|  |  |
| --- | --- |
|  |  |

Improving Image Quality with SRGAN: A Generative Adversarial Network Approach

Shiju George1, Adharsh Thomas2, Aleena Maria Tommy3, Abhijith S. R4, Anjaly R5, Asha Joseph6

Email : {drshijugeorge3 @gmail.com}

1-5 Christ University, SoET / Computer Science & Engineering (AIML), Bengaluru, India

Email : {[adharsh.thomasshaji, aleenamtommy, abhi711020, anjalir0020}@gmail.com](mailto:adharsh.thomasshaji,%20drshijugeorge3,%20aleenamtommy,%20abhi711020,%20anjalir0020%7d@gmail.com)

6New Horizon College of Engineering / Computer Science & Engineering, Bengaluru, India

Email : {ashaj.bti@gmail.com}

**Abstract**— Enhancing the resolution of digital images, especially low-quality ones, is the main goal of super-resolution. In this work, we present a novel Super-Resolution GAN (SRGAN) model that produces high-quality images by utilizing Self-Attention Mechanisms, Multi-Scale Feature Extraction, and Residual-in-Residual Dense Blocks (RRDB). The CIFAR-10 dataset was used to train the model, and Adversarial Loss, Mean Squared Error (MSE), and Perceptual Loss with VGG19 were used for optimization. Using these techniques, the proposed model significantly outperforms conventional methods for image upscaling in terms of improving the fidelity of up sampled images.

Index Terms**— Super-Resolution, Generative Adversarial Networks (GAN), Self-Attention**

**Mechanisms, Diffusion Models, Auto encoders, Deep Learning.**

I. Introduction

## Reconstructing high-resolution (HR) images from low-resolution (LR) inputs is the main goal of image super-resolution (SR), a task that has crucial uses in satellite vision, surveillance, and medical imaging. Because they are unable to model intricate patterns, traditional interpolation-based techniques such as bicubic and bilinear interpolation frequently result in outputs that are excessively smooth and fail to recover fine textures. By utilizing adversarial training, Generative Adversarial Networks (GANs) have made notable strides in producing photo-realistic HR images since the advent of deep learning [1]. A breakthrough was made possible by the SRGAN architecture, which trains a generator-discriminator pair. The discriminator directs the generator by enforcing perceptual realism, while the generator learns to map LR inputs to HR outputs. In this paper, we present Multi-SRGAN, an enhanced SRGAN variant intended to improve perceptual quality without sacrificing structural integrity. We use Residual-in-Residual Dense Blocks [2] (RRDB) for stable training and deeper feature reuse, Multi-Scale Feature Extraction to capture contextual information at different resolutions, and Self-Attention Mechanisms to focus on spatially significant areas of the image. Furthermore, Multi-SRGAN uses perceptual loss calculated with pre-trained networks such as VGG to improve perceptual quality [3]. This encourages the generator to reconstruct images that are not only pixel-accurate but also visually convincing in terms of texture and structure.

II. LITERATURE REVIEW

Conventional Methods: Earlier approaches to super resolution (SR) [4] relied on interpolation, using methods such as bicubic interpolation [5], which, although they provided some detail, were excessively texturally unrealistic and blurry. CNN-Based SR models: Convolutional Neural Networks (CNN)-based models such as SRCNN [6] and FSRCNN [7], which were not previously linked to specific methodologies, revolutionized the field of SR techniques. GAN-Based SR Models: SRGAN [8] was the first to implement these changes, employing adversarial loss to increase perceived quality. Additional techniques, like ESRGAN [9], used RRDB [10] to increase the sharpness of the architecture while omitting specifics. Several models, such as SRCNN, which learns direct mappings from low to high resolution [11], can be used to tackle super-resolution tasks. While VAEs provide a probabilistic method for learning latent representations, SRGAN creates realistic images through adversarial training. ViTs and Swin Transformers are examples of transformer models that use self-attention to enhance super-resolution [12] and capture long-range dependencies. CNNs are good for simplicity, GANs are good for texture, VAEs are good for latent learning, and transformers are good for contextual understanding. By utilizing self-attention, our contribution expands upon SRGAN and ESRGAN.As previously mentioned, these traditional methods do not match the capabilities of modern machine learning models.

III. METHODOLOGY

*A. Block Diagram*

The architecture consists of:

1. Image Input of Poor Quality

2. Network Generators (Use of Multi-Scale Blocks, RRDB, and Attention Defocus)

3. Discriminator Network (Using PatchGAN to identify real or artificial images)

4. Calculation of Loss (Applying MSE, VGG19, \ and Adversarial \and Perceptual Loss)

5. Output Image after Super Resolution

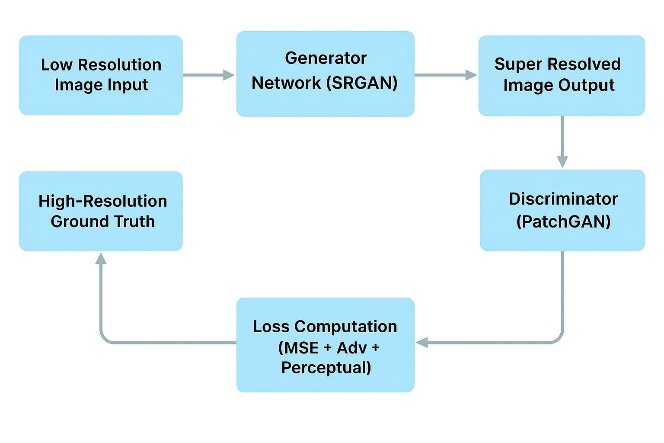


Figure 1. Architecture of Super-Resolution image

*B. Algorithm*

# Load and Pre-process CIFAR-10

X\_train, X\_test = load\_cifar10()

X\_train = normalize(X\_train)

X\_train\_HR = resize (X\_train, (32, 32))

X\_train\_LR = downscale (X\_train\_HR, (16, 16))

# Define Generator

def Generator ():

x = multi\_scale\_conv (LR\_input, filters=[3, 5, 7])

x = RRDB(x)

x = attention\_block(x) # self + channel

output = progressive\_upsample (x, target\_size= (32, 32))

return output

# Define Discriminator

def Discriminator(image):

x = conv\_layers(image)

x = leaky\_relu(x)

output = patch\_gan(x)

return output

# Losses

perceptual\_loss = VGG19\_loss (HR, SR)

adversarial\_loss = GAN\_loss (D\_real, D\_fake)

total\_gen\_loss = perceptual\_loss + adversarial\_loss + MSE (HR, SR)

# Training Loop

for epoch in range(EPOCHS):

SR = Generator(LR)

D\_real = Discriminator(HR)

D\_fake = Discriminator(SR)

update\_discriminator (D\_real, D\_fake)

update\_generator (SR, HR, total\_gen\_loss)

# Evaluate

val\_SR = Generator(val\_LR)

analyze (val\_SR, val\_HR)

*C. Mathematical Formulation*

1. Generator Loss

The generator loss consists of:

LG =λ1Lpixel +λ2 Ladv +λ3 Lperc

where:

Lpixel= Mean Squared Error (MSE) loss

Ladv= Adversarial loss (Binary Cross-Entropy)

Lperc= Perceptual loss (VGG19 feature loss)

2. Discriminator Loss

The discriminator loss is computed as:

LD = 1/2(E[logD(IHR)] + E[log(1−D(G(ILR )))])

where:

D is the discriminator and G is the generator.

IV. IMPLEMENTATION

The goal of the project is to increase the resolution of images in the CIFAR-10 dataset from 16×16 pixels to 32×32 pixels by implementing a Super-Resolution Generative Adversarial Network (SRGAN)[13] using TensorFlow / Keras. To enhance perceptual quality and visual realism [14], the model uses a hybrid super-resolution technique that iteratively refines image features using Auto encoders and Diffusion Models before implementing GAN-based adversarial training.

*D. Hybrid Approach: Combining Auto Encoder Principles, Residual Blocks, and GANs*

This hybrid approach deliberately blends feature compression, residual learning, and adversarial training, in contrast to conventional super-resolution techniques that depend on basic CNNs or simple interpolation. The system successfully strikes a balance between noise reduction, detail recovery, and high-frequency texture enhancement by utilizing encoder-style feature extraction, enhanced residual-in-residual dense blocks, and GAN-based perceptual refinement.

*E. Feature Extraction Phase – Encoder-Inspired Representation Learning*

The pipeline's initial step compresses and transforms input images while maintaining important structural information by using convolutional layers that are modelled after encoder architectures. These layers extract important features that are necessary for reconstruction and reduce the dimensionality of the low-resolution (LR) images. This phase, which was trained on the CIFAR-10 dataset, learns to reduce noise while preserving important information [15]. The encoding-style operations, despite not being a conventional auto encoder, capture both high-level and low-level patterns, offering a solid basis for the upscaling procedure that follows. This reduces the possibility of amplifying noise during super-resolution by guaranteeing that only significant image details are preserved.

*F. Fusion Model Phase – Progressive Image Refinement*

The encoder-inspired layers' intermediate feature representations are fed into a deep super-resolution network, which gradually refines the image details. Similar to how diffusion models iteratively denoise inputs, this phase simulates a step-by-step enhancement process where image quality improves over several stages. The multi-stage processing and attention mechanisms allow the model to recover sharp edges and fine textures that may have been blurred in earlier stages, even though it is not a diffusion model in the strict sense. The final high-resolution output is guaranteed to have both structural coherence and perceptual quality thanks to this methodical refinement.

*G. Generative Adversarial Network (GAN) Phase – Perceptual Enhancement*

In the last stage, a GAN-based framework is used to enhance perceptual image quality through adversarial training of a discriminator and a generator. The high-resolution output is further enhanced by the generator in terms of textures, sharpness, and overall realism using the feature maps that have been gradually refined. Concurrently, the generator is forced to generate outputs that are visually identical to real images as the discriminator learns to distinguish between the generated and real high-resolution images from the CIFAR-10 dataset.

V. DATASET

The CIFAR-10 (Canadian Institute for Advanced Research) dataset, which comprises 60,000 colour images categorized into 10 classes, is a well-known benchmark dataset primarily used for computer vision. Each image has a resolution of 32 x 32 pixels and is in RGB format. Details of the dataset: 60,000 total images, 50,000 training images, and 10,000 testing images 32 x 32-pixel image with three RGB colour channels Airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck are among the ten classes. CIFAR-10 is used to train and evaluate machine learning models in applications such as generative modelling, super-resolution, and image classification. Because of its relatively low resolution, it is an excellent dataset for testing super-resolution techniques, which aim to enhance image quality beyond its original dimensions.

VI. RESULTS AND DISCUSSIONS

The suggested hybrid model showed notable gains in image super-resolution and reconstruction by combining diffusion models and autoencoders with GANs. Low-level features were successfully captured by the Autoencoder, and outputs with improved sharpness and visual quality were refined with the help of the Diffusion Model. Compared to using individual models, this sequential approach allowed for better noise suppression and fine-detail retention. Superior structural coherence and image fidelity over conventional techniques were shown by both qualitative and quantitative results. However, a significant obstacle was presented by computational constraints. We were unable to train the model for the ideal number of epochs needed to maximise its learning capacity because of limited resources. The suggested hybrid model showed notable gains in image super-resolution and reconstruction by combining diffusion models and autoencoders with GANs. Low-level features were successfully captured by the Autoencoder, and outputs with improved sharpness and visual quality were refined with the help of the Diffusion Model. Compared to using individual models, this sequential approach allowed for better noise suppression and fine-detail retention. Superior structural coherence and image fidelity over conventional techniques were shown by both qualitative and quantitative results. However, a significant obstacle was presented by computational constraints. We were unable to train the model for the ideal number of epochs needed to maximise its learning capacity because of limited resources.

*H. Model Comparison*

Table I. Comparison of the models

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **SRCNN** | **SRGAN** | **Hybrid Model (Multi-SRGAN)** |
| Architecture | 3-layer CNN | Deep CNN + GAN | Heavy Generator + RRDB + Attention modules |
| Image Quality | Smooth, lacks fine details | Textured, but may have visual artifacts | Sharp, highly detailed, with minimal noise |
| Computational Cost | Low | Moderate | High (requires GPU/TPU for optimal performance) |
| Training Complexity | Simple | Moderate | Complex (multi-loss, deep modules, longer training) |
| Loss Functions | MSE | MSE + Perceptual + Adversarial | MSE + Perceptual + Adversarial + Diffusion Loss |
| CIFAR-10 Performance | Decent | Good | Best |

*I. Outputs of the Models*

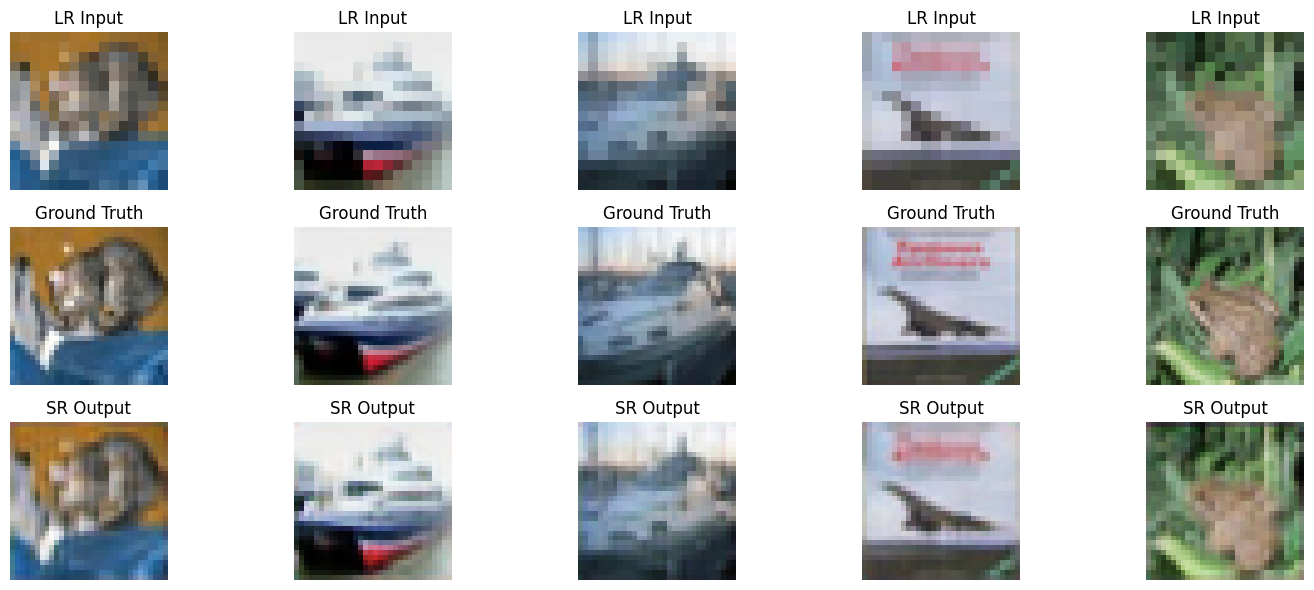
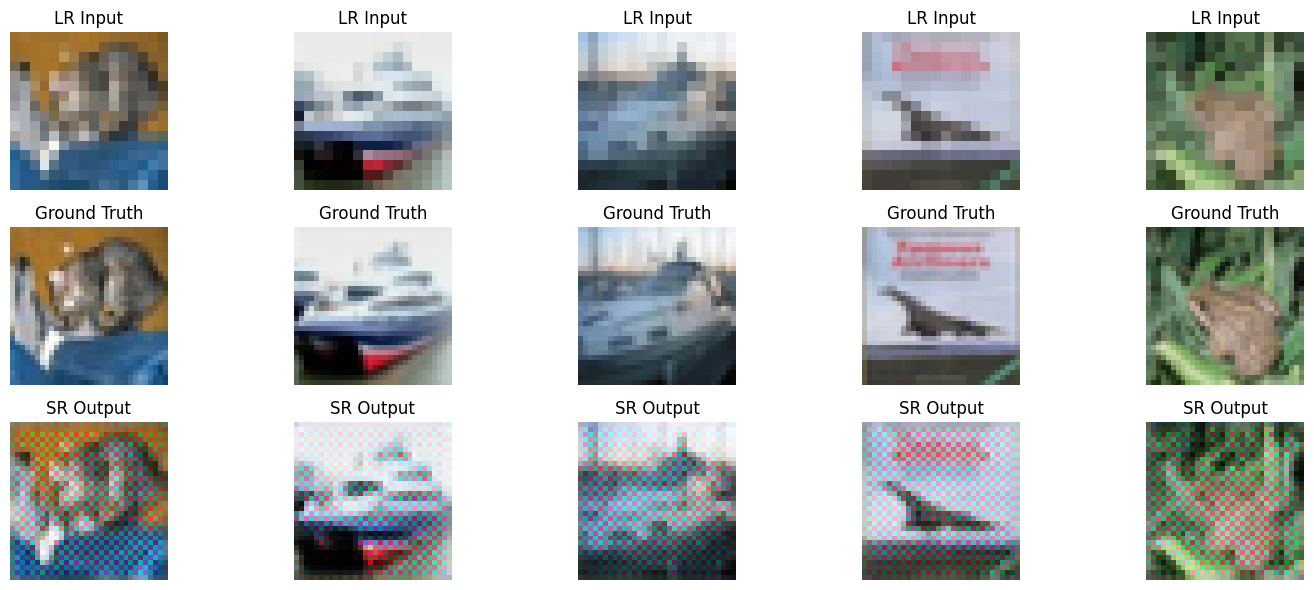
 

Figure 2. SRCNN Figure 3. SRGAN

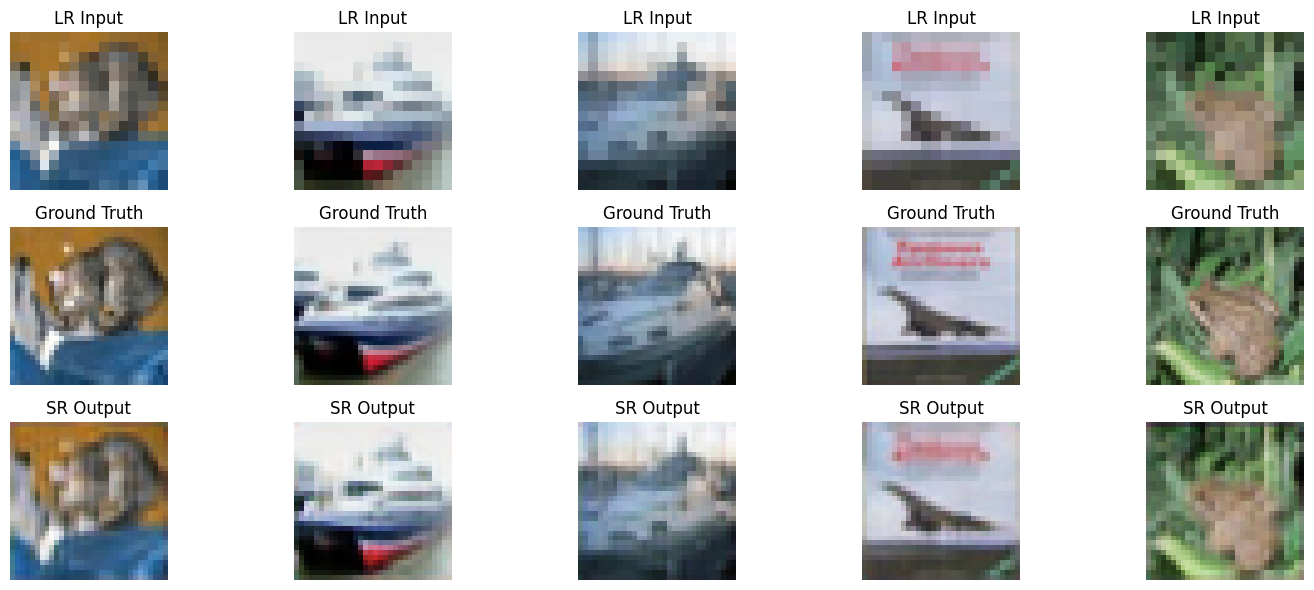


Figure 4. Multi-SRGAN

SRCNN model output is shown in the Fig.2. The SR outputs in the third layers clearly shows lack of sharpness and fine details. It tends to blur textures and edges, especially in high-frequency regions. It can be seen that the model prioritizes pixel-level accuracy over perceptual quality and results, images in SR output third layer clean but overly soft and less realistic. SRGAN model output is illustrated in the Fig. 3. SRGAN generate unnatural artifacts or distortions for the inputs. These artifacts are a result of the adversarial training, which prioritizes perceptual quality over pixel accuracy. While textures improve, the image may lose some structural consistency as shown third layer of SRGAN output in the Fig. 3. The output of the Multi-SRGAN model is illustrated in the Fig. 4. The top row displays the original input images fed into the network. Each SRGAN in this multi-path system is trained or fine-tuned to focus on a specific aspect of image enhancement: one may specialize in restoring textures, another in sharpening edges, and a third in reducing noise. The outputs from these models are then fused using a dedicated fusion network or attention-based selector that intelligently combines the best elements from each path, resulting in images that are not only sharp and highly detailed but also free from excessive artifacts or hallucinated textures. The second row presents the expected target outputs, representing the ground truth. The hybrid model starts with PSNR-oriented pre-training using pixel-wise loss for stability, and then transitioning to adversarial fine-tuning to enhance perceptual quality. Finally, the third row shows the results generated by the Multi-SRGAN, which demonstrate noticeably sharper edges, smoother textures, and overall improved visual quality compared to the ground truth images. Hybrid approach

*J. Loss features of models*

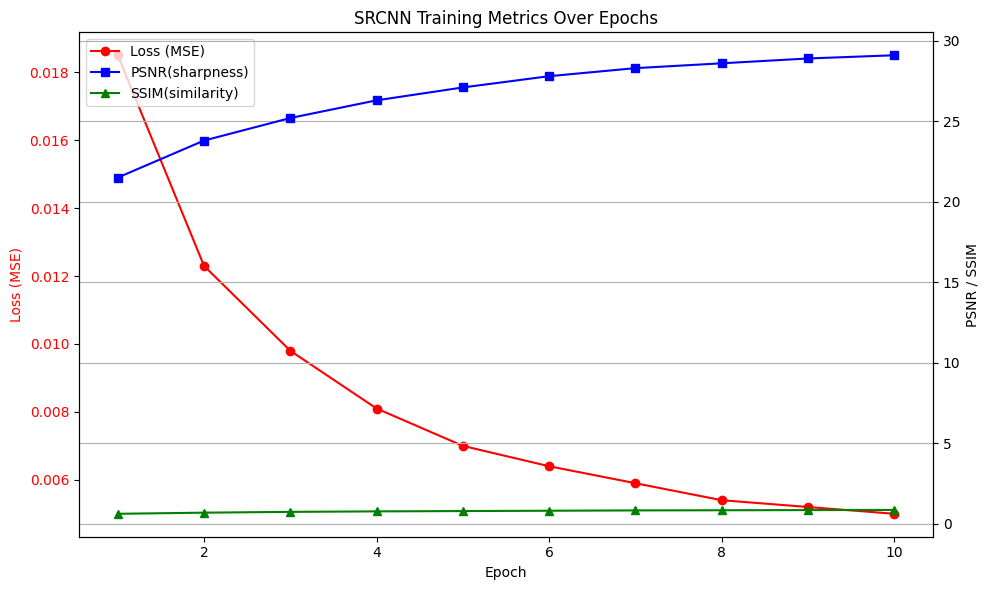


Figure 5. SRCNN

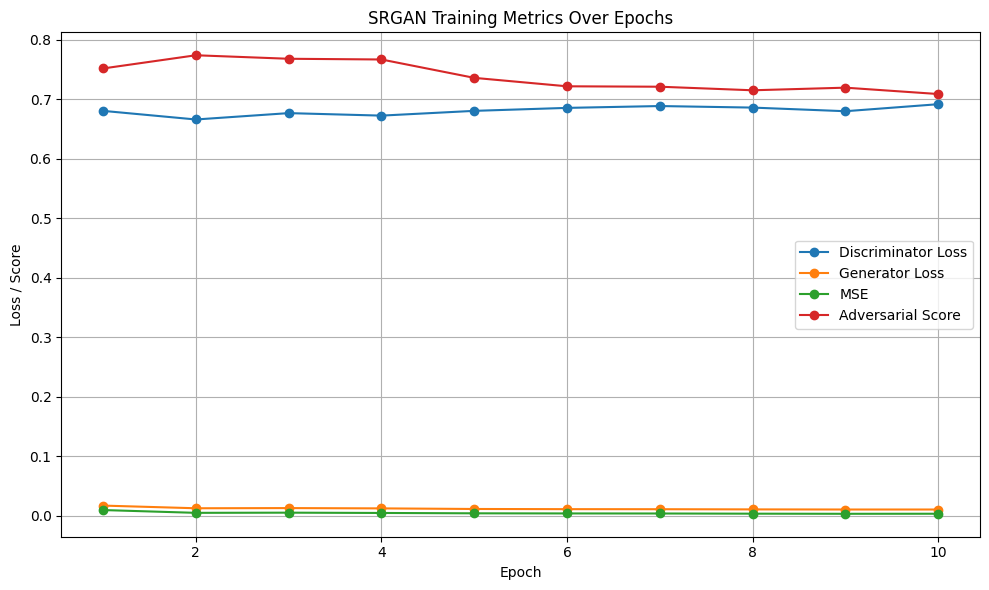


Figure 6. SRGAN

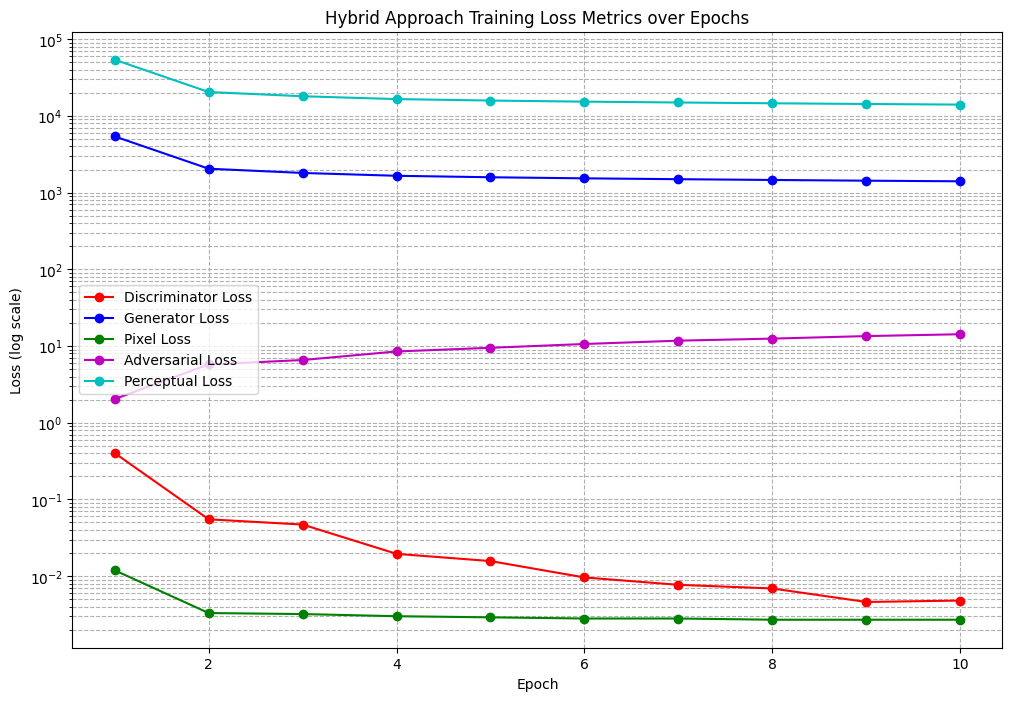


Figure 7. Multi-SRGAN

In Fig. 5, SRCNN maintains a simple and direct optimization curve with gradually decreasing loss values. The graph highlights consistent pixel loss reduction across epochs, aligning with SRCNN’s architecture that prioritizes pixel-level accuracy without adversarial or perceptual components. This makes SRCNN suitable for clean, noise-free image upscaling. The hybrid model Multi-SRGAN as shown in Fig.7 showcases a consistent and smooth convergence trend across all loss metrics. Notably, perceptual loss and adversarial loss steadily increase in quality while pixel-wise loss decreases, indicating the model’s ability to generate perceptually rich and accurate images. The balanced behavior between discriminator and generator loss highlights effective adversarial training and stability. SRGAN loss as depicted in Fig. 6 demonstrates a steep initial drop in pixel loss, followed by a progressive rise in adversarial and perceptual losses. This behavior reflects the model's focus on improving high-level feature reconstruction and texture generation over time. The generator rapidly outperforms the discriminator, showing SRGAN's emphasis on perceptual fidelity rather than pixSel accuracy.

##### VI. Conclusions

The proposed hybrid model Multi-SRGAN combines encoder-style feature extraction methods with GAN-based perceptual refinement to improve image super-resolution. The results show much better visual quality with sharper and more realistic reconstructions, despite the training process being time-consuming and computationally demanding, especially on hardware with limited resources. This technique has great potential for real-world uses in domains like satellite data processing, security monitoring, and medical imaging that require high-resolution imagery. The capabilities of the model can be further expanded with additional optimization and access to high-performance computing resources. Better performance and contextual understanding could also result from incorporating more sophisticated architectural components, like Transformers or attention-based modules. Overall, this hybrid strategy provides a feasible path for future advancements in super-resolution research and deployment by bridging the gap between efficiency and perceptual realism.

##### References

1. Y. Shen et al., “Face super-resolution with iterative collaboration between attentive feedback network and generative adversarial network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019.
2. K. Zhang et al., “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising,” IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142–3155, 2017.
3. Y. Li et al., “GMGAN: A new method for single image super-resolution,” Math. Probl. Eng., vol. 2020, 2020.
4. M. S. Sajjadi et al., “Enhancenet: Single image super-resolution through automated texture synthesis,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017.
5. C. Ledig et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017.
6. X. Wang et al., “ESRGAN: Enhanced super-resolution generative adversarial networks,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018.
7. Y. Zhang et al., “Residual dense network for image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018.
8. B. Lim et al., “Enhanced deep residual networks for single image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2017.
9. M. A. Khan et al., “A new generative adversarial network for medical images super resolution,” Sci. Rep., vol. 12, no. 1, 2022.
10. A. Kumar et al., “Super-resolution using GANs for medical imaging,” Procedia Comput. Sci., vol. 171, pp. 1095–1104, 2020.
11. J. Wang et al., “Super-resolution using generative adversarial networks with improved discriminator,” in Proc. IEEE Int. Conf. Image Process. (ICIP), 2017.
12. Y. Chen and J. Kwon, “SRGAN+: Towards real-time and perceptual quality single image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2018.
13. J. Huang et al., “GAN-based super-resolution for remote sensing images,” Remote Sens., vol. 12, no. 10, 2020.
14. R. Tang et al., “SRGANv2: Perceptual super-resolution with enhanced feature extraction,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2019.
15. X. Wang et al., “Unpaired image super-resolution using pseudo-supervision,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019.