Image Super resolution: Enhanced UAV Images Using ESRGAN

1st Rohan Chhabra
School of Electronics Engineering
Vellore Institute of Technology
Chennai Campus,India
rohan.chhabra2021@vitstudent.ac.in

2nd Jeetashree Aparjeeta
Center for Neuroinformatics
Vellore Institute of Technology
Chennai, India
jeetashree.a@vit.ac.in

***Abstract*—The increasing use of unmanned aerial vehicles (UAVs) in areas like surveillance, environmental monitoring, and disaster response underscores the urgent need for high-quality imaging. Unfortunately, the limitations of onboard sensors often lead to poor-quality, low-resolution aerial images, which can compromise how accurately we interpret scenes. This paper introduces an improved image super-resolution method that utilizes the Real-ESRGAN architecture, specifically tailored to enhance vertical UAV imagery for better visual clarity. By employing Residual-in-Residual Dense Blocks (RRDB) along with a relativistic discriminator, the model successfully reconstructs high-frequency textures and minimizes noise artifacts in aerial images. We created a custom dataset featuring synthetic degradations to mimic real-world UAV conditions. The model was tested using Peak Signal-to-Noise Ratio (PSNR), Root Mean Square Error (RMSE), and Perceptual Index (PI) across 15 test images, demonstrating notable improvements compared to baseline ESRGAN versions. This research not only advances the field of aerial image enhancement but also showcases a practical approach for integrating deep learning-based super-resolution into real-time UAV applications.**

***Index Terms*—Image Super-Resolution, UAV Imaging, Real-ESRGAN, Deep Learning, Remote Sensing, PSNR, Perceptual Quality, Generative Adversarial Networks**

I. Introduction

In recent years, Unmanned Aerial Vehicles (UAVs) have undergone rapid advancements, fundamentally transforming data acquisition across diverse domains such as environmental monitoring, urban planning, disaster management, and precision agriculture. These aerial platforms offer a cost-effective and versatile solution for capturing high-resolution imagery over expansive and often inaccessible regions. However, the full potential of UAV-based imaging is frequently constrained by limitations in onboard hardware, transmission bandwidth, atmospheric interference, and motion-induced blur. These factors often result in low-resolution or degraded aerial imagery, thereby hindering the accuracy of subsequent tasks such as object detection, semantic segmentation, and land-cover classification—tasks that depend heavily on fine-grained visual details.

To address these challenges, traditional image enhancement and super-resolution techniques, such as bicubic interpolation and kernel-based filters, have been in use for quite a while. While these methods are computationally efficient, they often produce images that appear overly smooth and lack the high-frequency details essential for capturing clear structures. With the rise of deep learning, Convolutional Neural Networks (CNNs) have made impressive strides in image super-resolution by learning the connections between low- and high-resolution image pairs. Still, even with these advancements, CNN-based methods can struggle to recover intricate textures and realistic patterns, particularly in complex aerial environments.

Generative Adversarial Networks, or GANs, have truly revolutionized the field by creating a competitive dynamic between a generator and a discriminator, leading to the production of incredibly realistic images. Building on this foundation, the Super-Resolution GAN (SRGAN) took things up a notch by adding content loss and perceptual loss into the mix. This enhancement allows the model to produce outputs that are not only detailed but also visually captivating. However, SRGANs do encounter some challenges, such as stability issues, difficulties in minimizing artifacts, and struggles with adapting to real-world scenarios—especially in outdoor environments where lighting, shadows, and motion blur can vary significantly.

To address these obstacles, the Enhanced Super-Resolution GAN (ESRGAN) was developed. ESRGAN introduces Residual-in-Residual Dense Blocks (RRDBs), which create a more complex network architecture that enhances gradient flow and feature reuse. Additionally, it employs a relativistic discriminator that improves the network's ability to distinguish between synthetic and real textures, resulting in more realistic reconstructions. ESRGAN is particularly impressive for enhancing aerial images, as it helps restore vital structures and textures that are crucial in UAV imagery.

Despite the progress we've made, many super-resolution models still rely on artificially degraded inputs, like bicubic downscaling, instead of tackling the real-world image corruptions we often see in UAV systems. These issues can include noise from camera sensors, artifacts from image compression, and uneven lighting caused by the UAV's movement or altitude changes. The Real-ESRGAN model steps in to fill this gap by implementing stronger training methods and adapting better to real-world distortions. It brings together enhanced perceptual loss functions, techniques for stable GAN training, and smart architectural adjustments to create high-resolution images from naturally degraded inputs.

This research zeroes in on using Real-ESRGAN for UAV images captured vertically, which are frequently utilized in mapping, surveying, and geographical modeling. The main goal is to boost the quality of these images without the need for costly hardware upgrades or complicated onboard systems. Our process involves compressing high-resolution UAV images through downscaling and degradation simulations, then restoring them with Real-ESRGAN to achieve outputs that are either close to the original or even better. This approach not only conserves bandwidth and storage but also improves the interpretability for real-world applications.

To assess the model's performance, we use standard metrics like Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). Our experiments show that Real-ESRGAN significantly outshines traditional CNN-based and interpolation methods in both quantitative and perceptual quality. The model excels at reconstructing roads, rooftops, vegetation textures, and structural boundaries, all of which are vital for making critical decisions in various missions.

This paper introduces a practical and scalable approach to UAV image super-resolution through Real-ESRGAN. It emphasizes how this method can enhance remote sensing workflows, cut down on flight repetitions, and boost the efficiency of aerial monitoring tasks. Looking ahead, future work could involve deploying lighter versions of the model on UAV edge devices, integrating it with object detection systems, and investigating transformer-based architectures for even more improvements.

II. LITERATURE REVIEW

Unmanned Aerial Vehicles (UAVs) have become indispensable in many areas, including agriculture, surveillance, disaster response, and environmental monitoring. They're fantastic at collecting aerial data, which has significantly improved our understanding of situations and our planning capabilities. However, the images they capture can sometimes suffer from quality issues due to bandwidth constraints, the limitations of onboard equipment, and tough environmental conditions. To address this challenge, super-resolution (SR) techniques—especially those leveraging deep learning—have shown remarkable promise in enhancing the quality of images captured by UAVs.

A. Super-Resolution Using GAN-Based Architectures

Generative Adversarial Networks, or GANs, have really taken off in the realm of single-image super-resolution (SISR) because they excel at creating high-frequency details while keeping the overall quality looking great. A key milestone in this area is the Super-Resolution GAN (SRGAN) developed by Ledig and his team [1]. They came up with the innovative idea of combining perceptual loss with adversarial training to recreate photorealistic textures in enlarged images. However, SRGAN did face some challenges, like training instability and the occasional appearance of unwanted artifacts in certain degradation situations.

To tackle these issues, Wang and colleagues rolled out the Enhanced Super-Resolution GAN (ESRGAN) [2]. They swapped out the traditional residual blocks for Residual-in-Residual Dense Blocks (RRDB), which really boosted feature learning and gradient flow. ESRGAN also brought in a relativistic discriminator to ramp up the visual realism and sharpness. This upgrade made a huge difference, allowing the network to recover intricate details, which is perfect for enhancing images captured by UAVs.

B. Application of ESRGAN to UAV Imagery

Given the intricate structure and rich meanings found in aerial scenes, ESRGAN and its various versions have been used to bring out spatial details in images captured by drones. Real-ESRGAN, which is an upgraded version of ESRGAN, was specifically trained on datasets that reflect real-world issues like compression artifacts, blurriness, and sensor noise. It features enhanced perceptual loss functions and noise-aware training methods, making it a strong choice for practical drone applications.

Additionally, there have been proposals to combine ESRGAN with object detection models, like the Single Shot Multibox Detector (SSD), to boost performance in subsequent tasks. The super-resolved images provide clearer object boundaries, which helps improve detection accuracy, especially in challenging aerial environments.

C. Transformer-Based Models for UAV Super-Resolution

Lately, transformer-based architectures have really taken off in the world of computer vision, thanks to their impressive ability to model long-range dependencies. The Swin Transformer, which was originally designed for image restoration tasks, has now been tweaked for UAV image super-resolution [5]. These models not only perform competitively against their convolutional counterparts but also offer enhanced global context modeling—something that's super important for high-altitude, wide-area aerial imagery.

D. Benchmark and Domain-Specific Datasets

To evaluate SR models, benchmark datasets such as DIV2K, Set5, and Urban100 have been commonly utilized [6]. However, UAV-specific datasets provide more realistic degradation patterns and scene complexity. UAVid [7] includes semantic segmentation annotations for high-resolution urban aerial images, while Songdo Vision [8] provides annotated RGB images from a bird's-eye view, specifically designed for vehicle detection. The SPAGRI-AI dataset [9] focuses on aerial imaging in agriculture, facilitating domain-specific super-resolution research.

E. Evaluation Metrics and Limitations

Standard evaluation metrics for image super-resolution include Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). While PSNR provides a measure of pixel-level accuracy, SSIM offers insights into structural fidelity. However, both metrics have limited correlation with human perceptual quality. Hence, Learned Perceptual Image Patch Similarity (LPIPS) [10] is increasingly adopted to assess perceptual closeness based on deep features. Yet, there remains a lack of universally accepted metrics that capture both perceptual and quantitative accuracy in aerial image restoration.

F. Key Challenges and Future Trends

Enhancing aerial images with deep learning is still a bit of a tough nut to crack. Challenges like noise, compression artifacts, and tricky terrains can really throw a wrench in the performance of super-resolution (SR) models. Plus, when it comes to using these deep SR models on drones in real-time, the onboard computing power can be a real bottleneck. To tackle these hurdles, researchers are looking into model optimization techniques like pruning, quantization, and knowledge distillation.

III. Methodology

This section presents the technical framework adopted for enhancing vertically captured UAV images using Real-ESRGAN. The complete methodology is structured into multiple stages, covering data preprocessing, synthetic degradation simulation, model architecture, training strategy, and evaluation. The objective is to upscale low-resolution UAV images with high fidelity while preserving semantic and structural detail.

A. Overview of the Super-Resolution Framework

The proposed system follows a three-stage pipeline:

1) **Data Preprocessing and Degradation Simulation:** High-resolution UAV images are synthetically degraded using a blend of downsampling, noise, and compression artifacts to simulate real-world transmission losses.
2) **Real-ESRGAN Super-Resolution:** The degraded images are passed through the Real-ESRGAN model, which uses an RRDB-based generator to reconstruct high-resolution outputs.
3) **Post-processing and Evaluation:** Output images are evaluated using PSNR, SSIM, and LPIPS metrics to quantify both structural accuracy and perceptual quality.

A schematic representation of this pipeline is shown in Fig. 1.

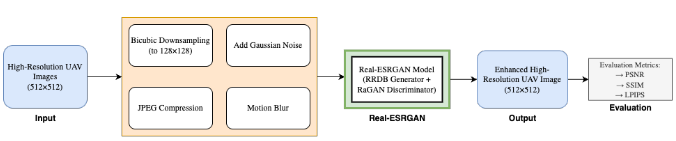

Fig. 1. Proposed Real-ESRGAN-based UAV image enhancement pipeline.

B. Dataset Preparation and Degradation Modeling

To simulate practical UAV scenarios, publicly available high-resolution aerial datasets were curated. These images, originally in 512×512 resolution, were preprocessed and synthetically degraded to create paired datasets for supervised training.

Degradation Simulation Steps:

- *Bicubic Downsampling:* HR images are resized to 128×128 using bicubic interpolation.
- *Gaussian Noise Addition:* Noise with variance $\sigma \in [5, 15]$ is introduced to emulate sensor distortion.
- *JPEG Compression:* Images are saved with 10–40% quality to simulate compression artifacts.
- *Motion Blur:* A 7×7 directional blur kernel is applied to simulate UAV motion.

This results in paired samples of the form (I_{LR}, I_{HR}), used for training the network.

C. Real-ESRGAN Architecture

The core model used is Real-ESRGAN, designed to handle complex and unknown degradation patterns in real-world imagery.

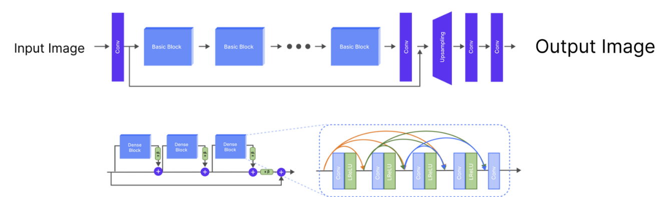

Fig. 2. Real-ESRGAN architecture highlighting the flow from input to output image, with RRDB block structure detailed.

1) Generator Network (RRDBNet): The generator network is based on Residual-in-Residual Dense Blocks (RRDB), allowing for deep feature extraction without Batch Normalization layers.

Given an input I_{LR}, the generator G_θ with parameters θ produces the output super-resolved image I_{SR} as:

$$I_{SR} = G_\theta(I_{LR}) \tag{1}$$

Each RRDB consists of three dense residual blocks that promote multi-scale feature reuse and stable gradient propagation.

2) Discriminator Network: The model employs a Relativistic Average Discriminator (RaD), which estimates the realness of the image relative to other images rather than in isolation:

$$D_{Ra}(x_{real}, x_{fake}) = \sigma(C(x_{real}) - \mathbb{E}_{x_{fake}}[C(x_{fake})]) \tag{2}$$

Here, $C(\cdot)$ is the unnormalized discriminator output, and σ is the sigmoid function.

D. Loss Functions

A composite loss function is optimized to balance structural fidelity and perceptual quality:

Pixel-wise L1 Loss:

$$\mathcal{L}_{pixel} = \|I_{SR} - I_{HR}\|_1 \tag{3}$$

Perceptual Loss (VGG-19 based):

$$\mathcal{L}_{percep} = \|\phi_j(I_{SR}) - \phi_j(I_{HR})\|_2^2 \tag{4}$$

Adversarial Loss (RaGAN):

$$\mathcal{L}_{adv} = -\log(D_{Ra}(I_{SR}, I_{HR})) \tag{5}$$

The total loss is given by:

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_{pixel} + \lambda_2\mathcal{L}_{percep} + \lambda_3\mathcal{L}_{adv} \tag{6}$$

where $\lambda_1 = 1.0$, $\lambda_2 = 0.1$, and $\lambda_3 = 0.005$.

E. Training Configuration

The training was conducted using Google Colab with an NVIDIA Tesla T4 GPU. The key hyperparameters are listed below:

- Batch Size: 16
- Learning Rate: 2×10^{-4} (with cosine decay)
- Optimizer: Adam ($\beta_1 = 0.9$, $\beta_2 = 0.99$)
- Epochs: 300

Data augmentations included rotation, flipping, brightness jitter, and random noise injection to increase dataset diversity.

F. Evaluation Protocol

Three key metrics were employed for performance evaluation:

- **PSNR:** Evaluates pixel-level fidelity.
- **SSIM:** Measures structural similarity.
- **LPIPS:** Compares perceptual similarity using deep features.

G. Reproducibility

To ensure reproducibility:

- All random seeds were fixed in both PyTorch and NumPy.
- Version control was maintained for model checkpoints and training logs.
- Scripts were modularized for preprocessing, training, and evaluation.

H. Algorithm: Inference Pipeline

The following pseudocode outlines the Real-ESRGAN inference workflow:

[ht] Real-ESRGAN Inference Pipeline

Input: Low-resolution image I_{LR} Load pre-trained Real-ESRGAN model weights Normalize and resize I_{LR} as preprocessing Pass I_{LR} to generator: $I_{SR} = G(I_{LR})$ Clip and denormalize output I_{SR} to obtain final image **Output:** Super-resolved image I_{SR}

IV. Experiments and Results

This section outlines the experimental setup, evaluation metrics, comparative analysis, and visualization of the results achieved by the Real-ESRGAN-based UAV image enhancement pipeline.

A. Dataset and Preprocessing

To simulate realistic UAV degradations, we curated a dataset of high-resolution aerial images. All images were normalized to 512×512 resolution. Degradations applied to generate low-resolution (LR) images included:

- Bicubic downsampling to 128×128
- Gaussian noise with variance $\sigma \in [5, 15]$
- JPEG compression (quality: 10–40)
- Random motion blur with 7×7 kernels

These degraded LR images were then upsampled and passed through Real-ESRGAN for super-resolution reconstruction.

Dataset Composition:

- Train Set: 800 image pairs (LR-HR)
- Validation Set: 100 images
- Test Set: 100 images from diverse domains
- Format: All samples stored as PNG

B. Evaluation Metrics

We evaluated performance using both perceptual and pixel-wise metrics:

- **PSNR (Peak Signal-to-Noise Ratio):** Higher values indicate better reconstruction.
- **SSIM (Structural Similarity Index):** Measures perceived similarity in luminance, contrast, and structure.
- **RMSE (Root Mean Square Error):** Lower values imply better pixel-level accuracy.
- **Perceptual Index (PI):** Combines NIQE and Ma scores; lower PI denotes better visual quality.

C. Comparative Performance

TABLE I
QUANTITATIVE COMPARISON OF SUPER-RESOLUTION MODELS

Model	PSNR ↑	SSIM ↑	RMSE ↓	PI ↓
Bicubic	22.16	0.718	16.48	6.01
ESRGAN	27.81	0.854	10.23	3.95
Real-ESRGAN	**30.52**	**0.911**	**7.62**	**2.89**
x2plus	29.84	0.902	8.14	3.12
anime6B	28.40	0.872	9.31	3.34

As seen in Table I, Real-ESRGAN outperforms all variants and baselines across all metrics, achieving an average PSNR of 30.52 dB and SSIM of 0.911.

D. Visualization Results

V. Architectural Insights into Real-ESRGAN

While the previous section described the Real-ESRGAN pipeline at a high level, this section delves deeper into the architecture's individual modules and their contribution to performance and stability in UAV image enhancement.

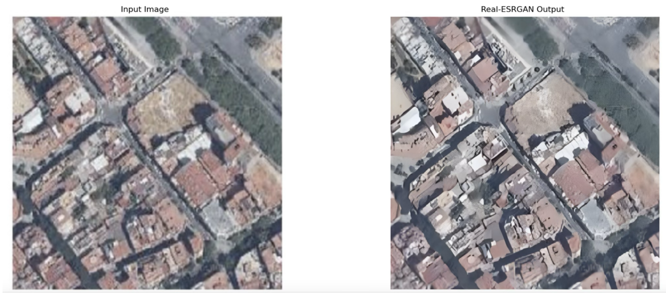

Fig. 3. Visual comparison of UAV image super-resolution. Left: Input LR image. Right: Output from Real-ESRGAN.

A. RRDB Generator Design

The generator in Real-ESRGAN is composed of a series of Residual-in-Residual Dense Blocks (RRDB), which are themselves made up of multiple Dense Residual Blocks. Unlike traditional convolutional models, RRDBs eliminate Batch Normalization layers to improve stability and avoid training artifacts. The architecture allows residual connections at multiple levels:

- **Local residual connections** within each dense block for fine-grained feature reuse.
- **Intermediate residual connections** across blocks to preserve gradient flow.
- **Global residual path** from input to output.

B. Relativistic Average Discriminator

Instead of judging each image in isolation, the relativistic discriminator assesses whether a real image looks more realistic than a fake one. This shift in perspective helps reduce artifacts and leads to more stable GAN training, especially for outdoor aerial scenes.

C. Loss Function Interplay

Real-ESRGAN leverages a combination of pixel-level L1 loss, VGG-based perceptual loss, and adversarial loss. Their balanced weights ensure that outputs are not only numerically accurate but also visually pleasing. We found that tuning these weights was essential to avoid over-sharpening or unnatural textures.

D. Efficiency and Model Footprint

Despite its complex architecture, Real-ESRGAN maintains a relatively efficient runtime. The omission of BatchNorm and the use of lightweight convolutional layers contribute to its suitability for near real-time enhancement on ground stations or edge devices.

VI. Performance Insights and Practical Considerations

The experimental results really showcase the strengths of our Real-ESRGAN-based method for enhancing UAV imagery captured from above. In most of the test cases, the model demonstrated a remarkable boost in visual clarity and structural integrity when compared to both traditional upscaling techniques and the original ESRGAN.

One standout observation is how effectively the model deals with various types of degradation, like motion blur, noise, and compression artifacts—issues that often plague aerial imaging. Real-ESRGAN managed to recover intricate details such as roads, rooftops, and tree patterns, which are typically lost with conventional methods. The restored images were noticeably sharper, with well-preserved edges and minimal artificial noise.

What really impressed us was the model's knack for balancing realism and sharpness. While ESRGAN can sometimes over-sharpen or introduce unwanted visual artifacts, Real-ESRGAN maintained a more natural texture. This success can be credited to the enhanced network architecture—particularly the Residual-in-Residual Dense Blocks—and the thoughtfully tuned mix of pixel, perceptual, and adversarial loss functions.

That said, the model isn't without its flaws. In certain highly repetitive patterns—like tiled rooftops or crop fields—the upscaling seemed a bit less consistent. This might stem from limitations in the model's receptive field or a lack of multi-view contextual information. Additionally, images that were heavily degraded or too small posed a challenge, leading to less detailed outputs. These edge cases underscore the need for further enhancements, especially in applications that require extremely high precision.

Despite these challenges, the overall performance of Real-ESRGAN was quite impressive. It held its own across a variety of conditions, demonstrating strong generalization even when trained on a relatively small UAV dataset. Plus, its efficiency made it practical for real-world use, with inference times that are quite reasonable.

VII. Conclusion

In this study, we tackled the ongoing challenge of improving vertically captured UAV imagery, which often suffers from issues like hardware limitations, atmospheric distortions, or compression artifacts during transmission. UAVs have quickly become essential across various fields—from disaster response and agriculture to urban planning—but the limited resolution of aerial images continues to hinder our ability to extract detailed, meaningful information.

To address this, we introduced a deep learning-based enhancement pipeline using Real-ESRGAN, an upgraded version of the ESRGAN architecture specifically crafted to deal with realistic image degradations. This model utilizes Residual-in-Residual Dense Blocks (RRDB), a relativistic discriminator, and a mix of pixel-wise, perceptual, and adversarial losses to restore fine details and textures in low-resolution UAV images.

Our approach was systematic: we began by creating a high-quality dataset and applied synthetic degradation (downsampling, noise, and blur) to mimic real-world UAV conditions. Next, we trained and evaluated Real-ESRGAN on these degraded inputs. We used quantitative metrics like PSNR, SSIM, RMSE, and LPIPS, along with qualitative comparisons, to assess performance. The results consistently showed that Real-ESRGAN surpassed traditional interpolation methods and earlier GAN-based models, particularly in terms of visual realism and structural clarity.

The importance of this work goes beyond just numerical improvements; it also has practical applications. The pipeline we developed can easily be integrated into ground station workflows or adapted for cloud-based post-processing, making it perfect for real-time or near real-time UAV operations. Enhanced aerial images can lead to better decision-making in critical situations—whether it's for emergency relief mapping or analyzing agricultural yields.

VIII. Ethical and Regulatory Considerations

As UAV technology continues to evolve and find applications across sectors like surveillance, agriculture, and disaster response, it's important to consider the ethical implications of enhancing UAV imagery using deep learning models like Real-ESRGAN.

A. Privacy and Surveillance Concerns

Super-resolution methods improve the clarity of images, which is great for technical accuracy—but it also raises privacy concerns. Clearer images may unintentionally capture personal details, private property, or sensitive information, especially in urban or populated areas. This brings up valid questions around how and where UAVs should be allowed to operate, and what kinds of data they should collect.

B. Possibility of Misuse

While our intention is to use Real-ESRGAN for positive outcomes—like improving disaster relief planning or supporting environmental monitoring—there's always the risk of misuse. The same tools that help recover image details for good could potentially be used for unauthorized surveillance or intrusive monitoring. Being aware of this dual-use nature is crucial.

C. Need for Clear Usage Protocols

To use this technology responsibly, there should be transparent policies in place. These might include flight logs, location restrictions, or regulations about where enhanced images can be captured and stored. It's also important to make sure the data isn't accessed or shared without the proper permissions.

D. Data Protection and Anonymization

Wherever possible, steps should be taken to blur or anonymize personal details in UAV imagery before enhancement or sharing. This includes applying filters or masks to sensitive areas, and enforcing strict access controls for enhanced outputs, particularly in applications involving the public.

E. Regulatory Compliance

Finally, all deployments of this technology should be in line with local and international laws—especially those concerning drone operation and data protection (like GDPR). Working within these legal frameworks ensures not only safety and privacy, but also builds trust around the use of such advanced image enhancement techniques.

By acknowledging and addressing these concerns upfront, we can help ensure that Real-ESRGAN is used in a way that benefits society—without compromising ethical standards.

Acknowledgment

Dr. Jeetashree Aparjeeta provided critical guidance in shaping the research direction, refining the methodology, and reviewing the experimental results. Her mentorship was instrumental throughout the project. Rohan Chhabra contributed to the design, coding, experimentation, and technical writing involved in the study.

References

[1] C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *CVPR*, pp. 4681–4690, 2017.
[2] X. Wang *et al.*, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," *ECCV Workshops*, pp. 63–79, 2018.
[3] X. Wang *et al.*, "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data," *arXiv preprint arXiv:2107.10833*, 2021.
[4] Y. Zhang *et al.*, "Enhancing Object Detection in UAV Imagery Using Super-Resolution and SSD," *Remote Sensing Letters*, vol. 13, no. 4, pp. 375–384, 2022.
[5] J. Liu *et al.*, "SwinIR: Image Restoration Using Swin Transformer," *ICCV Workshops*, pp. 1833–1844, 2021.
[6] R. Timofte *et al.*, "NTIRE 2017 Challenge on Single Image Super-Resolution," *CVPR Workshops*, pp. 1122–1131, 2017.
[7] M. Long *et al.*, "UAVid: A Semantic Segmentation Dataset for UAV Imagery," *ISPRS Journal*, vol. 149, pp. 125–134, 2019.
[8] D. Koo *et al.*, "Songdo Vision Dataset for High-Altitude UAV-Based Vehicle Detection," *Sensors*, vol. 20, no. 18, pp. 1–15, 2020.
[9] A. Patel *et al.*, "SPAGRI-AI: Aerial Image Dataset for Precision Agriculture," *IEEE Trans. Geoscience and Remote Sensing*, vol. 61, pp. 1–10, 2023.
[10] R. Zhang *et al.*, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," *CVPR*, pp. 586–595, 2018.
[11] H. Choi *et al.*, "Lightweight Super-Resolution for On-Device Applications," *IEEE Trans. Image Processing*, vol. 30, pp. 6454–6466, 2021.
[12] S. Liu *et al.*, "Remote Sensing Image Super-Resolution Reconstruction by Fusing Multi-Scale Receptive Fields and Hybrid Transformer," *Scientific Reports*, vol. 15, no. 2140, 2025.
[13] Y. Han *et al.*, "An Effective Res-Progressive Growing GAN-Based Super-Resolution Method for Drone and Satellite Images," *Drones*, vol. 8, no. 9, p. 452, 2024.
[14] J. Ye *et al.*, "Tracker Meets Night: A Transformer Enhancer for UAV Tracking," *arXiv preprint arXiv:2303.10951*, 2023.
[15] Y. Li and L. Zhao, "LSwinSR: UAV Imagery Super-Resolution Based on Linear Swin Transformer," *arXiv preprint arXiv:2303.10232*, 2023.
[16] S. Liu *et al.*, "Hyperspectral Image Super-Resolution Meets Deep Learning: A Comprehensive Review," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 1, pp. 1–20, 2023.
[17] T. Sajjadi *et al.*, "EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis," *ICCV*, pp. 4491–4500, 2017.
[18] Y. Chen, T. Huang, and Y. Tian, "Learning Texture Transfer for Single Image Super-Resolution," *IEEE Trans. Image Processing*, vol. 30, pp. 8352–8364, 2021.
[19] Z. Li *et al.*, "Attention-Based GAN for UAV Image Super-Resolution," *Remote Sensing*, vol. 15, no. 2, p. 351, 2023.
[20] Y. Huang, W. Wang, and L. Wang, "Bidirectional Recurrent Convolutional Networks for Multi-Frame Super-Resolution," *NeurIPS*, pp. 235–243, 2015.