Background

Video frame interpolation has a variety of applications, from allowing for smooth slow-motion playback, to compressing video files by only storing a subset of frames and interpolating the remaining frames. However, a basic per-pixel interpolation is poor, since to properly interpolate, one must intelligently infer the motion of objects in the scene.

Data and Features

To test our implementation, we used videos taken on a smartphone of various scenes, which were rescaled to 384x384px square videos. This video was processed to create the input/output pairs.

We preprocess the video by computing a rolling median over the frames used as input to the process. This is then subtracted from the inputs to remove the background.

We use an approximate L1 loss over the RGB channel errors:

\[ L(y, \hat{y}) = \sum_c \sqrt{(y_c - \hat{y}_c)^2 + \epsilon^2} \quad c = \{r, g, b\} \]

Problem Statement

We pose the interpolation task as a regression task, where the input features are frames immediately before and after the frame to interpolate, and the output is the reconstructed frame.

Neural Network Architecture

We use a convolutional neural network, as they can be trained efficiently for image processing tasks. As shown in the figure above, the outputs of each convolutional layer are linked directly to the corresponding layer in the deconvolution process to aid in detail reconstruction.

Results

Output of process for two examples, one from the training video, and one from a similar test video. We see that the model outperforms simple mean interpolation when there is motion in the frames, and does well on the training data, making it suitable for compression applications. The poor generalization could be addressed by varying network architecture.

Future Work

We suspect that with a better loss function, or different feature representation (e.g. HSV rather than RGB), the noise in the process output can be reduced. Experimentation with different network architectures could improve generalization.

References

Haitam Ben Yahia, Frame Interpolation using Convolutional Neural Networks on 2D Animation, 2016.