Digital Image Processing Term Project
- Video Enhancer

R14942141 張佑禕 R13921002 林祉瑩 R14945026 張詠哲

Abstract

USB cameras have become increasingly popular in portable, embedded, and cost-sensitive devices due to their compact form factor, low power consumption, and ease of integration. Despite these advantages, the image quality produced by USB cameras is often inferior to that of industrial or professional imaging systems. Common issues include inaccurate exposure control under varying illumination conditions, noticeable barrel distortion caused by low-cost wide-angle lenses, and overall low spatial resolution and detail degradation resulting from sensor limitations and aggressive on-device compression. These factors significantly restrict the applicability of USB cameras in scenarios where reliable visual perception and image fidelity are required.

To address these challenges, this project proposes an integrated digital image processing pipeline designed to enhance the visual quality of images captured by USB cameras. The pipeline consists of three major stages. First, a light enhancement module is employed to compensate for exposure errors and insufficient illumination by improving global brightness distribution and local contrast, thereby recovering visually meaningful details in underexposed or unevenly lit regions. Second, a distortion correction stage is introduced to mitigate barrel distortion effects by geometrically rectifying the captured images, ensuring that straight lines and object shapes are preserved for subsequent processing and analysis. Finally, a deep learning–based super-resolution module built upon the Real-ESRGAN framework is applied to enhance spatial resolution and perceptual quality by reconstructing high-frequency details that are typically lost in low-quality USB camera outputs.

The proposed pipeline aims to provide a practical and computationally feasible solution for improving image quality in resource-constrained imaging systems. By combining traditional image processing techniques with modern deep learning–based super-resolution methods, the system effectively addresses both optical and sensor-induced degradations. Experimental results demonstrate

that the enhanced images exhibit improved visual clarity, reduced geometric distortion, and higher perceptual quality, making the proposed approach suitable for embedded vision, low-cost monitoring, and portable imaging applications.

I. Introduction

With the rapid adoption of USB cameras in embedded systems and low-cost vision applications, the limitations of such devices have become increasingly apparent, particularly in challenging imaging environments. Compared to industrial-grade cameras, USB cameras typically employ inexpensive sensors and wide-angle lenses, which results in unstable exposure behavior, significant geometric distortion, and limited spatial resolution. These deficiencies not only degrade visual quality but also negatively affect the reliability of downstream computer vision tasks. Consequently, there is a growing demand for an efficient and modular image enhancement framework that can compensate for hardware constraints while remaining suitable for real-time or near–real-time deployment.

In this work, a multi-stage image processing pipeline is introduced to systematically address the major sources of degradation present in USB camera imagery. The first stage focuses on adaptive light enhancement, where the average brightness of each frame is analyzed to determine the appropriate enhancement strategy. Under low-light conditions, white balance correction based on the Gray World assumption is applied to mitigate color bias, followed by Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve local contrast without excessively amplifying noise. To further suppress noise commonly observed in dark scenes, median filtering is employed. In addition, temporal luma blending is introduced by combining the luminance channel of the current frame with that of previous frames, which effectively smooths brightness fluctuations and reduces visible flickering while preserving chrominance information to maintain motion-related color consistency.

For scenes affected by overexposure, the proposed system incorporates adaptive recovery mechanisms. When sudden flashes or transient lighting changes are detected, the system blends the affected frame with the most recent well-exposed frame to conceal washed-out regions. In cases where overexposure persists, global darkening and contrast stretching are applied to recover suppressed details, while gamma correction is used to compress highlights and improve tonal balance. Furthermore, in extreme low-illumination scenarios approaching pitch-black conditions, the enhancement process is deliberately constrained by blending toward a black reference frame, thereby preventing excessive noise amplification that would otherwise dominate the image.

The second stage of the pipeline addresses geometric distortions introduced by low-cost wide-angle optics. A fisheye distortion correction module is implemented using camera-specific intrinsic parameters and distortion coefficients obtained through calibration. By employing OpenCV's fisheye undistortion and remapping functions, the captured frames are rectified to restore geometric consistency, ensuring that straight lines and object proportions are preserved. This step is critical not only for improving visual realism but also for maintaining spatial accuracy in subsequent processing stages.

Finally, a deep learning–based super-resolution module is applied to enhance spatial detail and perceptual quality. The Real-ESRGAN framework is adopted due to its robustness in handling real-world degradations commonly present in low-quality video streams. To achieve computational efficiency, the input video is segmented into chunks and processed in parallel using multiprocessing, with support for both multi-GPU and optimized single-GPU execution. The enhanced video segments are then merged using a video encoding pipeline, producing a high-resolution output with improved sharpness and reduced artifacts.

By integrating adaptive light enhancement, distortion correction, and deep learning–based super resolution into a unified framework, the proposed approach effectively mitigates the inherent limitations of USB cameras. The modular design of the pipeline allows for flexible deployment across a wide range of embedded and portable imaging systems, making it suitable for applications such as low-cost surveillance, robotic vision, and mobile imaging platforms.

II. Methodology

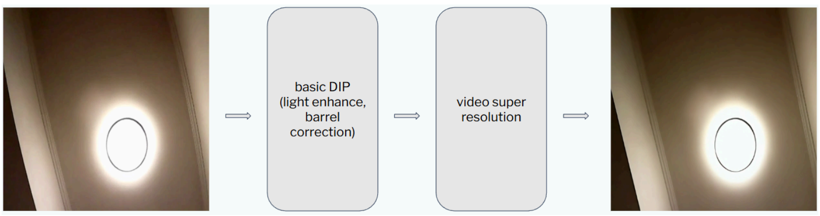


Fig 1. Our video enhancer pipeline

A. Light Enhancing

The light enhancement stage constitutes the first component of the proposed image processing pipeline and is designed to adaptively improve visual quality under a wide range of illumination conditions. Due to the limited dynamic range and unstable exposure control of USB cameras, consecutive frames may exhibit significant brightness fluctuations, noise amplification, or detail loss caused by clipping. To address these issues, the light enhancer integrates multiple image processing techniques, including frame blending, gamma correction, gray-world white balance, CLAHE equalization, luminance-domain blending, RGB-domain blending, and contrast stretching. These operations are selectively activated according to the estimated illumination characteristics of each input frame, enabling scene-aware enhancement rather than uniform global processing.

At the beginning of the process, the average illuminance of the input frame is computed and compared against predefined thresholds to classify the lighting condition. If the mean illuminance exceeds the overexposure threshold, the frame is considered overexposed, indicating potential detail loss due to white clipping in bright regions. In this case, the frame is first darkened to suppress saturated intensities and recover partially clipped details. To further stabilize the visual output, the processed frame is blended with the most recent "good frame" that exhibits acceptable exposure, if such a reference frame is available. This temporal blending strategy helps conceal sudden exposure spikes caused by flashes or abrupt lighting changes. After blending, contrast stretching is applied to re-expand the effective dynamic range and improve overall visual clarity. If no suitable reference frame exists, such as during prolonged overexposure, gamma correction is employed as an alternative strategy to compress highlights and rebalance tonal distribution.

In scenarios where the input frame is extremely underexposed, with mean illuminance falling below a predefined threshold THRESH_BLACK_HOLE, the frame is treated as carrying little to no meaningful visual information. Such frames are often dominated by sensor noise rather than valid scene content. To prevent aggressive enhancement from amplifying noise, the system blends the current frame toward a pure black reference frame while maintaining temporal similarity with the preceding frame. This approach produces a visually stable and noise-free output, avoids abrupt transitions, and preserves continuity in video sequences where the scene is intentionally or temporarily dark.

For mildly underexposed frames, where the mean illuminance lies between THRESH_BLACK_HOLE and THRESH_LOW_LIGHT, a more detailed enhancement procedure is applied. First, gray-world white balance is used to correct color casts introduced by non-uniform or low-intensity lighting, under the assumption that the average reflectance of the scene is achromatic. Next, CLAHE is applied to improve local contrast and redistribute brightness more evenly across the frame while limiting excessive noise amplification. To further suppress noise that becomes prominent in low-light conditions, median blurring is employed. Finally, luminance-domain blending is performed with respect to the previous frame, where only the LUMA channel is temporally smoothed. This strategy reduces flickering and noise-freezing artifacts while preserving chrominance information from the current frame, thereby preventing ghosting effects on moving objects.

Through this adaptive, threshold-driven design, the light enhancement module effectively stabilizes exposure, suppresses noise, and improves perceptual quality across diverse lighting conditions. By combining spatial enhancement techniques with temporal blending mechanisms, the module provides a robust preprocessing foundation for subsequent distortion correction and super-resolution stages in the pipeline.

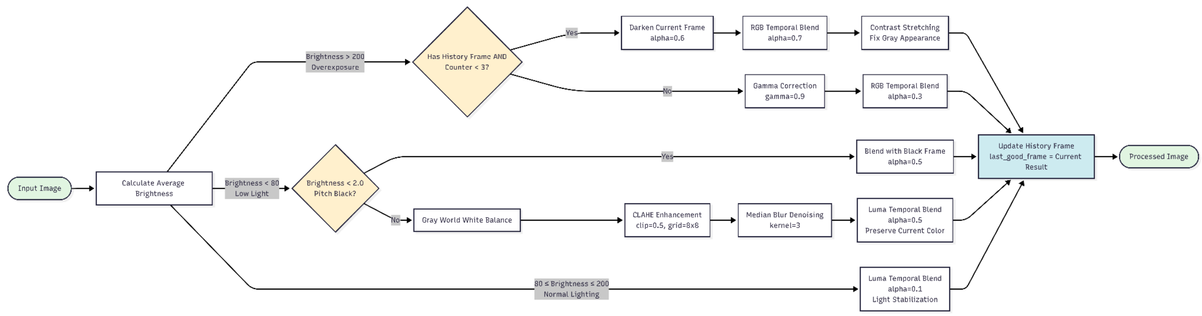


Fig 2. The light enhancing flowchart

B. Barrel Distortion Recovering

Although the ultimate goal of this project is to perform image correction on the captured camera images, fisheye distortion cannot be effectively addressed using a purely image-based transformation. Fisheye distortion originates from the inherent imaging geometry of wide-angle lenses, where the mapping from three-dimensional scene rays to two-dimensional image coordinates is highly nonlinear. As a result, distortion correction requires an explicit understanding of the camera's internal imaging parameters rather than treating the distortion as an arbitrary visual deformation.

Camera calibration provides a principled way to model the geometric imaging process of the camera by estimating both the intrinsic matrix and the nonlinear distortion parameters. By recovering these parameters through calibration, the fisheye distortion can be removed in a geometrically consistent manner, ensuring that straight lines in the real world are correctly preserved after image rectification.

In this work, we adopt Zhang's camera calibration method, a widely used and flexible calibration technique that relies on a planar calibration target. Compared to other calibration approaches that require specialized three-dimensional calibration setups, Zhang's method only needs a flat calibration pattern observed from multiple viewpoints, making it both practical and robust for real-world applications. This method estimates the camera intrinsic matrix together with nonlinear distortion parameters through observations of the calibration target captured under different orientations and positions.

The technical pipeline for fisheye distortion correction consists of a sequence of well-defined processing steps. First, a set of two-dimensional image points is extracted by detecting checkerboard corner locations in the captured images. These image points are associated with known three-dimensional coordinates defined on a planar calibration target. Based on the resulting 2D–3D correspondences across multiple views, camera calibration is performed to jointly estimate the camera intrinsic matrix and fisheye distortion parameters.

To provide the required calibration data, a checkerboard calibration pattern consisting of 7×9 blocks is used as the planar target. Instead of using individual images, we record a video sequence of the checkerboard while varying its position and orientation relative to the camera. Frames are sampled from the video to ensure a diverse set of viewpoints, which improves the numerical stability and accuracy of the calibration process. This approach allows us to efficiently collect a large number of calibration samples while maintaining consistent image resolution and camera settings.

Once calibration is completed, the estimated parameters are used to construct a pixel-wise undistortion mapping. For each pixel in the rectified image, the corresponding location in the distorted image is computed by applying the inverse fisheye projection defined by the calibrated camera model. This mapping subsequently applied to incoming images through an efficient remapping operation, enabling consistent and repeatable fisheye correction.

In practice, the calibration process is implemented using the fisheye calibration module provided by OpenCV. This module integrates Zhang's planar calibration framework with a fisheye camera model and performs nonlinear optimization to jointly estimate the camera intrinsic matrix and fisheye distortion coefficients. The resulting parameters fully characterize the camera's imaging geometry under the fisheye projection model.

Within this model, the camera intrinsic matrix and the fisheye distortion coefficients jointly define the projection from three-dimensional rays to two-dimensional pixels. These parameters are subsequently used to construct a model-based undistortion mapping , enabling pixel-wise rectification of the distorted images.

This approach ensures that the undistortion process remains consistent with the physical imaging geometry of the camera, allowing straight lines and spatial structures in the real world to be accurately preserved after correction.

C. Super Resolution with Real-ESRGAN

The final stage of the proposed pipeline addresses the limitations of spatial resolution and detail loss. Images captured by low-cost USB cameras often suffer from a combination of aggressive on-chip compression, sensor noise, and optical blurring. Standard upscaling methods, such as bicubic interpolation, fail to recover high-frequency details and often exacerbate existing artifacts. To overcome this, we employ Real-ESRGAN [1], a deep learning framework specifically designed for blind super-resolution.

1. High-Order Degradation Modeling

Traditional super-resolution (SR) methods usually train on synthetic data generated by simple bicubic downsampling. However, real-world degradations are complex and cannot be modeled by a single downsampling operation. Real-ESRGAN addresses this mismatch by introducing a high-order degradation modeling process.

As described by Wang et al. [1], the degradation process is modeled as a chain of operations repeated multiple times. The synthetic data generation pipeline simulates:

- **Blur:** Using both isotropic and anisotropic Gaussian kernels, as well as generalized Gaussian kernels, to mimic camera defocus and motion blur.
- **Noise:** Injecting Gaussian and Poisson noise to simulate sensor thermal noise and photon shot noise.
- **Resize:** Utilizing random resizing operations (area, bilinear, bicubic) to simulate various scaling artifacts.
- **JPEG Compression:** Applying compression to mimic the block artifacts introduced by the USB transmission bandwidth limits.
- **Sinc Filter:** A critical addition in Real-ESRGAN is the use of sinc filters to simulate ringing and overshoot artifacts, which are common in images that have undergone in-camera sharpening—a standard feature in many commercial USB cameras.

By training on pure synthetic data generated through this rigorous "second-order" degradation process, the network learns to restore details while suppressing the specific artifacts (ringing, blockiness, and noise) inherent to our target hardware.

2. Implementation and Video Optimization

While the original Real-ESRGAN architecture utilizes a heavy Residual-in-Residual Dense Block (RRDB) network, our implementation prioritizes a balance between perceptual quality and inference speed. We utilize the SRVGGNetCompact architecture (realesr-general-x4v3), which provides a 4× upscaling factor. This model is optimized for general-purpose restoration, offering a significant reduction in computational cost compared to the standard ESRGAN generator while maintaining superior visual fidelity.

To handle high-definition video streams efficiently, we developed a robust processing pipeline that addresses Input/Output (I/O) bottlenecks and GPU memory constraints:

- **FFmpeg Streaming and Piping:** Instead of decoding the entire video into individual frames on the disk—which creates massive I/O overhead, we utilize ffmpeg-python to pipe raw video frames (BGR24 format) directly from the video container into the system memory. Similarly, processed frames are piped directly back into an FFmpeg encoder (libx264) to generate the final output video.
- **Tiling Strategy:** Super-resolution on high-resolution inputs consumes significant VRAM. To prevent CUDA Out-Of-Memory (OOM) errors, the input frames are dynamically divided into smaller tiles (e.g., with a tile size of 0, indicating auto-tiling, or specific dimensions based on available memory) with overlapping padding. These tiles are processed independently and then stitched back together, ensuring seamless restoration without boundary artifacts.

III. Experimental Result

A. Light enhancing

To evaluate the effectiveness of the proposed light enhancement module, we conducted experiments under both underexposed and overexposed lighting conditions.

Underexposed condition

In underexposed scenes, the original frames exhibit noticeable color bias, low overall brightness, and poor contrast, which together lead to significant loss of detail in dark regions. Details are often indistinguishable from noise, and the visual appearance is dominated by color shifts caused by insufficient and uneven illumination.

As shown in the experimental results (Fig. 3), after applying the proposed light enhancement pipeline, the enhanced images show a substantial improvement in brightness distribution and local contrast. Color bias is mitigated through gray-world white balance, which corrects channel-wise intensity imbalance and restores more natural color appearance. Subsequently, CLAHE enhances local contrast by redistributing intensity values within local regions, making fine structures in previously dark areas more distinguishable.

To prevent noise amplification introduced by contrast enhancement, median filtering is applied to suppress impulsive noise while preserving edge information. In addition, luminance-domain temporal blending further stabilizes visual output across consecutive frames by smoothing brightness fluctuations without introducing color ghosting.

Compared to naive global brightness adjustment, the proposed method produces more natural-looking results by combining spatial enhancement with temporal blending.

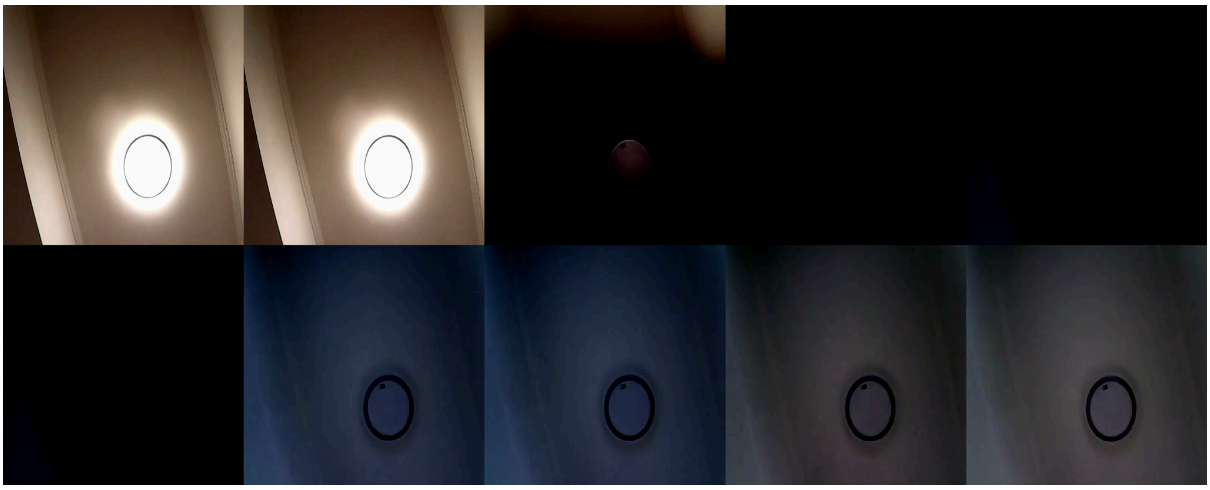
Overexposed Conditions

For overexposed scenes, the original images suffer from saturated highlights and loss of detail in bright regions. Sudden lighting changes or flash-like illumination often introduce visually disturbing artifacts and abrupt brightness fluctuations. As shown in Fig.4, the proposed enhancement module successfully mitigates these effects by adaptively darkening the input frames and blending them with previously well-exposed reference frames when available.

After enhancement, details in bright regions are partially recovered, and the overall brightness becomes more balanced. Contrast stretching further improves visual clarity by redistributing intensity values, while gamma correction helps compress highlights in cases of persistent overexposure. As a result, the enhanced frames exhibit smoother brightness transitions and improved visual consistency compared to the original input. In addition, the enhanced frames exhibit improved tonal continuity in bright regions, with reduced saturation artifacts and smoother intensity transitions. This prevents abrupt visual discontinuities and produces a more visually coherent appearance across consecutive frames.

The experimental results indicate that the proposed light enhancement module effectively stabilizes exposure, improves perceptual quality, and preserves scene details under challenging illumination conditions. This adaptive preprocessing stage provides a robust foundation for subsequent distortion correction and super-resolution stages in the pipeline.

Before enhance



After enhance

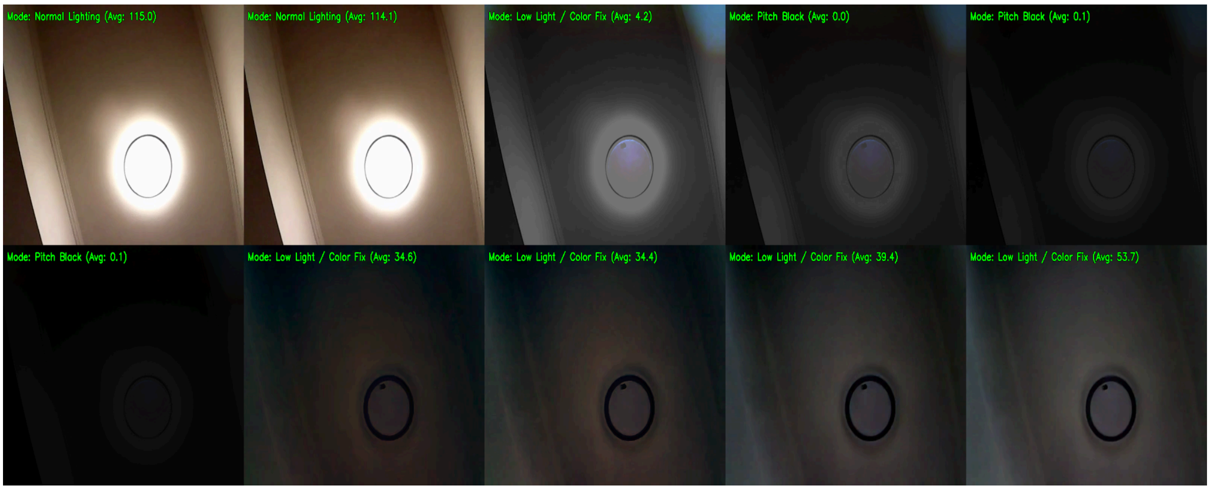


Fig 3. On under exposure condition

Before enhance

After enhance

Fig 4. On over exposure condition

B. Barrel Distortion Recovering

To evaluate the effectiveness of the proposed barrel distortion recovery module, we compare representative frames before and after fisheye distortion correction. The original images captured by the USB camera exhibit noticeable barrel distortion, where straight lines near the image boundaries appear curved and object proportions are visibly warped.

After applying the proposed fisheye distortion correction, as shown in Fig. 5, the rectified images demonstrate a clear improvement in geometric consistency. Straight structures in the scene, such as object edges and background boundaries, are correctly restored as straight lines, indicating that the nonlinear lens distortion has been effectively compensated. By enforcing consistency with the calibrated camera imaging geometry, the corrected images provide a reliable and visually coherent representation of the scene, forming a stable foundation for the subsequent super-resolution stage.

Before

After

Fig 5. Barrel Distortion Recovering

C. Super Resolution with Real-ESRGAN

To evaluate the effectiveness of the super-resolution module, we compare the visual quality of the original input frames with the enhanced outputs produced by the Real-ESRGAN–based approach. Due to sensor limitations, optical blur, and compression artifacts, the original USB camera images exhibit noticeable loss of fine details and blurred edges.

After applying the proposed super-resolution module, as shown in Fig. 6, the enhanced images demonstrate a clear improvement in spatial resolution and perceptual sharpness. Fine structures that are barely distinguishable in the original frames, such as edges, textures, and small-scale patterns, become more visually prominent. Importantly, this enhancement is achieved without introducing significant ringing or overshoot artifacts, which are commonly observed in aggressive sharpening-based methods.

Overall, the experimental results confirm that the Real-ESRGAN–based super-resolution module substantially enhances the perceptual quality of USB camera imagery. When combined with the preceding light enhancement and distortion correction stages, the proposed pipeline produces images with improved clarity, geometric consistency, and spatial detail, making it suitable for practical applications such as low-cost monitoring, embedded vision, and portable imaging systems.

Before

After (Basic DIP + SR)

Fig 6. Super Resolution

IV. Conclusion

In this project, we have presented an integrated video enhancement pipeline specifically designed to address the inherent hardware limitations of low-cost USB cameras. By systematically combining adaptive light enhancement, calibration-based distortion correction, and deep learning–driven super-resolution, the proposed system effectively mitigates common image degradations including unstable exposure, geometric warping, and loss of high-frequency details.

Experimental results confirm that the light enhancement module successfully balances global illumination and recovers local details in both underexposed and overexposed conditions, while the distortion correction stage restores geometric fidelity essential for reliable visual perception. Furthermore, the integration of the Real-ESRGAN framework, optimized with efficient video piping and tiling strategies, significantly elevates spatial resolution and perceptual sharpness without introducing prohibitive computational overhead.

In conclusion, this work demonstrates that a hybrid approach, merging traditional image processing techniques with modern neural networks, can transform low-quality video streams into high-fidelity outputs. The proposed pipeline offers a robust, modular, and computationally feasible solution suitable.

V. Reference

[1] Wang, Xintao, et al. "Real-esrgan: Training real-world blind super-resolution with pure synthetic data." *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.