NCE Journal of Science and Engineering (NJSE), Vol. 2, June 2025

Video Enhancement using SRGAN

Aayush Shrestha[1], Ishita Chalise[1], Pradish Tamrakar[1], and Sharmila Bista[1,*]
[1]Department of Electronics and Computer Engineering, National College of Engineering, Lalitpur, Nepal
[*]Correspondence: sharmila@nce.edu.np
Manuscript received January 23, 2025; accepted April 15, 2025

***Abstract*—The advancement of deep learning techniques, particularly Generative Adversarial Networks (GANs), has revolutionized the field of video enhancement. Improving video quality is essential across multiple domains, including entertainment, media production, and surveillance. This project introduces a video enhancement system based on Super-Resolution GAN (SRGAN) to upscale and restore low-resolution, degraded videos. The project includes the design and implementation of advanced models utilizing Convolutional Neural Networks (CNNs) and SRGAN, with a focus on optimizing these models for efficient training and inference. A user-friendly interface has been developed to facilitate seamless interaction, making the technology accessible to non-technical users. For training the model, the REDS dataset was carefully preprocessed to generate blurred, downsampled, and compressed video frames, enhancing the diversity and complexity of the data. This approach enables the model to effectively reconstruct high-quality video frames from low-quality inputs. The model achieved a Peak Signal-to-Noise Ratio (PSNR) of 28.67 and a Structural Similarity Index (SSIM) of 0.87, demonstrating its ability to significantly improve video resolution and perceptual quality.**

***Keywords*—Convolutional neural network (CNN), generative adversarial network (GAN), super resolution generative adversarial network (SRGAN), video super resolution, and video enhancement.**

I. INTRODUCTION

VIDEO enhancement has evolved significantly with the emergence of deep learning techniques, particularly Generative Adversarial Networks (GANs). Beginning with basic signal processing methods, the field progressed to embrace more sophisticated algorithms as computational power increased. Machine learning, through models like Support Vector Machines and Random Forests, demonstrated notable improvements over traditional approaches by learning mappings between low- and high-quality video pairs. However, it was the emergence of deep learning, notably Convolutional Neural Networks (CNNs), that started a fundamental change in video enhancement. CNNs, with their ability to learn hierarchical features directly from raw pixel data, became instrumental in tasks such as super-resolution, as demonstrated by the Super Resolution Convolutional Neural Network (SRCNN). The introduction of GANs by Goodfellow et al. [1] in 2014 entered in a new era of video enhancement. GANs, comprising a generator and a discriminator trained adversarially, demonstrated remarkable capabilities in generating high-quality, realistic images. The Super-Resolution GAN (SRGAN), introduced by Ledig et al. [2] in 2017, further advanced the field by combining adversarial and content losses to produce photorealistic high-resolution images from low-resolution inputs. The application of GANs to video enhancement demonstrated impressive results in improving visual quality and reducing noise.

II. RELATED THEORY

The field of generative adversarial networks (GANs) has seen remarkable advancements, enabling significant improvements in various domains, including video enhancement and medical imaging. Recent studies have leveraged the unique capabilities of GANs to address complex challenges, such as improving compressed video quality and augmenting medical imaging datasets for enhanced diagnostic accuracy. Below, we review notable works that highlight the diverse applications of GAN-based architectures.

CVEGAN: A Perceptually-inspired GAN for Compressed Video Enhancement" by Ma et al. [3] proposes a novel GAN architecture to enhance compressed video frames. The generator uses a Multi-Resolution block (Mul2Res) with multiple residual learning branches, an Enhanced Residual NonLocal Block (ERNB), and an Enhanced Convolutional Block Attention Module (ECBAM). The training strategy employs a relativistic sphere GAN (ReSphereGAN) and new perceptual loss functions. Evaluated within MPEG HEVC and VVC test models, CVEGAN achieves significant coding gains, with up to 28% improvement in post-processing (PP) and 38% in spatial resolution adaptation (SRA) for HM 16.20, and up to 8.0% and 20.3% respectively for VTM 7.0.

"GANs-Based Intracoronary Optical Coherence Tomography Image Augmentation for Improved Plaques Characterization Using Deep Neural Networks" by Nikroo et

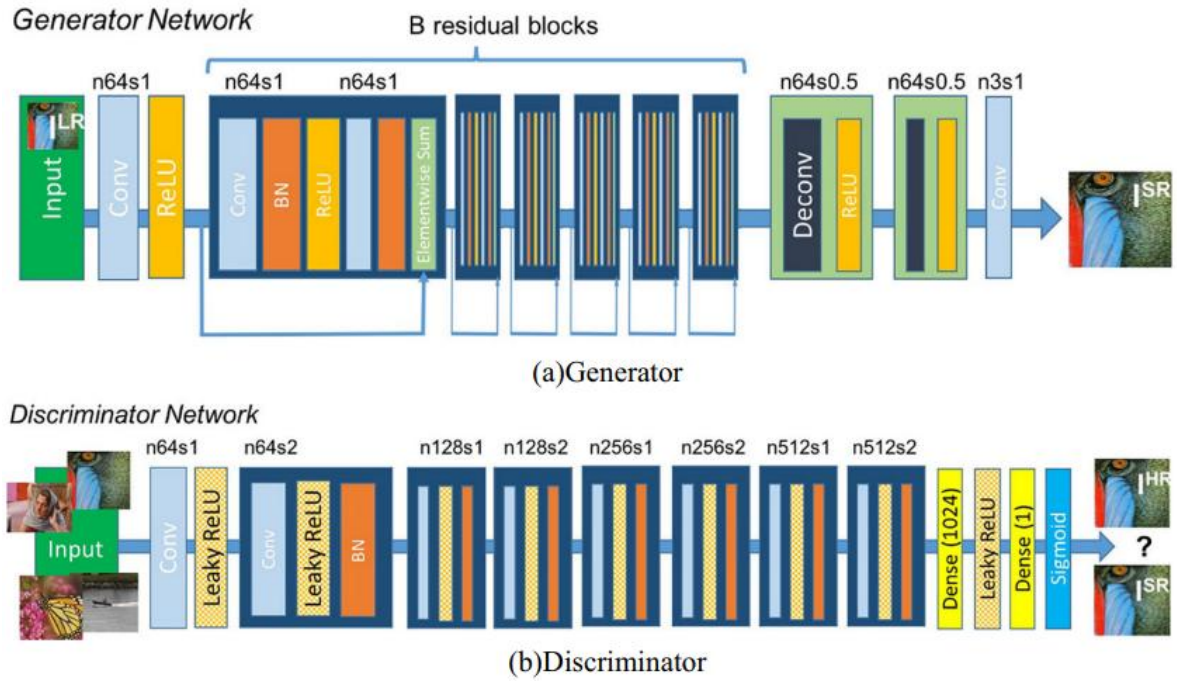

Fig. 1. SRGAN architecture [7].

al. [4] The study presents a method for augmenting a dataset of intracoronary optical coherence tomography (OCT) images using conditional generative adversarial networks (cGANs). The goal is to enhance the classification of coronary plaques. The dataset consists of OCT images from 51 patients, which were augmented by factors of 5×, 10×, 50×, and 100× using cGANs. The augmented images were used to train an AlexNet model, and it was found that augmenting the dataset by a factor of 50× improved classification accuracy by 15.8%. The study demonstrates that synthetic images generated by cGANs can effectively complement real images in training deep learning models, resulting in better classification performance.

”Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network” by Ledig et al. [2] This paper introduces SRGAN, a Generative Adversarial Network (GAN) for single image super-resolution. The SRGAN aims to produce photo-realistic high-resolution images from low-resolution inputs by employing a perceptual loss function that combines an adversarial loss with a content loss. The adver8sarial loss helps the generator network produce images that are indistinguishable from real high-resolution images, while the content loss, based on VGG network feature maps, focuses on perceptual similarity rather than pixel-wise accuracy. The SRGAN significantly improves the visual quality of super-resolved images, particularly for high upscaling factors like 4×, surpassing traditional methods that optimize for mean squared error (MSE) and peak signal-to-noise ratio (PSNR)

A Comparative Analysis of SRGAN Models” by Zafar et al. [5] evaluates the performance of several state-of-the-art Super-Resolution Generative Adversarial Network (SRGAN) models on real-world images. The Enhanced Deep Super Resolution (EDSR) model is highlighted for its effectiveness, employing deep convolutional neural networks to achieve high-quality image reconstructions. Similarly, the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) and its upgraded version, Real-ESRGAN, are noted for their ability to generate realistic textures and high-resolution images through advanced architectural modifications like Residual-in-Residual Dense Blocks (RRDB). The study also emphasizes the importance of GANs in enhancing image quality, particularly for applications involving Optical Character Recognition (OCR), where improved visual fidelity directly correlates with better text recognition accuracy.

III. Methodology

A. Dataset Preparation

For this study, the dataset was derived from the REDS dataset [6], a widely used benchmark in video enhancement tasks. The REDS dataset contains 300 high-quality videos, which served as the ground truth for training and evaluation. To simulate low-resolution conditions and create input-output pairs for the model, the high-resolution frames were processed using the following techniques:

1) Downscaling: The frames were resized to lower resolutions (94,94) to simulate standard degradation caused by resolution loss.

2) Downscaling with Blurring: In addition to downscaling, a Gaussian blur was applied to the frames, mimicking the effects of motion blur and optical distortions.

3) **Downscaling with Compression:** After downscaling, the frames were compressed to replicate the degradation caused by video compression.

These preprocessing steps ensured the creation of diverse low-resolution inputs that reflect real-world scenarios, enabling the model to learn to restore high-quality frames from various

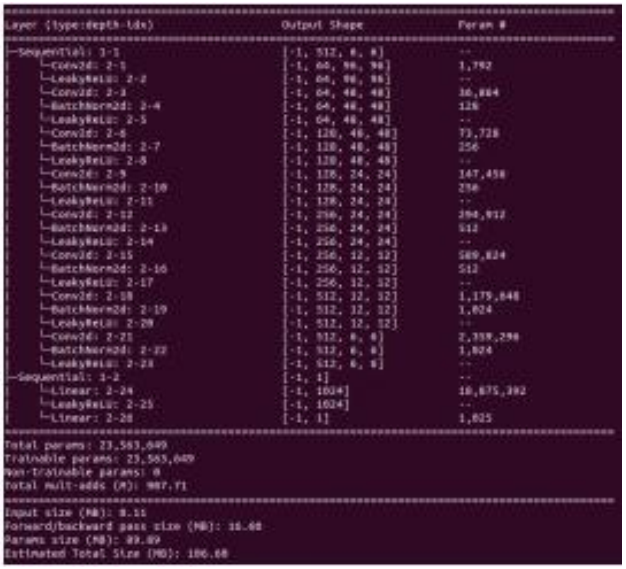

Fig. 2. Generator network.

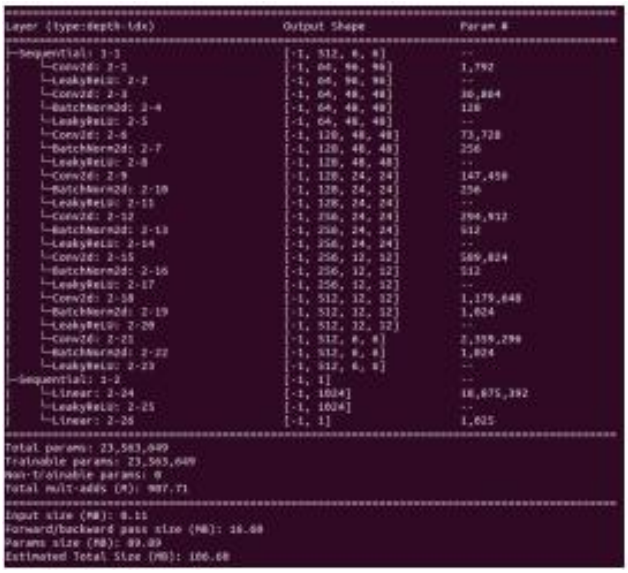

Fig. 3. Generator network.

types of degraded inputs. By applying preprocessing techniques, a robust dataset was created to train the model.

B. Algorithm

SRGAN introduced by Ledig et al. [2] is a generative adversarial network for single image super-resolution. It uses a perceptual loss function which consists of an adversarial loss and a content loss. The adversarial loss pushes the solution to the natural image manifold using a discriminator network that is trained to differentiate between the super-resolved images and original photo-realistic images. In addition, the authors use a content loss motivated by perceptual similarity instead of similarity in pixel space. The actual networks - depicted in the Figure 1 consist mainly of residual blocks for feature extraction. Formally we write the perceptual loss function as a weighted sum of a content loss lX SR and an adversarial loss component lGEN SR:

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR} \quad (1)$$

The figure illustrates the architecture of a Super-Resolution Generative Adversarial Network (SRGAN), which consists of two main parts: the Generator Network and the Discriminator Network. The Generator Network begins with an input image processed through an initial convolutional layer and a PReLU activation. It then passes through a series of residual blocks, each containing convolutional layers, batch normalization (BN), and PReLU activations. These blocks capture complex features from the input image. The output 14from the residual blocks is combined with the initial input through a skip connection, preserving the original details. After the residual blocks, the network includes more convolutional layers, batch normalization, and an element-wise sum operation. Finally, the image is upscaled using pixel shufflers and ReLU activations to produce the super-resolved image.

C. Model Architecture

The architecture consists of a generator network and a discriminator network, each designed to complement the generative adversarial framework.

a) Generator Network: The generator employs a deep convolutional neural network (CNN) to map a lowdimensional input latent space to a high-resolution output. The architecture primarily comprises residual convolutional blocks, upsampling layers, and activations.

Input layer: The generator take input of size [3, 96, 96].

Convolutional Layer: The input tensor is processed through an initial convolutional layer, followed by PReLU activation, producing feature maps of dimensions [64, 96, 96]. • Residual Blocks: The network integrates multiple ResidualConvBlocks, each containing two convolutional layers with skip connections, enhancing feature representation and gradient flow. Each block outputs feature maps with consistent spatial dimensions.

Upsampling Mechanism: To achieve spatial resolution enhancement, UpsampleBlocks are used, which integrate learnable interpolation layers followed by convolution and activation, increasing the resolution by factors of 2 (from [64, 96, 96] to [64, 384, 384]).

Output Layer: A final convolutional layer maps the upsampled feature maps to the desired output dimensions, [3, 384, 384], corresponding to the RGB image output.

b) Discriminator Network: The discriminator is a CNN-based classifier that evaluates the authenticity of the generated samples against the real data. The architecture is structured as follows:

Convolutional Feature Extraction: A sequence of convolutional layers progressively reduces spatial dimensions while increasing feature channels, starting from 64 to a maximum of 512. Each convolutional layer is followed by LeakyReLU activations and BatchNorm layers for enhanced convergence.

Fully Connected Layers: Flattened features from the final convolutional block are fed into a fully connected layer of size 1024 with LeakyReLU activation. The final output layer maps

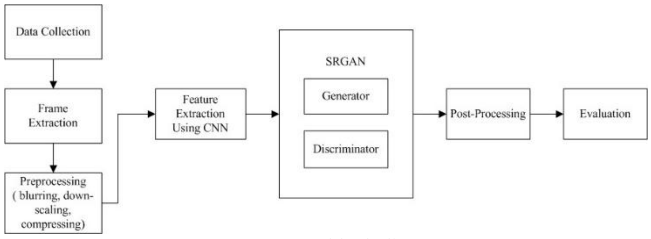

Fig. 4. System block diagram.

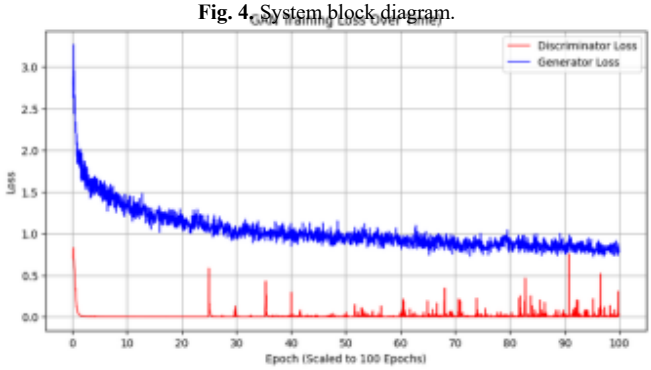

Fig. 5. Discriminator and generator loss over epochs.

to a single neuron using a linear activation, providing a scalar score for adversarial loss.

Discriminative Power: The network's hierarchical design allows it to effectively differentiate between high frequency details and contextual features in generated vs. real samples.

D. System Block Diagram

This diagram illustrates the system workflow for video enhancement using SRGAN.

1) Data Collection: The videos sourced from REDs serves as our dataset.

2) Frame Extraction: Videos are split into 100 individual frames to create static images that can be processed by the model.

3) Preprocessing: The extracted frames are degraded through blurring, downscaling, and compression to simulate real-world low-resolution inputs. This step ensures the creation of low-quality data to train the model effectively.

4) Feature Extraction Using CNN: Convolutional Neural Network (CNNs) is used to extract features from the low-resolution frames, which serve as input to the SRGAN model.

5) SRGAN (Super-Resolution Generative Adversarial Network):

• Generator: The generator learns to transform lowresolution frames into high-resolution outputs by synthesizing realistic details.

• Discriminator: The discriminator evaluates the outputs of the generator, distinguishing between the generated high-resolution frames and the original ground truth to improve the generator's performance.

6) Post-Processing: Once the generator produces high-resolution frames, they are recombined to make video.

7) Evaluation: The final enhanced frames are compared to the ground truth frames to evaluate the model's performance using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity

Fig. 6. Sample image.

Fig. 7. Enhanced image.

Index (SSIM).

E. Performance Metrices

Peak Signal-to-Noise Ratio (PSNR)

Measures the ratio between the maximum possible value of a signal and the power of corrupting noise that affects the fidelity of its representation. It is represented as:

$$PSNR = 10log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (2)$$

where MSE is mean squared error.

Structural Similarity Index (SSIM)

Measures the similarity between two images. It considers changes in structural information, luminance, and contrast.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_2)} \quad (3)$$

III. Results and Discussion

A. Generator and Discriminator Loss Over Epochs

This plot shows how the generator and discriminator losses evolve during training. Early on, the generator has a high loss because it starts out producing poor-quality images and the discriminator quickly learns to distinguish real from fake, so its loss is relatively low. Over time, the generator improves and its loss settles around 0.9, indicating it is producing convincing images. The discriminator's spikes suggest moments when the generator briefly fools it by producing high-quality images,

causing the discriminator to temporarily misclassify. The plot shows a stable adversarial balance: the generator steadily improves while the discriminator mostly maintains the edge but occasionally gets fooled by the generator.

B. Qualitative Analysis

Figures 6 and 7 illustrate a difference between a low-resolution input frame and its corresponding enhanced output. Figure 6 represents the original low-resolution frame, which exhibits visible pixelation, blurring, and a lack of fine details. In contrast, Figure 7 showcases the enhanced frame produced by the model, demonstrating significant improvements in sharpness, texture restoration, and overall visual fidelity. To quantitatively assess the enhancement, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) metrics are used. The model achieves a PSNR of 28.67 dB and an SSIM of 0.87, indicating a substantial improvement in image quality while preserving structural details. The high SSIM score further confirms that the enhanced frames maintain perceptual similarity to high-resolution references, ensuring a realistic and visually appealing output. The visual and numerical results demonstrate the capability of the model in producing high-quality video frames, making it suitable for real-world applications where video clarity is crucial.

V. CONCLUSION

In this work, a video enhancement model based on Super Resolution GAN (SRGAN) was developed to improve the resolution and perceptual quality of low-resolution videos. Trained on the REDS dataset, the model achieved a PSNR of 28.67 dB and an SSIM of 0.87, showcasing its effectiveness in generating high-quality video frames. The integration of a user-friendly interface allows for easy interaction, making the technology accessible to both technical and non-technical users. This model demonstrates its potential for real-world applications, including video restoration, surveillance footage enhancement, and media production, where high-resolution video is critical. In future work, optimizing the model to improve PSNR and SSIM scores can further enhance the quality of generated frames. Real-time processing for efficient inference on mobile devices and embedded systems would also make the model more applicable to a wider range of industries. Additionally, expanding the model's capability to handle various video formats and resolutions, along with generalizing across diverse datasets, would improve robustness and ensure better performance in real-world scenarios.

REFERENCES

[1] I. J. Goodfellow and et al., "Generative adversarial networks," ArXiv.org, 2014. [Online]. Available: https : //arxiv.org/abs/1406.2661.
[2] C. Ledig and et al., "Photo-realistic single image super-resolution using a generative adversarial network," ArXiv.org, 2016. [Online]. Available: https://arxiv.org/ abs/1609.04802.
[3] D. Ma and et al., "Cvegan: A perceptually-inspired gan for compressed video enhancement," ArXiv.org, 2020. [Online]. Available: https://arxiv.org/abs/2011.09190.
[4] F. R. Nikroo and et al., "A comparative analysis of srgan models," ArXiv.org, 2023. [Online]. Available: https:// arxiv.org/abs/2307.09456.
[5] H. Zafar and et al., "Gans-based intracoronary optical coherence tomography image augmentation for improved plaques characterization using deep neural networks," Optics, vol. 4, no. 2, pp. 288–299, 2023.
[6] S. Nah et al., "Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019.
[7] J. Huang, "Image super-resolution reconstruction based on generative adversarial network model with double discriminators," Multimedia Tools and Applications, 2020. [Online]. Available: https : / / www . researchgate. net / publication/343602746.