

## On the limits of feed-forward processing in visual object recognition

Gabriel Kreiman<sup>1,2</sup>, Thomas Serre<sup>3</sup>, and Tomaso Poggio<sup>3</sup>

<sup>1</sup>Division of Neuroscience and Ophthalmology, Children's Hospital Boston, Harvard Medical School

<sup>2</sup>Center for Brain Science, Harvard University

<sup>3</sup>McGovern Institute for Brain Research, MIT

The primate visual system shows a high degree of selectivity while at the same time being robust to large transformations of the images<sup>1</sup>. This remarkable degree of invariance and specificity can be achieved quite rapidly as evidenced by the short latency of selective spike responses in higher levels of visual cortex and also by psychophysics performance in rapid visual presentation tasks and backward masking tasks<sup>2-4</sup>. The speed of processing in the visual system has led to the notion that feed-forward processing can account for several aspects of visual object recognition. Indeed, a purely feedforward computational model is very successful in explaining physiological responses in several areas of visual cortex from primary visual cortex through inferior temporal cortex and also several psychophysical observations where humans need to rapidly identify or categorize objects<sup>5-7</sup>.

Here we quantitatively explore the limits of feedforward processing by examining how well a purely feedforward model can perform identification and categorization tasks in images containing multiple objects or other scenarios where we parametrically vary the amount of clutter in the image. We first show that the performance of the model in response to isolated objects matches the electrophysiological properties of IT neurons in terms of their accuracy as well the robustness to changes in object scale and position. We then show that performance degrades with increasing number of objects or with increasing degrees of clutter in natural scenes. These results point to the limits of feedforward processing in visual object recognition and emphasize the role of attention in processing and interpreting complex natural scenes.

### Acknowledgments

We thank Minjoon Kouh for helpful discussions. This work was supported by a McGovern Institute Fellowship and the Children's Hospital Boston.

### References

1. Logothetis, N. K. & Sheinberg, D. L. Visual object recognition. *Annual Review of Neuroscience* 19, 577-621 (1996).
2. Potter, M. & Levy, E. Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology* 81, 10-15 (1969).
3. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *Nature* 381, 520-522 (1996).
4. Desimone, R., Albright, T., Gross, C. & Bruce, C. Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience* 4, 2051-2062 (1984).
5. Serre, T. et al. CBCL Paper #259/AI Memo #2005-036 (MIT, Boston, 2005).
6. Rolls, E. T. Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron* 27, 205-18 (2000).
7. Riesenhuber, M. & Poggio, T. Neural mechanisms of object recognition. *Curr Opin Neurobiol* 12, 162-8 (2002).