# A Deep Learning Model of the Retina

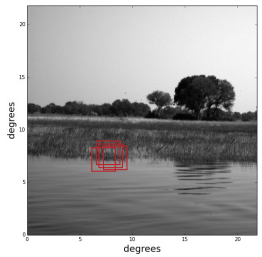
Lane McIntosh and Niru Maheswaranathan
Neurosciences Program, Stanford University School of Medicine

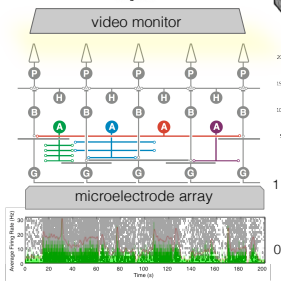
## Introduction

- The retina is just 3 layers of cells that encode the visual world into a binary code of action potentials conveying information about motion, object edges, direction, and even predictions
- Can convolutional neural networks provide a sufficient description of the retina?
- We fit 3- and 5-layer convolutional neural networks to real and generated data
- General problem of predicting time series data from movies

## Dataset

Experimental Setup



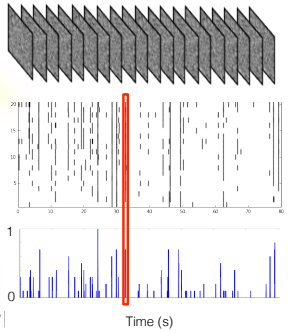




Data Pairs

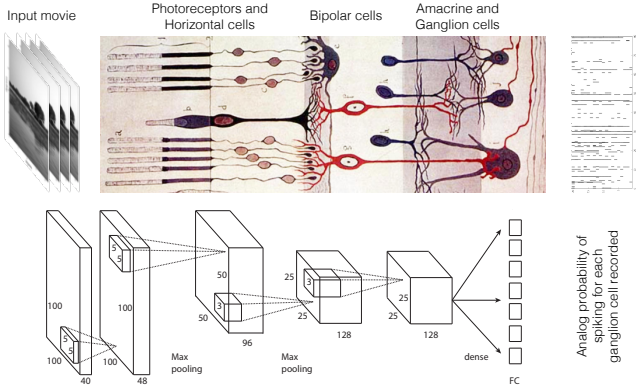
$$\left( \quad , \; p_{\text{spike}}(t) \right)$$

40 grayscale frames spanning 400 ms

Probability of spiking in the last frame's time bin




## Architecture

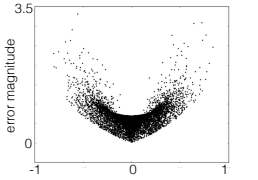



Cross Entropy Loss

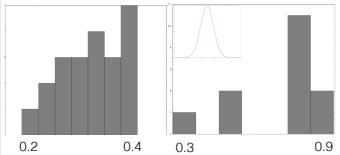
$$\text{Loss} = -\frac{1}{T}\sum_{t=0}^{T}\left[y_t \log \hat{y}_t + (1-y_t)\log(1-\hat{y}_t)\right]$$

Benchmarks

| Study | Stimulus | Model | Metric | Performance |
|---|---|---|---|---|
| Keat *et al.* 2001 [11] | full-field Gaussian white noise | LN | Average # spikes | 0.47 |
| Lesica and Stanley 2004 [12] | Indiana Jones + noise | LNP | Correlation coeff. | 0.6 |
| Pillow *et al.* 2005 [14] | binary noise | generalized IF | Explained variance | 0.91 |
| David and Gallant 2005 [6] | natural images | LN | Explained variance | 0.4 |
| Pillow *et al.* 2008 [15] | binary noise | GLM | Explained variance | 0.9 |




Oracle




## Results

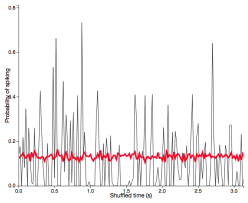
Hyperparameters

Typical heuristics recommend weight scale = 2/N

where N is inputs to first layer

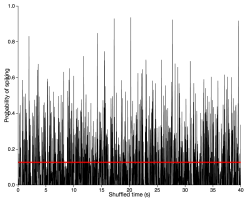
But actual weight scale for temporal prediction was ~0.2

Noise in the data was a sufficient regularizer



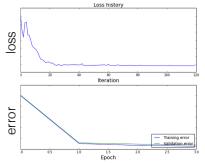
Choosing Model Complexity

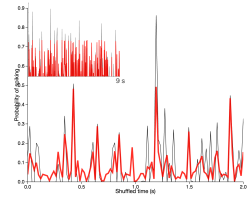
Even with optimizing hyperparameters, 5 layer networks couldn't learn more than the mean firing rate



Performance

Correlation coefficient of ~0.8



## Conclusions

- The difficulty of fitting simple LN models with spatiotemporal features increases with the convolutional neural network architecture complexity
- Noise in the biological system prevents overfitting in absence of regularization on the filter weights
- Weight scales can be order of magnitude larger when the underlying models are sparse with respect to the original data
- Multiple filters were important, even when true model has only a single filter

Thanks to Ben Poole for all the helpful discussions!