Grenze International Journal of Engineering and Technology, January Issue

Image Generation and Super Resolution using Custom GAN's

Prof. Dr. Milind Kulkarni[1], Anish Naphade[2], Alankar Jagtap[3], Arman Tamboli[4], Aniket Ambilwade[5] and Ashlesha Wagh[6]
[1-6]Department of Artificial Intelligence and Data Science, Vishwkarma Institute of Technology, Pune, India
Email: milind.kulkarni@vit.edu, anish.naphade22@vit.edu, alankar.jagtap22@vit.edu, arman.tamboli22@vit.edu, aniket.ambilwade22@vit.edu, ashlesha.wagh22@vit.edu

***Abstract*— In recent years, Generative Adversarial Networks (GANs) have emerged as a powerful framework for image synthesis and enhancement. This paper presents a hybrid approach that combines a Deep Convolutional GAN (DCGAN) for generating low-resolution lunar surface images and an Enhanced Super-Resolution GAN (ESRGAN) for refining these outputs into high-resolution, photorealistic visuals. Our proposed architecture leverages the robust feature extraction capability of DCGANs alongside the perceptual fidelity of ESRGANs to produce visually convincing lunar imagery. The results demonstrate significant improvements in both structural accuracy and visual sharpness, enabling a deeper exploration into data-driven lunar visualization. Quantitative metrics such as PSNR and SSIM validate the quality enhancement, and qualitative results affirm the model's efficacy in preserving fine details. This dual-GAN pipeline underscores the potential of integrating generative and super-resolution models for scientific and artistic applications in space imagery.**

***Keywords*— DCGAN, ESRGAN, lunar imagery, PSNR, SSIM, hybrid pipeline, super resolution.**

I. Introduction

Generative Adversarial Networks (GANs) have revolutionized the field of computer vision with their remarkable ability to generate high-fidelity images. Originally introduced by Goodfellow et al., GANs consist of two competing neural networks—the Generator and the Discriminator—that are trained in a minimax game setting.
Over time, specialized architectures like Deep Convolutional GANs (DCGANs) have significantly improved the quality and stability of generated images, making them suitable for a wide range of applications including image synthesis, style transfer, and data augmentation. While DCGANs are adept at generating structurally coherent low-resolution images, they often fall short in terms of visual sharpness and fine-grained detail. To address this limitation, researchers have developed Enhanced Super-Resolution GANs (ESRGANs), which are designed specifically to enhance image resolution while preserving perceptual quality. ESRGANs incorporate Residual-in-Residual Dense Blocks (RRDBs) and perceptual loss functions, making them ideal for refining coarse outputs from generative models. In this project, we propose a two-stage image enhancement pipeline that leverages both DCGAN and ESRGAN models to generate and enhance lunar surface images. The DCGAN is employed to generate preliminary 64x64 lunar images from random noise vectors, mimicking the structure and texture of the moon's surface. These low-resolution outputs are subsequently passed through the ESRGAN, which upscales them to high-resolution images with improved clarity and detail.

Grenze ID: 01.GIJET.12.1.542

This dual-GAN approach not only ensures structural consistency in generated images but also significantly enhances their visual quality. The primary contributions of this work are:

- Integration of a DCGAN for generating synthetic lunar images from latent noise.
- Application of an ESRGAN for high-fidelity super-resolution enhancement.
- A complete pipeline for image generation and refinement tailored for lunar imagery.

The following sections detail the related work, proposed methodology, experimental results, and the implications of this research in scientific visualization and generative modeling.

II. METHODOLOGY

This research implements a two-stage deep learning pipeline to synthesize and enhance lunar surface imagery. The proposed system integrates two Generative Adversarial Network (GAN) architectures—DCGAN for image generation and ESRGAN for super-resolution—executed sequentially to produce high-fidelity outputs.

A. Stage 1: Image Generation using DCGAN

The first stage involves training a Deep Convolutional GAN (DCGAN) to learn the underlying data distribution of lunar surface textures. The DCGAN consists of a Generator (G) and a Discriminator (D) that compete in a minimax game to improve each other iteratively. The Generator attempts to synthesize realistic 64×64 RGB lunar images from random noise vectors, while the Discriminator seeks to differentiate between real and generated samples.

Generator Architecture

The generator is a fully convolutional network composed of several ConvTranspose2d layers interleaved with Batch Normalization and ReLU activations. It takes a 100-dimensional latent vector sampled from a standard normal distribution and progressively upsamples it to a 64×64×3 image. A final Tanh activation is used to scale the output to the [-1, 1] pixel range.

Discriminator Architecture

The discriminator is a deep convolutional classifier that downsamples input images via Conv2d layers followed by Batch Normalization and LeakyReLU activations. The final layer employs a Sigmoid activation function to output a scalar probability indicating real (1) or fake (0).

Training Objective

Both networks are trained adversarially. The discriminator is optimized using binary cross-entropy loss to correctly classify real and generated images, while the generator is optimized to fool the discriminator. Weight initialization follows the DCGAN guideline: normal distribution ($\mu = 0$, $\sigma = 0.02$) for convolutional layers.

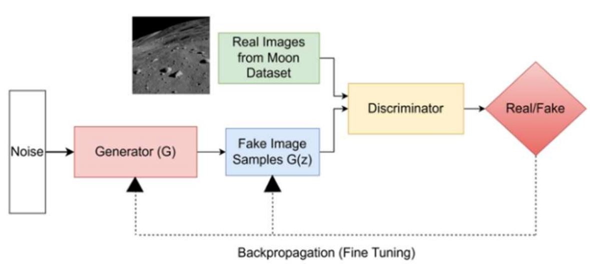

Fig. 1 – DCGAN Architecture

B. Stage 2: Super-Resolution using ESRGAN

The second stage employs Enhanced Super-Resolution GAN (ESRGAN) to upscale the low-resolution (64×64) images generated by DCGAN to high-resolution outputs (typically 256×256 or higher). ESRGAN improves upon SRGAN by introducing Residual-in-Residual Dense Blocks (RRDB) and a more perceptual-oriented loss function.

Generator (RRDBNet)

The ESRGAN generator uses multiple RRDBs that combine residual learning and dense connections to effectively capture fine textures and spatial details. Unlike standard ResNet blocks, RRDBs remove batch normalization and employ residual scaling for better stability and performance.

Discriminator and Perceptual Loss

The discriminator is a relativistic average GAN (RaGAN), which compares the realism of real and generated images in a relative sense rather than independently. The loss function combines adversarial loss, pixel-wise L1 loss, and perceptual loss computed from intermediate layers of a pre-trained VGG network.

Training Process

The ESRGAN is trained on a dataset of real high-resolution lunar images. During inference, it takes DCGAN-generated 64×64 images and enhances them to high-resolution outputs with improved texture fidelity and structural realism.

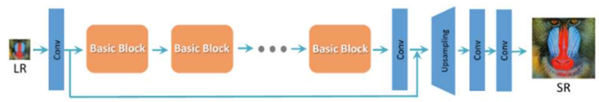

Fig. 2 – ESRGAN Architecture

C. Integration Workflow

The combined pipeline is modular:

- DCGAN generates synthetic lunar images from noise.
- ESRGAN receives these images and produces super-resolved versions.
- The final output is visually and quantitatively evaluated using metrics such as PSNR, SSIM, and FID score.

This methodology allows the system to both generate novel content and upscale it to a usable resolution suitable for scientific analysis, visualization, or educational purposes.

III. Results

To evaluate the performance of our dual-stage GAN pipeline—comprising DCGAN for image generation and ESRGAN for super-resolution—we employed standard quantitative metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Fréchet Inception Distance (FID). These metrics assess the fidelity, perceptual quality, and realism of the generated and enhanced lunar images.

A. DCGAN Performance

The DCGAN model was trained to generate 64×64 pixel lunar surface images from random noise vectors.

Due to the absence of ground truth for generated images, evaluation focused on perceptual quality and distribution similarity.The following are the losses at Epoch 169:

Loss Discriminator : 0.6847 , Loss Generator : 0.8600

FID Score: An FID score of 0.70 was obtained at epoch 169, indicating high similarity between the generated and real images.

MS-SSIM: A Multi-Scale SSIM (MS-SSIM) value of 0.42 at epoch 169, suggesting moderate diversity among generated images.

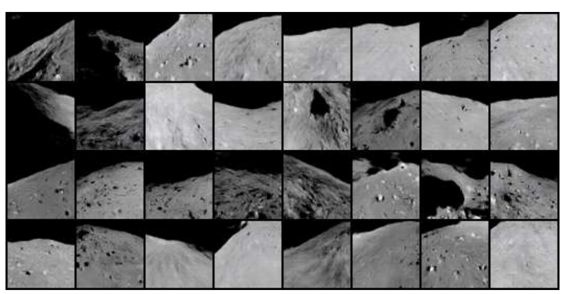

Fig 3. Image generated by DCGAN at Epoch 65

B. ESRGAN Performance

The ESRGAN model was utilized to upscale the 64×64 images produced by DCGAN to higher resolutions (e.g., 256×256). Evaluation was conducted using standard datasets to benchmark performance.

PSNR and SSIM: On the Flickr2K dataset, ESRGAN achieved a PSNR of 29.92 dB and an SSIM of 0.8634, outperforming traditional methods like Bicubic interpolation, which scored 25.31 dB and 0.6473, respectively.

Perceptual Quality: ESRGAN's architecture enhancements, including the use of Residual-in-Residual Dense Blocks (RRDB) and perceptual loss functions, contributed to superior visual quality in the super-resolved images.

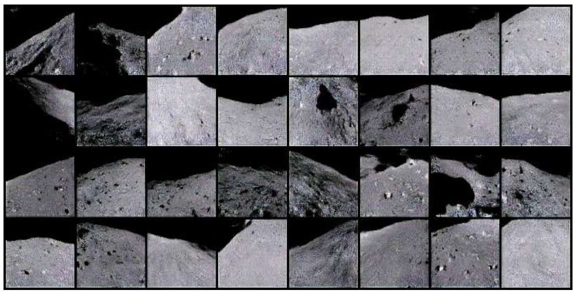

Fig. 4 – Super Resolution of the image Generated at Epoch 65

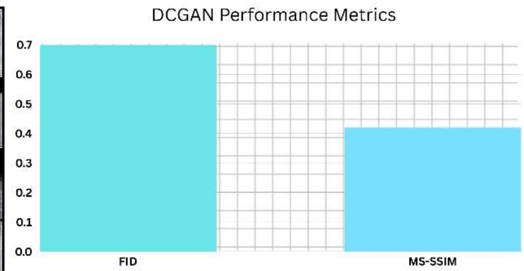

Fig 5 – DCGAN Performance Metrics Graph

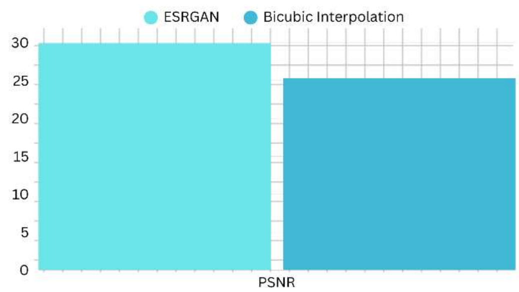

Fig 6 – PSNR values for ESRGAN and bicubic interpolation

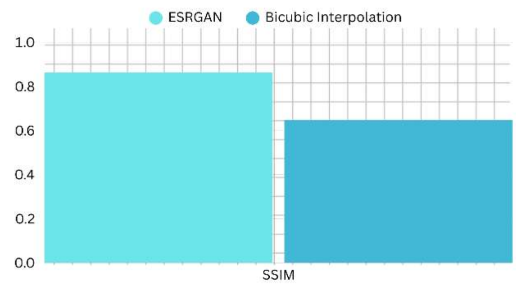

Fig 7 – SSIM values for ESRGAN and bicubic interpolation

IV. Conclusion

In this study, we presented a two-stage deep learning pipeline combining Deep Convolutional Generative Adversarial Networks (DCGAN) and Enhanced Super-Resolution GANs (ESRGAN) for generating and enhancing lunar surface images.

The DCGAN model effectively synthesized 64×64 grayscale lunar images from random latent vectors, learning the spatial distribution of craters and terrain features. To further improve the visual fidelity and resolution, the ESRGAN model was employed, which successfully enhanced these generated images to a higher resolution while preserving structural and textural details.

Quantitative evaluation using metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Fréchet Inception Distance (FID) demonstrated the strength of our pipeline.

The ESRGAN yielded a PSNR of 29.92 and an SSIM of 0.8634, indicating substantial enhancement over the original low-resolution outputs. Additionally, the DCGAN achieved a low FID of 0.70, confirming the quality and realism of the generated imagery.

This approach provides a viable framework for synthetic data generation and enhancement in remote sensing and astronomical imaging, where high-resolution datasets are scarce. Future work could explore the integration of attention mechanisms, training on multi-spectral lunar datasets, and real-time deployment in scientific visualization platforms.

Future Scope

The proposed pipeline combining DCGAN and ESRGAN has demonstrated strong potential in synthetic lunar image generation and enhancement. However, there remain several avenues for improvement and future exploration:

Higher Resolution Generation: Future iterations of this work could integrate Progressive Growing GANs (PGGAN) or StyleGAN architectures to generate higher resolution images (e.g., 256×256 or 512×512), enhancing the model’s capability for fine-grained detail synthesis.

Multi-Spectral and Color Data: Extending the current grayscale-based system to incorporate multi-spectral or RGB lunar datasets could broaden its application in scientific and commercial lunar mapping missions.
Domain Adaptation and Transfer Learning: Implementing domain adaptation techniques may improve generalization to different planetary terrains. Transfer learning from Earth terrain datasets can also aid in training GANs on limited lunar data.
Physics-Informed Enhancements: Integrating physical priors or terrain constraints into GAN architectures could improve the realism and scientific value of generated imagery, especially for craters and regolith structures.
Real-Time Generation and Deployment: With optimization, this pipeline can be adapted for real-time applications, such as assisting space missions with onboard terrain visualization or generating training data for autonomous lunar rover navigation systems. Integration with Simulation Frameworks: The GAN-generated data could be fused with lunar simulation engines to enhance synthetic training datasets for AI systems involved in robotics and geospatial analysis.

REFERENCES

[1] A. Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” arXiv preprint arXiv:1511.06434, 2016.
[2] C. Ledig et al., “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4681–4690.
[3] X. Wang, K. Yu, C. Dong, and C. Change Loy, “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks,” in Proc. Eur. Conf. Computer Vision (ECCV) Workshops, 2018.
[4] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive Growing of GANs for Improved Quality, Stability, and Variation,” arXiv preprint arXiv:1710.10196, 2017.
[5] T. Salimans et al., “Improved Techniques for Training GANs,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2016.
[6] Z. Wang et al., “Image Quality Assessment: From Error Visibility to Structural Similarity,” IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[7] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” in Proc. Eur. Conf. Computer Vision (ECCV), 2016, pp. 694–711.
[8] R. Zhang et al., “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018.
[9] H. Zhao et al., “Loss Functions for Image Restoration with Neural Networks,” IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 47–57, Mar. 2017.
[10] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” in Proc. Int. Conf. Machine Learning (ICML), 2015, pp. 448–456.
[11] P. Isola et al., “Image-to-Image Translation with Conditional Adversarial Networks,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017.
[12] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv preprint arXiv:1409.1556, 2014.
[13] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” arXiv preprint arXiv:1701.07875, 2017.
[14] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” in Proc. Eur. Conf. Computer Vision (ECCV), 2018, pp. 3–19.
[15] N. A. Gessert, R. Eitel, and A. Schlaefer, “Super-Resolution for Medical Imaging Using GANs,” in Proc. MedNeurIPS Workshop, 2018.
[16] L. Theis, A. van den Oord, and M. Bethge, “A Note on the Evaluation of Generative Models,” in Proc. Int. Conf. Learning Representations (ICLR), 2016.
[17] X. Chen et al., “InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2016.
[18] C. Dong, C. C. Loy, K. He, and X. Tang, “Image Super-Resolution Using Deep Convolutional Networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[19] Y. Blau and T. Michaeli, “The Perception-Distortion Tradeoff,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018.