Recurrent networks fitting neural temporal responses to natural images exhibit contextual modulation

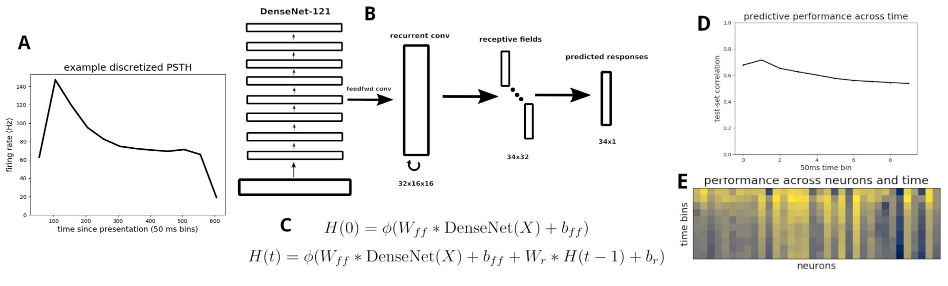
Harold Rockwell[1,3], Yimeng Zhang[2,3], Gayathri Mohankumar[3], Stephen Tsou[3], and Tai Sing Lee[2,3]

Carnegie Mellon University Department of Biological Sciences[1], CMU Computer Science Department[2], Center for the Neural Basis of Cognition[3]

Introduction

We propose a convolutional neural network with a layer of lateral recurrent connections to predict the observed full temporal responses of a neuronal population. The bottom-up receptive fields are obtained through transfer learning from an intermediate layer of a standard Imagenet-trained DenseNet-121 and the recurrent kernels are learned to fit measured neuronal responses, producing high predictive performance. We performed two standard neurophysiological V1 experiments on the hidden units and found that the hidden units exhibit quintessential contextual modulation effects observed in V1, namely longitudinal facilitation and lateral suppression of oriented bars as in association field (Kapadia et. al., 1999), as well as contextual modulation resulting in dynamic reduction in orientation bandwidths and spatial frequency bandwidths over time (Ringach et. al., 1997, 2002). Notably, we find that the early layers of a recurrent CNN trained for object recognition on the Ecoset dataset (Spoerer et. al., 2020) does not replicate these features in its recurrent connections, implying that the neural data prediction objective leads to more realistic learned kernels.

Our results demonstrate that deep learning models with appropriately structured recurrent circuits, trained end-to-end for neural response prediction, can be meaningfully analyzed and reproduce neurophysiological phenomena, therefore potentially providing a computational approach to investigate the mechanisms and circuits of early visual cortex.



$$H(0) = \phi(W_{ff} * \mathrm{DenseNet}(X) + b_{ff})$$

$$H(t) = \phi(W_{ff} * \mathrm{DenseNet}(X) + b_{ff} + W_r * H(t-1) + b_r)$$

Figure 1: Data, model, and performance. **A**: example discretized temporal response of one neuron. **B**: diagram of the predictive CNN model's architecture. **C**: update equation of the recurrent convolutional layer. **D** and **E**: performance across time (measured by Pearson correlation with the real neural response on a test set) averaged across neurons (top) and not (bottom).

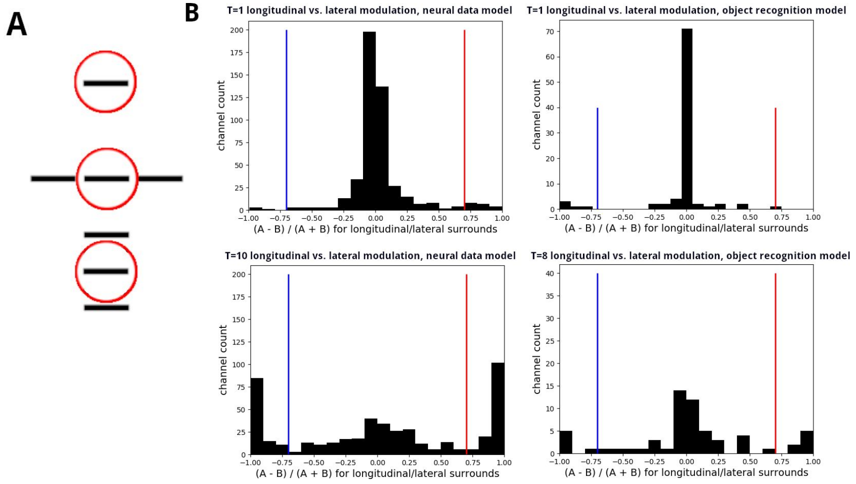
Methods and Results

We recorded spiking activity from 34 neurons in V1, V2 and V4 simultaneously, during a passive fixation task in which 2,250 images were presented for 500 ms in each of 8-10 trials. We discretized the temporal response of the neurons with average spike counts in 50ms time bins, starting 30ms after the inital presentation of the image (Fig 1A).

The recurrent neural network model (Fig 1B) adopts a transfer learning approach, taking as bottom-up input the response of an intermediate layer of a DenseNet-121 to the input image, condensed into 32 hidden units by a 1x1 feedforward convolution, which are processed by a 3x3 lateral recurrent kernel over a number of time steps. The recurrent layer updates according to the equation in Figure 1C, as in (Liang et. al., 2015), and at each timestep, a learned readout compute the predicted output for each modeled neuron, which we train to predict the neural response via a standard Poisson loss (Cadena et. al., 2019) averaged over time. Performance is fairly high and degrades slightly over time (Fig 1D/E).

We run two physiological experiments on the recurrent hidden layer units of our model, and the first layer of the object recognition-trained recurrent CNN from Spoerer et. al., 2020. The first (Fig 2) probes the nature of surround modulation, finding that after recurrent processing, units cluster into longitudinally- and laterally- facilitated groups (the former corresponding to the association field as in Kapadia et. al., 1997). The second (Fig 3) probes the dynamics of tuning across time to sine-wave gratings of varying spatial frequency and orientation. Our model shows sharpening of tuning to both features across time, matching real neural data from Bredfeldt et. al. 2002 and Ringach et. al., 1997, while the object recognition model does not.

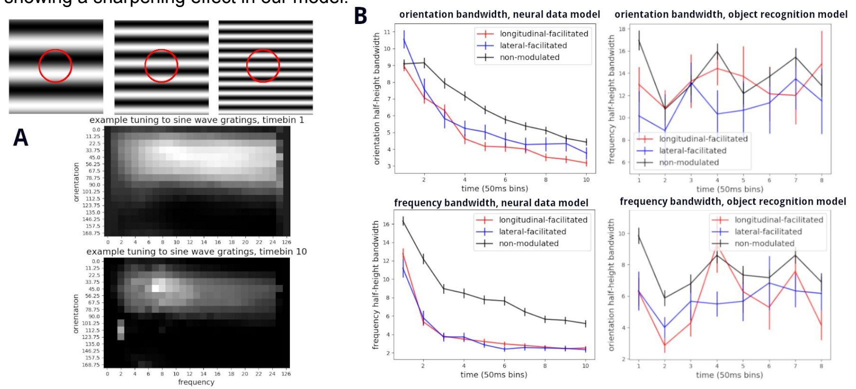
Figure 2: Surround modulation experiment. **A**: example stimuli, base (top), longitudinal (middle) and lateral (bottom) surrounds. **B**: distributions of the difference between the responses to longitudinal and lateral surround stimuli in the first (top) and last (bottom) time steps, for our neural data prediction model (left) and the object recognition model (right).



Discussion

We find that a recurrent CNN model trained end-to-end to predict neural response in early visual cortex both succeeds at that task and learns recurrent circuits that recreate known dynamic tuning properties of cortical cells, similar to previous work in retina (Tanuka et. al., 2019). In addition, we show that a model trained on object recognition alone does not recreate those properties to the same extent, implying that they are not universal properties of functional visual systems, but more unique to the brain. This suggests that deep learning models can be used to infer certain properties of neural circuits, and that the neural data prediction objective is vital for that purpose.

Figure 3: Spatial frequency and orientation tuning across time. **A**: example sine-wave grating stimuli (top) and tuning to both features in the first (middle) and last (bottom) time bins of a single model unit. **B:** average half-height bandwidth of tuning to orientation (top) and frequency (bottom) of our neural data prediction model's units (left) and the object recognition model's units (right) across time, showing a sharpening effect in our model.



References

Bredfeldt et. al. (2002). *Journal of Neuroscience*.
Cadena et. al. (2019). *PLOS Computational Biology*.
Kapadia et. al. (1999). *PNAS*.
Klindt et. al. (2017). *Advances in Neural Information Processing Systems*.
Liang et. al. (2015). *Conference on Computer Vision and Pattern Recognition*.
Ringach et. al. (1997). *Nature*.
Spoerer et. al. (2020). *PLOS Computational Biology*.
Tanuka et. al. (2019). *Advances in Neural Information Processing Systems..*