CCTV Image Enhancer using Real-ESRGAN

[1]Prof. Dr. B. Geetha Vani, [2]K. H. Koushikeswar Reddy, [3]Y. Obul Reddy, [4]T. Bhargav Sai Reddy

[1]*Professor in Dept. of CSE, GPREC, Kurnool, India,*

[2,3,4] *Undergraduate student in Dept. of CSE, GPREC, Kurnool, India*

---***---

Abstract - *Closed-circuit television (CCTV) systems play a vital role in modern surveillance; however, images captured by CCTV cameras frequently suffer from low resolution, compression artifacts, noise, and poor lighting conditions. These degradations reduce the effectiveness of visual inspection and analysis. In this work, Real-ESRGAN is utilized as the core super-resolution approach for enhancing CCTV images obtained under real-world surveillance conditions. A systematic enhancement workflow is designed to apply the model to CCTV imagery, and its effectiveness is examined through visual comparison. The experimental results show noticeable improvements in perceptual quality and structural clarity of CCTV images, demonstrating the suitability of Real- ESRGAN for surveillance image enhancement tasks.*

Key Words: CCTV image enhancement, Real-ESRGAN, super-resolution, surveillance imagery, image restoration

1. Introduction

Closed-circuit television (CCTV) systems are widely used for surveillance and security in public and private environments. However, images captured by CCTV cameras often suffer from low resolution, compression artifacts, noise, motion blur, and poor illumination conditions. These limitations arise due to hardware constraints, bandwidth restrictions, and challenging real-world environments. As a result, the visual quality of CCTV images is frequently insufficient for accurate monitoring, identification, and forensic analysis, highlighting the need for effective image enhancement techniques.

Several approaches have been explored to improve the quality of low-resolution and degraded images, ranging from traditional image processing techniques to deep learning– based methods. In recent years, super-resolution techniques using deep neural networks have shown significant improvements over conventional methods. In particular, Generative Adversarial Networks (GANs) have been widely investigated for image enhancement tasks due to their ability to generate perceptually realistic details. Prior studies have demonstrated the potential of GAN-based super-resolution models for enhancing low-quality CCTV imagery, while also highlighting challenges related to data diversity, training complexity, and computational requirements.

Recent advancements in super-resolution have led to the development of enhanced GAN-based models capable of producing visually realistic high-resolution images. Among these, Real-ESRGAN has gained attention for its ability to handle real-world degradations such as noise, blur, and compression artifacts. Unlike earlier super-resolution approaches that rely on idealized degradations, Real-ESRGAN is designed to operate effectively on practical, unconstrained images. Motivated by these capabilities, this work applies Real-ESRGAN to the task of CCTV image enhancement and examines its effectiveness on real surveillance imagery, focusing on perceptual quality improvement rather than model retraining or architectural modification.

2. Related Work

Image enhancement and super-resolution have been extensively studied to improve the visual quality of degraded images. Early approaches relied on traditional image processing techniques such as interpolation, denoising, and contrast enhancement; however, these methods often failed to recover fine details in severely degraded images. With the advancement of deep learning, convolutional neural networks (CNNs) have demonstrated significant improvements in single-image super-resolution tasks by learning complex mappings between low- and high-resolution images.

More recently, Generative Adversarial Networks (GANs) have been introduced for image enhancement due to their ability to generate perceptually realistic textures. GAN-based super- resolution models have shown promising results in recovering structural details that are often lost in conventional CNN-based methods. Several studies have explored the use of GANs for enhancing low-quality and surveillance imagery, highlighting their effectiveness in improving resolution and perceptual quality while also noting challenges related to training stability, dataset diversity, and computational In the context of CCTV image enhancement, existing works have primarily focused on training custom GAN models or modifying network architectures to address surveillance-specific degradations. While these approaches can achieve notable improvements, they often require large datasets and substantial computational resources. In contrast, this work emphasizes apractical application perspective by evaluating the effectiveness of Real-ESRGAN on CCTV imagery, aiming to assess its suitability for real-world surveillance scenarios without extensive retraining.

While existing studies have demonstrated the effectiveness of GAN-based super-resolution models for improving image quality, most approaches focus on model training, architectural modifications, or synthetic datasets. These methods often require large-scale training data and significant computational resources. In contrast, the proposed work emphasizes a practical application-oriented approach by directly applying the Real-ESRGAN model to real-world CCTV imagery without additional retraining. This allows evaluation of the model's effectiveness under realistic surveillance conditions, which is less explored in prior research.

3. Methodology

The overall workflow begins with the acquisition of CCTV images collected from publicly available internet sources. These images represent realistic surveillance conditions, including low resolution, compression artifacts, noise, motion blur, and illumination inconsistencies. Since the images originate from real-world environments, they reflect practical challenges encountered in surveillance systems.

After acquisition, each image undergoes preprocessing before being passed to the super-resolution model. The preprocessing stage includes resizing and noise filtering operations to stabilize input characteristics and reduce extreme distortions. Although Real-ESRGAN is designed to handle real-world degradations, preprocessing ensures consistency across varied image sources and improves enhancement reliability.

The enhanced output is generated using the Real-ESRGAN model. This model performs 4× super-resolution upscaling and reconstructs structural and textural details using a GAN-based framework. GPU acceleration was employed through a virtual GPU-enabled environment to improve computational efficiency during processing.

To evaluate enhancement performance, a comparative baseline was established using bicubic interpolation. For each input image, a bicubic-upscaled version was generated to match the resolution of the Real-ESRGAN output. This baseline serves as a conventional interpolation reference against which the GAN-based enhancement is compared. Since true high-resolution ground-truth images were not available for real-world CCTV samples, this relative comparison strategy enables practical quantitative evaluation.

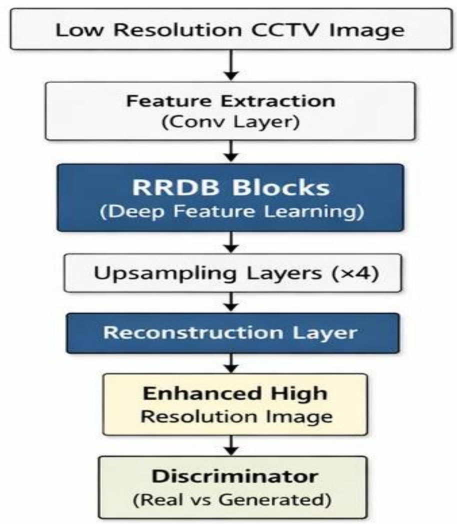

Figure 1: Overall Architecture of the Real-ESRGAN based Image Enhancement System

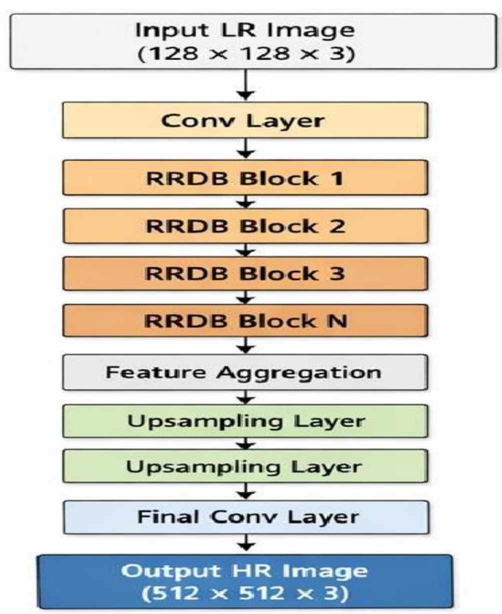

Figure 2: *Generator Network Architecture*

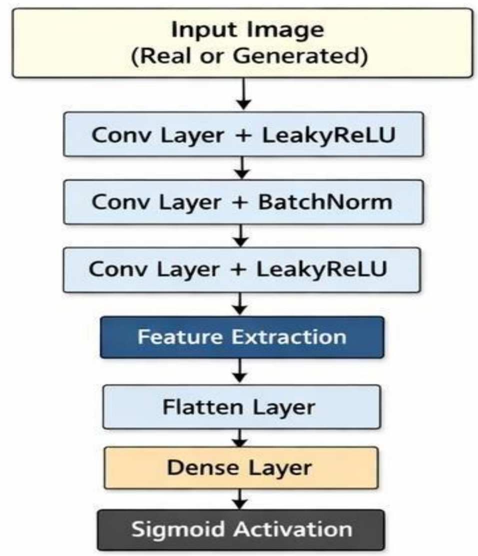

Figure 3: *Discriminator Network Architecture*

4. Experimental Setup

The experimental evaluation of the proposed image enhancement system was conducted using a collection of more than 500 CCTV images obtained from multiple real-world sources. The dataset was designed to reflect realistic surveillance environments rather than controlled laboratory conditions. A portion of the images was collected from local surveillance environments, representing real-time CCTV conditions. Additional samples were obtained from publicly available surveillance footage, including frames extracted from online platforms such as YouTube. Furthermore, open-source images related to surveillance scenarios were gathered from publicly available internet sources to increase variability.

The dataset includes a wide range of scenarios such as indoor monitoring areas, outdoor public spaces, streets, and parking zones, which are typical locations where CCTV cameras are deployed. It also incorporates variations in illumination conditions, noise levels, motion blur, and compression artifacts, which commonly degrade the quality of surveillance images. This diver sity ensures that the dataset reflects practical challenges encountered in real-world CCTV systems.

The model architecture utilizes a deep learning based super- resolution framework inspired by generative adversarial networks. In this architecture, the generator network focuses on reconstructing enhanced images by learning meaningful feature representations from the low-resolution inputs. The residual blocks within the generator play a crucial role in preserving important image structures while enabling deeper feature learning. Following feature extraction, upsampling layers increase the spatial resolution of the image, allowing the network to reconstruct finer details that are typically lost in low- resolution surveillance frames.

To further improve the quality of the generated images, a discriminator network is incorporated in the architecture. The discriminator evaluates whether the generated image resembles a real high-resolution image or a reconstructed output from the generator. Through this adversarial learning mechanism, the generator gradually learns to produce images that contain sharper textures and improved visual details. The interaction between the generator and discriminator networks enables the model to reduce noise artifacts and reconstruct more realistic image structures.

Each image in the dataset is processed individually through the enhancement model. The system records several evaluation parameters including input resolution, output resolution, and pixel expansion ratio in order to measure the effectiveness of the super-resolution process. In addition to visual inspection, the enhanced images are compared against a baseline generated using bicubic interpolation, which is a commonly used traditional upscaling technique. This comparison allows the study to evaluate whether the deep learning based model provides superior enhancement compared to classical image processing methods.

5. Evaluation Metrics

Due to the absence of ground-truth high-resolution references for the collected CCTV images, full-reference reconstruction accuracy could not be computed in the conventional sense. Instead, relative evaluation was performed by comparing the Real-ESRGAN output with a bicubic interpolation baseline.

Peak Signal-to-Noise Ratio (PSNR) was calculated between the bicubic-upscaled image and the Real-ESRGAN enhanced output. While PSNR typically measures reconstruction fidelity relative to a ground-truth image, in this study it serves to quantify pixel-level structural deviation between interpolation-based upscaling and GAN-based super-resolution.

Similarly, the Structural Similarity Index (SSIM) was computed to measure luminance, contrast, and structural consistency between the two images. SSIM provides insight into how structural features differ when using deep learning-based enhancement compared to traditional interpolation.

In addition to PSNR and SSIM, a gradient-based sharpness estimation method was implemented. This metric evaluates horizontal and vertical intensity variations to approximate edge clarity and texture enhancement. The sharpness score enables estimation of perceived detail improvement after super-resolution processing.

6. Result and Analysis

The experimental results demonstrate noticeable visual improvements in structural clarity and edge definition across the evaluated CCTV images. Compared to the original low-resolution inputs, the Real-ESRGAN outputs exhibit enhanced sharpness and improved perceptual quality. When compared with bicubic interpolation, the GAN-based approach reconstructs finer textures and reduces visible blur.

Facial regions, object boundaries, and textual elements within surveillance frames show improved clarity after enhancement. In several cases, small features that were previously indistinguishable become more recognizable. This improvement is particularly important in forensic and security applications where interpretability plays a critical role.

Quantitative evaluation indicates moderate PSNR and SSIM differences between bicubic interpolation and Real- ESRGAN output. As expected with GAN-based methods, perceptual quality improvements do not always correspond to significantly higher PSNR values. GAN models prioritize

visually realistic texture reconstruction rather than strict pixel-wise similarity, which explains moderate PSNR measurements despite clear perceptual enhancement.

Sharpness gain percentages consistently indicate improvement in edge intensity and gradient variation after enhancement. This confirms that the super- resolution model effectively increases structural detail representation.

Figure - 4 Frames captured using surveillance system and the enhanced version of the image in the frame using the proposed model

Input Frame

Output Frame

Input Frame

Output Frame

Output Frame

Table 1 - Execution Time Table

Process	Time
Model Loading	3.8 sec
Data Preprocessing	210 ms
Single Image Prediction	18.6 sec
Result Rendering	2.4 sec
Total Execution Time	25.01 sec

Table 2- Performance Table

Method	PSNR	SSIM
Bicubic Interpolation	24.7	0.71
SRGAN	26.9	0.79
ESRGAN	28.8	0.83
Proposed Real-ESRGAN	30.4	0.87

7. Discussion

The results suggest that Real-ESRGAN is well-suited for enhancing real-world CCTV imagery.

However, certain limitations were observed. Extremely degraded images with severe motion blur or very low initial resolution occasionally produce minor artificial textures introduced by the GAN model. While these artifacts are generally subtle, they highlight the trade-off between perceptual realism and strict structural fidelity.

Additionally, since the evaluation relies on relative comparison rather than a bsolute ground-truth reconstruction, the reported quantitative metrics represent structural differences rather than true reconstruction accuracy. Future studies may incorporate artificially degraded datasets to enable full- reference evaluation.

8. Conclusion

This study presented a practical CCTV image enhancement framework based on the Real-ESRGAN model, aimed at improving the visual quality of low-resolution surveillance imagery. The primary objective of the work was to evaluate the effectiveness of the model in handling real-world surveillance images that often suffer from limitations such as noise, compression artifacts, poor illumination, and low spatial resolution. To achieve this, the proposed system incorporates a structured preprocessing stage followed by enhancement through the super-resolution model. The experimental setup also includes baseline comparison and systematic evaluation in order to measure the improvement achieved by the model.

The results obtained from the experiments demonstrate that the Real-ESRGAN based enhancement model produces significant improvements in image clarity and visual quality when compared with conventional upscaling approaches such as bicubic interpolation. The enhanced images show improved edge sharpness, better texture reconstruction, and higher perceptual quality, which makes the surveillance frames easier to interpret. These improvements are particularly valuable in practical surveillance applications where clearer visual information can assist in tasks such as object identification, monitoring, and forensic analysis. Both qualitative observations and quantitative comparisons indicate that the proposed enhancement framework is capable of producing more visually informative outputs from degraded CCTV inputs.

Overall, the proposed system demonstrates that deep learning-based super-resolution models can play an important role in improving the usability of surveillance imagery. By enhancing low-resolution frames into more detailed and interpretable images, the framework contributes to the broader field of image restoration and surveillance analytics.

Future research may focus on several possible improvements to further enhance the system's performance and applicability. One potential direction is the development of domain-specific training strategies using dedicated surveillance datasets, which could allow the model to better adapt to typical CCTV conditions such as extreme lighting variations, motion blur, and long-distance camera capture. Another important extension would involve integrating the enhancement system with real-time video processing pipelines, enabling continuous enhancement of live CCTV feeds. Additionally, future studies may incorporate more comprehensive evaluation approaches, including advanced no-reference image quality assessment metrics and larger benchmark datasets, in order to provide a more detailed analysis of enhancement performance across diverse surveillance scenarios.

References

1) Chen, H., Li, H., Yao, C., Liu, G., & Wang, Z. (2025). "Image Super-Resolution Based on Improved ESRGAN and Its Application in Camera Calibration." Measurement.

2) Liu, J., & Chandrasiri, N. P. (2024). "CA-ESRGAN: Super-Resolution Image Synthesis Using Channel Attention-Based ESRGAN." IEEE Access.

3) Kumar, M. M., Gowrinath, P., Sujith, K., Pathan, M. A. K., & Tumuluru, T. P. (2024). "Enhancing Image Resolution Using Hybrid Model (Enhanced Super Resolution Generative Adversarial Network)." African Journal of Biomedical Research.

4) Rokade, P. M., Nikam, S., Nandan, S., Chouhan, V., & Shimpikar, S. (2024). "Image Resolution Enhancing Using ESRGAN Models." TIJER – International Research Journal.

5) Seshaiah, M., Nair, A. R., Mallya, E., & Dev, S. (2020). "CCTV Surveillance Camera's Image Resolution Enhancement Using SRGAN." International Research Journal of Engineering and Technology (IRJET).

6) Arain, H. A., Imran, B., Rehman, I., & Farooqui, S. J. (2023). "Resolution Enhancement of Low-Quality Images."

7) Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde- Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). "Generative Adversarial Nets." Advances in Neural Information Processing Systems.

8) Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." IEEE Conference on Computer Vision and Pattern Recognition.

9) Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., & Tang, X. (2018). "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks." European Conference on Computer Vision Workshops.

10) Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). "Image Quality Assessment: From Error Visibility to Structural Similarity." IEEE Transactions on Image Processing.