

AI Powered Frame Interpolation

Narayanan AK (nk1581), Carlos Fernandez(csf5856), Sharjeel Ansari(sa2676) | GCCIS | Department of Computer Science.

Introduction

Motion Interpolation is a technique to generate intermediate frames with the purpose of increasing the number of frames per second or deal with artificial slowing techniques or visual artifacts. We will be exploring the idea of using convolutional neural networks to tackle this task.

Motivation

- Increasing the frame rate of media can improve the viewing experience.
- The higher the framerate, the smother the video will appear to the user.

Current Methods

- Linear Interpolation
- Motion Flow
- PhaseNet [3]
- Depth-Aware Video Frame Interpolation [4]

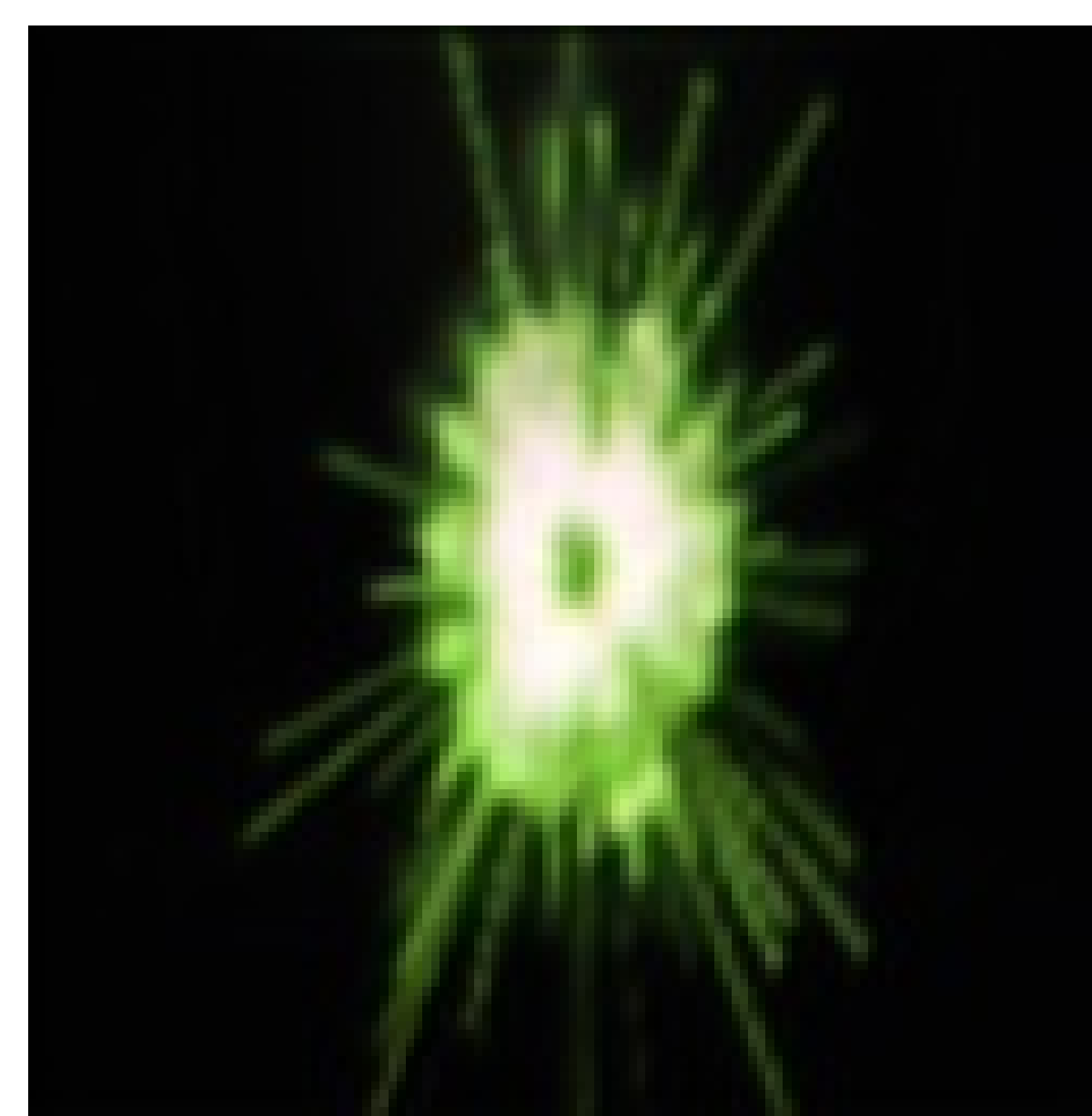
Our method

- First, we collected 60-fps videos from YouTube and decomposed them into individual frames.
- Then we arranged the frames into triplets and we called them f_i , f_j and f_k where f_i and f_k are the inputs and f_j is the frame we will be trying to predict.
- This data is then fed to a CNN for training.
- After training we used this model to generate an interpolation between the frames of a 30-fps video in order to convert it to a 60-fps video.

CNN Architecture

- Since we will be working with small images of size $128 \times 128 \times 3$, the input to the CNN is of size $128 \times 128 \times 6$. This is because we need to feed in both f_i and f_k at the same time.
- The architecture used as of now is two 2×2 convolutional layers, with one 2×2 max-pool layer followed by a 2×2 up-sampling layer and two 2×2 convolutional layers.
- The final output layer is a 1×1 convolutional layer with 3 filters, so we get an output of size $128 \times 128 \times 3$.
- This is a condensed architecture found in [1].

Preliminary Results



Predicted Image



Original Image



Predicted Image



Original Image

Result Discussion

- Currently, the model has few layers and is trained over a very low number of epochs.
- We can see that the predicted frame appears accurate however, the resulting image is blurred.
- This is due to the fact that our convolutional and up-sampling layers are really small.
- But even with these limitation the model was able to give a somewhat accurate prediction of the frame.
- This shows that this approach has potential.

Works Cited

- [1] Neil Joshi, Duncan Woodbury "Deep Motion: A Convolutional Neural Network for Frame Interpolation" 25 Jan 2017
- [2] "Motion Interpolation." Wikipedia, Wikimedia Foundation, 31 Jan. 2020, en.wikipedia.org/wiki/Motion_interpolation.
- [3] Simone Meyer, Abdelaziz Djelouah, Brian McWilliams, Alexander Sorkine-Hornung, Markus Gross, Christopher Schroers "PhaseNet for Video Frame Interpolation" 18 Jun 2018
- [4] Wenbo Bao, Wei-Sheng Lai, Chao Ma, Xiaoyun Zhang, Zhiyong Gao, Ming-Hsuan Yang "Depth-Aware Video Frame Interpolation" 1 Apr 2019

