Implementing and Extending Inception-inspired LSTM for Next-frame Video Prediction

Krishna Palle

pallekc@bu.edu

1. Task

Understanding and Implementing a paper, Inception-inspired LSTM for Next-frame Video Prediction[1] and testing it on KITTY[2] dataset and extending the paper's implementation to multiple frames and try and video generation and if possible adopt the architecture from another paper, TwoStreamVAN: Improving Motion Modelling in Video Generation[3] to see if it could improve the performance in video prediction.

2. Related Work

The problem of video frame prediction has received much interest due to its relevance to in many computer vision applications such as autonomous vehicles or robotics. The state of the art in video prediction is with MIM(Memory in Memory)[4] model. There are other famous models like PredNN[5], FRNN[6] etc. which were the state of their art during their time which were tested on moving MNIST[7] dataset. The paper intending to implement[1] was not tested on the moving MNIST[7] dataset.

3. Approach

The plan is to implement the inception inspired LSTM network[1] for video prediction and test it against the moving MNIST dataset[7] to compete against the current state of the art(This would be the MVP of the project). Make architectural changes like number of layers, height and depth of layers, change in hyperparameters to see if we can improve the performance. If time permits, perform literature review to see if the architecture of TwoStreamVANs[3] can fit our use-case and implement it and check to see if it improves our results. The plan is to use the PyTorch Library for the development and implementation of model architectures and use Google Colab as the development interface.

4. Datasets and Metrics

For the above project, we plan to use KITTY dataset, the computer vision benchmark suite. It is a video dataset where each frame has a resolution of 1392x512 pixels. We also plan to use the moving MNIST dataset with size 7 82Mb] contains 10,000 sequences each of length 20 showing 2 digits moving in a 64 x 64 frame

http://www.cvlibs.net/datasets/kitti/

http://www.cs.toronto.edu/~nitish/unsupervised_video/

5. Approximate Timeline

Task	Predicted Deadline
Literature Review of the state of the art	03/17/2020
Implementation of inception inspired LSTM	03/27/2020
Extending the above to improve results	04/13/2020
Incorporating design of TwoStreamVANs to improve results	04/27/2020

References

[1] Hosseini, M., Maida, A., Hosseini, M. and Raju, G. (2019). *Inception-inspired LSTM for Next-frame Video Prediction*. [online] arXiv.org. Available at: https://arxiv.org/abs/1909.05622 [Accessed 1 Mar. 2020].

[2] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Re- search*, 32(11):1231–1237, 2013.

[3] Sun, X., Xu, H. and Saenko, K. (2018). *TwoStreamVAN: Improving Motion Modeling in Video Generation*. [online] arXiv.org. Available at: https://arxiv.org/abs/1812.01037 [Accessed 1 Mar. 2020].

[4] Wang, Y., Zhang, J., Zhu, H., Long, M., Wang, J. and Yu, P. (2018). *Memory In Memory: A Predictive Neural Network for Learning Higher-Order Non-Stationarity from Spatiotemporal Dynamics*. [online] arXiv.org. Available at: https://arxiv.org/abs/1811.07490 [Accessed 1 Mar. 2020].

[5] Wang, Y., Long, M., Wang, J., Gao, Z. and Yu, P. (2017). *PredRNN: Recurrent Neural Networks for Predictive Learning using Spatiotemporal LSTMs*. [online] Papers.nips.cc. Available at: http://papers.nips.cc/paper/6689-predrnn-recurrent-neural-networks-for-video-prediction-using-spatiotemporal-lstms [Accessed 1 Mar. 2020].

[6] Oliu, M., Selva, J. and Escalera, S. (2018). *Folded Recurrent Neural Networks for Future Video Prediction*. [online] Openaccess.thecvf.com. Available at: http://openaccess.thecvf.com/content_ECCV_2018/html/Marc_Oliu_Folded_Recurrent_Neural_ECCV_2018_paper.html [Accessed 1 Mar. 2020].

[7] Cs.toronto.edu. (n.d.). [online] Available at: http://www.cs.toronto.edu/~nitish/unsupervised_video/ [Accessed 1 Mar. 2020].