Deblurring for Video Frame prediction

Team 2

May 4, 2023

Introduction

- Problem of video frame prediction has become highly intriguing due to its importance in computer vision applications like autonomous vehicles.
- Various deep learning techniques exist which extract spatio-temporal correlations from video data in a self-supervised fashion.
- We explore image deblurring techniques for the predicted frames to improve the performance of existing models.

Problem Statement

- Next frame prediction models predict the immediate next frame with a very good accuracy.
- Predicting a series of frames with the given initial position and movement returns blurry frames.
- Came up with an approach to use denoising and/or motion-deblurring on the predicted frames to fix this issue.

Problem Statement

(a)

(b)

Figure: Demonstration of the cascading effect of the next frame prediction models

ConvLSTM

- Input video is segmented into frames.
- Each frame sent to ConvLSTM, which extracts the features and stores it in the LSTM memory cells. This is used to make the prediction.
- LSTM Encoder extracts a fixed-length motion vector and a predicted frame with the help of previous reconstructed frame.
- LSTM Decoder takes the residual frame and motion vectors and a frame is predicted using these vectors and the previous reconstructed frame.
- The final output is the sum of residual frames and predicted frame.

ConvLSTM model

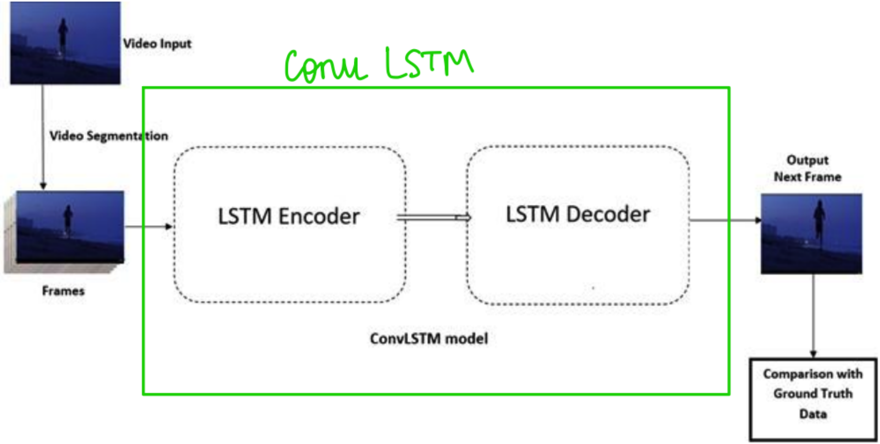


Figure: Architecture of the ConvLSTM model

ConvLSTM model

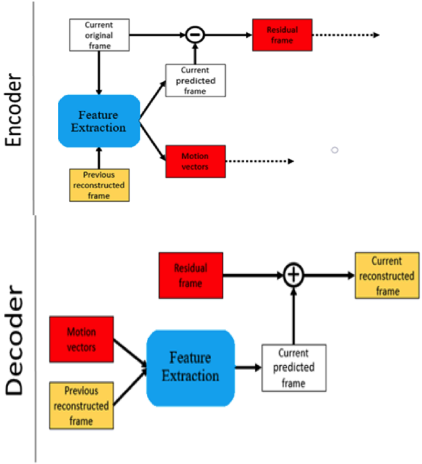


Figure: Architecture of the ConvLSTM model

Restormer

- Efficient Transformer model which handles high-resolution images for restoration tasks.
- Firstly, applies a convolution to obtain low-level feature embeddings.
- The features pass through a 4-level symmetric encoder-decoder layers and transformed into deep features.
- Encoder hierarchically reduces spatial size and increases channel capacity.
- Decoder takes low-resolution latent features and recovers the high-resolution representations.
- Few more convolutions are applied to obtain a refined image.

Restormer model

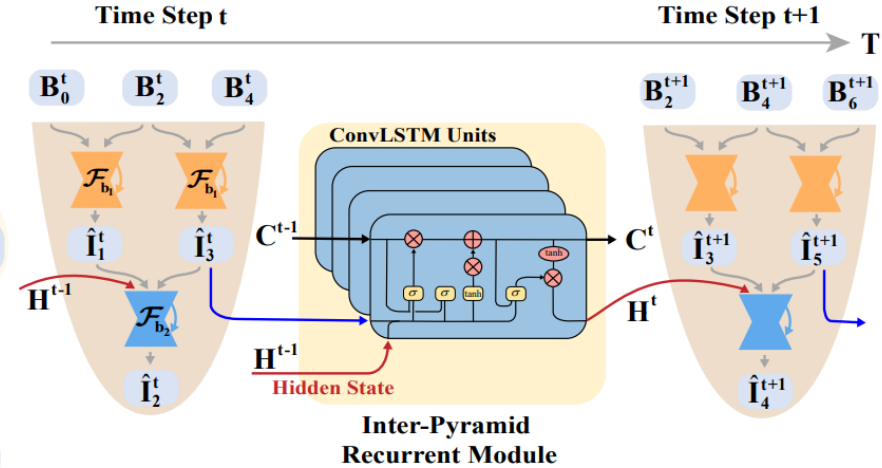


Figure: Architecture of Restormer model

ConvLSTM Training Procedure

- Trained on the entire dataset which gave slightly poor results because of mixing of data.
- Segregated the dataset into two based on where the videos were shot.
- Used the weights obtained from training on the entire dataset and performed transfer learning.

ConvLSTM Training Procedure

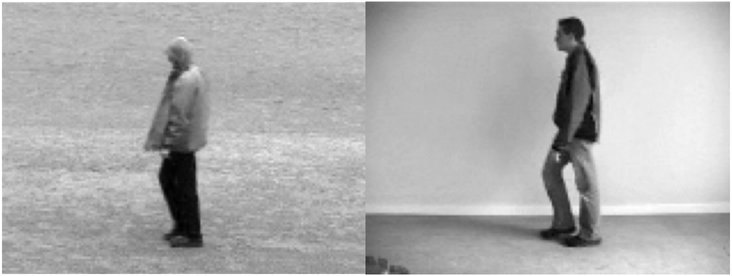


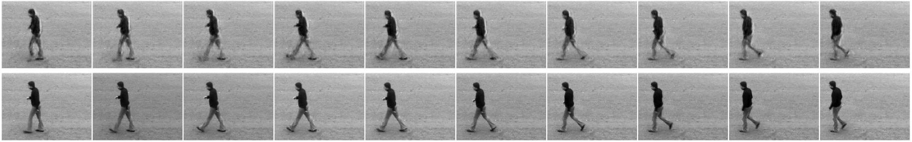
Figure: Two types of videos present in the dataset

ConvLSTM Training Procedure

Figure: Side effect of training on the entire dataset.

ConvLSTM Training Procedure

(a)



(b)

Figure: Next frame prediction of ConvLSTM

Motion Deblurring Training Procedure

- We noticed that the frames had undergone blurring in the subject's leg.
- Can approximately say that the legs have blurred due to motion.
- To train the model, we generated multiple pairs of next frame predicted images for different time instances of a video.
- Fed the predicted and ground truth frames to a pre-trained Restormer model.

Denoising Training Procedure

- After applying motion-deblur, it was observed that the noise in the background was amplified.
- Hence denoised the output image.
- Used pre-trained Restormer model for denoising.
- Training procedure similar to motion-deblurring procedure.

Overall pipeline of the approach

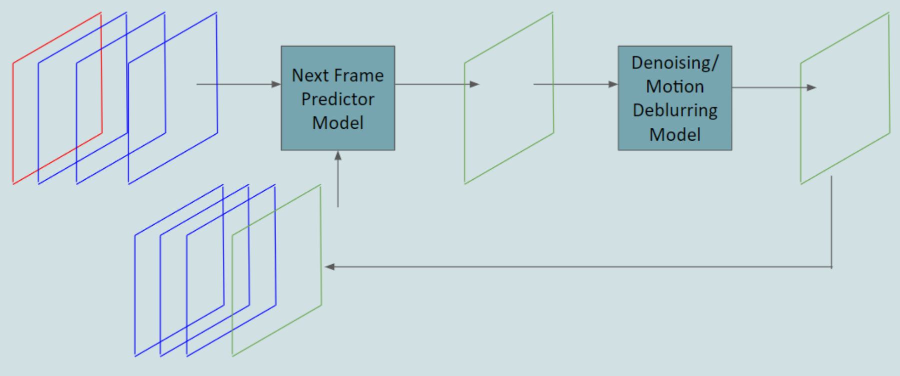


Figure: Overall pipeline

Final Results

Figure: Final predictions. First row images are the ground truths. Second row is using next frame predictions. Third row is using deblurring.

Final Results

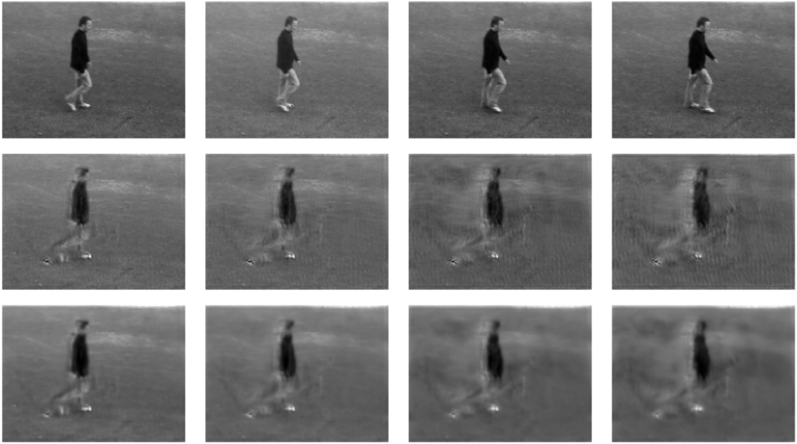


Figure: Final predictions. First row images are the ground truths. Second row is using next frame predictions. Third row is using deblurring.

Future work

These are certain approaches which could be tried out if more computational power were present:

- Ensembling the denoising and the motion-deblurring models
- Trying out different training methodologies for the denoising and deblurring separately
- Train the models for more number of epochs