Can Sparse Coding Help Account for Non-Classical Receptive Field Effects in V1 Simple Cells?

Michael Teti[1], Emily Meyer[2], and Garrett Kenyon[3]

[1]Center for Complex Systems and Brain Sciences, Florida Atlantic University
[2]Center for Nonlinear Studies and [3]CCS-3, Los Alamos National Laboratory

FAU Center for Complex Systems and Brain Sciences
Charles E. Schmidt College of Science
Florida Atlantic University

Models of V1 Simple Cells to Natural Videos

Recent models designed to predict the responses of V1 simple cells to natural videos given estimated spatio-temporal receptive fields (STRFs) have been able to account for only a moderate fraction of the explainable response variance [1] [3]. This may be due to the fact that these models are based only on the classical receptive field and ignore well-established non-classical computations like lateral inhibition. In this preliminary research, we compared the responses of the Locally Competitive Algorithm (LCA) [2], a biologically-motivated model which incorporates lateral inhibition, after convolution and threshold vs. after convolution, threshold and lateral inhibition when presented with previously unseen natural movies and initialized with the same STRFs. In doing so, we observed the extent to which the LCA responses are correlated with the classical STRF while controlling for different thresholds and how this correlation changed as lateral inhibition progressed.

LCA Dictionary Learning

Here, we use the Locally Competitive Algorithm, which is a sparse coding model that incorporates lateral inhibition between neurons to obtain the sparse representation (i.e. where most of the responses are zero for any given input). Given an input vector and an overcomplete basis, the LCA finds an optimal sparse vector of basis coefficients by minimizing the following cost function with respect to the coefficient vector:

$$E = \frac{1}{2}||\mathbf{I} - \mathbf{\Phi} * \mathbf{a}||_2^2 + \frac{1}{2}\lambda^2||\mathbf{a}||_0 \tag{1}$$

where $\mathbf{I}$ is the input vector, $\mathbf{\Phi}$ is the basis, and $\mathbf{a}$ is a sparse vector of basis coefficients. The $||\cdot||_0$ operator is the ℓ_0-pseudonorm, which is typically approximated by the ℓ_1-norm and computes the number of non-zero entries in $\mathbf{a}$, and λ controls the weighting between reconstruction performance and the sparsity of the coefficient vector.

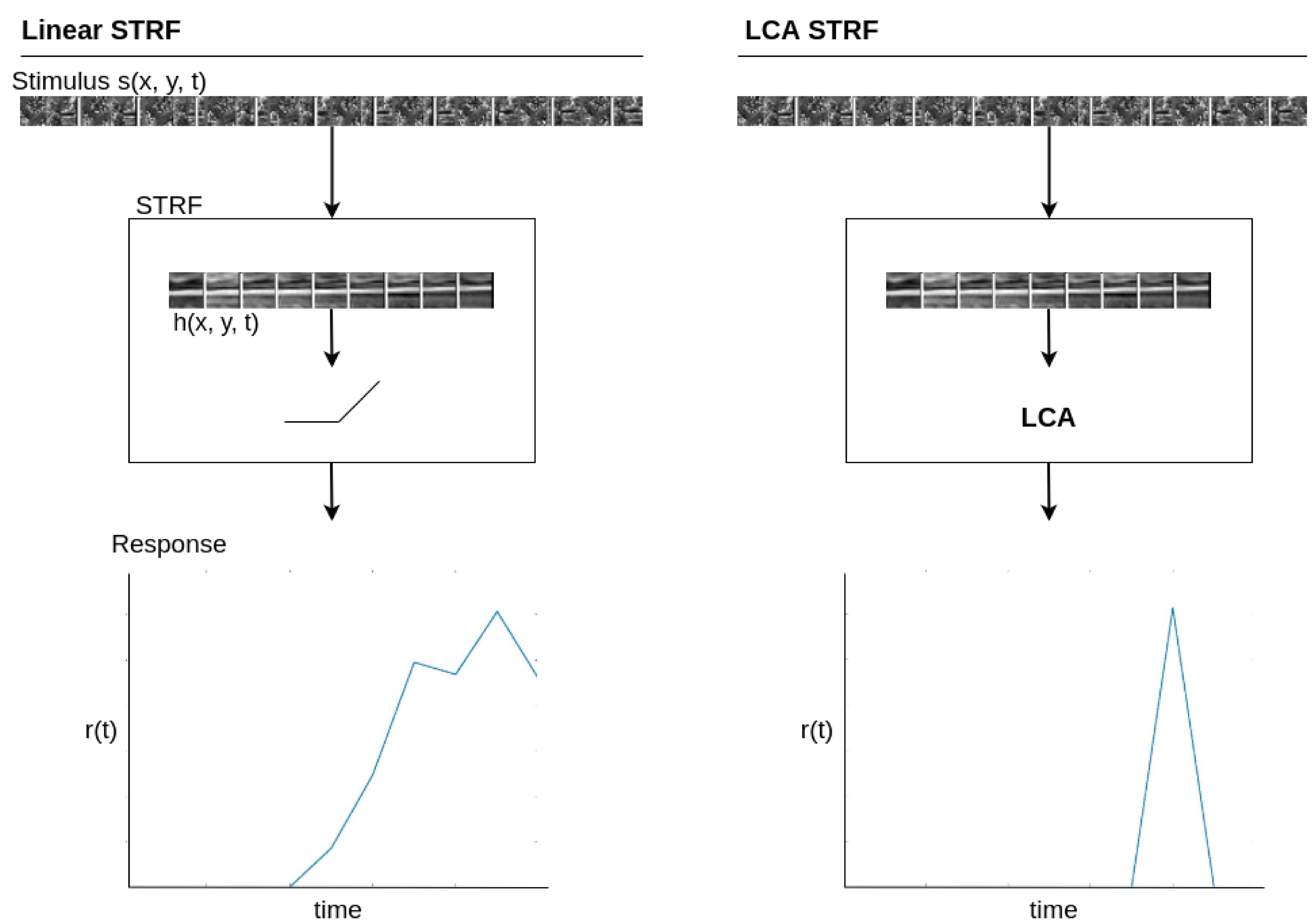
When sparse coding is performed to find a sparse representation of a given input with a fixed dictionary (or batch of inputs). Next these activations are used to update the dictionary by minimizing an energy function with respect to the dictionary via iterative stochastic gradient descent and a using a local Hebbian learning rule with momentum. Once the dictionary is updated, it is once again fixed and used to find a sparse representation given the next input(s), and the process is repeated.

Methods

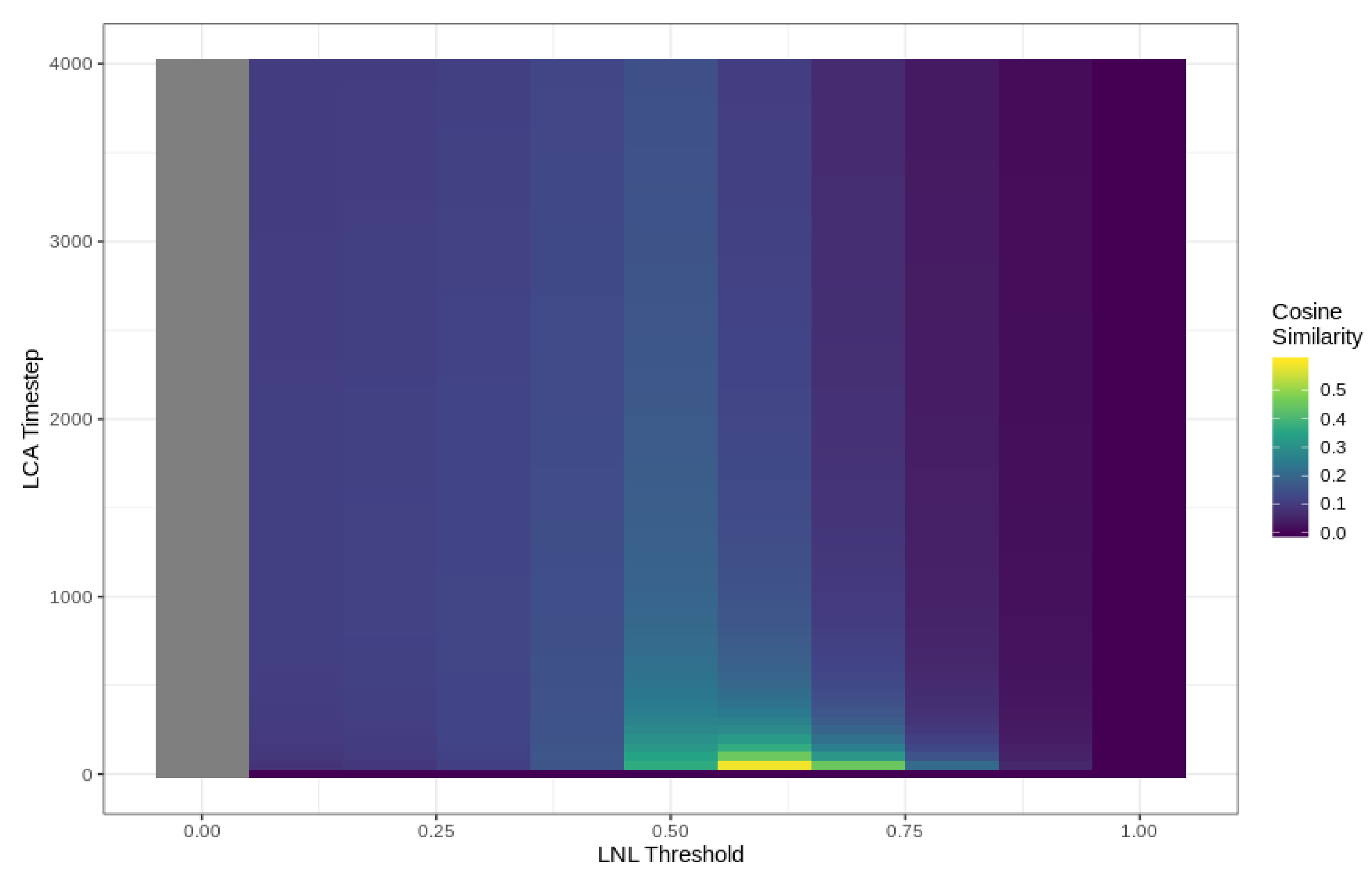
1. Learn a dictionary of 1000 convolutional STRFs (17 x 17 pixels) via the sparse dictionary learning procedure above (with a threshold of 0.125) by training the network to reconstruct nine consecutive 32 x 64 grayscale video frames (from the ImageNet video dataset) at a time.
2. Fine-tune the learned STRFs in step 1 under three separate threshold values – 0.075 (low sparsity), 0.125 (mid sparsity), and 0.175 (high sparsity) – to observe whether the threshold used during step 1 had a large effect on these results or if the results generalize beyond this hyper-parameter.
3. Initialize each of the fine-tuned LCA models from step 2 (one for each threshold value used in fine-tuning) with the learned STRFs and record the responses of each one to consecutive frames from the ImageNet test set before, during, and after 4000 iterations of lateral inhibition.

Responses Before and After Lateral Inhibition

An example of the responses generated from a representative LCA neuron before and after lateral inhibition using the same input frames and threshold can be seen in the figure below. Whereas the neuron is active during most of the video frames before lateral inhibition, it is only active for one frame after lateral inhibition has been performed.

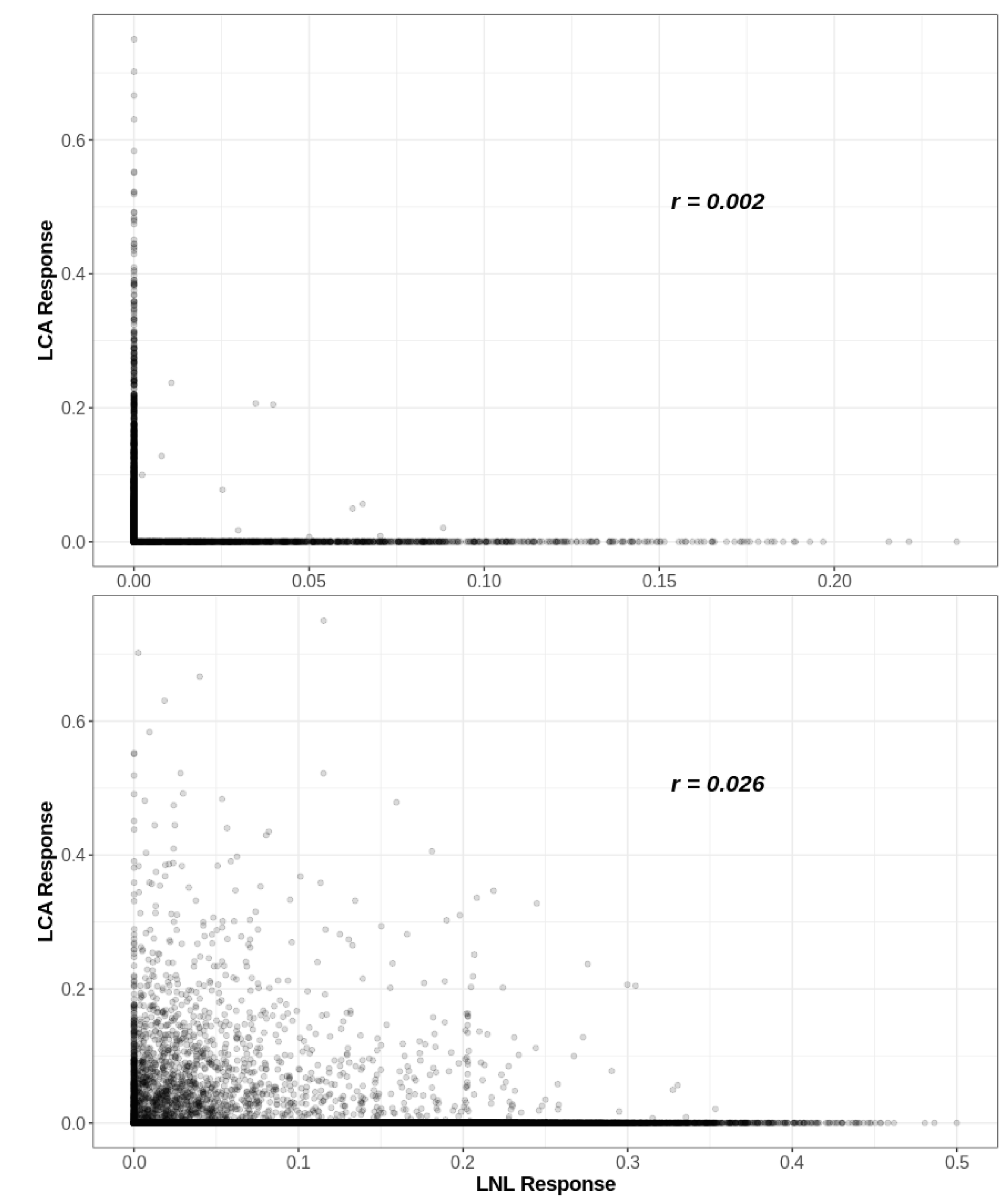


The cosine similarity between the responses before and after lateral inhibition was the highest at the first LCA timestep and decreased rapidly thereafter. Because it requires time for the lateral inhibition to develop, it makes sense that the responses should be in maximal alignment early on when presented with a new sequence of video frames before progressively diverging over time.



Correlation Analysis

The correlation between normalized responses of LCA neurons beffore and after lateral inhibition is extremely weak, even when using the optimal threshold on the pre-lateral-inhibition responses to minimize the difference in sparsity (top) and maximize the cosine similarity (bottom).



Conclusion

We explored the relationship between the LCA model's responses under computations based only on the classical STRF (as in typical encoding models) vs. when modulation from surrounding neurons was enacted. Our results indicate that responses before and after lateral inhibition are strikingly different given the exact same STRF. Given the fact that lateral inhibition is a known computation of visual cortical neurons, this may help explain why current encoding models are not able to predict responses to natural videos with a high accuracy, but more experimentation is needed to confirm this.

References

[1] Stephen V David and Jack L Gallant. "Predicting neuronal responses during natural vision". In: *Network: Computation in Neural Systems* 16.2-3 (2005), pp. 239–260.

[2] Christopher Rozell et al. "Locally competitive algorithms for sparse approximation". In: *2007 IEEE International Conference on Image Processing*. Vol. 4. IEEE. 2007, pp. IV–169.

[3] Saskia EJ de Vries et al. "A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex". In: *Nature Neuroscience* 23.1 (2020), pp. 138–151.