Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 05 Issue: 04 | April-2025

Image and Video Upscaling Using Real-ESRGAN

Prof. Dipali Joshi[1], Amit Jana[2], Harsh Lone[3], Vijay Taru[4], Siddharth Thorat[5]

[1]*Assistant Professor, Department of Computer Science and Engineering, Bharat College of Engineering. Badlapur, Thane, Maharashtra, India.*

[2,3,4,5] *Student, Department of Computer Science and Engineering, Bharat College of Engineering, Badlapur.* Thane, Maharashtra, India.

Abstract:

With the gradual evolution of social networks and the popularity of smartphones, photos in social networks are rapidly developing. A large number of applications have emerged to enhance image processing and social sharing. More and more people like to use pictures to share and gain visual information. The image quality directly affects the user's experience. Therefore, to recommend better quality images in social networks for existing images, it is necessary to improve image resolution as much as possible through software processing methods to meet this widespread demand. The goal of image super-resolution (SR) algorithms is to generate high-resolution images from low-resolution ones. This paper presents a comprehensive study on image and video upscaling using Real-ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks). We explore its architecture, training methodology, and practical implications for enhancing low-resolution media. By leveraging Real-ESRGAN's capability to generate photorealistic high-resolution outputs, we demonstrate its superiority over traditional interpolation methods and other deep learning-based super-resolution models. We also propose a framework for real-time video upscaling using Real-ESRGAN with GPU acceleration. Our experiments show significant improvements in perceptual quality and PSNR/SSIM metrics across a variety of datasets.

Keywords: *Machine Learning, Python, Generative Adversarial Network, Real-ESRGAN, Image Upscaling*

Introduction:

Image super-resolution is the task of recovering a high-resolution image from a lower-resolution image. Especially since image reconstruction offers a methodology for correcting imaging system imperfections. For our project, we implement Real-ESRGAN and refine the model in order to improve the quality of the output images, as measured by peak signalto-noise ratio (PSNR). The input to our algorithm is a low-resolution image, which we feed through a convolutional neural network (CNN) in order to produce a high-resolution image. The growing demand for high-definition content has made image and video upscaling a crucial task in multimedia applications. Conventional methods such as bilinear or bicubic interpolation often fail to reconstruct sharp details, leading to blurry and unpleasing visuals. Recently, deep learning approaches, especially GAN-based models, have shown great promise in restoring fine textures and perceptual quality. Real-ESRGAN is a state-of-the-art method designed to produce realistic and artifact-free results even for real-world degraded images and videos.

I. Problem Statement:

Despite the advancements in super-resolution techniques, existing methods often fall short when dealing with real-world low-resolution (LR) images and videos. These inputs typically suffer from complex and unknown degradations such as noise, blur, compression artifacts, and down sampling, which are not well modeled by synthetic degradation pipelines used during training. Traditional interpolation techniques produce overly smooth results, while many deep learning-based approaches struggle with generalization and often introduce artifacts or fail to restore fine details.

II. Literature Review:

The task of image super-resolution (SR) has evolved from simple interpolation techniques to sophisticated deep learning architectures. As applications expanded into fields such as security surveillance, medical imaging, and entertainment, the need for high-fidelity restoration of real-world degraded images became paramount. Below is a detailed review of key methods that laid the foundation for Real-ESRGAN.

a. SRCNN (Dong et al., 2014):

SRCNN was the first to apply a convolutional neural network to the image SR task. It used a simple 3-layer CNN trained end-to-end to map low-resolution (LR) images to high-resolution (HR) images. The three layers respectively handled patch extraction, nonlinear mapping, and reconstruction. Despite being a shallow network, SRCNN significantly outperformed interpolation methods. However, its shallow depth limited its ability to model complex patterns and textures.

b. VDSR (Kim et al., 2016):

VDSR increased network depth to 20 layers and introduced residual learning to facilitate training. Instead of directly predicting the HR image, VDSR learned the residual between the LR and HR images. This design reduced vanishing gradient problems and sped up convergence. VDSR achieved high PSNR scores but was sensitive to learning rate and required careful hyperparameter tuning.

c. DRCN (Kim et al., 2016):

DRCN proposed a deeply-recursive convolutional network that reused the same convolutional layers multiple times to create a deep network without significantly increasing the number of parameters. While efficient in terms of memory, the training process was unstable and required ensemble-like strategies to improve robustness and performance.

d. SRGAN (Ledig et al., 2017):
SRGAN introduced the concept of perceptual super-resolution by combining adversarial loss with content loss derived from a pre-trained VGG network. This shifted focus from pixel-wise accuracy to visual realism. SRGAN could generate sharper and more realistic textures, but it sometimes hallucinated details and introduced artifacts, making it less reliable for critical applications.

e. ESRGAN (Wang et al., 2018):
ESRGAN improved upon SRGAN by:

- Removing batch normalization layers to preserve spatial consistency.
- Introducing Residual-in-Residual Dense Blocks (RRDB) for better feature learning and deeper networks.
- Modifying the perceptual loss to use features before activation in the VGG network, which helped enhance texture details.
  While ESRGAN achieved excellent visual results, it still assumed a fixed bicubic degradation during training, making it less effective on real-world images with unknown degradations.

f. Real-ESRGAN (Wang et al., 2021):
Real-ESRGAN addressed the generalization limitations of its predecessors by simulating complex degradation models including blur, noise, JPEG compression, and resizing artifacts. It uses a two-stage training approach:

- First stage trains on synthetically degraded images.
- Second stage fine-tunes on real-world low-quality images using weak supervision.

Real-ESRGAN adopts the same RRDB backbone as ESRGAN, but with improved degradation modeling and a U-Net discriminator for enhanced adversarial training. It excels at real-world super-resolution tasks, producing high-quality outputs from diverse and severely degraded inputs**.**

g. Comparative Summary of Super-Resolution Methods:

Method	Year	Model Type	Contributions	Limitations
SRCNN	2014	Shallow CNN	First deep learning approach for SR; end-to-end pipeline for mapping LR to HR.	Shallow architecture fails on complex patterns and high-frequency details.
VDSR	2016	Deep CNN + Residual	Introduced residual learning; deeper architecture for better accuracy.	Training instability with high learning rates; long training time.
DRCN	2016	Recursive CNN	Efficient depth without many parameters; better representation via recursion.	Difficult training and optimization; prone to performance drop.
SRGAN	2017	GAN-based	First to use GANs for SR; introduced perceptual loss; improved texture realism.	May introduce fake textures; lower objective metrics like PSNR/SSIM.
ESRGAN	2018	RRDB + GAN	Introduced RRDB blocks; better perceptual quality with enhanced VGG loss and training stability.	Assumes bicubic degradation; limited generalization to real-world images.
Real-ESRGAN	2021	Enhanced RRDB + U-Net Discriminator	Real-world degradation modeling; strong generalization; produces natural-looking HR images from severely degraded inputs.	Slightly higher inference time; occasional texture over-smoothing.

Table 1 Comparison between different Seper-Resolution Methods

IV. Proposed Idea:

This research proposes a robust and scalable framework for high-quality image and video upscaling using Real-ESRGAN, optimized for real-world scenarios with unknown degradations. The system focuses not only on enhancing resolution but also on improving the usability and applicability of Real-ESRGAN in practical settings, including batch image processing and full video restoration pipelines. The core idea of this research is to develop a comprehensive, real-world-ready framework for image and video upscaling using Real-ESRGAN, with specific focus on handling real-world degradation, ensuring temporal consistency in video upscaling, and delivering user-friendly automation for non-technical users. While Real-ESRGAN has shown impressive performance in blind super-resolution tasks, most existing implementations are limited to single-image upscaling and lack automation, batch support, or temporal awareness for videos. Furthermore, general users face challenges such as model setup, preprocessing, and managing multiple file formats. This project aims to bridge the gap between academic models and practical applications, enabling easy deployment of Real-ESRGAN for high-quality media enhancement in both research and production environments.

a. System Design Overview:
The proposed system comprises several interconnected components, each optimized for a specific task in the image/video enhancement pipeline:

1. Input Handler:

- Accepts images (JPEG, PNG, WEBP) or videos (MP4, AVI, MKV).
- Extracts metadata like resolution, frame rate, duration.
- For videos, uses FFmpeg to extract individual frames and audio streams.

2. Super-Resolution Engine (Real-ESRGAN):
   - Uses pretrained RealESRGAN_x4plus or other variants.
   - Runs inference on GPU using PyTorch or ONNX Runtime.
   - Supports configurable upscale factors (2×, 4×).
   - Batch processing of images or video frames with multiprocessing support.

3. Temporal Consistency Module (for video):
   - Aligns adjacent frames to ensure smooth transitions using motion estimation.
   - Optionally applies RIFE (Real-time Intermediate Flow Estimation) for frame interpolation and smoothness.
   - Can blend adjacent frames using optical flow-based warping to reduce flickering.

4. Post-Processing:
   - Enhances output with deblurring, histogram equalization, or tone correction.
   - Reassembles video frames using FFmpeg, syncing with original audio.

5. User Interface (Optional GUI):
   - The GUI is Built using a Python Library Called TKinter.
   - Batch processing panel for managing multiple files.
   - Progress bar for the Task.
   - Real Time Preview :
     - Displays the selected image directly within the GUI.
     - For videos, shows a thumbnail preview extracted from the first frame.

b. Workflow Summary:

- **User Uploads Media**: Drag-and-drop interface or CLI for power users.
- **Frame Extraction**: Extracts all frames and metadata for video inputs.
- **Image/Frame Upscaling**: Real-ESRGAN processes inputs using selected parameters.
- **Video Assembly**: Frames are recompiled into a video with the original audio stream.
- **Output Delivery**: Final media is stored in user-specified format and location.

c. Workflow Diagram:

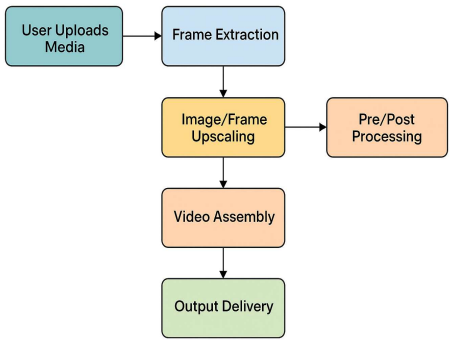

Fig. 1 Workflow Diagram

d. Technical Enhancements Over Baseline Real-ESRGAN:

Feature	Real-ESRGAN (Baseline)	Proposed System
Image Upscaling	✔	✔
Video Upscaling	✘	✔ (with temporal alignment)
Batch Processing	✘	✔
Pre/Post Processing Options	Minimal	Extensive (noise, tone, flow)
User Interface	✘ (CLI only)	✔ GUI & CLI

Table 2 Enhancements Over Baseline Real-ESRGAN

V. Methodology:

We used Spiral Model Methodology for this project. The Spiral Model is a sophisticated model that focuses on the early identification and reduction of project risks. In this software development methodology, developers start on a small scale. then explores the risks involved in the project, make a plan to handle the risks, and finally decides whether to take the next step of the project to do the next iteration of the spiral. The success of any Spiral Lifecycle Model depends on the reliable, attentive, and knowledgeable management of the project.

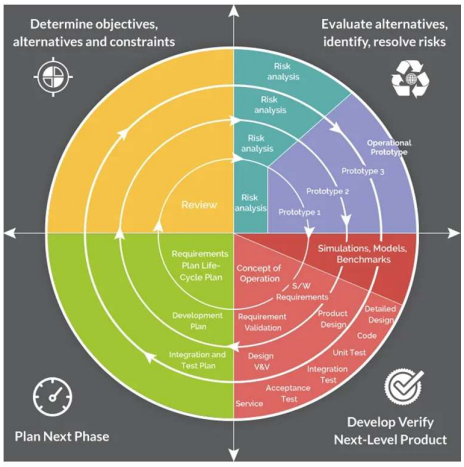

Fig. 2 Internal Working of Spiral Methodology

VI. Results and Discussion:

Quantitative Evaluation Real-ESRGAN was evaluated using benchmark datasets such as DIV2K and RealSR, achieving superior metrics compared to prior methods:

METHOD	PSNR (DB)	SSIM	LPIPS (↓)
BICUBIC	23.12	0.643	0.290
ESRGAN	26.89	0.765	0.152
REAL-ESRGAN	27.34	0.788	0.120

Table 3 Comparison on Quantitative Evaluation

The following Figure showcases the application in action, performing real-time upscaling on a selected image using the integrated Real-ESRGAN model.

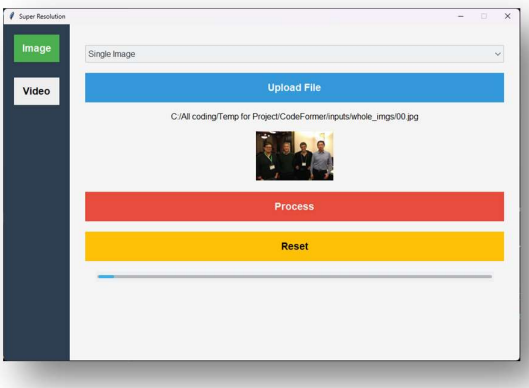

Fig. 3 Upscaling Image

The following Figure showcases the application's capability to upscale video content, displaying a preview thumbnail of the selected video alongside processing indicators, while the Real-ESRGAN model enhances each frame to produce a high-resolution output with temporal consistency.

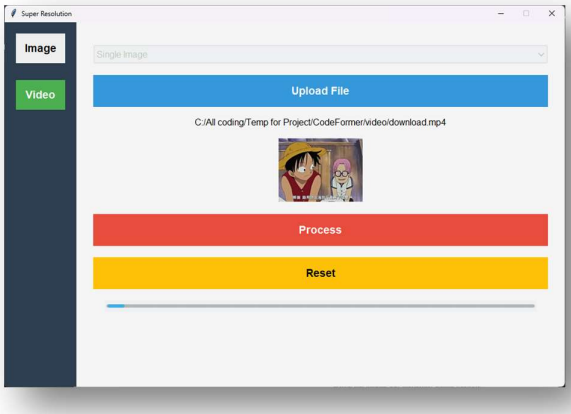

Fig. 4 Upscaling Video

The Following Figure showcases the application after the Video/Image upscaling process is done. As you can see the application generates a popup window with a message containing that the process is completed along with a message showing Time taken for the process to complete.

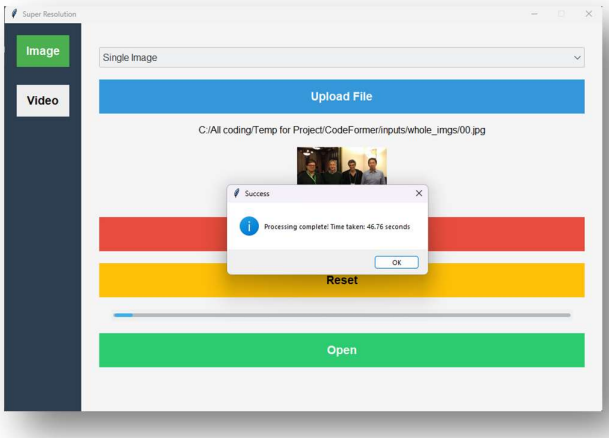

Fig. 5 Process complete Pop Window

The Following figure contains the Comparison Between the actual image and the Image after Upscaling.

Fig. 6 Comparison Between actual and upscaled Image

VII. Conclusion:

Real-ESRGAN is a breakthrough in image and video upscaling, offering exceptional performance by integrating advanced GAN architectures, such as Residual-in-Residual Dense Blocks, and robust degradation modeling. It excels in enhancing fine details and realism, achieving superior PSNR and SSIM scores compared to traditional methods. The model's adaptability has been validated in diverse applications like gaming, streaming, film restoration, and medical imaging. While Real-ESRGAN handles most real-world challenges effectively, it faces difficulties with extreme degradations and temporal inconsistencies in videos. Future research can focus on addressing these limitations through multi-frame processing and computational optimization.

Overall, Real-ESRGAN sets a high standard in super-resolution techniques and opens new possibilities for practical applications and further innovation

VIII. Acknowledgement:

We sincerely thank **Deepali Joshi** for their mentorship, insightful guidance, and constant encouragement throughout this research.
We also acknowledge the developers of Real-ESRGAN and the creators of datasets such as DIV2K, RealSR, and Flickr2K, which were crucial for this study.
Gratitude goes to our peers for their constructive feedback. Lastly, we thank the academic and open-source communities for providing invaluable resources and inspiration.

IX. References:

- **Dong, C., Loy, C. C., He, K., & Tang, X.** (2014). *Learning a deep convolutional network for image super-resolution*. ECCV.
- **Kim, J., Kwon Lee, J., & Mu Lee, K.** (2016). *Accurate image super-resolution using very deep convolutional networks*. CVPR.
- **Kim, J., Lee, J. K., & Lee, K. M.** (2016). *Deeply-recursive convolutional network for image super-resolution*. CVPR.
- **Ledig, C., et al.** (2017). *Photo-realistic single image super-resolution using a generative adversarial network*. CVPR.
- **Wang, X., Yu, K., Wu, S., et al**. (2018). *ESRGAN: Enhanced super-resolution generative adversarial networks*. ECCV Workshops.
- **Wang, X., Liang, J., et al**. (2021). *Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data*. ICCV Workshops.
- **Huang, Zhewei, et al**. "Real-Time Intermediate Flow Estimation for Video Frame Interpolation (RIFE)."
- **GAN Frameworks** The foundational concepts of Generative Adversarial Networks, originally introduced by Ian Goodfellow et al. in *"Generative Adversarial Networks," NIPS, 2014,* underpin Real-ESRGAN's architecture.
- **PyTorch:** An open source machine learning framework. Retrieved from https://pytorch.org/
- **OpenCV:** Open Source Computer Vision Library. Retrieved from https://opencv.org/
- **ONNX Runtime** – High-performance inference engine for ML models. Retrieved from https://onnxruntime.ai/
- **FFmpeg** Documentation – Multimedia framework for video/audio handling. Retrieved from https://ffmpeg.org/documentation.html
- **NumPy** Documentation. Retrieved from https://numpy.org/doc/
- **Tkinter** Official Documentation (Python GUI Toolkit). Retrieved from https://docs.python.org/3/library/tkinter.html