Aman Chadha
amanc@stanford.edu

CS230: Deep Learning

Project Proposal

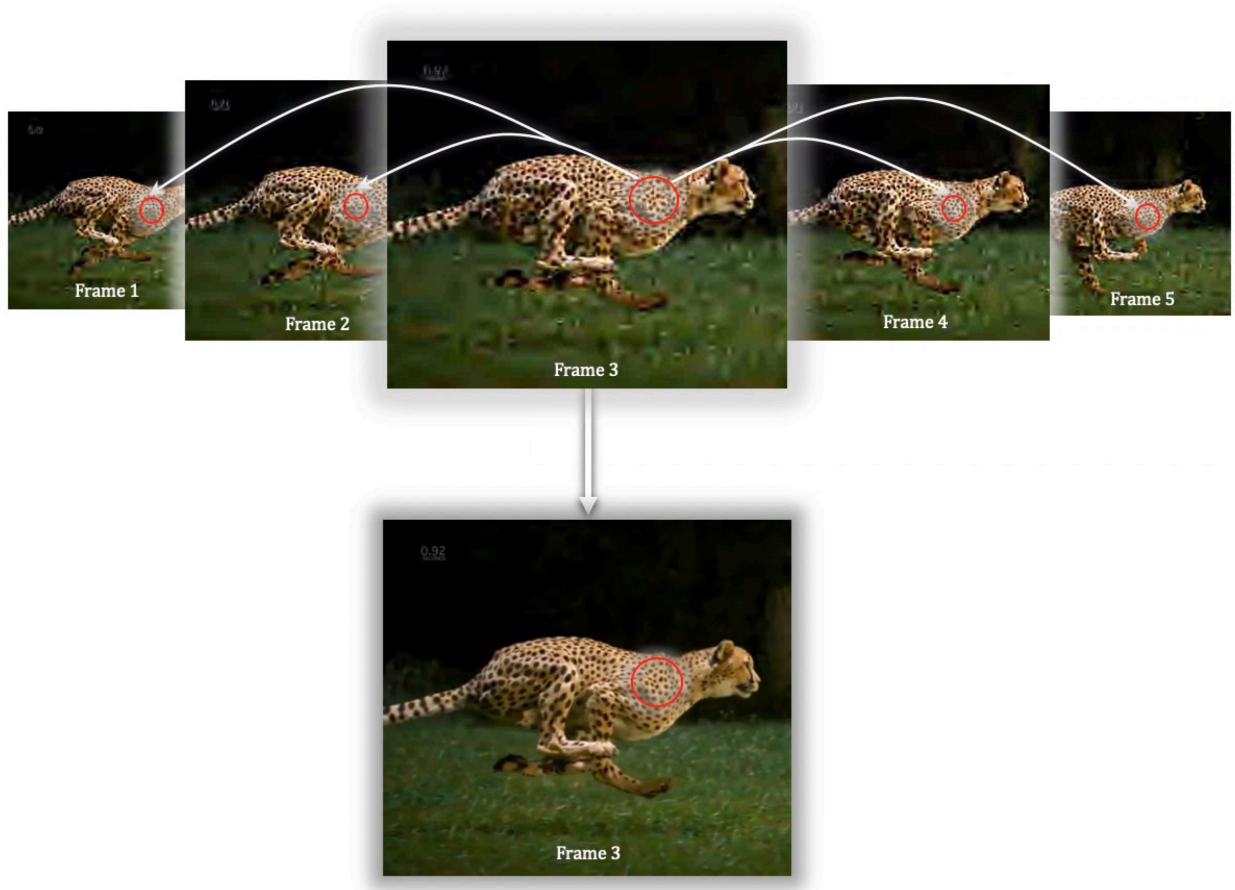
# iSeeBetter: A Novel Approach to Video Super-Resolution using GANs and Adjacent Frame Similarity

Generative Modeling and Computer Vision

## *Problem statement*

Traditional Video Super-Resolution (VSR) methods based on deep neural networks upscale based on a single degradation model (usually bicubic interpolation), followed by reconstruction. This is sub-optimal and adds computational complexity [1]. Further, the ability of mean square error, which these studies utilize to capture high texture details based on pixel-wise frame differences, is very limited causing the resulting video frames to be too smooth [2].

## *Proposed idea*



This project presents a novel implementation of VSR through Generative Adversarial Networks (GANs).

GANs have recently been used for solving the video super-resolution (VSR) problem. Results with GAN show better Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) compared to traditional methods. GANs can learn from large datasets and automatically add high-frequency details and features to frames while traditional methods can't. The generator discriminator architecture in GAN pushes it to generate more realistic and appealing frames and eliminates artifacts seen with traditional algorithms.

Our technique aims to train a GAN guided by an adaptive loss function. This loss function ***not only takes into account data from neighboring pixels but also pixel values at the same location from n forward/backward frames***. We use a distance function to figure out a similarity metric between the adjacent frames. If the similarity between *m* adjacent frames (where 0 <= m <= n) is beyond a certain threshold, we qualify the pixel values at the same location from these frames as usable, and use them for prediction.

## *Applications*

VSR is not only useful in improving the quality of videos obtained at lower resolutions, but also useful to save storage space by storing low resolution videos and upscaling them during playback.

## *Project challenges*

The challenges would mostly be associated with the novel aspects of the project, i.e., to figure out how our prediction algorithm would dynamically react to inputs of different sizes. Our prediction algorithm should be able to resiliently handle cases where different numbers of adjacent frames can be correlated with the current frame (for e.g., cases where all *n* forward/backward frames can be correlated, cases where only *n/2* forward/backward frames can be correlated, cases where none of the adjacent frames can be correlated etc.).

In addition, as an aggressive target, once we have achieved functional convergence, we would like to improve the performance of the algorithm by focusing on the "generation" runtime, i.e., by profiling the algorithm's runtime and optimizing it, we can make the algorithm run at real-time during playback, thereby letting us perform VSR on-the-fly while playing the original video. This would require us to be able to perform frame generation within a couple of milliseconds to subsume processing within the 16.67ms intervals between successive VSYNCs (for smooth 60FPS playback).

### *Dataset*

From [3], we have seen that GAN-based VSR models significantly benefit from training with a dataset with more diverse scenes and motions. We will thus use a dataset created from a subset of videos from the YouTube-8M dataset [3] to obtain a significant increase in perceptual quality.

We also plan to experiment with the commonly used Myanmar dataset, however, this dataset is known to have limited variation of both scene types and motions [3].

### *Results and evaluation*

- Peak Signal-to-Noise Ratio (PSNR) is a common measure used to evaluate super resolution algorithms.
- The structural similarity (SSIM) index is also a common method for predicting the perceived quality of digital television and cinematic pictures. We will also try applying the SSIM for each channel (R, G, B) of a color image, separately.
- Histogram is another method to find similarities among images.
- Chi-Square distance can also be used to compare two histograms.

We plan to use the above metrics not only to evaluate our algorithm, but also for identifying if *n* forward/backward images are similar and can be fed into the VSR prediction model.

We also plan to do a range scan (run a sweep of *n*) to identify the maximum value of *n* that yields the best results on the training set.

We expect our results to demonstrate that the learned model achieves improved results with sharper images, fewer artifacts and less noise compared to current VSR methods.

### *References*

[1] Shi, Wenzhe, et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[2] M. Cheng, N. Lin, K. Hwang, and J. Jeng, “Fast video super-resolution using artificial neural networks,” in Proc. 8th Int. Symp. Commun. Syst. Netw. Digital Signal Process. (CSNDSP), 2012, pp. 1–4.

[3] López-Tapia, Santiago, et al. "A Single Video Super-Resolution GAN for Multiple Downsampling Operators based on Pseudo-Inverse Image Formation Models." arXiv preprint arXiv:1907.01399 (2019).

[4] Gopan, K. Gopika and G A Sathish Kumar. “Video Super Resolution with Generative Adversarial Network.” 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI) (2018): 1489-1493.