AI video & audio enhancement

By: Abdallah Ashraf, Seif Ahmed, Mohamed Ahmed,
Nour El Deen, Seif El Deen Ahmed, Abdelrahman Emad
Supervised by: Prof. Dr. Sally Saad,
TA. Mohamed Essam
Faculty of Computer and Information Sciences - Ain Shams University

Introduction

In an age dominated by digital communication and content sharing, the quality of multimedia especially videos and audio plays a critical role in delivering clear, engaging and professional educational content experiences, unfortunately many videos suffer from poor resolution and noisy audio this becomes a real significant issue as clarity is essential resulting in a very bad user experience.
to tackle these challenges, our project introduces an intelligent enhancement system that leverages AI techniques to improve both video and audio quality by integrating advanced deep learning models for video super-resolution and audio noise suppression

Methods

1.Video Super-Resolution (VSR) :
We used the RVRT model to enhance low-resolution videos by restoring frame consistency and detail. The video is processed in tiles to reduce memory usage and enable faster & higher quality performance.

2.Model Enhancement and Tiling Optimization :
We integrated additional residual blocks into the RVRT architecture to better preserve textures and fine details in the upscaled video. Our tiling-based optimization strategy helps in avoiding memory bottlenecks by processing video patches individually and then seamlessly merging them, resulting in high-quality output with minimal artifacts.

3.Audio Noise Suppression : For audio enhancement, the system utilizes a fine-tuned Demucs model, a U-Net-based deep learning architecture well-suited for denoising and source separation.

Results

Architecture	Avg. PSNR (dB)	SSIM	PSNR_Y (dB)	SSIM_Y
Our Model	31.76	0.8881	33.14	0.9006
Baseline	29.91	0.8398	31.26	0.8562

Our system demonstrated notable performance gains across both video and audio enhancement tasks:

1. Video Enhancement :

- Enhanced model architecture by adding a layer and achieved a 1.85 dB PSNR improvement and a 0.0483 SSIM boost on the REDS benchmark by enhancing .
- Reduced GPU memory usage by 50% through a custom tile-based optimization approach.

2. Audio Enhancement :

- Delivered a PESQ score of 3.15 on the Valentini-noise dataset and 2.91 on VoiceBank+DEMAND.
- Maintained high speech intelligibility and naturalness while removing diverse noise types.

3. System Integration :

- Unified video and audio enhancement into a seamless pipeline.
- Real-time capable and accessible via an intuitive web-based interface.

Conclusions

This project presents an AI-based system for enhancing video and audio quality through deep learning. By integrating advanced models like RVRT for video super-resolution and Demucs for audio denoising, A web-based interface was developed to make the tool accessible for non-technical users

Bibliography

1.Wang et al., EvTexture: Event-driven Texture Enhancement for Video Super-Resolution, IEEE TPAMI, 2024.
2.Liang et al., *VRT: A Video Restoration Transformer*, CVPR, 2022.
3.Wang et al., *Real-ESRGAN: Blind Super-Resolution with Synthetic Data*, ICCV, 2021.
4. Defossez et al., *Demucs: Music Source Separation in the Waveform Domain*, arXiv, 2019.
5. Valentini-Botinhao et al., RNN-based Speech Enhancement for TTS, SSW, 2016.
6. Chen et al., Image Super-Resolution Transformer, CVPR, 2023.