Toward Human Pose Prediction Using The Encoder-Decoder LSTM

Nima Fathi
Sharif University of Technology

Armin Saadat
Sharif University of Technology

Saeed Saadatnejad
Swiss Federal Institute of Technology Lausanne

Social Motion Forcasting:

- Process of observing human activities for a portion of time and predicting future poses for short-term or long-term.
- It has numerous applications in various fields, such as traffic control and autonomous driving.
- This task can be seen as fine-grained task while bounding box and trajectory prediction deal with more coarse-grained information.

Observed Poses Through Time

Future Poses Through Time

Pose Prediction Autoencoder:

- Proposed model is a sequence to sequence LSTM based autoencoder.
- It takes as inputs velocities and observed positions of past joints, from which the future positions will be computed.

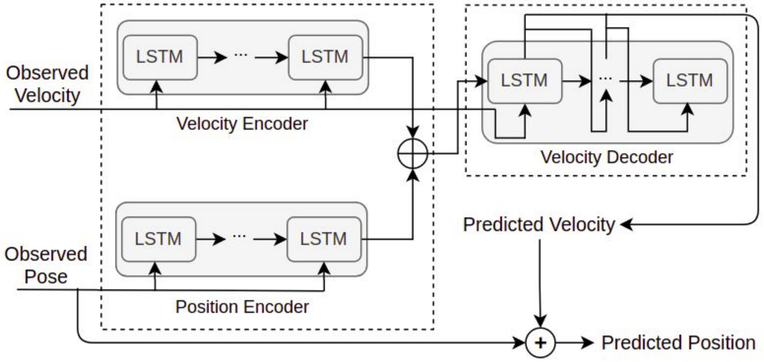

Fig 1: Model Architecture

- As Fig 1 shows, the model encodes the position and the velocity of each person into a hidden layer.
- The hidden layer will be used as the trigger for the decoder.
  - Using the encoded state, decoder exploits the velocity of the last observed pose and predicts the next future velocity.
  - Generated velocity will be used as the input for next LSTM cell and other future velocities will be generated likewise.
- To train this model, we leverage L1 loss between predicted and ground-truth velocities.
- This model works upon keypoints only, making it time-efficient and lightweight in comparison with other image-based models.

Results:

- We evaluated our model on VIM (Visibility-Ignored Metric) and VAM(Visibility-Aware Metric) which are defined in TRiPOD[1] paper.
- We leveraged TRiPOD, SC-MPF,[2] and Zero-Vel (simply replicate last observed pose for future poses) as our baselines.
- The quantitative results show a significant improvement in comparison with available baselines.

Method	80 ms	160 ms	320 ms	400 ms	560 ms
SC-MPF	22.0/78.3	37.9/99.8	64.6/124.3	75.8/138.5	93.5/147.9
TRiPOD	15.2/30.0	26.7/49.6	48.1/80.3	58.6/93.3	71.1/110.3
Zero-Vel	13.1/26.5	24.0/45.1	43.3/72.9	52.1/83.8	65.6/97.3
Ours	**9.2/20.8**	**17.6/36.3**	**36.2/62.9**	**44.8/74.7**	**59.3/91.5**

Fig 2: Results On PoseTrack

Method	100 ms	240 ms	500 ms	640 ms	900 ms
SC-MPF	46.28	73.88	130.23	160.83	208.44
TRiPOD	30.26	51.84	**85.08**	**104.78**	146.33
Zero-Vel	29.35	53.56	94.52	112.68	**143.10**
Ours	**25.89**	**47.57**	86.39	106.65	148.28

Fig 3: Results On 3DPW

Conclusion:

We have presented a recurrent autoencoder method for pose prediction of 2D and 3D data. Our method achieves better performance on mentioned metrics compare to aforementioned baselines. We believe for better performance, interactions should be considered and that could be further investigated in future tasks.

Code Repository:

You can find our source code in these ways:

- GitHub Repository URL: https://github.com/Armin-Saadat/pose-prediction-autoencoder
- GitHub Repository QR Code: