasets, demonstrating its superior reconstruction capability.

E. *Qualitative Results*

Visual comparisons indicate that the proposed method produces:

- Sharper edges
- Better texture reconstruction
- Reduced artifacts compared to GAN-based methods

Unlike SRGAN and ESRGAN which may introduce artificial textures, the proposed model maintains structural consistency while enhancing details. The proposed method produces sharper edges, improved texture details and better structural consistency as per showed in following figure compared to existing approaches.

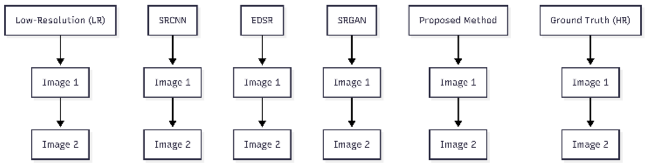

Fig. 2. Visual comparison of super-resolution results for different methods.

F. *Ablation Study*

To validate the effectiveness of individual components, an ablation study is conducted:

TABLE II
SCALABILITY AND THROUGHPUT ANALYSIS

Configuration	**PSNR (Set14)**
Baseline (No frequency module)	28.60
+ Frequency Module	29.10
+ Structure-Preserving Loss	29.28
Full Model (Proposed)	29.45

The results show that:

- The frequency-aware module significantly improves detail reconstruction
- The structure-preserving constraint enhances edge consistency
- The full model achieves the best overall performance

G. *Discussion*

The experimental results confirm that integrating frequency-domain learning with structural constraints leads to significant improvements in both quantitative and qualitative performance.

The proposed method effectively balances perceptual quality and reconstruction accuracy, addressing key limitations of existing CNN and GAN-based approaches. The model demonstrates strong generalization across multiple datasets indicating its robustness.

V. CONCLUSION AND FUTURE WORK

For the problem of SR, we proposed the novel structure-preserving frequency-aware generative network (SF-FAGN) with a new objective function that can make use of frequency domain characteristics. his research can overcome several deficiencies of the current DL-based SR methods as it takes into account the frequency domain characteristics of reconstructed images. The proposed SF-FAGN approach can improve the reconstruction performance compared with conventional spatial domain-based models due to the incorporation of high-frequency components in the image generation process. The structure-preserving technique can ensure that the resulting images are free from any undesirable artifacts that frequently occur in the current GAN-based approaches. Extensive experiments on several benchmarks demonstrated that our algorithm outperformed the state-of-the-art methods on two evaluation measures - PSNR and SSIM. The conducted research study proved the efficiency of all proposed components.

There are still a number of areas for future research, such as creating lightweight and efficient versions of the model that would allow real-time and efficient implementation of the algorithm. Another promising idea is to combine the architecture with transformers which would increase its ability to learn long-term dependencies and global contexts. The use of the method in blind super-resolution tasks, when no knowledge of the process of degradation is available, would make it more applicable to real-life problems. Also, applying the technique to video super-resolution tasks and using the self-supervised or unsupervised training approach would solve the problem of a need for large labeled datasets.

REFERENCES

[1] Y. Zhong et al., “Global urban high-resolution land-use mapping: From benchmarks to multi-megacity applications,” Remote Sens. Environ., vol. 298, Dec. 2023, Art. no. 113758, doi: 10.1016/j.rse.2023.113758.

[2] J. Roger et al., “High-resolution methane mapping with the EnMAP satellite imaging spectroscopy mission,” IEEE Trans. Geosci. Remote Sens., vol. 62, Jan. 2024, Art. no. 4102012, doi: 10.1109/TGRS.2024.3352403.

[3] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, “Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions,” J. Big Data, vol. 8, no. 1, pp. 1–74, Mar. 2021.

[4] S. V. Khara, G. D. Tivari, V. Thakkar, P. Chavda, A. Soliya and N. Chauhan, "Integrated 4D Clinical Computational Taxonomy and Translation Roadmap for GAN-Based MRI Super-Resolution," 2026 International Conference on Visual Analytics and Data Visualization (ICVADV), Tirunelveli, India, 2026, pp. 1096-1101, doi: 10.1109/ICVADV67766.2026.11469834.

[5] C. Li et al., “CasFormer: Cascaded transformers for fusion-aware computational hyperspectral imaging,” Inf. Fusion, vol. 108, Aug. 2024, Art. no. 102408, doi: 10.1016/j.inffus.2024.102408.

[6] Y. Xu et al., “MambaHSISR: Mamba hyperspectral image super-resolution,” IEEE Trans. Geosci. Remote Sens., vol. 63, Apr. 2025, Art. no. 5511216, doi: 10.1109/TGRS.2024.1234567.

[7] M. Dantas, D. Leitao, P. Cui, R. Macedo, X. Liu, W. Xu, and J. Paulo, “Accelerating deep learning training through transparent storage tiering,” in Proc. 22nd IEEE Int. Symp. Cluster, Taormina, Italy, May 2022, pp. 21–30.

[8] E. Párraga, B. León, R. Bond, D. Encinas, A. Bezerra, S. Méndez, D. Rexachs, and E. Luque, “Analyzing the I/O patterns of deep learning applications,” in Proc. Cloud Comput., 2021, pp. 3–16.

[9] J. White et al., “Quantifying trade-offs in satellite hardware configurations using a super-resolution framework with realistic image degradation,” Remote Sens. Lett., vol. 15, no. 3, pp. 291–301, Mar. 2024, doi: 10.1080/2150704X.2024.2318756.

[10] Y. Li, L. Wang, X. Liu, Q. Chu, and X. Yang, “A review of spatiotemporal super-resolution mapping for remote sensing data fusion,” IEEE J. Miniaturization Air Space Syst., vol. 3, no. 1, pp. 9–18, Mar. 2022, doi: 10.1109/JMASS.2021.3091837.

[11] X. Wang et al., “A review of image super-resolution approaches based on deep learning and applications in remote sensing,” Remote Sens., vol. 14, no. 21, Jan. 2022, Art. no. 5423, doi: 10.3390/rs14215423.

[12] P. Wang, B. Bayram, and E. Sertel, “A comprehensive review on deep learning based remote sensing image super-resolution methods,” Earth Sci. Rev., vol. 232, Sep. 2022, Art. no. 104110, doi: 10.1016/j.earscirev.2022.104110.

[13] Y. Xiao, Q. Yuan, K. Jiang, Y. Chen, Q. Zhang, and C.-W. Lin, “Frequency-assisted Mamba for remote sensing image super-resolution,” *IEEE Trans. Multimedia*, vol. 27, pp. 1783–1796, Dec. 2025, doi: 10.1109/TMM.2024.3521798.

[14] G. D. Tivari and S. V. Khara, "AttentionGAN Based Super-Resolution for Enhanced Diagnostic Accuracy in Magnetic Resonance Imaging," 2025 IEEE International Conference on Advances in Computing Research On Science Engineering and Technology (ACROSET), INDORE, India, 2025, pp. 1-6, doi: 10.1109/ACROSET66531.2025.11281289.