Super-resolution applied to Chrono Cross videos

Anonymous CVPR submission

Paper ID *****

Abstract

Chrono Cross is a video game originally released in 1999 that recently received a new version in modern game console systems. As the game had poor evaluations of its remastered visuals, this work proposes to evaluate the application of common super-resolution solutions to its videos. Our methods include applying bilinear and bicubic filters as well as the learning-based ESPCN and evaluating the changes through the PSNR and SSIM metrics. To facilitate test reproducibility, we share our code in a public repository on GitHub: `https://github.com/misaelrezende/Computer-Vision-Project`.

1. Introduction

Super-resolution is a type of image processing technique used to improve the resolution of images and videos in computer vision. It aims to generate a high resolution image (HR) from a low resolution image (LR). It has several applications in surveillance, aerial images, medicine, etc. While image super-resolution is focused in estimating HR images from LR, video super-resolution, on the other hand, aims to obtain a high-resolution video from low-resolution frames. Additionally, video super-resolution can benefit from using information from neighboring frames.

In this work, we aim to test some super-resolution methods to improve Chrono Cross videos. Chrono Cross is a video game that was originally released in 1999 by Japanese developer SquareSoft. It was recently re-released in a remastered version for modern game systems but various users complained the work done to the game was below expectations. As some artifacts were still left in the image, there were even comments that a barebones AI upscaler would do the job better. Our purpose is to super-resolve videos from the original version and try to see if we can get better results than the remaster with our method.

In section 2 we present the works in literature that addresses this subject. In section 3 we present our method applied to achieve our target. Then, in section 4 we show our results... To finish, in section 5 close our conclusion and brings to light some problems that could be further improved in future works.

2. Related Works

Super-resolution is an ill-posed problem with approaches varying from interpolation methods like bilinear and bicubic, to newly more widespread learning-based methods. At the beginning, methods concentrated on frequency domain [7]. According to [9] These "algorithms can make use of the relationship between the HR image and the LR observations based on a simple theoretical basis, and have high computational efficiency". But these methods had limitations, such as the sensibility to model errors and the difficulty to deal with complicated motion, which slowed down research in the area. To fix these issues, algorithms concentrated on the spatial domain were proposed and this approach became the main trend [3].

Recently, learning-based algorithms are more prominent due to a few factors, such as fast computation, optimal speed and greater results. These methods usually utilize machine learning algorithms to analyze statistical relationships between the LR and its corresponding HR counterpart from substantial training examples. Neighbor embedding methods [1] proposed by Chang et al. took advantage of similar local geometry between LR and HR to restore HR image patches. Lately, random forest [5] has also been used to achieve improvement in the reconstruction performance.

Very recently, deep learning based methods are the usual choice of the research community when tackling the issue of super-resolution. These algorithms rely on data-driven approaches focused on reconstructing the required details for accurate super-resolution.

Unlike image super-resolution, video super-resolution can benefit from using information from neighboring frames. FRVSR [4] employed an end-to-end framework that makes use of previously inferred HR to super-resolve further frames. The authors used the temporal information between frames to improve the quality of inferred frames and also reduce computational cost.

3. Method

Our intention is to use the original video as an input to a super-resolution method, get the output and compare it with the remastered video. We used the color images to be compared, i.e., we used the three RGB channels in our experiments, therefore, we report the results based on the three channels. To compare the results, we used two metrics, peak signal-noise ratio (PSNR) and structural similarity index (SSIM).

PSNR is a popular and well-documented metric in literature. It is used to measure reconstruction quality in image compression, and does so by rating the differences between the corresponding pixels. On the other hand, SSIM evaluates the image reconstruction quality based on the human visual system and does so by ”measuring the structural similarity between images, based on independent comparisons in terms of luminance, contrast, and structures” [8]. In PSNR, usually the greater the value, the better the reconstruction. Good values are found above 30. In SSIM, which ranges from 0 to 1, the closer to 1, more the images are similar (equal).

4. Experiments

The first experiments aimed to compare a single low resolution frame - which were downsampled by a scale factor from the original frame and then upsampled using a specific method we wanted to evaluate - with the original frame. We used interpolation methods like bilinear and bicubic, as well as learning-based methods like ESPCN [6].

In figure 1, we show the comparison of the original image and our super-resolved version using the bilinear method in a 2x scale factor. The difference between the versions isn't so noticeable by the naked eye, though it's possible to see some blurring when magnifying it. The same happened to the bicubic filter and both versions on a 4x scale factor.

In table 1 we show the metrics that evaluate the filters used in the first experiments. Bilinear and bicubic versions didn't differ at all in both PSNR and SSIM, so we conclude they can be considered of similar quality.

In table 2 and figure 2, we compare the bilinear and ESPCN filter applications using PSNR and SSIM. The results are averaged from all frames, and demonstrated that ESPCN can achieve a slight improvement over the bilinear interpolation.

As surveys in the literature have discussed [2], [8], the use of PSNR and SSIM in evaluating the quality of the reconstruction are susceptible to failures. Their values do not capture the human visual perception accurately. Some work still need to be done to create a better image quality assessment.

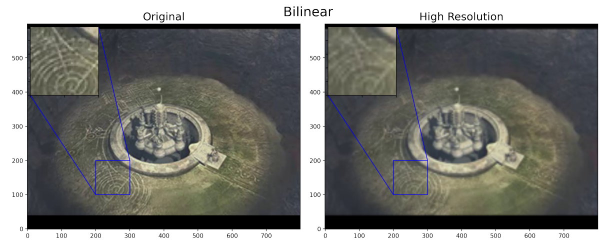

Figure 1. Bilinear vs Bicubic. 2x scale factor.

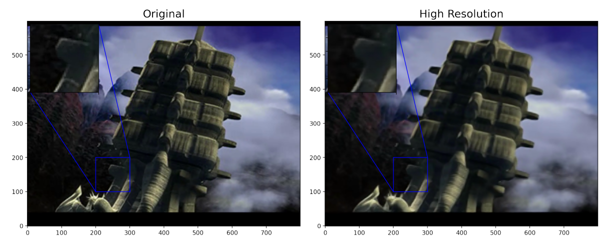

Figure 2. Bilinear vs ESPCN. 4x scale factor.

Scale factor	Bilinear	Bicubic	
2x	40.12	40.12	PSNR
	0.97	0.97	SSIM
4x	33.39	33.39	PSNR
	0.89	0.89	SSIM

Table 1. Bilinear vs Bicubic. Comparison between PSNR and SSIM metrics.

Scale factor	Bilinear	ESPCN	
2x	36.80	39.15	PSNR
	0.97	0.97	SSIM
4x	29.68	31.15	PSNR
	0.88	0.90	SSIM

Table 2. Bilinear vs ESPCN. Comparison between PSNR and SSIM metrics.

5. Conclusion

In this work, we showed some tests using video super-resolution to Chrono Cross videos.

One drawback of our approach is that we used learning-based models that were not trained in our videos, so we expected these models to be generalized enough to demonstrate good results in other distributions. Another disadvantage of our approach to evaluate the super-resolution experiments is that we downsampled the original video frames, upsampled them and tried to compare them with the remastered version. As a result, we could not see improvements to the image besides some pixelated aspects of the image being smoothed, which wasn't our original intention. When looking for some kind of objective quality metric for comparison in our experiments, we came across PSNR and SSIM. While both of them can be helpful, they indicate when images are changed or similar but PSNR's numerical magnitudes can be a little complicated to understand and the direct comparison with the remastered edition had extra complications, probably due to a misalignment of timestamps. In the end, we didn't find a way to make SR versions of the videos that would be adequate substitutes to the poorly evaluated remastered ones.

References

[1] Hong Chang, Dit-Yan Yeung, and Yimin Xiong. Super-resolution through neighbor embedding. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2004.*, volume 1, pages 275–282, 2004.

[2] Hongying Liu, Zhubo Ruan, Peng Zhao, Chao Dong, Fanhua Shang, Yuanyuan Liu, Linlin Yang, and Radu Timofte. Video super-resolution based on deep learning: a comprehensive survey. *Artificial Intelligence Review*, pages 1–55, 2022.

[3] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36, 2003.

[4] Mehdi S. M. Sajjadi, Raviteja Vemulapalli, and Matthew Brown. Frame-Recurrent Video Super-Resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[5] Samuel Schulter, Christian Leistner, and Horst Bischof. Fast and accurate image upscaling with super-resolution forests. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3791–3799, 2015.

[6] Huszr F Totz J Aitken AP Bishop R Rueckert D Wang Z Shi W, Caballero J. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[7] R Tsai. Multiframe image restoration and registration. *Advance Computer Visual and Image Processing*, 1:317–339, 1984.

[8] Zhihao Wang, Jian Chen, and Steven CH Hoi. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3365–3387, 2020.

[9] Linwei Yue, Huanfeng Shen, Jie Li, Qiangqiang Yuan, Hongyan Zhang, and Liangpei Zhang. Image super-resolution: The techniques, applications, and future. *Signal Processing*, 128:389–408, 2016.