Video Upscaling Using Deep Neural Networks

Team Members: Aditya Rathi | Ruchita Sinha | Jaydutt Kulkarni | Vatsal Joshi	TA Advisor: Yilang Liu

In this project we plan to investigate the upscaling of a lower resolution input video stream to one that is at a significantly higher resolution while maintaining visual fidelity using deep learning networks. The aim is to have the network contextually interpolate the video frame producing a larger output of a higher resolution. The key enhancements offered are superior image quality and lower storage requirements. For applications that involve rendering frames such as video games, the network will offer image quality comparable to native resolution while only having to render one-quarter to one-half of the pixels. Resulting in significantly higher performance without loss in visual fidelity. For applications that do not involve rendering such as network-based video streaming, this can result in decreased video usage as the footage can be transmitted at a lower resolution and then locally upscaled on the client device. Video files can also be stored in lower resolutions, saving space.

There have been several works published recently that use deep learning for the super-sampling of input images. [1] showcases comparisons of deep-learning-based networks with traditional techniques such as bicubic interpolation. [2] shows how humans perceive these super-resolution outputs and shows a clear preference of people towards CNN-based approaches. [3] shows a practical application of these techniques as used by the Nvidia DLSS network for super-resolution in video games. Finally [4] introduces a modern reference architecture that showcases very promising results on low-resolution input images. These readings showcase clearly that CNN-based approaches are significantly better than traditional upscaling approaches both quantitatively and qualitatively; these approaches are being used by real-world applications and also provide reference models to base our strategies.

There needs to be training data, where low-resolution images are provided to the deep learning model, while comparing the output of the model to ground truth, i.e., higher resolution versions of the same images. There also needs to be validation and test data sets for benchmarking of the neural network. We plan to generate video footage of our own and compare it to an existing dataset [5], with respect to the performance of the model achieved after training. We also plan to test our model to understand its generalisability in terms of domain-specific training sets and performance outside the domain.

There are some existing algorithms for upscaling video. For example, Super Resolution CNN (SRCNN) extracts feature maps which are then mapped nonlinearly to high-resolution patch representations. The features are extracted using convolution layers with different layer sizes. Multiple convolution layers are stacked in conjunction with a non-linear activation function to extract features at multiple scales from the original input image. The features are then mapped to create a high-resolution map of the features from the original image. The final layer averages the high-resolution map and combines it with the spatial neighborhood to generate the final image. We plan to change the layer sizes of the convolution layers and tweak the number of layers present in the network to achieve a good balance between visual fidelity and the performance of the network. We also plan to tweak the hyper-parameters of the model to see what affects the performance the most and find an optimum solution in the Pareto set.

We plan to compare the trained model with the results obtained using bicubic interpolation. The videos obtained from our deep learning architecture will be displayed alongside the video obtained from the original method. Furthermore, the results obtained from our model will be benchmarked using standard performance techniques such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). [6]

References:

[1] S. Huang and J. Xie, "Pearl: A Fast Deep Learning Driven Compression Framework for UHD Video Delivery," ICC 2021 - IEEE International Conference on Communications, 2021, pp. 1-6, doi: 10.1109/ICC42927.2021.9500754.

[2] Lundkvist, Fredrik. "Deep upscaling for video streaming: a case evaluation at SVT." (2021).

[3] Watson, Alexander. "Deep learning techniques for super-resolution in video games." arXiv preprint arXiv:2012.09810 (2020).

[4] Dong, Chao, et al. "Image super-resolution using deep convolutional networks." IEEE transactions on pattern analysis and machine intelligence 38.2 (2015): 295-307.

[5] D. Martin, C. Fowlkes, D. Tal and J. Malik, "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics", ICCV, July 2001.

[6] Ignatov, Andrey, et al. "Real-time video super-resolution on smartphones with deep learning, mobile ai 2021 challenge: Report." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021