Milestone

Marek Nevole

CTU FIT

nevolmar@fit.cvut.cz

December 2, 2021

1 Introduction

The task of this semester assignment was to research, implement and evaluate neural network models that are capable of video frame interpolation which results in frame rate upscaling. Link to repository: `https://gitlab.fit.cvut.cz/nevolmar/mvi-sp`.

2 Research

The problem of video frame interpolation is a classic problem in image and video processing. There are many methods that achieve the wanted results. Popular approaches are based either on finding optical flow for perfect motion description or using kernel based methods to keep structural alignment of the frames. However, in recent years, with the rise of deep learning, in popularity, many new approaches have been proposed. The most suited neural networks for this problem are convolutional neural networks and generative adversarial networks. Which I will use.

2.1 CNN

Convolutional neural networks are very popular choice when it comes to image and video processing. These networks are based on convolutional layers, which apply filters to input to create feature maps.

Fully convolutional networks are often used to interpolate frames. Fully convolutional network is a network that uses only convolution, pooling, upscaling and misc layers. One architecture that shows promising results is a U-Net[3]. This network is mostly based on two blocks, pooling and upscaling. Pooling block consists of 2 consecutive convolution layers followed by pooling layer. The upscaling block first upscales output from previous convolution layer and merges it with output from pooling block at the same height. This layer is followed by another 2 convolution layers. U-Net is then a chain made firstly from pooling blocks followed by chain of upscaling blocks.[3] Other, simpler, approaches just use convolution and pooling layers[2, 5].

2.2 GAN

GANs consist of 2 neural networks, generator and discriminator. Generator generates desired output usually from high dimensional random noise data and discriminator classifies whether the output is fake from generator or real from training data. These 2 networks are trained simultaneously.[1]

For video frame interpolation GAN is often made from 2 convolutional neural networks where generator is fully convolutional network and discriminator is again convolutional network used as binary classifier. These GANs are often called deep convolutional GANs (DCGAN).[4, 5]

3 Data

2 hours of footage were downloaded from the official F1 YouTube channel[1]for training and testing purposes. Links to these videos were scrapped using the *Selenium*[2]library, and the videos themselves were downloaded via the *pytube*[3]library. The videos are in 360p resolution with a frequency of 25 fps. Thus, approximately 180,000 images can be extracted from the 2 hours of video. From which the input data is created, so that out of 3 consecutive frames, frames 1 and 3 are used as input to predict frame 2. Thus $x_i = (frame_1, frame_3)$, $y_i = frame_2$. All individual frames were downscaled to 144p resolution using Lanczos method for faster training.

4 Future work

In the upcoming weeks the U-NET and GAN models will be finely tuned and trained to the best of my ability to give best results. The output of this work should be a python script that takes video at the input and, using a trained neural network that will give the best results, returns a video with a doubled frame rate.

[3] Formula 1 YT channel

[3] selenium.dev

[3] pytube.io

References

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[2] Neil Joshi and Duncan Woodbury. Deep motion: A convolutional neural network for frame interpolation. `https://github.com/neil454/deep-motion/blob/master/deep-motion_paper.pdf`, 2017.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[4] Quang Tran. Efficient video frame interpolation using generative adversarial networks. *Applied Sciences*, 10, 09 2020.

[5] T. Xiao. Frame rate upscaling with deep neural networks. 2016.