Dynamic Zoom – Real-time, Focused Super-Resolution for Videos

1 Team Members

Name	Email	Campus ID
Shantanu Vichare	vichare2@wisc.edu	9085918648
Nico Ranabhat	nranabhat@wisc.edu	9080936215
Dario Placencio	placenciohid@wisc.edu	9072846018
Hemanth Sridhar Nakshatri	nakshatri@wisc.edu	9085807346

2 Introduction

The quest for high-quality video content in today's digital age often confronts the challenge of image degradation when zooming into specific areas of interest within video feeds. This project aims to introduce an innovative solution: a real-time super-resolution method focused exclusively on enhancing the resolution of cropped regions of interest, drawing inspiration from advancements in dynamic-range enhancement and super-resolution algorithms optimized for real-time processing [1]. Unlike traditional approaches that process entire video frames, our method targets dynamic regions of interest, resulting in significantly reduced inference time and memory consumption. Inspired by fast-inference models like Swift-SRGAN, this project aims to have a significant impact on digital zoom by maintaining high image quality without the usual performance penalties.

3 Problem Statement

Traditional super-resolution techniques demand substantial computational resources when applied across entire video frames, limiting their utility in real-time applications. This limitation is particularly acute in scenarios where enhancing specific areas of interest could suffice. The need for a super-resolution method that combines efficiency with high-quality output, especially for dynamically selected cropped regions, is evident. This project seeks to address this problem, offering users the ability to focus into video sections of interest by improving the spatial resolution.

4 Motivation

Our project addresses these limitations by offering a real-time solution for high-quality zoomed-in views in videos. This opens up various exciting applications:

- **Enhanced Sports Analysis**: Referees and viewers can zoom in on crucial plays with exceptional clarity.
- **Sharpened Security Monitoring**: Users can focus on specific areas of concern in surveillance footage, improving security and safety.
- **Streamlined Content Creation**: Video editors can achieve high-quality close-up effects without expensive equipment, facilitating creative freedom.
- **Improved Accessibility**: Users with impaired vision can benefit from clearer zoomed-in views, promoting inclusivity.
- **Immersive User Experiences**: Combined with features like eye-tracking and object-tracking, this can significantly enhance AR/VR and video streaming experiences for users.

This innovation is supported by the potential demonstrated in systems like STRUCT++ for real-time video spatial resolution upconversion, which offers functions such as batch processing and local region zooming [2].

5 Methodology

Focused on real-time super-resolution, our project emphasizes:

- **Efficient Model Selection**: Adopting lightweight, fast inference deep learning models like Swift-SRGAN as a foundation.
- **Adaptive Region-of-Interest (ROI)**: Incorporating pre-processing steps like tweaked model encodings and dynamic scaling to accommodate for varying window sizes of ROIs in real-time.
- **Targeted Processing**: Modifying the network architecture or incorporating attention mechanisms to focus computational resources on the cropped ROI, inspired by adaptive encoding based on user access patterns for zoomable video streams [3]. This would significantly reduce processing time and memory footprint compared to processing the entire frame.
- **Optimized Training**: Training the model on diverse video datasets, prioritizing fast inference speeds while maintaining exceptional image quality within the ROI.
- **Performance Evaluation**: Rigorously evaluating the model's effectiveness using standard metrics like PSNR and SSIM, incorporating frameworks like FAST for accelerated super-resolution processing on compressed videos [4].

6 Timeline and Resource Requirements

Dates	Objective	Resources
Feb 16 - Mar 1	Conduct literature review, finalize project scope, and begin initial model development.	Hardware: High-performance computing resources for model training and testing.
Mar 2 - Mar 15	Continue model development, start dataset preparation, and start working on model pipeline.	Software: Deep learning frameworks (e.g., TensorFlow, PyTorch), web development tools.
Mar 16 - Mar 29	Complete model pipeline, start model testing and identify optimization scopes, prepare midterm report..	Datasets: Access to high-quality video datasets for training and evaluation.
Mar 30 - Apr 12	Finalize optimizations, finalize testing protocols, prepare presentation, and initiate web page design and content creation.	Cloud services if necessary for computing, like AWS or G-Cloud.
Apr 13 - Apr 27	Document execution and testing processes, complete web page development and launch, finalize project documentation.	

7 Conclusion

This project aims to set a new benchmark for performance and quality in video processing technologies by introducing an efficient, real-time super-resolution method for cropped regions of interest. By focusing our efforts on optimizing digital zoom capabilities, we anticipate making a significant contribution for performance and quality in video processing technologies. Through this innovative approach, we aim to advance the field of video enhancement and potentially address the pressing demands of modern digital media consumption.

References

[1] H. Okuhata, R. Imai, M. Ise, R. Y. Omaki, H. Nakamura, S. Hara, and I. Shirakawa, "Implementation of dynamic-range enhancement and super-resolution algorithms for medical image processing," in *2014 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2014, pp. 181–184.

[2] W. Yang, S. Deng, Y. Hu, J. Xing, and J. Liu, "Real-time deep video spatial resolution upconversion system (struct++ demo)," in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, pp. 1255–1256.

[3] K. Q. M. Ngo, R. Guntur, and W. T. Ooi, "Adaptive encoding of zoomable video streams based on user access pattern," in *Proceedings of the second annual ACM conference on Multimedia systems*, 2011, pp. 211–222.

[4] Z. Zhang and V. Sze, "Fast: A framework to accelerate super-resolution processing on compressed videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 19–28.