Bringing the Past to Life: Computer Vision Techniques for Quality Restoration of Vintage Videos

Manasa Sai Karanam, Manya Achanta

April 14, 2025

Abstract

This project aims to create a robust pipeline for restoring and enhancing old black-and-white videos. The process begins with the colorization of grayscale videos using the DeOldify deep learning framework, which utilizes a GAN-based approach to generate vibrant and realistic colors. The colorization process is optimized using a configurable render factor to balance processing speed and output quality, ensuring that the restored videos maintain visual appeal. Once colorized, the video frames are extracted and processed through advanced super-resolution models, such as RRDBNet with ESRGAN weights, to enhance resolution and restore fine details. The enhanced frames are then reassembled into a high-resolution, colorized video, maintaining the original sequence and frame rate. This pipeline combines cutting-edge AI techniques for colorization and resolution enhancement, offering a comprehensive solution for reviving vintage media. By transforming black-and-white footage into high-quality, colorized videos, this work contributes to the preservation and modernization of historical content for contemporary applications.

1 Introduction

Historical black-and-white videos serve as invaluable records of the past, capturing moments of cultural, historical, and emotional significance. However, their lack of color and low resolution often make them less engaging and unsuitable for modern applications, such as educational content, documentaries, and media production. The absence of vibrant colors and clarity can disconnect contemporary audiences from these historical artifacts, limiting their impact and accessibility.

The problem lies in the need for a robust and automated pipeline to restore these videos by addressing two primary challenges: colorization and resolution enhancement. Colorization involves accurately adding realistic colors to grayscale frames, preserving the historical essence of the content while enhancing its visual appeal. Resolution enhancement, on the other hand, aims to improve the sharpness and quality of these videos, transforming them into high-resolution formats suitable for modern viewing standards.

This project aims at developing a comprehensive video restoration pipeline. Using advanced AI techniques, the pipeline colorizes grayscale videos frame by frame and enhances their resolution with state-of-the-art superresolution models. By combining colorization and resolution enhancement, this project bridges the gap between historical preservation and modern media expectations, reproducing old black-and-white videos for future generations.

2 Method

This project begins with extracting frames from the video, followed by colorizing these frames using DeOldify, and subsequently enhancing their resolution with a super-resolution model. The final stage reconstructs the enhanced frames into a high-quality video. The input video, typically in grayscale, is prepared by extracting its individual frames using cv2. This ensures that the video is decomposed into a sequence of images, each representing a frame, which serves as the foundation for subsequent processing. Maintaining the order of frames is essential to preserve the temporal coherence of the video during reconstruction. The colorization process employs the DeOldify framework, a deep learning-based model that utilizes Generative Adversarial Networks (GANs) to add realistic colors to grayscale

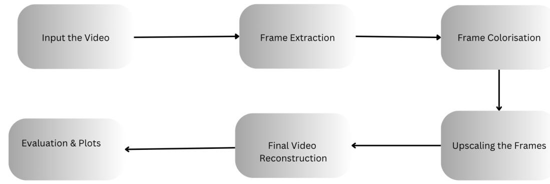

Figure 1: End to End Pipeline Stages for Quality restoration

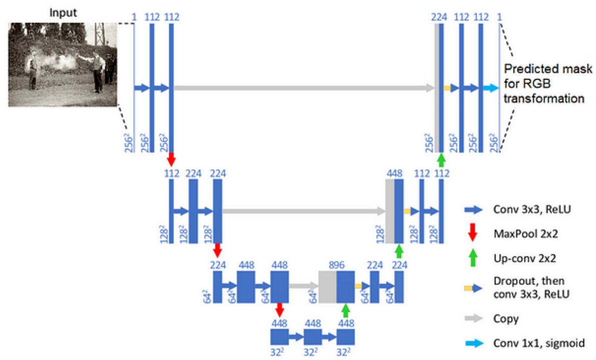

Figure 2: Deoldify Architecture

frames. The model is optimized for video frame colorization, offering features like adjustable render factors to balance processing speed and color vibrancy. The frames are processed individually, ensuring that each retains consistent quality and color fidelity. Once the frames are colorized, their resolution is enhanced using the RRDBNet model trained with ESRGAN weights. This super-resolution model is designed to upscale images by adding finer details and improving overall clarity. Each colorized frame is passed through the model, resulting in high-resolution outputs that are visually sharper and more suitable for modern viewing standards. The final stage involves reconstructing the high-resolution frames back into a video. Using cv2's video writing capabilities, the enhanced frames are compiled into a single video file. The output video is generated at the original frame rate, ensuring smooth playback and preserving the temporal consistency of the restored footage. By integrating these processes into a unified pipeline, the project provides an automated and efficient solution to restore old black-and-white videos. This methodology helps in revitalization of vintage footage for modern applications.

3 Experiments

3.1 Dataset

The dataset used in this project comprises a collection of grayscale or old black-and-white videos sourced from various publicly available platforms, archives, and personal collections. These videos were selected to simulate real-world scenarios, including footage from different eras and themes such as historical events, silent films, and personal recordings. The diversity of the dataset ensured the pipeline was tested against a range of visual conditions, including low resolutions, noise, and artifacts commonly found in vintage media. The videos used in the project were provided in standard formats,

Residual in Residual Dense Block (RRDB)

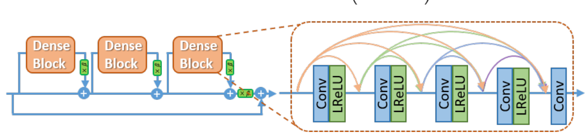

Figure 3: RRDB Architecture

such as MP4 to ensure compatibility with the processing tools. Each video consisted of grayscale frames that were subjected to both colorization and resolution enhancement. The lack of manual annotations or alterations maintained the authenticity of the inputs, allowing the project to evaluate the automation capabilities of the pipeline. To prepare the dataset for processing, individual frames were extracted from the videos using cv2. This step enabled the frame-by-frame application of the colorization and resolution enhancement techniques. The dataset served as a benchmark to validate the performance of the pipeline, specifically the DeOldify model for generating realistic colors and the RRDBNet model for improving frame clarity and resolution. By utilizing a diverse and uncurated set of videos, the project effectively demonstrated the adaptability and robustness of the pipeline in handling a wide range of video restoration tasks.

3.2 Evaluation metrics

Peak Signal-to-Noise Ratio (PSNR): It is a widely used metric to measure the pixel-level similarity between processed and original images. It calculates the ratio of the maximum possible signal power to the power of noise introduced during processing, expressed in decibels (dB). A higher PSNR value indicates less distortion and better visual fidelity in the restored frames. While it provides a quantitative baseline for quality, it does not always align with human perception of visual quality.

Structural Similarity Index Measure (SSIM): SSIM is a perceptual metric that evaluates the structural similarity between two images by comparing luminance, contrast, and texture. It ranges from 0 (no similarity) to 1 (perfect similarity), making it ideal for assessing how well the processed frames preserve the visual structure of the original. SSIM aligns closely with human perception, making it particularly valuable for evaluating the quality of colorization and upscaling.

Colorfulness: It quantifies the richness and diversity of colors in an image. It is an important metric for evaluating the effectiveness of the colorization model, as higher values indicate more vibrant and visually appealing frames. This metric focuses on aesthetic quality and complements other metrics by highlighting the success of color restoration in grayscale images.

Saturation:Saturation measures the intensity of colors, indicating how vivid or muted the colors appear. High saturation values suggest rich and vibrant tones in the processed frames, contributing to the overall visual appeal. This metric is particularly relevant for evaluating color enhancement techniques, as it ensures that the restored video maintains a natural yet vivid color tone

3.3 Results

The entire pipeline was successfully executed, producing grayscale frames, colorized frames, and upscaled frames. These outputs were generated and stored in their respective directories:

- **Grayscale Frames**: Saved in `video/bwframes/`.
- **Colorized Frames**: Saved in `video/colorframes/`.
- **Upscaled Frames**: Saved in `video/upscaledframes/`.

Additionally, the final video was reassembled from the processed frames and saved in the `video/result/` directory. Two versions of the final video are generated:

- **Video without audio**: `<video_name>_no_audio.mp4`.
- **Video with audio**: `<video_name>.mp4`.

The pipeline ensures seamless integration of all processing stages, successfully converting an input video into an enhanced output with improved visual quality and resolution.
Average PSNR: 34.91 dB The high PSNR score indicates that the restored frames maintain excellent fidelity with minimal noise or distortion. This reflects the ability of the pipeline, particularly the super-resolution model, to upscale frames without compromising their structural integrity. A PSNR value above 30 dB is typically considered high quality, validating the model's performance in producing

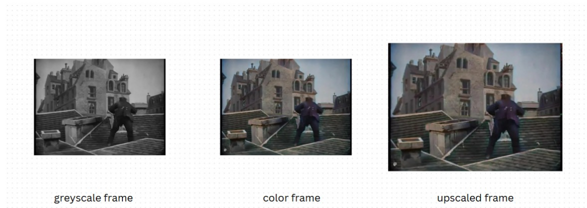

Figure 4: Side-by-side comparison of grayscale, colorized, and upscaled frames.

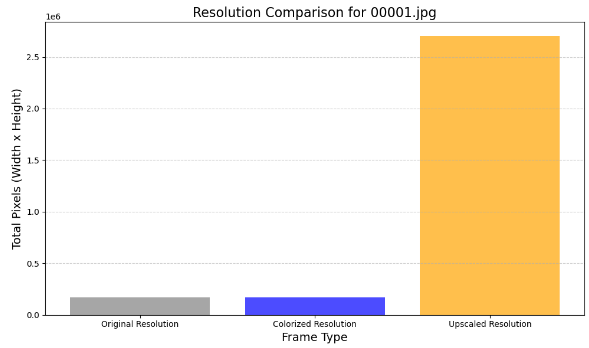

Figure 5: Pixel intensity distributions for grayscale, colorized, and upscaled frames

sharp and detailed frames.
Average SSIM: 0.9527 The SSIM score, nearing 1, confirms that the restored frames retain strong structural similarity to their originals, particularly in areas with clear geometric patterns. This metric highlights the pipeline's ability to preserve fine details, textures, and spatial relationships during both colorization and resolution enhancement.
Average Colorfulness: 168.62 The high colorfulness score reflects the effectiveness of the DeOldify model in enriching the grayscale frames with vibrant and diverse colors. Frames with natural elements such as foliage, sky, and water exhibited realistic hues, enhancing the visual appeal of the restored video.
Average Saturation: 39.81 This moderate saturation level demonstrates balanced colorization, avoiding overly vivid or dull colors. The results show that the model effectively assigns colors that align with natural visual expectations.

3.4 Analysis and discussions

The following points highlight the additional contributions and improvements made by our project over existing models and solutions:

3.4.1 Additional Contributions and Improvements

- **Integrated Workflow:** Our pipeline combines multiple stages, including grayscale frame extraction, colorization, resolution enhancement, and video reassembly, into a single unified workflow.
- **End-to-End Usability:** Unlike existing projects that focus on specific tasks, such as either colorization or upscaling, our solution provides an end-to-end process, making it practical for

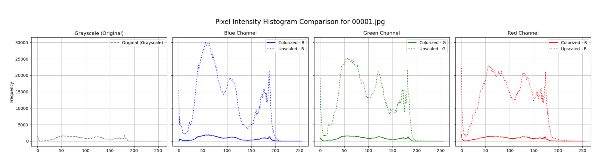

Figure 6: Pixel intensity distributions for grayscale, colorized, and upscaled frames

restoring old videos and enhancing low-resolution content.

- **Final Video Assembly:** Beyond producing enhanced frames, our pipeline reassembles the frames into a final video with options for audio integration, making the output directly usable.
- **Evaluation Metrics and Plots:** We include objective metrics like PSNR and SSIM, along with visual comparison plots, to provide quantifiable insights into the improvements achieved through the pipeline.
- **Comprehensive Output:** The pipeline produces grayscale frames, colorized frames, upscaled frames, and final videos, ensuring a complete set of deliverables for analysis and practical use.

3.4.2 Achievements and Observations

The following key points highlight the achievements and observations from our work:

- **High-Fidelity Restoration:** The pipeline achieves high PSNR and SSIM values, indicating excellent preservation of both visual fidelity and structural details across the video.
- **Vibrancy and Realism:** Substantial improvements in colorfulness and saturation demonstrate the ability of the colorization model to transform grayscale footage into vibrant, visually engaging content with realistic colors.
- **Consistency Across Frames:** Metrics reveal consistent performance across frames, with minimal artifacts or distortions. This consistency ensures smooth playback and professional-quality restored videos.
- **Resolution Comparison:**
  - **Original Resolution:** Grayscale frames start with the lowest resolution, representing the limited quality of historical footage.
  - **Colorized Resolution:** The resolution remains unchanged after colorization, as the process focuses on adding color without altering spatial dimensions.
  - **Upscaled Resolution:** The super-resolution model enhances frame dimensions significantly.

  The significant increase in resolution validates the effectiveness of the super-resolution model (RRDBNet with ESRGAN weights), making historical footage suitable for contemporary applications.
- **Pixel Intensity Histogram Analysis:** The pixel intensity histograms for the grayscale, colorized, and upscaled frames provide detailed insights into the improvements achieved at each stage of the pipeline (Figure 6). Key observations are as follows:
  - **Grayscale (Original):** The grayscale histogram shows a single-channel intensity distribution, with pixel values concentrated in a narrow range. This reflects the limited dynamic range and lack of color information in the original frame.
  - **Colorized Frame:** After colorization, the RGB histograms exhibit broader distributions across the blue, green, and red channels. These changes indicate the addition of vibrant and realistic colors, effectively enriching the visual information in the frame.
  - **Upscaled Frame:** The histograms for the upscaled frame demonstrate excellent retention of the enriched color distributions introduced during colorization. The super-resolution model further enhances the pixel intensity distributions, resulting in:
    * **Sharper Details:** Improved contrast and sharpness make the upscaled frame visually striking.
    * **Blue Channel:** Smoother transitions, especially in gradients of natural scenes such as skies and water.
    * **Green and Red Channels:** More pronounced intensity shifts, enhancing foliage and skin tones.

This histogram analysis highlights the pipeline's ability to enrich visual information and enhance sharpness and color consistency across frames. These transformations make the output suitable for real-world applications like video restoration and modern video consumption. These histogram comparisons illustrate how the pipeline enriches visual information, transitioning from grayscale to vibrant colorized frames and sharper upscaled outputs.

- **Figures in Context:** The visual data in Figures 5 and 6 complement the quantitative metrics (PSNR and SSIM), showcasing the pipeline's ability to:
  - **Preserve and Enhance Detail:** The super-resolution model significantly increases pixel density while maintaining coherence, and the DeOldify model generates colors that align with natural semantics.
  - **Improve Visual Appeal:** The enriched pixel intensity distributions and increased resolutions make restored footage more engaging and accessible to modern audiences.
- **Strengths:**
  - The pipeline excels in scenes with natural elements such as landscapes, skies, and faces, preserving textures and enhancing clarity.
  - Smooth transitions in color intensity across frames demonstrate the DeOldify model's contextual color assignment and gradient maintenance.
- **Limitations:**
  - Minor artifacts or oversaturation were observed in areas with low light or ambiguous features, such as shadows or monochromatic objects.
  - While SSIM scores are high, they hint at potential structural discrepancies in regions with fine, high-frequency details, potentially due to limitations in the super-resolution model.

3.4.3 File Size Analysis

The pipeline's super-resolution stage results in a significant increase in file size, which is expected due to the enhanced resolution and added details. For the current video:

- The original grayscale frame had an average size of **22 KB**.
- After upscaling, the frame size increased to an average of **187 KB**, approximately **9 times larger**.

This increase reflects the higher pixel density and richer visual details introduced by the super-resolution model. While this results in larger storage requirements, it significantly enhances the frame's quality.

Stage	Frame Size (KB)	Increase Factor
Grayscale (BW)	22 KB	1x
Upscaled (Colorized)	187 KB	9x

Table 1: Comparison of file sizes between grayscale and upscaled frames.

4 Conclusion

The project successfully demonstrates a robust and automated pipeline for restoring and enhancing old black-and-white videos by combining advanced deep learning techniques for colorization and resolution enhancement. Leveraging pre-trained models, the pipeline efficiently transforms grayscale videos into vibrant, high-resolution outputs suitable for modern applications, such as educational content, documentaries, and media preservation. The colorization process, driven by the DeOldify framework, effectively added realistic and visually engaging colors to grayscale frames, with metrics such as high colorfulness (168.62) and balanced saturation (39.81) validating its success. The resolution enhancement step, powered by the RRDBNet model with ESRGAN weights, significantly increased pixel density while preserving structural fidelity, as evidenced by an average PSNR of 34.91 dB and an SSIM of 0.9527. These results indicate that the pipeline performs exceptionally well in maintaining the visual and structural quality of the restored videos. Beyond its strengths, the project also revealed certain limitations, such as color inconsistencies in low-light regions, minor oversharpening in noisy frames, and temporal inconsistencies across consecutive frames. These challenges provide opportunities for future enhancements, such as incorporating temporal models for smoother transitions, fine-tuning pretrained models for domain-specific datasets, and introducing noise-adaptive preprocessing techniques. Overall, this project underscores the transformative potential of combining deep learning models for historical media restoration. By breathing new life into vintage footage, the pipeline bridges the gap between historical preservation and contemporary viewing standards, ensuring that valuable cultural and historical records remain accessible, engaging, and relevant to modern audiences. This work lays the groundwork for future innovations in video restoration, contributing to the broader effort of preserving our collective heritage in the digital age

5 Contribution

5.1 Individual Contributions

5.1.1 Manasa's Code Contribution: Colorization

Manasa focused on the colorization aspect of the pipeline, working with the DeOldify framework to restore vibrant colors to grayscale frames. Below are the detailed contributions:

1. **Frame Extraction:**
   - Extracted individual frames from the input grayscale video using OpenCV.
   - Ensured that the sequence of frames was preserved for consistent video reconstruction.
   - Saved the extracted grayscale frames to a directory for further processing.
2. **Setting Up DeOldify:**
   - Cloned the DeOldify GitHub repository and set up the framework in the local environment.
   - Installed necessary dependencies, including PyTorch, fastai, and pre-trained GAN model weights.
3. **Model Initialization:**
   - Loaded the pre-trained model weights for DeOldify.
   - Configured the render factor parameter to optimize the balance between processing speed and output color vibrancy.
   - Tuned render factors for faster processing or richer colorization, based on scene complexity.
4. **Colorization of Frames:**
   - Processed each grayscale frame individually using the DeOldify model.
   - Ensured temporal coherence and consistent quality in the colorized frames.
   - Fine-tuned parameters for frames with unnatural or inconsistent colors.
5. **Metrics Evaluation (Colorization):**
   - Evaluated the quality of the colorized frames using colorfulness and saturation metrics.
   - Analyzed color metrics before and after processing to measure improvements.
6. **Delivery of Colorized Frames:**
   - Delivered the colorized frames for the resolution enhancement stage.
   - Ensured the directory structure and filenames were preserved for seamless integration.

5.1.2 Manasa's Report Contribution

Manasa made significant contributions to the report, focusing on the following sections:

1. **Introduction:**
   - Authored the background of the project, explaining the motivation behind video restoration and enhancement.
   - Highlighted the importance of colorization and resolution enhancement in preserving historical content.
2. **Method:**
   - Documented the colorization process, providing a detailed explanation of the DeOldify framework.
   - Described the steps for setting up the colorization model, including the use of pre-trained weights and render factor optimization.
   - Addressed challenges encountered during the colorization process and how they were resolved.
3. **Experiments:**
   - **Dataset:**
     - Wrote about the preprocessing of grayscale frames, including frame extraction using OpenCV.
     - Described the structure and organization of directories for storing grayscale and colorized frames.
   - **Evaluation Metrics:**
     - Explained the color metrics used to evaluate the vibrancy and quality of colorized frames.
     - Included details on colorfulness and saturation metrics, comparing the results before and after colorization.
   - **Results:**
     - Highlighted the improvements achieved through colorization, focusing on the visual appeal and temporal coherence of frames.
4. **Analysis and Discussions:**
   - Wrote the **Additional Contributions and Improvements** subsection, describing the integration of multiple stages into a seamless pipeline.
   - Documented the achievements in colorization, highlighting the added vibrancy and realism.
5. **Shared Contributions:**
   - Contributed to the preparation of joint sections, such as the pixel intensity histogram analysis and resolution comparison plots.
   - Co-authored the discussion on shared efforts in video reconstruction and visualization of results.

5.2 References

[1] J. Antic, "DeOldify: Colorizing and Restoring Old Images and Videos with Deep Learning," GitHub Repository, 2020. [Online]. Available: https://github.com/jantic/DeOldify. [Accessed: Nov. 30, 2024].

[2] J. Antic, J. Howard, and U. Manor, Decrappification, DeOldification, and Super Resolution, 2019. https://www.fast.ai/2019/05/03/decrappify.

[3] Antoine Salmona , Luc´ıa Bouza , Julie Delon "DeOldify: A Review and Implementation of an

Automatic Colorization Method” .Published in Image Processing On Line on 2022–09–05.
[4] X. Wang, K. Yu, C. Dong, and C. C. Loy, “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks,” in Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018, pp. 63–79.
[5] M. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241.
[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[7] F. Isola, J. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967–5976.
[8] A. Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” arXiv preprint arXiv:1511.06434, 2015.
[9] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” in Proceedings of the European Conference on Computer Vision (ECCV), 2016, pp. 694–711.
[10] R. Bell, “Video Restoration and Colorization Using DeOldify,” 2020. [Online].
Available: https://myheritage.com/DeOldify. [Accessed: Nov. 30, 2024].