The Intelligent Enhancement System using the ESRGAN with Quality Assessment and Automated Filtering

Deeksha D[1], Suresha D[2], Spoorthi B[3]
[1]*Computer Science and Engineering, Srinivas Institute of Technology, Mangaluru, India*
[2]*Professor, Computer Science and Engineering, Srinivas Institute of Technology, Mangaluru, India*
[3]*Assistant Professor, Computer Science and Engineering, Srinivas Institute of Technology, Mangaluru, India*
doi.org/10.64643/IJIRTV12I12-206754-459

***Abstract*—Enhancing the low-quality data is important in applications like medical imaging, surveillance and the remote sensing. Traditional ways of doing this like making the images bigger or using filters do not always work well and can make the images look blurry. This project is about a system that makes low quality images look better using ESRGAN, a kind of computer program that can make images look more real and detailed. The system combines ways of making images look better with old ways like making the image clearer making the colors more vivid and using special filters. The system also has a part that checks how blurry, noisy or bright a image is and then chooses the way to make it look better so people do not have to do it themselves. An interface that makes it easy for people to use the system. They can upload their images or video make them look better compare the results and see how well the system worked. The system can also work with videos making each frame look better one by one. It measures how well it did use things like PSNR, SSIM, SNR, MSE, RMSE and entropy and it shows the results in graphs, which shows that the Intelligent Enhancement System makes pictures look sharper and more real than old ways of doing it. The Intelligent Enhancement System is really good, at making low quality images look better.**

***Index Terms*—Automated filtering, Deblur Filter, Enhanced Super Resolution Generative Adversarial Network, ESRGAN, Entropy, Generative Adversarial Network, High Pass Filter, Low Pass Filter, Mean Squared Error, Peak Signal to Noise Ratio, Residual in Residual Dense Block, Root Mean Squared Error, Signal to Noise Ratio, Structural Similarity Index, Super Resolution, Quality Assessment, Upscaling, Deep Learning, Night Enhancement, Video Enhancement, Image Enhancement.**

I. INTRODUCTION

High quality visual data is essential in many modern computational applications. The real-world acquisition is affected by the limitations like hardware constraints, poor lighting conditions, motion induced blur, and compression artifacts. These factors lead to degraded or low-resolution outputs which significantly impact the both automated processing systems and human interpretation. Super resolution techniques aim to reconstruct high-resolution outputs from the low-resolution inputs. Traditional methods like bicubic interpolation are good for computers to handle. They often do not do a good job of keeping the details and textures of the images. On the hand methods that use convolutional neural networks can do a better job of fixing images but they can make the images look too smooth. Generative Adversarial Networks or GANs for short are better at making images look real by focusing on how they look to people than just making the numbers match. The Enhanced Super-Resolution Generative Adversarial Network or ESRGAN is a type of GAN that does a better job by using a special generator and a discriminator that compares things in a relative way. ESRGAN uses something called Residual-in- Dense Blocks or RRDB which allows it to look deeper into the images and get more details all while staying stable.

The system we are talking about uses ESRGAN as its part and adds some extra features like fixing blurry images making the contrast better and filtering out certain frequencies. It also has a way to check the quality of the images and figure out what is wrong with them like if they're blurry or noisy. Then it can use the tools to fix the problems without needing someone to do it by hand. To make the system easier to use we made a graphical user interface that lets people interact with it and see the results. They can also compare the before and after images. See how well the system did. We use measures, like PSNR, SSIM, SNR, Mean Square Error, Root Mean Square Error and entropy to see how well the system is doing. We also use graphs to show the results. This way the system can. Adapt, but it is also practical to use.

II. LITERATURE REVIEW

Super-resolution was first used to improve low resolution images. The images that came out of this process were not very good. They did not have a lot of details and texture in them. The final output of super-resolution lacked details and texture. After that neural networks were invented to help with super-resolution. Convolutional neural networks were then used to improve super-resolution.

Significantly improved performance, as shown by C. Dong et al. [4] Where they used something called SRCNN to learn how to enhance resolution from start to finish. Similarly, K. Zhang et al. [1] Proposed a model that could handle different types of degradation which is important for real-world scenarios. The development of Generative Adversarial Networks by I. Goodfellow et al. [5] introduced a way to enhance images by using something called adversarial training. Building on this idea SRGAN was proposed by C. Ledig et al. [3] which combined this training with something called perceptual loss [6] resulting in very realistic images. However, there were some problems with this method, such as it being hard to train and having some artifacts. The ESRGAN was introduced by X. Wang et al. [2] which improved upon SRGAN by using something called Residual-in-Residual Dense Blocks [12] and removing some layers leading to feature learning and texture reconstruction.

There were improvements in deep learning based super-resolution with architectures like Enhanced Deep Residual Networks [7] which made the network design simpler while making it deeper and Residual Channel Attention Networks [8] which helped the network focus on the important features. The idea of balancing how accurate the image is with how good it looks, discussed by Y. Blau and T. Michaeli [11] is still a challenge in making these systems better. Recent work like using learning to improve the quality of images and using non-local attention-based models [13] [14] has helped to recover more details and context in images.

Recently diffusion-based super-resolution techniques were proposed by C. Saharia et al. [9] which refine the images using probabilistic models to get high-quality results. Although these methods produce results, they often need a lot of computational power and are not suitable for real-time applications.

In addition to making single images look better recent studies have explored ways to handle data. BasicVSR introduced a framework that uses the relationships between frames to enhance the images. IconVSR [16] extended this idea by incorporating temporal modeling for long sequences while EDVR [17] used deformable convolution techniques for better alignment and reconstruction in dynamic scenarios.

Despite being effective these methods are computationally intensive. Do not work well with user-interactive systems. Recent research has explored systems that can automatically identify issues with images and apply the enhancement techniques. However, many existing approaches do not combine learning models with traditional filtering methods and user-friendly interfaces.

Although there have been improvements in model design most studies focus on making the architectures better rather than making the system more integrated, usable and automated. The proposed work overcomes these limitations by combining ESRGAN with filtering techniques, quality assessment metrics and an automated enhancement mechanism within a unified framework. The system also has an interface that allows the user to interact with the system. The user can see the results.

Compare them. The system provides support for processing image and video data. It also helps with evaluating the performance of image and video enhancement. This approach is really good, for image and video enhancement because it does everything in one place.

III. PROBLEM STATEMENT

A variety of quality degradation types such as blurring, noise, low contrast, loss of fine structural information is inherent in data acquired in unconstrained or limited hardware capabilities scenarios. Sensor’s limitations, motion blur, compression losses, and adverse environmental factors are the main causes of image degradation. The above-mentioned issues compromise not only visual clarity but also the automated analysis systems that depend on good quality images. Conventional interpolation-based and cheap filtering-based enhancement algorithms provide marginal enhancement only, and are incapable of yielding fine details and natural textures.

Although deep learning based super-resolution techniques have demonstrated better reconstruction ability, most of the existing approaches are standalone models that do not support other enhancement features, adaptive processing, or user interaction. Also, most systems lack automatic quality assessment and do not have the necessary mechanisms for choosing appropriate actions, leaving this task to the user through the identification of suitable enhancement methods. They also lack integrated frameworks for applying multiple operations, comparing their results in real time, and assessing the results using common metrics.

The limited works have been done in extending such systems to work with sequential data and analyze their performance over multiple frames. Need for a comprehensive enhancement system that integrates deep learning-based image reconstruction with traditional filtering, automatic quality assessment, and adaptive decision making to overcome the identified shortcomings.

It should also include an interactive interface for visualization and comparison, as well as the support of metric-based evaluation, and the ability to process images and videos efficiently in one unified system.

IV. PROPOSED METHODOLOGY

The proposed system is a framework that uses deep learning and classic filtering techniques to enhance images. It also has an automated quality assessment and an interactive user interface.

The system can analyze one image or a series of images. It gives both numbers and visuals to show the quality of the original and the enhanced image.

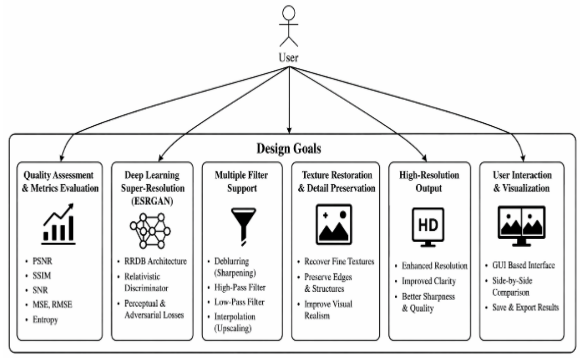


Fig.1 Design Goal Diagram of ESRGAN System.

A. ESRGAN-Based Super-Resolution

The main part of the system uses the Enhanced Super-Resolution Generative Adversarial Network. The generator architecture is made of Residual-in- Dense Blocks. This helps with feature extraction and stable training. Removing batch normalization layers helps keep brightness consistent and improves how well the system works.

The system up samples images using interpolation. This allows for reconstruction of high-resolution images while keeping structural and textural details.

B. Relativistic Discriminator

The discriminator in the ESRGAN uses an approach. It predicts if a real image is more real than a generated image. This makes the generator produce images that are closer to real-world images.

C. Perceptual and Adversarial Losses

The system uses loss functions to balance visual quality and reconstruction accuracy. Perceptual loss helps keep textures and edges. Adversarial loss makes the generated images look more real. Content-based loss ensures similarity.

D. Classical Filters

The system combines filtering with deep learning. The filters used are:

- Deblurring (sharpening based convolution filters)
- Edge enhancement, pass filtering
- Gaussian-based low-pass filtering for noise reduction
- Resolution improvement (interpolation)

These filters can be used separately as one or combined together.

E. Quality Assessment Module

The system has a quality assessment module. It checks the input data using perceptual measures like blur, contrast, brightness and noise. It also evaluates performance using metrics like PSNR, SSIM, SNR, MSE, RMSE and entropy.

F. Intelligent Automated Enhancement

The system has a decision-making mechanism. It automatically selects the technique based on quality assessment results. For example, if an image is blurry, it applies deblurring. If an image is noisy, it applies smoothing filters.

G. Sequential Data Processing

The system can process a sequence of images. It applies enhancement operations consistently. Averages performance indicators.

H. Graphical User Interface

The system has a user- graphical interface made with Tkinter. It allows users to upload data apply enhancement techniques and compare outputs. It also shows performance metrics. Allows saving of processed results. The original and enhanced images are displayed side by side.

V. SYSTEM ARCHITECTURE

The proposed system is designed in an organized way. It combines machine learning with methods to make images look better. The system is also easy to maintain. Can be improved without disrupting the current workflow.

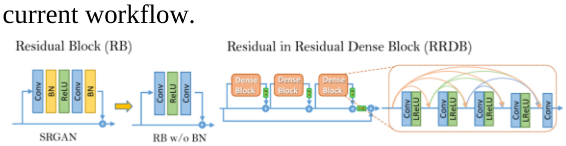


Fig. 2. ESRGAN Model Architecture [2].

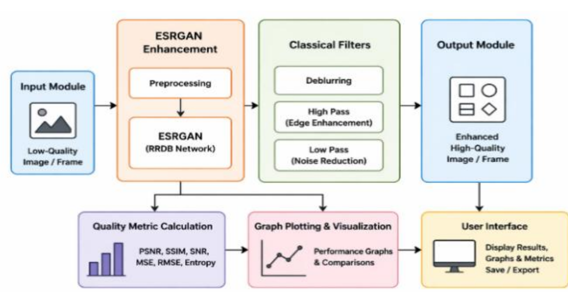


Fig. 3. Proposed System Architecture

This system can process data one by one or many at the time. It can also show how well it is performing and create graphs. Fig. 3.

The system architecture consists of the following modules:

- Input and Validation
- Preprocessing Module
- ESRGAN-Based Enhancement Module
- Filtering Module
- Quality Assessment and Metric Evaluation Module
- Intelligent Decision Module
- Display and Comparison Module
- Export Module

A. Input and Validation Module

This part of the system gets the input data. Makes sure it is in the right format. The system can handle formats like PNG, JPG and JPEG. It uses OpenCV to read the data and stores it as a -channel array.

To make sure the data is consistent the system checks a thing, including:

- Loading the image with OpenCV
- Format and integrity verification
- Checking the channels

B. Preprocessing Module

This part of the system gets the data ready for processing. It transforms the data into the format and converts the color space if needed.

The main tasks that are done are as follows:

- Image resizing using interpolation
- Color space conversion (BGR to RGB / grayscale where required)
- Normalization and formatting

This stage ensures consistent input representation across all modules.

C. ESRGAN Super-Resolution Module

This part of the system uses the ESRGAN method to make the images look better. It uses a kind of architecture called RRDBNet.

The system can capture the structure of the image and the details making the images look better with improved texture and edges.

The main things it does are:

- Feature extraction using RRDB blocks
- Residual learning
- Multi-stage up sampling
- High-resolution reconstruction

D. Filtering Module

This part of the system uses filters to make the images look even better. It can use these filters on their own or with the ESRGAN method.

The main tasks it does are as follows:

- Deblurring using sharpening filters
- Edge enhancement using high-pass filtering
- Noise reduction using Gaussian low-pass filtering
- Resolution scaling using interpolation

These filters enable targeted correction based on specific degradation types.

E. Quality Assessment and Metric Evaluation Module

This part of the system checks the quality of the input and output images. It uses perceptual metrics to do this. The operations performed by this part are as follows:

- Blur detection using Laplacian variance
- Contrast, brightness, and noise estimation
- Metric computation:
- MSE, RMSE (Error Measurement)
- PSNR, SNR (Signal Quality)
- SSIM (Structural Similarity)
- Entropy (Information Content)

F. Intelligent Decision Module

This part of the system makes decisions about how to process the images. It uses the results from the quality check to decide what to do.

For example, if the image is blurry, it will apply a deblurring filter. If the contrast is low, it will apply an enhancement filter. If there is noise it will apply a smoothing filter. This part of the system makes it easier to process the images without having to do it.

G. Display and Comparison Module

This part of the system shows the results. Compares them. It can display the enhanced images side by side and show the metrics. It allows users to compare original and enhanced outputs side by side and analyze performance metrics. The operations done by this part are:

- Image visualization
- Metric display
- Graph generation (bar charts, radar plots) using Matplotlib

H. Sequential Processing and Analysis

The system supports the sequential data processing by applying enhancement operations framewise. Metrics are computed periodically and aggregated to provide overall performance evaluation.

Graphical representations are used to visualize performance trends across the sequence.

I. Export Module

This module enables saving of the processed outputs in standard formats such as PNG. It allows users to store enhanced results for further use. The key operations are:

- File saving using OpenCV
- Path and format selection

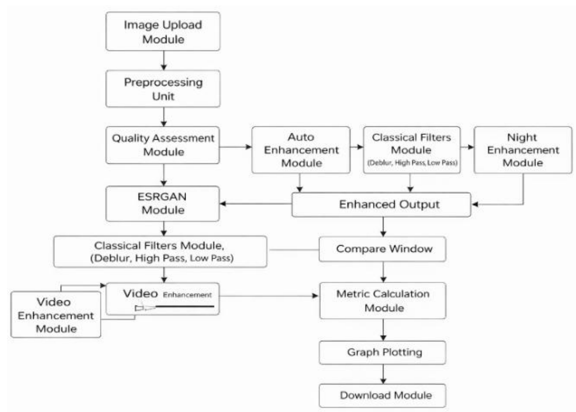


Fig. 4. Block Diagram of a Proposed ESRGAN System.

VI. RESULTS AND DISCUSSION

The ESRGAN based enhancement system shows improvements in image and video quality. This system is better than enhancement techniques. It uses learning and old image processing methods to make the visual output better. The ESRGAN model makes high-frequency details sharper and more appealing. Deblurring filters fix edges in images. High pass filtering makes structural details clearer. Low pass filtering reduces. Makes the image a bit smoother. The system can enhance images and videos automatically. This makes it more useful for real-world applications.

A. GUI Implementation

The system has a user- interface made with Tkinter. This interface lets users upload and display images. They can also see the enhanced images in time. The system supports filters and video enhancement. Users can compare the enhanced images easily.

The GUI has following features:

- Image upload and display
- Real-time enhancement preview
- Multiple filter selection
- Video enhancement functionality
- Metric visualization and comparison

The system is designed to handle multiple operations efficiently, ensuring smooth performance during processing. There is a separate panel which is used to display original and the enhanced image outputs for the comparison.

Fig. 5. ESRGAN Image Enhancement Graphical User Interface

B. Image Enhancement Results

The system supports enhancement techniques. Each technique gives results:

- ESRGAN Enhancement: makes high-resolution images with better texture and clarity
- Deblur Filter: Recovers sharpness from blurred images
- High-Pass Filter: Enhances edges and fine details
- Low-Pass Filter: reduces noise and smooths the image
- Upscaling: Increases resolution using interpolation

Fig. 6. Comparison of Original and ESRGAN Enhanced Image.

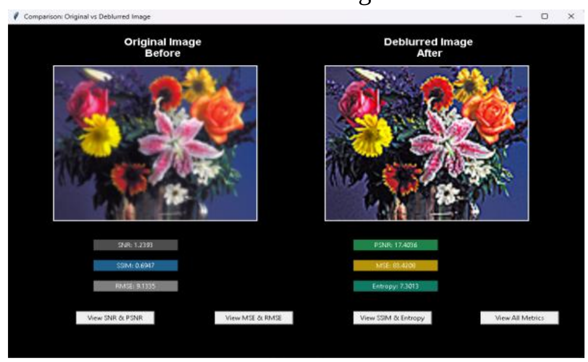


Fig. 7. Comparison of Original and Deblurred Image.

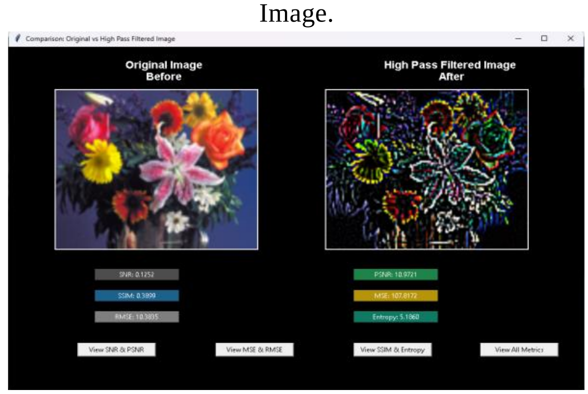


Fig. 8. Comparison of Original and High Pass Image.

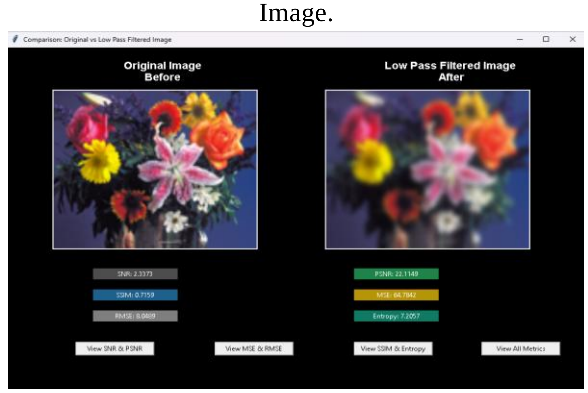


Fig. 9. Comparison of Original and Low Pass Filtered Image.

Fig. 10. Comparison of Original and Resolution Upscaled Image.

C. Low-Light (Night) Enhancement

The system can enhance low-light images. This is done using histogram equalization. The process is:

- Convert the image to LAB color space
- Enhance brightness using CLAHE
- Reconstruct the image

The observations are as follows:

- Improved visibility in dark regions
- Better contrast without excessive brightness
- Preservation of color information

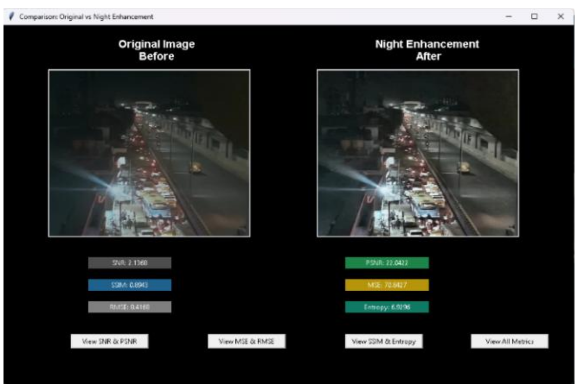


Fig. 11. Comparison of Original and Night Enhanced Image.

This feature is useful when images are taken in lighting.

D. Image Quality Assessment

The system checks the image quality before enhancement. This helps identify the type of degradation. The system checks:

- Blur level (Using Laplacian variance)
- Contrast (Intensity variation)
- Brightness (average pixel intensity)
- Noise level (Pixel variance)

Based on these values, the system classifies the image condition and provides a descriptive result such as:

- Blurry image
- Low contrast image
- High noise image
- Acceptable quality image

Fig. 12. Image Quality Assessment Window

This analysis helps the system make processing decisions.

E. Quality Assessment and Metrics

The system uses metrics to evaluate performance:

- PSNR (Peak Signal-to-Noise Ratio): Measures reconstruction quality
- SSIM (Structural Similarity Index): Evaluates structural similarity
- SNR (Signal-to-Noise Ratio): Indicates signal strength over noise
- MSE (Mean Squared Error): Measures average error
- RMSE (Root Mean Square Error): Square root of MSE
- Entropy: Represents information richness

These metrics compare the original and enhanced images. The results show that ESRGAN is better at:

- Higher PSNR values
- Improved SSIM scores
- Increased entropy (better detail representation)
- Lower MSE and RMSE

F. Auto Enhancement System

The system has an auto-enhancement module. This module analyzes image quality. Applies the best processing technique. The system checks:

- Blur (measured using Laplacian variance)
- Contrast (standard deviation of pixel intensity)
- Noise Level (variance of the image)

Based on these computed values, the system identifies the dominant issue and selects an appropriate enhancement method without user intervention.

The decision logic is:

- Blur detected implies Deblur filter applied
- Low contrast implies ESRGAN enhancement applied
- High noise implies Low-pass filter applied
- Slight blur implies High-pass filter applied

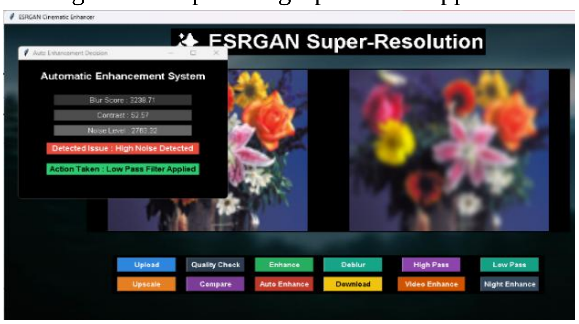


Fig. 13. Automatic Enhancement Decision Window

The system displays the analyzed values and the action taken.

G. Video Enhancement Results

The system enhances videos by applying frame-by-frame techniques.

Each frame is improved for noise reduction, contrast adjustment and sharpening.

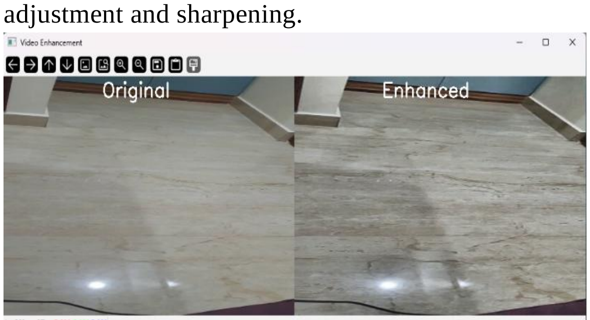


Fig. 14. Original and Enhanced Video Frame Comparison

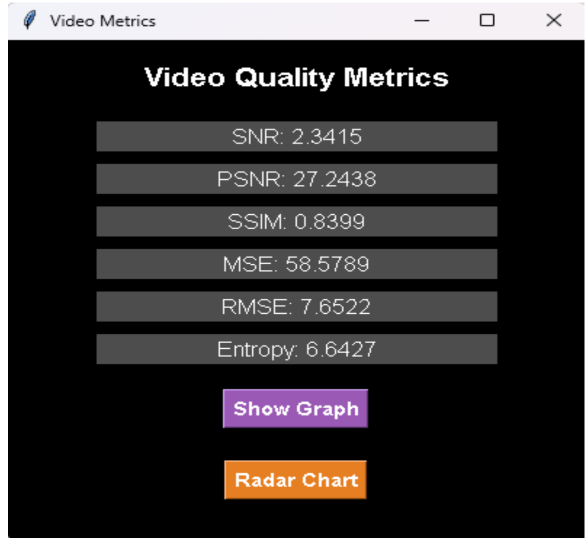


Fig. 15. Video Quality Metrics Display

The results show clarity, brightness and texture details.

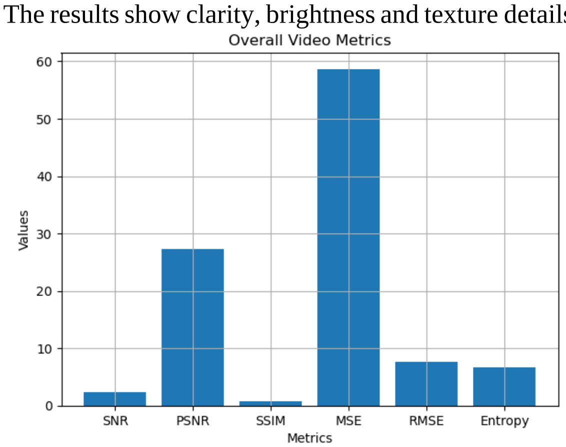


Fig. 16. Overall Video Metrics Bar Graph

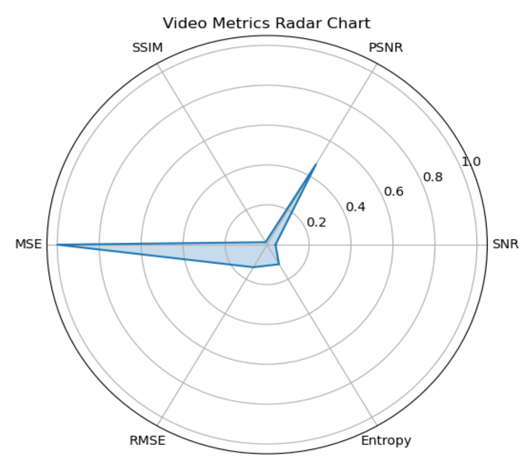


Fig. 17. Video Metrics Radar Chart

These results ensure the working of the video enhancement.

H. Metric Visualization

The system shows values using bar charts and radar charts. An input image is taken as first component for the measurement which is compared with the enhanced image for error computation.

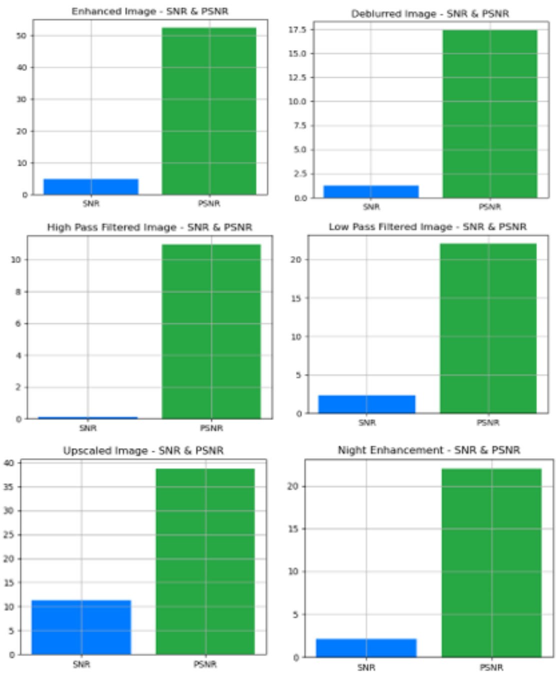


Fig. 18. PSNR and SNR Comparison for ESRGAN, Classical Filters, Upscale and Night Enhancement

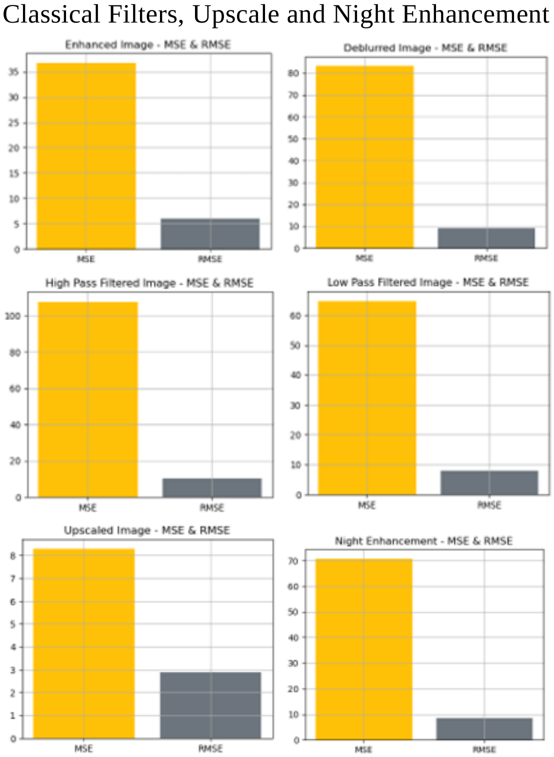


Fig. 19. MSE and RMSE Evaluation for ESRGAN, Classical Filters, Upscale and Night Enhancement

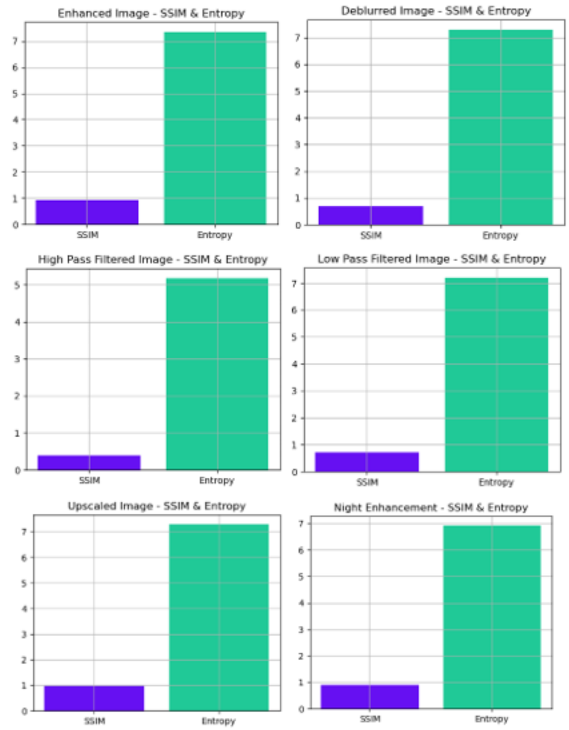
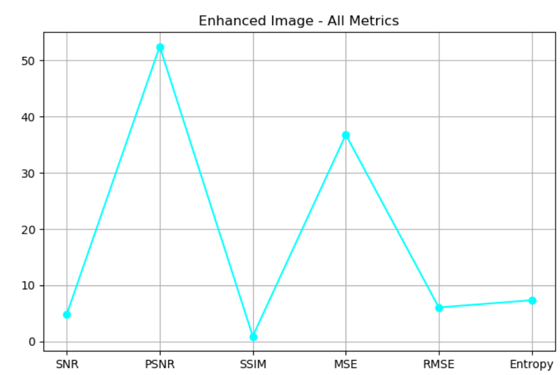
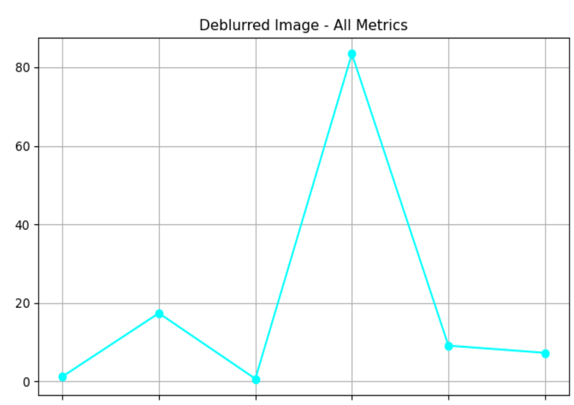


Fig. 20. SSIM and Entropy Metrics for ESRGAN, Classical Filters, Upscale and Night Enhancement





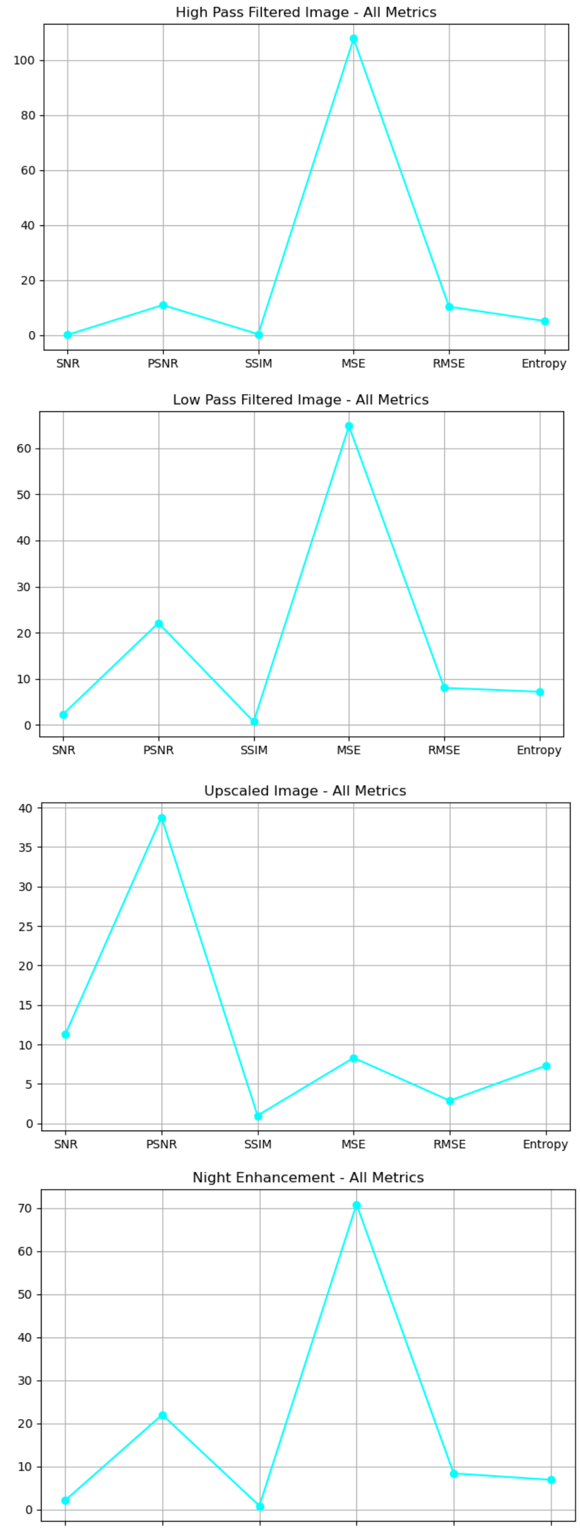


Fig. 21. Consolidated Performance Metrics for ESRGAN, Classical Filters, Upscale and Night Enhancement

I. Discussion

The results show the effectiveness of the proposed system:

- ESRGAN significantly improves visual quality by enhancing fine textures, sharpness, and structural similarity compared to traditional enhancement methods.
- Classical image processing technique such as deblurring, high-pass, and low-pass filtering effectively address specific issues like blur, noise, and edge enhancement.
- The auto-enhancement module introduces intelligent decision-making by analyzing image characteristics and selecting appropriate enhancement methods automatically.
- The video enhancement module demonstrates consistent performance across frames, improving clarity and maintaining visual continuity in dynamic content.
- Quantitative evaluation using the metrics such as PSNR, SSIM, SNR, MSE, RMSE, and entropy ensures an objective and reliable assessment of enhancement performance.
- Graphical visualizations, including bar charts and radar charts, provide clear insights into comparative performance across different enhancement techniques.
- The integrations of the deep learning, traditional image processing, and visualization tools results in robust and efficient multimedia enhancement system.

The system is efficient and suitable for real-world multimedia enhancement applications.

VII. CONCLUSION

The proposed system integrates ESRGAN-based super-resolution with filtering techniques, quality assessment and automated decision-making. The system improves resolution, clarity and adapts to different types of degradation. The inclusion of performance evaluation metrics and graphical visualization enables analysis and comparison of results. The system shows performance in terms of structural preservation, noise reduction and perceptual quality. The interactive interface enhances usability. The framework provides a combination of deep learning capabilities, adaptive processing and user-oriented design.

Future work can focus on extending the system for real-time processing scenarios and improving efficiency. Additional enhancements include support for data formats, incorporation of domain-specific learning strategies and further optimization of enhancement techniques, for challenging conditions. Expanding the training dataset and integrating advanced architectures will improve the robustness and generalization capability of the system.

REFERENCES

[1] K. Zhang, W. Zuo, and L. Zhang, "Learning a Single Convolutional Super-Resolution Network for Multiple Degradations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 3262–3271.

[2] X. Wang *et al*., "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, 2018.

[3] 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, pp. 4681–4690.

[4] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2016.

[5] I. Goodfellow *et al*., "Generative Adversarial Nets," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2014, pp. 2672–2680.

[6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.

[7] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2017, pp. 136–144.

[8] Y. Zhang *et al*., "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.

[9] C. Saharia *et al*., "Image Super-Resolution via Iterative Refinement Using Diffusion Models," *arXiv preprint* arXiv:2304.14820, 2023.

[10] S. Anwar and N. Barnes, "Densely Residual Laplacian Super-Resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1192–1204, 2020.

[11] Y. Blau and T. Michaeli, "The Perception-Distortion Tradeoff," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 6228–6237.

[12] Y. Zhang *et al*., "Residual Dense Network for Image Super-Resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 2472–2481.

[13] S. Park *et al*., "Improving Perceptual Quality by Adversarial Training for Real-World Image Super-Resolution," *IEEE Trans. Image Process.*, vol. 32, pp. 103–115, 2023.

[14] Y. Mei *et al*., "Image Super-Resolution with Non-Local Sparse Attention," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 3517–3526.

[15] K. Chan *et al*., "BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 4947–4956.

[16] K. Chan *et al*., "IconVSR: Learning Temporal Information for Video Super-Resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 13214–13223.

[17] X. Wang *et al*., "EDVR: Video Restoration with Enhanced Deformable Convolutional Networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2019.