Neurobiologically Inspired Encoding and Transfer Learning

Stanford University

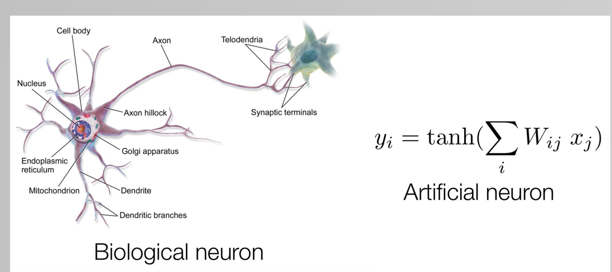
George Sivulka (gsivulka)

Satchel Grant; Josh Melander; The Baccus Lab at Stanford University

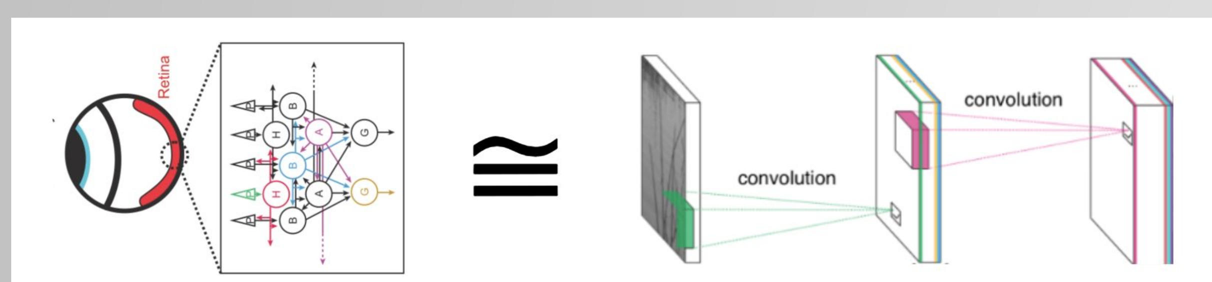
Predicting

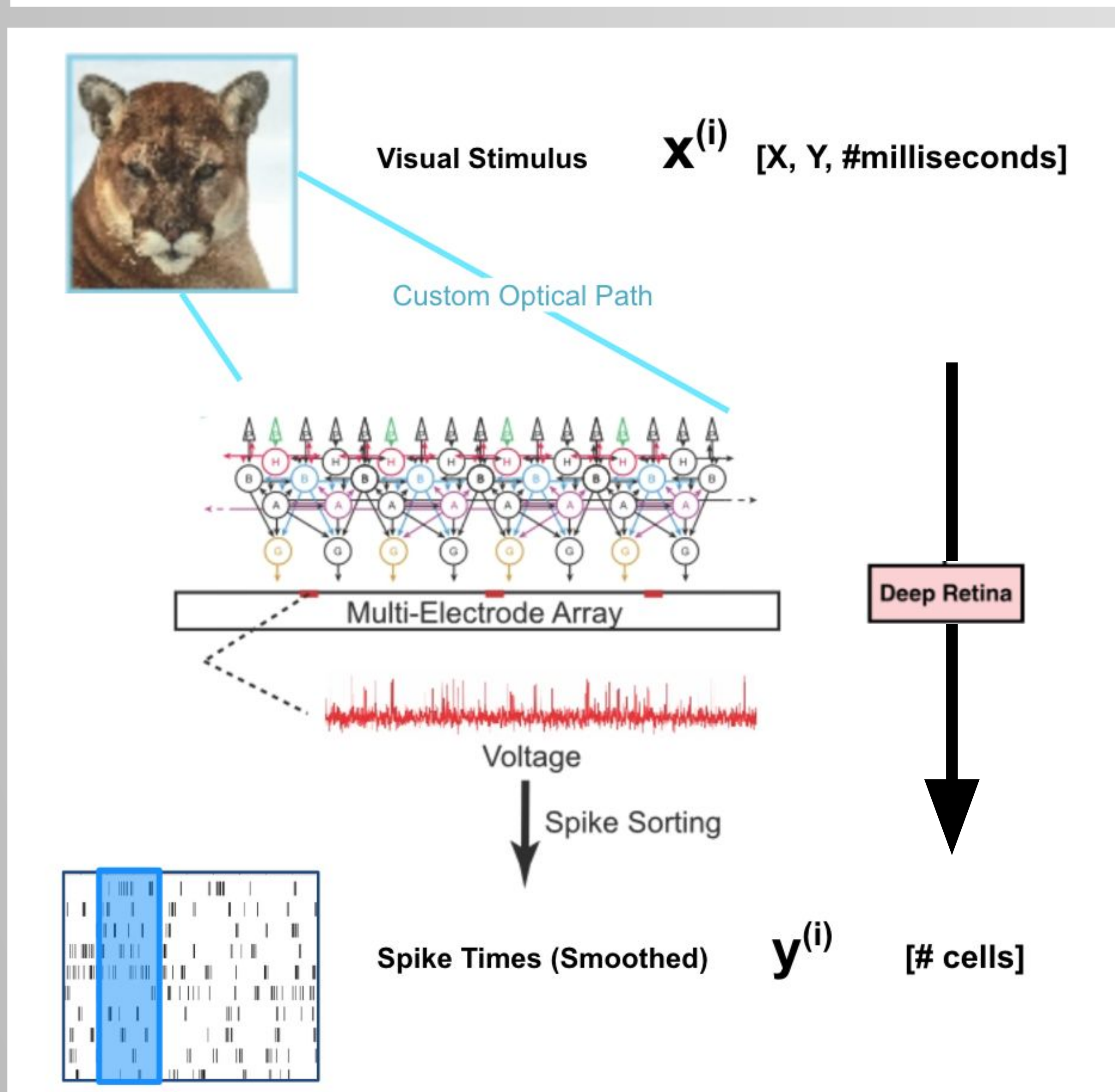
Computational Neuroscience: the feedback loop between neuroscience discovery and machine learning breakthroughs

Present Research: a transfer learning method employing the computations and mechanisms of the human retina as an encoder for time-dependent visual stimulus --analyzing their efficacy in video classification tasks.



Background: recently, deep convolutional neural networks (CNNs) have proven to be the most successful models of the nervous system's sensory processing paradigms to date. [1]





Methods

Fully Convolutional Loss: As part of the effort to make the Deep Retina model Fully Convolutional (see right pane) a "semantic loss" regularizer [2] was employed. This ensured that a selection matrix during training was "one-hot," selecting only one cell per learned filter, so the filter could capture its dynamics.

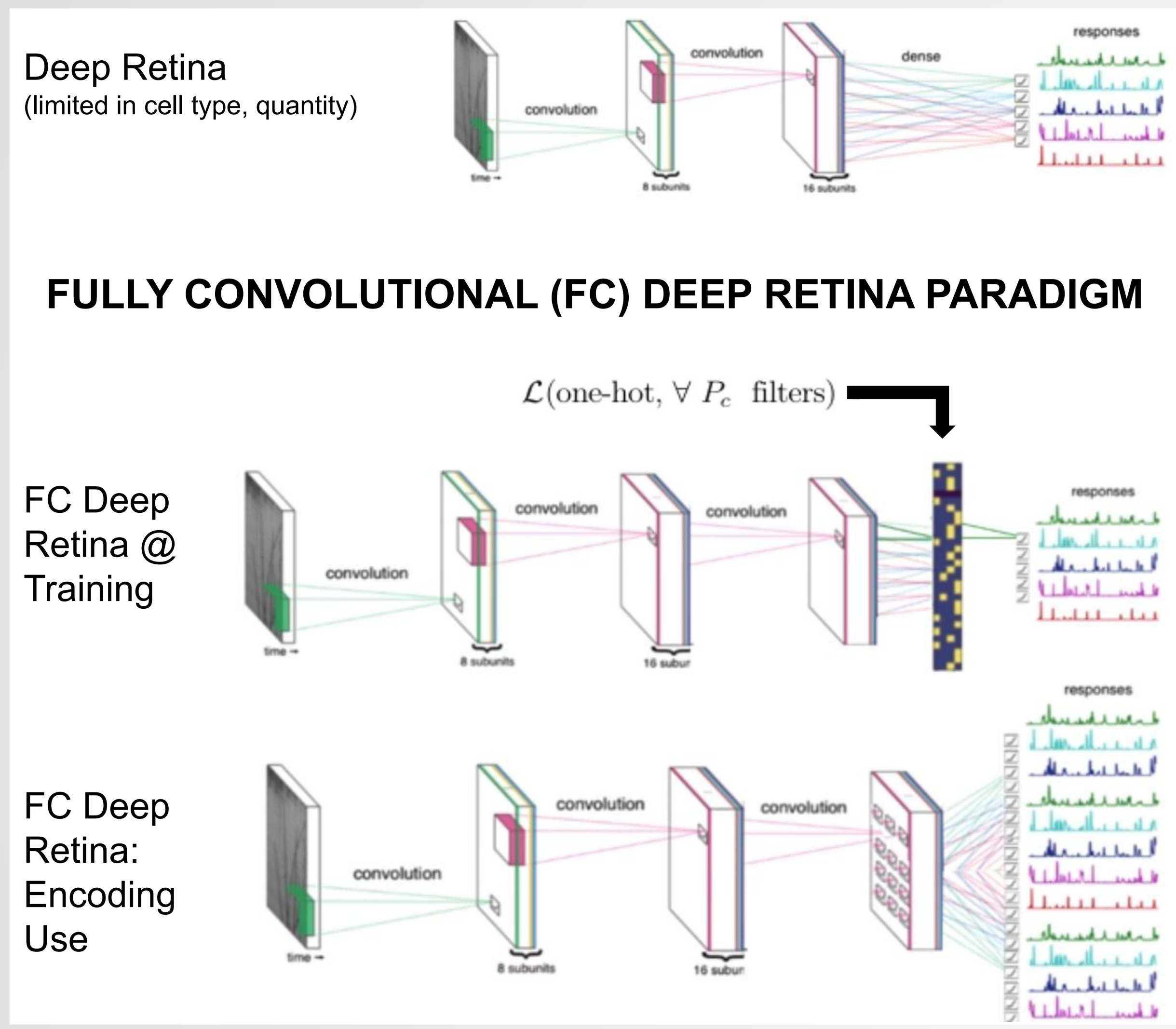
$$\mathcal{L}(\text{one-hot}, \forall P_c \text{ filters}) = -\beta \sum_{i=1}^{c} \log \sum_{j=1}^{n} p_j \prod_{k=1, k\neq j}^{n} (1-p_k)$$

$$p_i = \frac{|w_i|}{\sum_{j=1}^{n} w_j} \quad \forall w_i \subseteq W_c \text{ of each } c \text{ channel}$$

Biological Encoding (CNN)

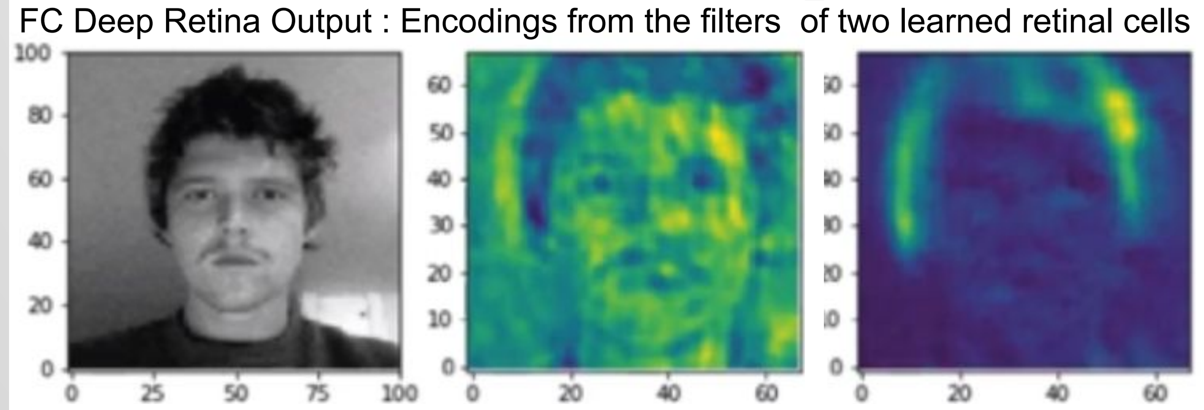
Aa CNN model of retinal spiking: (**Deep Retina**) trained on 40 binned 10ms frames of a natural movie labeled with experimentally measured neural spiking.

Fully Convolutional: to investigate encoding problems the model was changed to be fully convolutional, generalizing a learned filter from one cell type to all locations on an input.



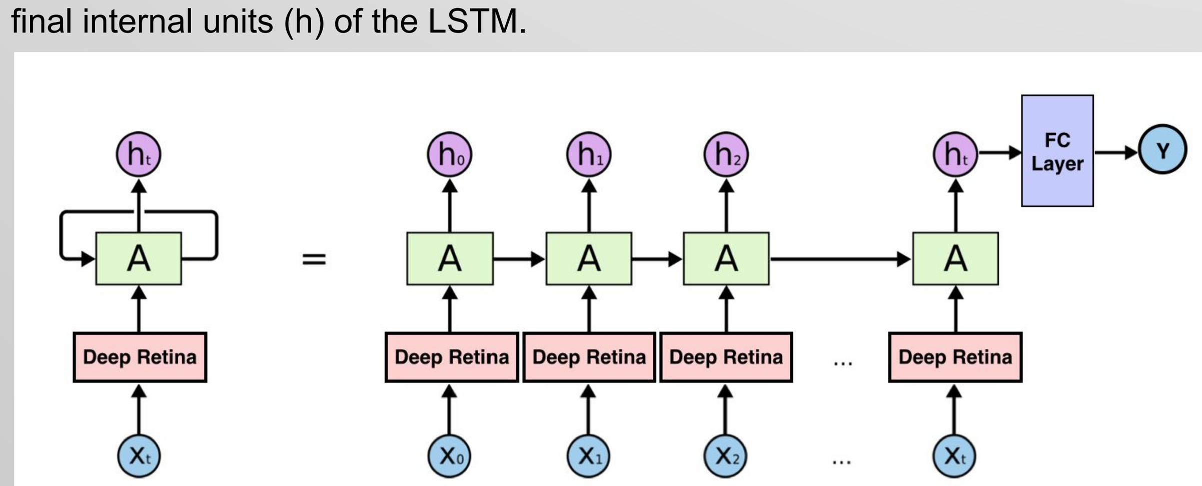
Visual Encodings (Features)

FC Deep Retina Output : Encodings from the filters of two learned retinal cells



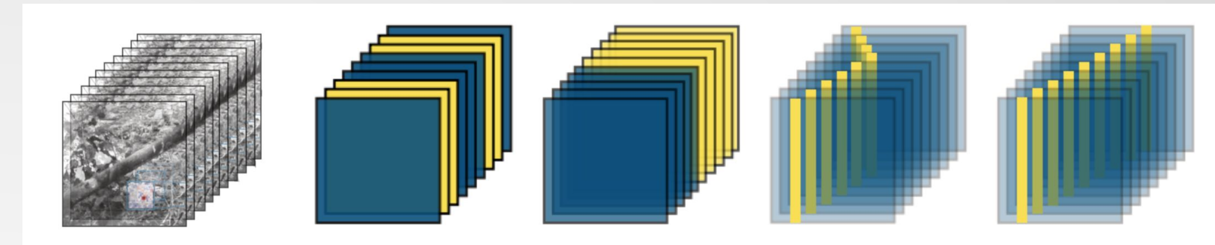
Encoding Classification (RNN)

Schematic of the LSTM architecture with input (X) frame bins first encoded by Deep Retina. Video classification (y) is performed on a FC-layer's output of the final internal units (h) of the LSTM.

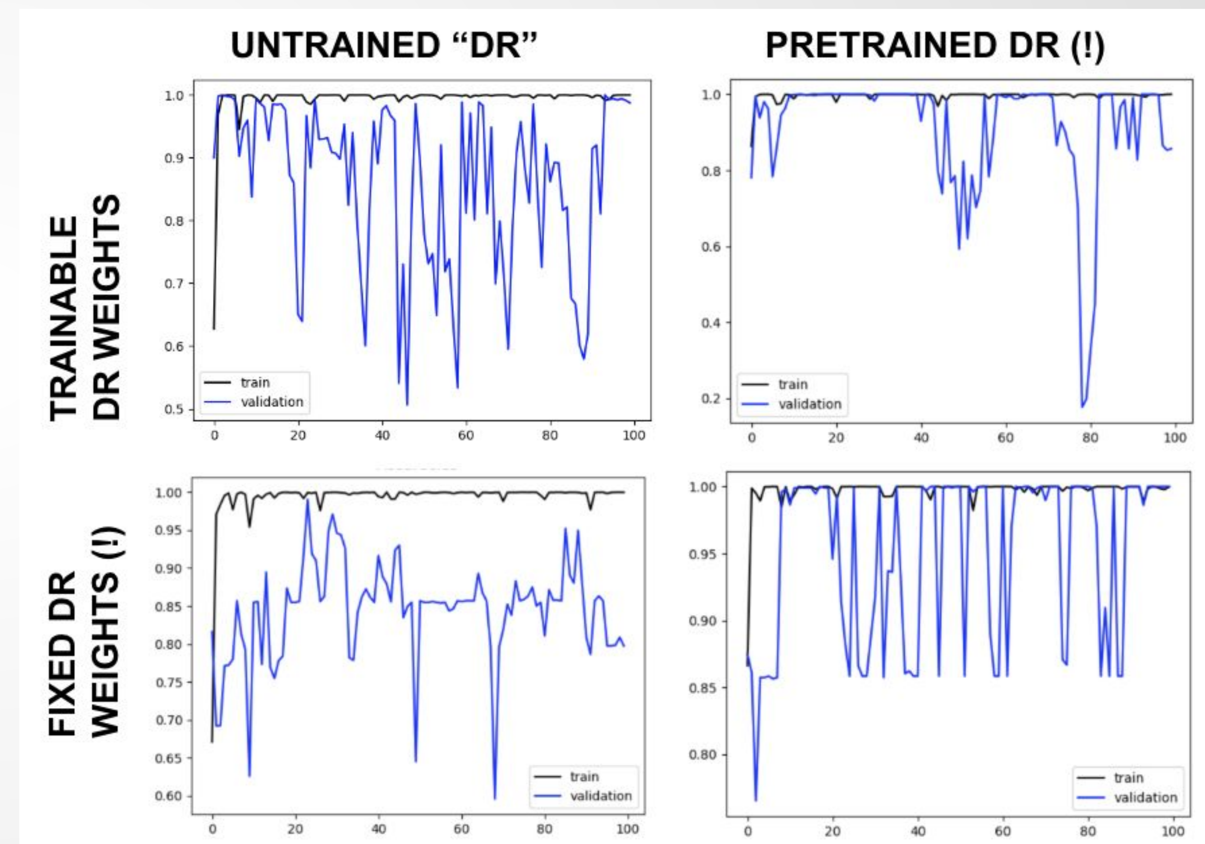


Synthetic Video Data

Synthesized videos of simple artificial stimuli known to evoke retinal response--allowing for a proof of concept first pass at this transfer learning system



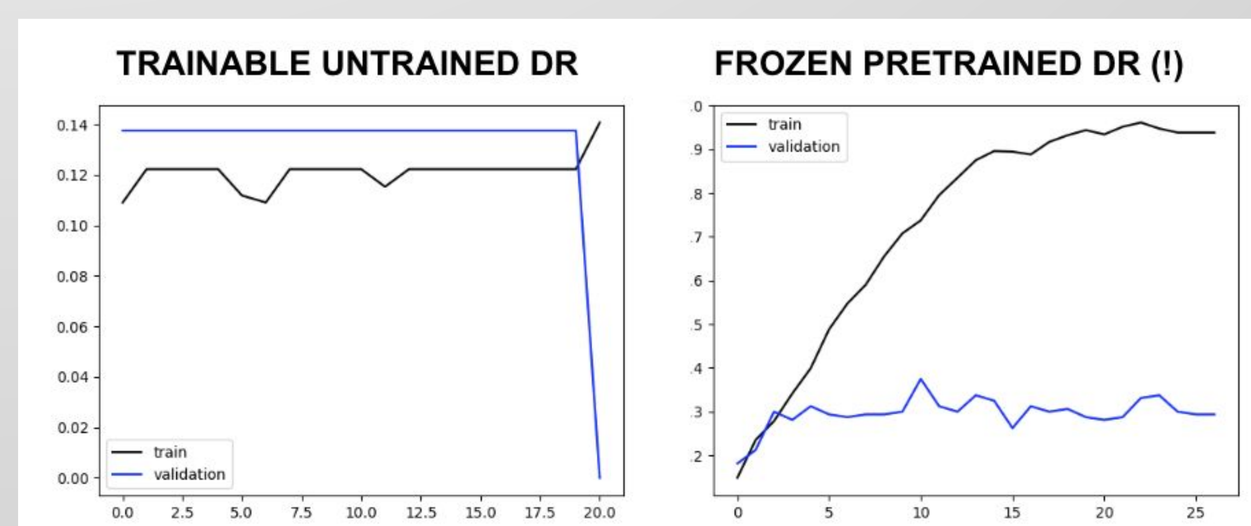
Synthetic Classification Results



UCF-11 Video Data

A curated dataset of action videos was used as a harder test for the system.

UCF-11 Classification Results



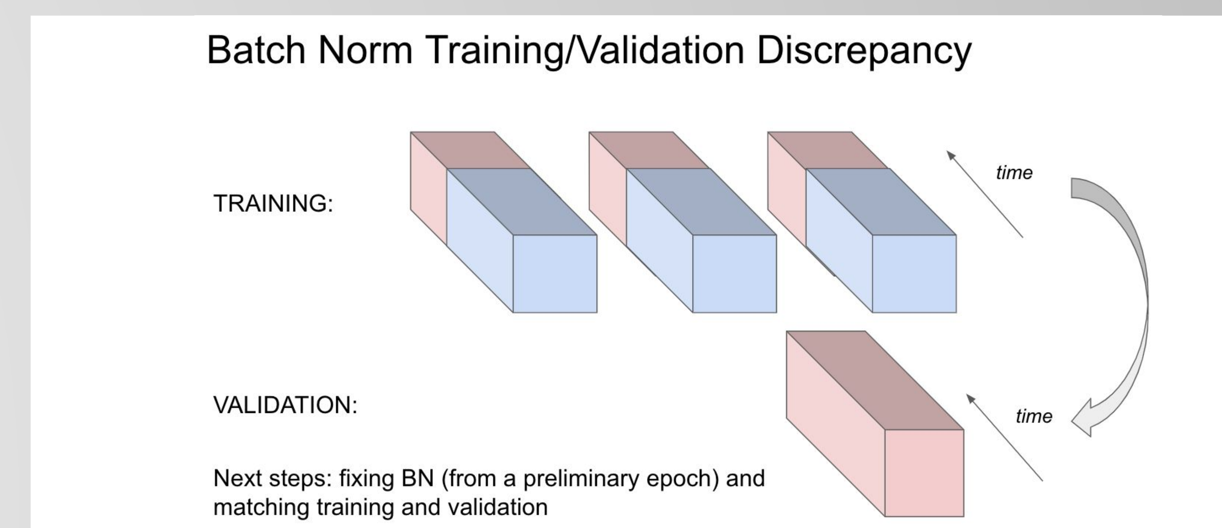
Discussion

Synthetic classification yields almost perfect accuracy across the board, as expected

- **Pretrained DR learns faster** and both trainable and untrainable DR models have **the least stochasticity**

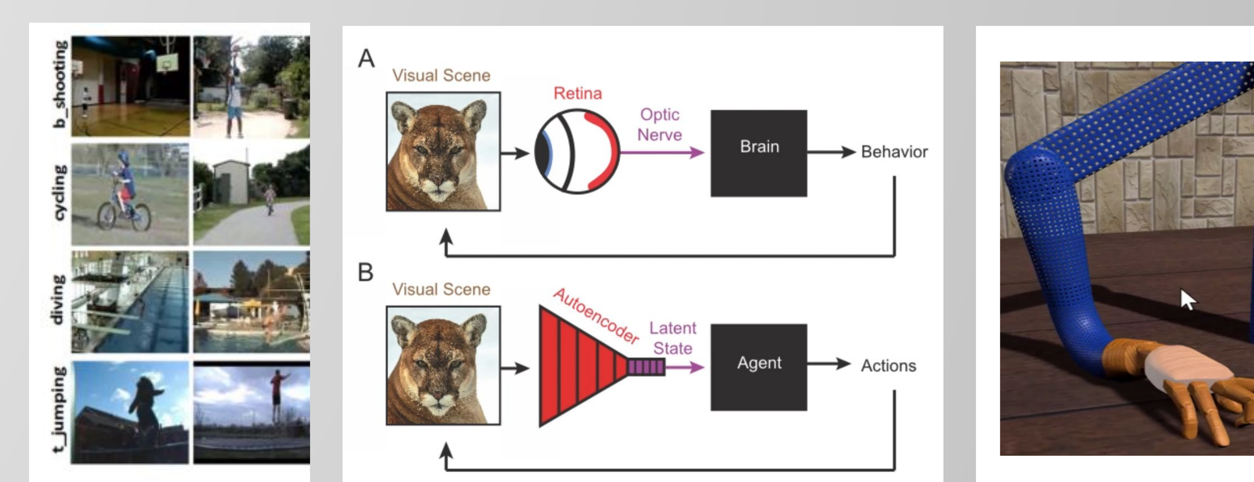
UCF-11 classification is more difficult, yet DR outperforms

- **Frozen DR weights have higher accuracy, quicker learning**
- Trainable random init doesn't seem to learn at all



Future Work

Future work efforts will focus on rectifying validation and training discrepancies, further analyzing UCF-11 performance, and applying these encodings to RL and Meta-RL tasks.



Acknowledgements

This research would not be possible without the assistance of Satchel Grant and Josh Melander of The Baccus Lab at Stanford.

Additionally, this research is indebted to the teaching staff of Stanford's CS 230: "*Deep Learning*" to which it was submitted as coursework.

References

[1] McIntosh, L., Maheswaranathan, N., Nayebi, A., Ganguli, S., & Baccus, S. (2016). Deep learning models of the retinal response to natural scenes. In Advances in neural information processing systems (pp. 1369-1377).

[2] Xu, J., Zhang, Z., Friedman, T., Liang, Y., & Broeck, G. V. D. (2017). A semantic loss function for deep learning with symbolic knowledge. arXiv preprint arXiv:1711.11157.