Modeling Deep Temporal Dependencies with Recurrent "Grammar Cells" (Supplementary Material)

Vincent Michalski
Goethe University Frankfurt, Germany
vmichals@rz.uni-frankfurt.de

Roland Memisevic
University of Montreal, Canada
roland.memisevic@umontreal.ca

Kishore Konda
Goethe University Frankfurt, Germany
konda.kishorereddy@gmail.com

1 Bouncing Balls Prediction

Figure 1 shows two bouncing ball sequences [1]. generated by the predictive gating pyramid (PGP) after seeing 5 frames from the training set[2]. Note that the training sequences were generated independently and are only 20 frames long, much shorter than the generated sequences shown here.

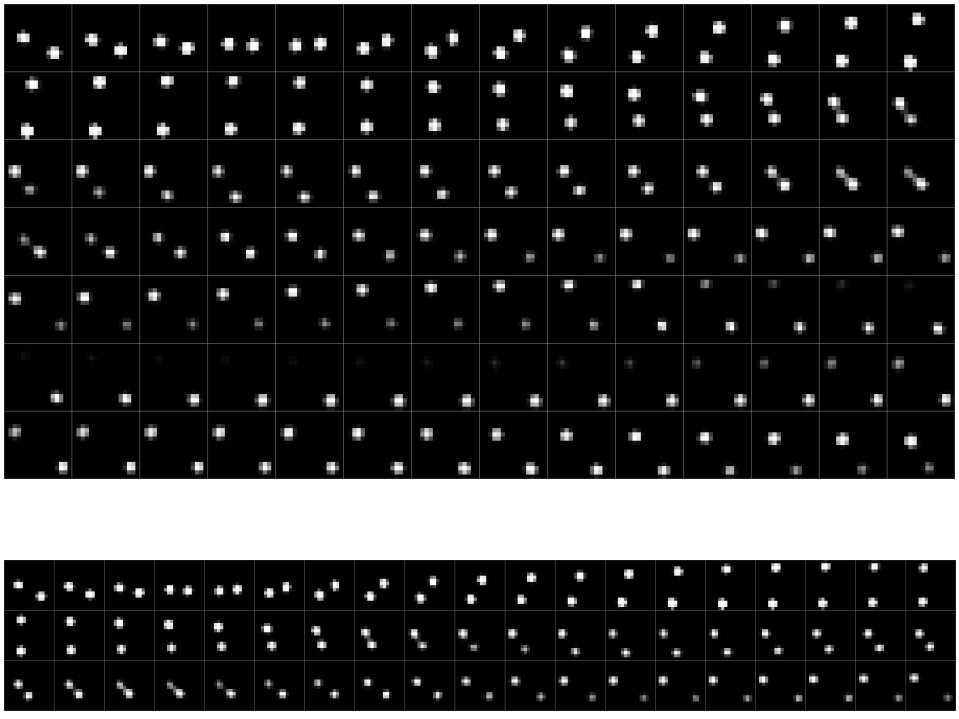

Figure 1: Long sequence prediction of bouncing balls videos.

[1] The training and test sequences were generated using the script released with [1].
[2] An animated version of these can be found in the supplementary zip archive.

2 Chirps Prediction

Figure 2 shows eight more test cases from the chirps data set, together with predictions (beyond the ground truth horizon of 160) of the PGP, the standard recurrent neural network (RNN) and the Conditional Restricted Boltzmann Machine (CRBM) [2]. We used the Theano [3] implementations of the RNN and CRBM written by Graham Taylor, available at `http://www.uoguelph.ca/~gwtaylor/code/`.

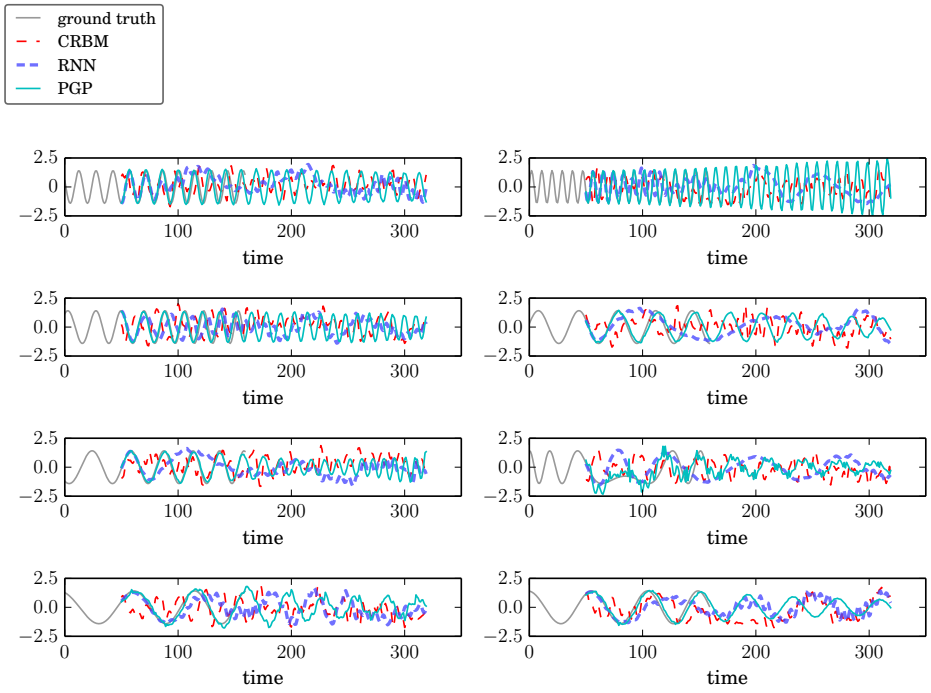

Figure 2: Some test predictions on chirp data.

3 NorbVideos

Figure 3 shows left and right receptive fields of the bottom layer which developed during predictive finetuning of the PGP. Due to the large input dimensionality (frame size 96×96) and the low number of training samples a few of the filters seem to be overfitting on the training data while many others are localized Gabor-like features.

4 Description of the Shift and Rotation Data Sets

4.1 Constant-Velocity Transformations

13×13 patches were cropped from the Berkeley Segmentation data set (BSDS300) [5]. Two data sets with videos featuring constant velocity shifts (CONSTSHIFT) and rotations (CONSTROT) were generated. Elements of the shift vectors for CONSTSHIFT were sampled uniformly from the interval $[-3, 3]$ (in pixels), and rotation angles from $(-\pi, \pi)$. Eight labels for CONSTSHIFT were assigned by partitioning shift angles into four quadrants and the shift magnitude into two bins. For CONSTROT rotation angles were divided into 8 equally-sized bins. Both data sets were partitioned into training, validation and testing set of size 100 000, 20 000 and 50 000, respectively.

4.2 Accelerated Transformations

Patches were again cropped from BSDS300 and artificially transformed with initial (angular) velocity and constant (angular) acceleration. Scalar angular accelerations were sampled uniformly from the interval $[-\frac{\pi}{12}, \frac{\pi}{12}]$ degrees. Initial angular velocities were sampled from the same interval. Angular accelerations were divided into 8 equally sized bins. For shifts, elements of the velocity and acceleration vectors were sampled from the interval $[-3, 3]$ (in pixels). Acceleration vectors were discretized in the same way as shift vectors in CONSTSHIFT.

References

[1] I. Sutskever, G. E. Hinton, and G. W. Taylor. The recurrent temporal restricted boltzmann machine. In *Advances in Neural Information Processing Systems 21*, pages 1601–1608, 2008.

[2] G. W. Taylor, G. E. Hinton, and S. T. Roweis. Modeling human motion using binary latent variables. In *Advances in Neural Information Processing Systems 20*, pages 1345–1352, 2007.

[3] James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: A cpu and gpu math compiler in python. In *Proceedings of 9th Python in Science Conference (SCIPY)*, 2010.

[4] R. Memisevic and G. Exarchakis. Learning invariant features by harnessing the aperture problem. In *Proceedings of the 30th International Conference on Machine Learning*, 2013.

[5] D. Martin, Fowlkes C., D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the Eigth IEEE International Conference on Computer Vision*, volume 2, pages 416–423, July 2001.

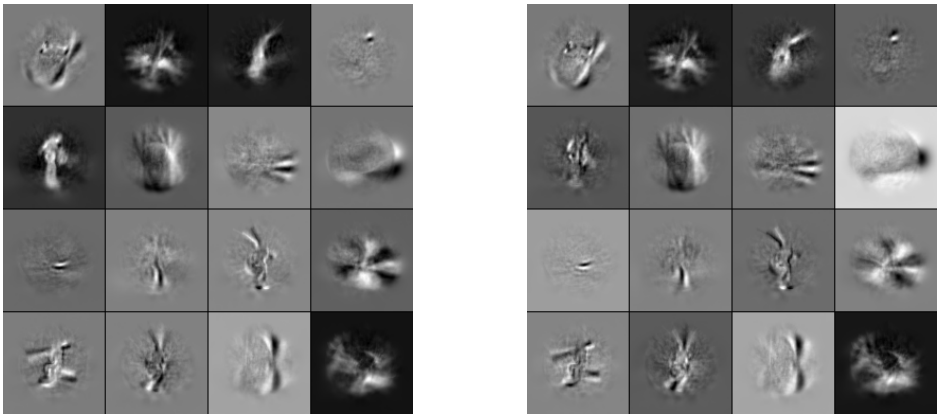

Figure 3: Receptive fields of the PGP trained on NORBvideos [4]