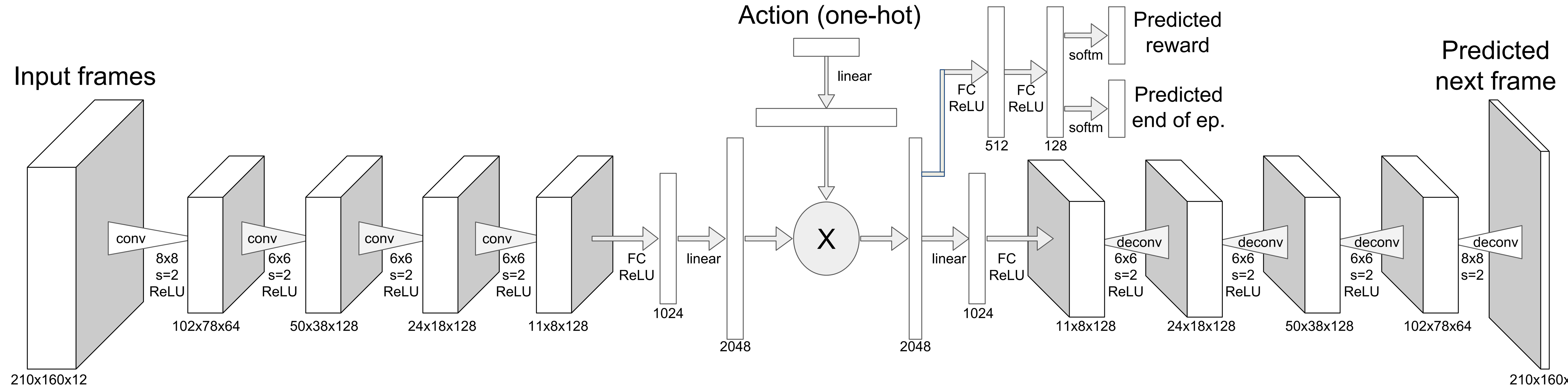
Deep Action Conditional Neural Network for Frame Prediction in Atari Games

Elias Wang, Atli Kosson, Tong Mu
Department of Electrical Engineering, Stanford University

Introduction and Problem Statement

In many applications of video prediction, future frames are not only dependent on previous frames, but external features as well. We study this problem of video prediction that is also conditioned on actions by looking at video frame prediction in the context of Atari 2600 video games where future frames are dependent on past frames as well as actions performed by the players. Previous work such as [1] and [2] have models for predicting future frames and rewards, but from our experiments, we have found different initializations changes the results greatly. To reduce this variance, we attempt to use an autoencoder.

Models



Similar to an autoencoder in structure. Has a predictive middle that uses multiplicative interactions to condition on an action. Based off models by [1] and [2].

Future Work

Curriculum training for faster and better results in model.

Balanced data better to improve prediction of rewards and end flags.

Test on games with more stochasticity such as pacman.

Use reward and frame prediction from model to improve training time and quality of DQN.

Data Generation and Post Processing

Data generated with a Deep Q Network (DQN) agent

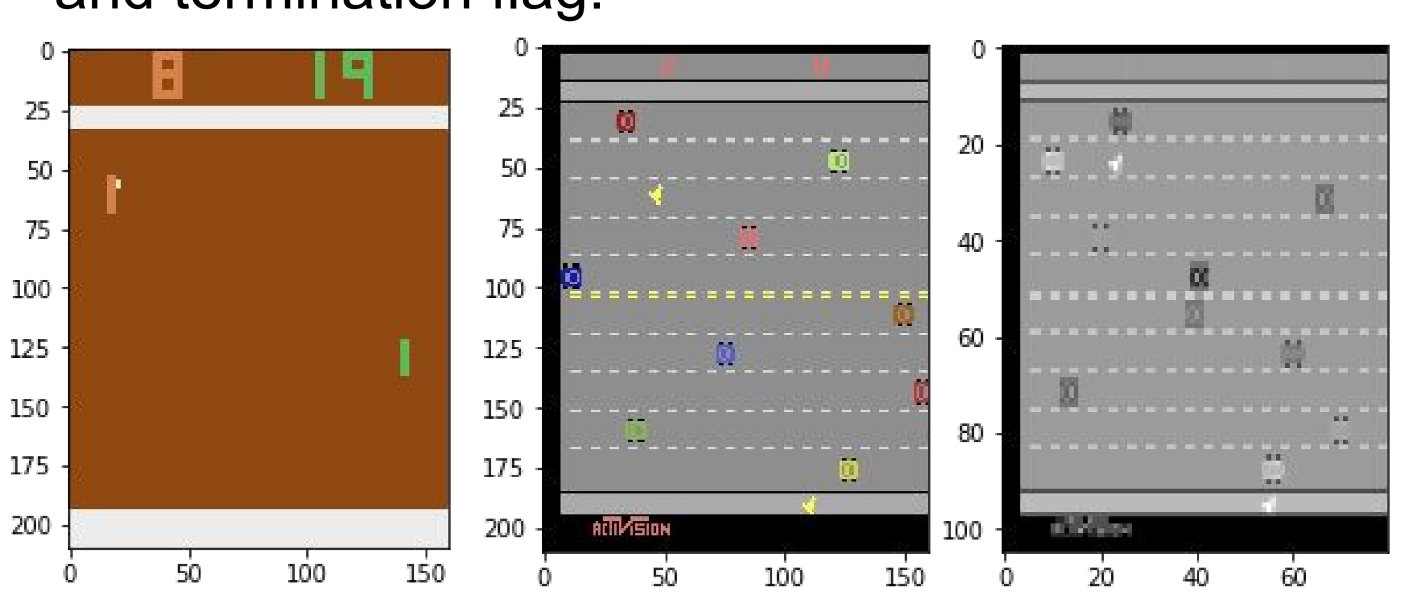
- Trained over 5 million steps
- Epsilon greedy exploration strategy
- Epsilon decays from 1 to 0.1 in first million steps

Collected full resolution frames with actions and rewards for the following Atari Games:

- Pong
- Freeway

Speed: Grayscale and Downsample

Computationally expensive to gather and difficulties in balancing dataset frames, rewards and termination flag.



Left: One frame of Pong, Middle: One frame of Freeway, Right: Freeway with grayscale and downsampled by 2

Initial Results and Conclusion

Evaluation metrics:

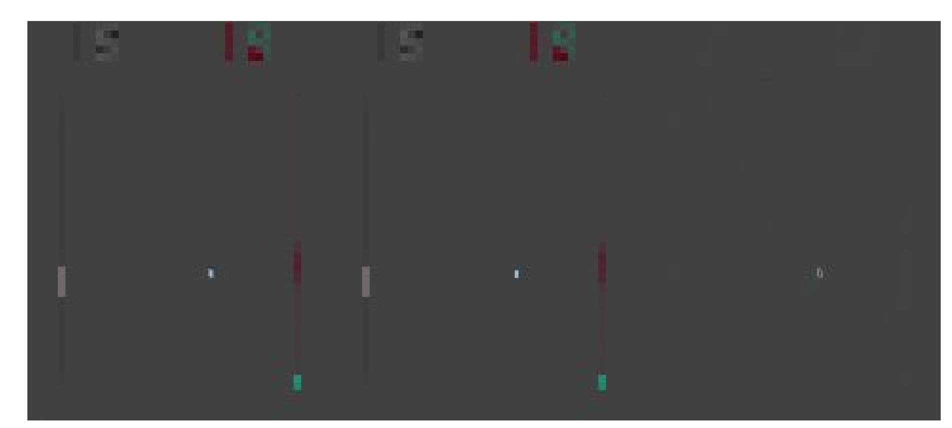
$$\mathrm{SSIM}(u^*,u^0)=\frac{(2\mu_{u^*}\mu_{u^0}+c_1)(2\sigma_{u^*u^0}+c_2)}{(\mu_{u^*}^2+\mu_{u^0}^2+c_1)(\sigma_{u^*}^2+\sigma_{u^0}^2+c_2)}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum\left(\hat{Y}_i-Y_i\right)^2}$$

We train our model on Pong along with a multi-layer perceptron model (MLP) as baseline.

Initial Conclusions:

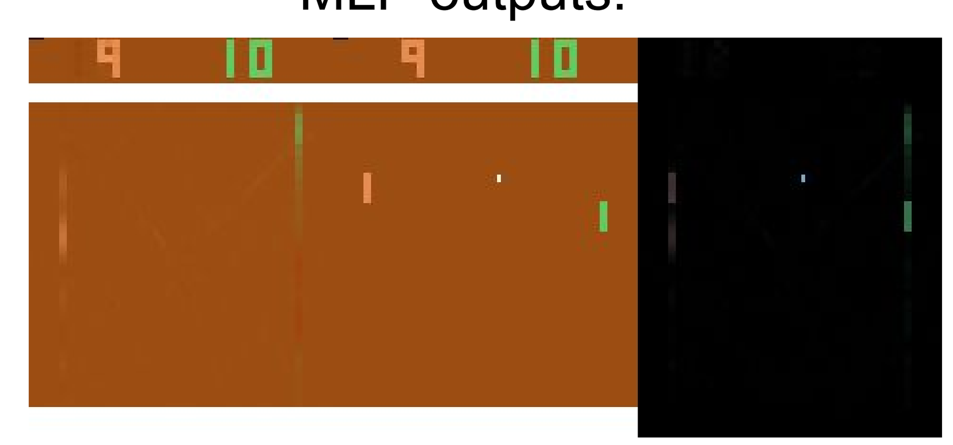
- The model produces better results qualitative results than the MLP
- The model produces useful predictions over multiple timesteps.
- The training process can likely be improved by incorporating curricular learning.

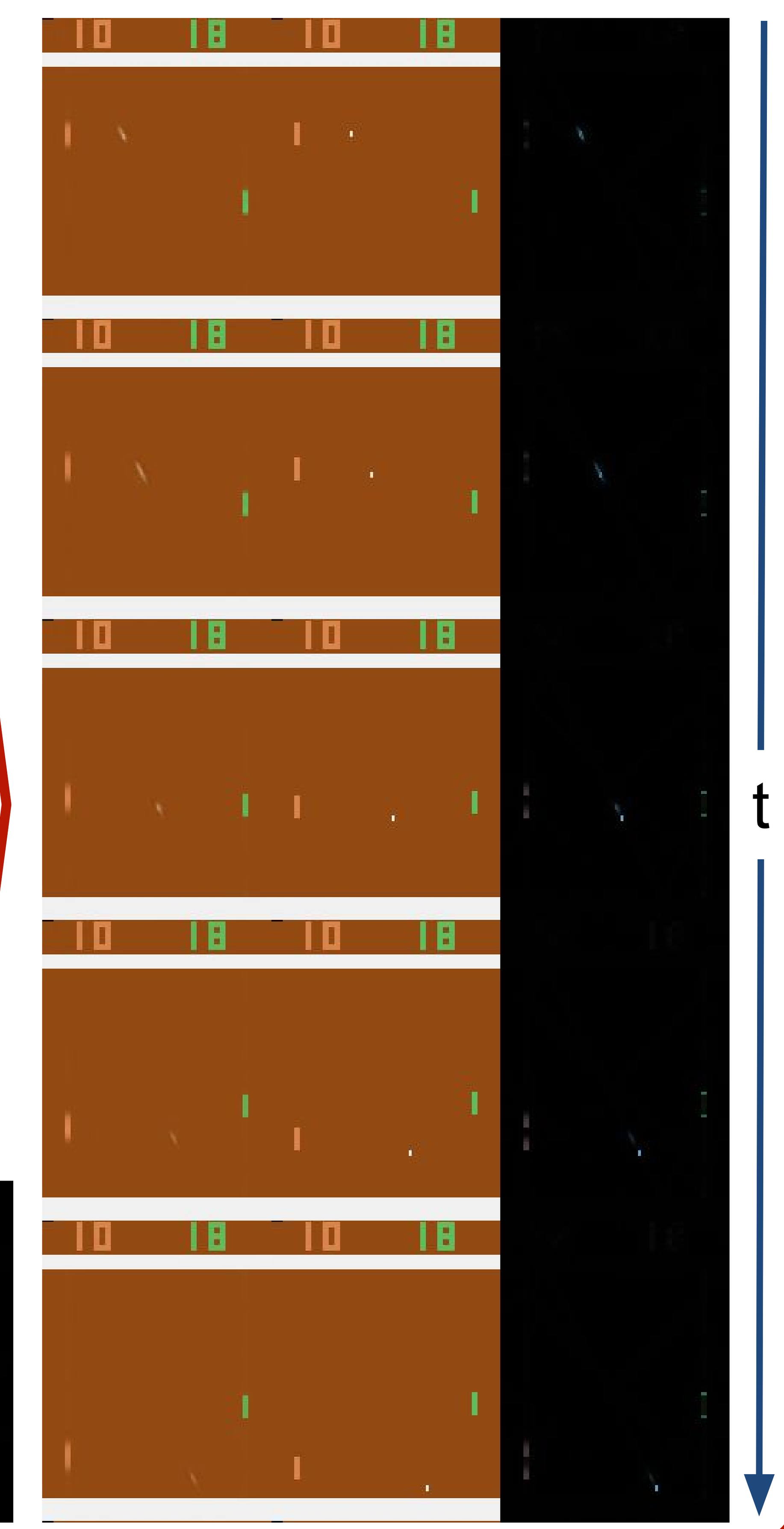


Autoencoder outputs. Doesn't include the mean

Model outputs. Three images are stacked horizontally. The image on the left a prediction, the middle ground truth and the right difference. Each vertical image corresponds to a single timestep. The output at each time is fed back in and used to predict the next frame.

MLP outputs.





References

[1] Oh, Junhyuk, et al. "Action-conditional video prediction using deep networks in atari games." *Advances in Neural Information Processing Systems*. 2015.

[2] Leibfried, Felix, Nate Kushman, and Katja Hofmann. "A Deep Learning Approach for Joint Video Frame and Reward Prediction in Atari Games." *arXiv preprint arXiv:1611.07078* (2016).