



Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2010-062
CBCL-295

December 31, 2010

**Neurons That Confuse Mirror-Symmetric
Object Views**

Jim Mutch, Joel Z Leibo, Steve Smale, Lorenzo
Rosasco, and Tomaso Poggio

Neurons That Confuse Mirror-Symmetric Object Views

Jim Mutch^{1,2}, Joel Z Leibo^{1,2}, Steve Smale^{2,3}, Lorenzo Rosasco^{1,2,4}, and Tomaso Poggio^{1,2}

¹*Center for Biological and Computational Learning and the McGovern Institute for Brain Research, Cambridge MA, USA*

²*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge MA, USA*

³*City University of Hong Kong, Kowloon Hong Kong*

⁴*Italian Institute of Technology, Genova Italy*

December 31, 2010

Abstract

Neurons in inferotemporal cortex that respond similarly to many pairs of mirror-symmetric images – for example, 45° and -45° views of the same face – have often been reported. The phenomenon seemed to be an interesting oddity. However, the same phenomenon has also emerged in simple hierarchical models of the ventral stream. Here we state a theorem characterizing sufficient conditions for this curious invariance to occur in a rather large class of hierarchical networks and demonstrate it with simulations.

Neurons in inferotemporal (IT) cortex – the highest visual area in the ventral visual pathway – respond with varying degrees of selectivity to a specific pattern such as one particular view of an object. In some cases, single cells responding in a similar way to many pairs of images, each representing an object view and its mirror-symmetric counterpart, have been reported (Logothetis et al., 1995), as shown in fig. 1. Similar tuning has also been observed for some units in a hierarchical model of cortex, as shown in fig. 2.

This property seems puzzling and so far appeared to originate from an accidental tuning of neurons in IT to features in common between views and their mirror-symmetric versions. Without discounting this possibility, we focus here on establishing an alternative – and quite possibly, complementary – explanation. We argue that it may (also) reflect a property of a large class of hierarchical architectures. Recently, we have been developing a general theory (Smale et al., 2010) of hierarchical feedforward architectures such as the model of visual cortex shown in fig. 3. Within this theoretical framework we have established

a proposition (see appendix) which says that if a unit in a hierarchical, balanced architecture is optimally tuned to a mirror-symmetric template (such as a frontal face or a symmetric patch from it), then it will respond equally to any image and its mirror-symmetric counterpart. The “balanced” condition means that the tunings of the cells in lower layers feeding into a higher-level unit must be a symmetric set – such as a set of simple cells tuned to vertical, horizontal, left-right oblique and right-left oblique edges; see fig. 3. This condition is relatively weak since most hierarchical architectures will satisfy it. The first condition on the other hand – optimal tuning to a symmetric pattern – is much more demanding and cannot be expected to hold for all units. The implication of this result is that the observation of tuning curves similar to the one shown in fig. 1 in electrophysiology experiments can be accounted for within the hypothesis of a hierarchical architecture if the recorded cell is optimally tuned to a symmetric template.

It is intriguing to speculate about the development of mirror-symmetric tuning in some IT cells. In our hierarchical model of visual cortex – which achieves recognition performance similar to humans in rapid scene categorization tasks (Serre et al., 2007) – we assume (Serre et al., 2005) that the tuning of cells in the higher levels is imprinted from parts of natural images during a developmental stage by a temporal association mechanism (Földiák, 1991). In particular, we have conjectured (Leibo et al., 2010) that the initial tuning of some of the cells may be due to image patches from objects to which the baby human or monkey is exposed early in development. Since a large part of early visual experience consists of faces, it is possible that symmetric templates could be acquired from frontal views of faces. Alternatively, symmetric tuning could result from the failure of a temporal association mechanism (e.g. the trace rule) leading to tuning to the symmetric superposition of image traces determined by rotation of a symmetric object (see figs 4 and 5).

In summary, we have established an alternative mechanism that may account for the observation of cells with equal tuning to an object and its mirror-symmetric pair: cells with these tuning curves may be optimally tuned to a symmetric template. In future work we plan to investigate this mechanism as well as its alternative – tuning to features that are similar under reflection – in the context of the recent experimental observation of a large number of cells with this tuning property located in the macaque anterior lateral face patch (Freiwald and Tsao, 2010).

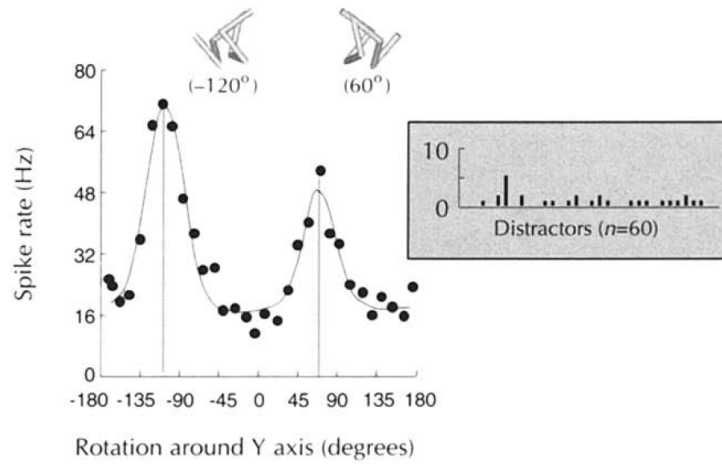


Figure 1: A neuron in IT tuned to a view and its mirror-symmetric version. (From (Logothetis et al., 1995), fig. 4b.)

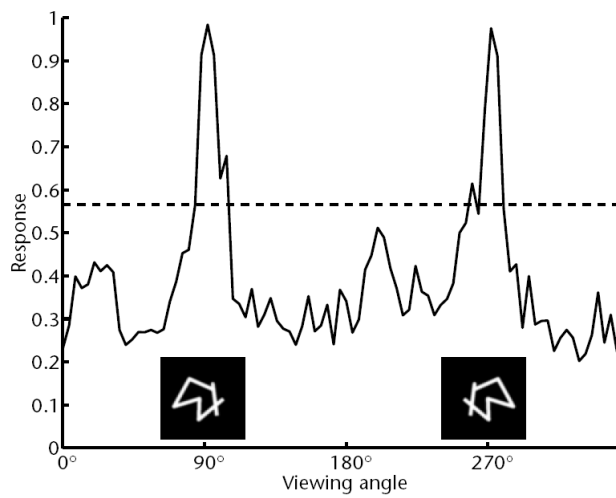


Figure 2: A model unit tuned to a view and its mirror-symmetric version. (From (Riesenhuber and Poggio, 1999), fig. 4d.)

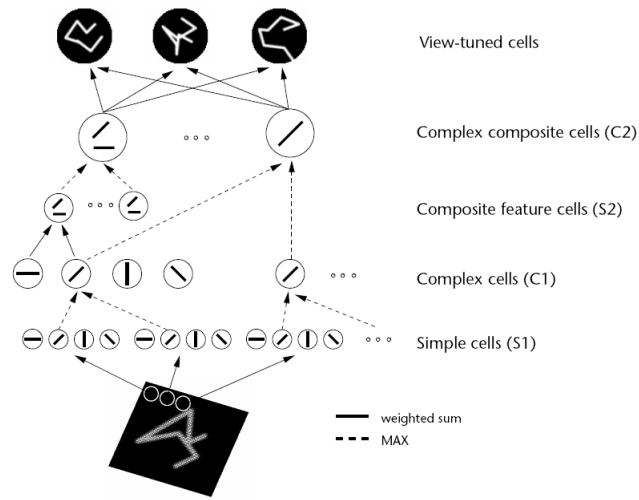


Figure 3: A hierarchical model of visual cortex, similar to Fukushima's. (From (Riesenhuber and Poggio, 1999), fig. 2.)

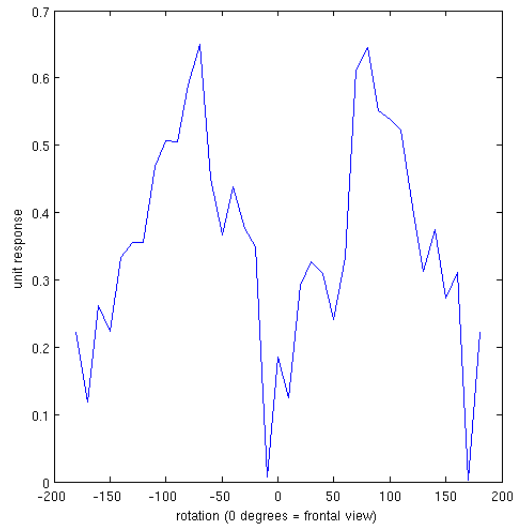


Figure 4: The response of a model unit – optimally tuned for a symmetric stimulus – to an object undergoing rotation around the vertical axis. The optimal stimulus is shown in fig. 5.

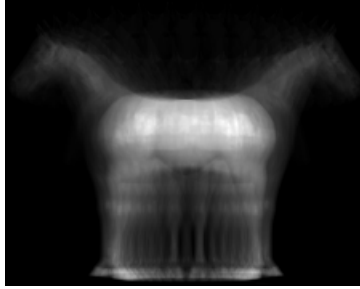


Figure 5: *The optimal stimulus of a model unit tuned to the superposition of images of a symmetric object undergoing rotation around the vertical axis.*

Appendix: A Theorem on Symmetry Invariance in Hierarchical Learning Architectures

In this appendix we discuss a theorem establishing invariance to mirror symmetry of a large class of hierarchical learning architectures. To do so, we first need to recall some basic concepts. We refer to (Bouvier et al., 2010) for a detailed proof and further technical details. A hierarchical learning architecture defines recursively *local* feature maps that map image patches to finite dimensional vector representations. At each layer a hierarchical learning machine provides a representation of image patches of a given receptive field size. The receptive field size increases as we proceed to higher layers. A global representation of the image can be obtained by concatenating such local representations (or eventually pooling over the whole image). The local representations are obtained by first measuring the similarity of an image patch of a given receptive field size to different *templates* (of the same size) by means of a suitable tuning function (simple cells) and then pooling the response of each template over a larger receptive field (complex cells). Such basic computations are repeated recursively as we proceed higher in the hierarchy. The templates are visual patterns to which the simple cells are responsive. We say that a template set on one layer is *balanced*, if for any given template in the set, its symmetric version is also in the set (and in particular if a template itself is symmetric). If we denote by $f, f \circ s$ an image and its symmetric version respectively, and by $N_n(f)(t)$ the response of a template t to an image f at layer n , we can state the theorem about symmetry invariance as follows.

Theorem 0.1. *Under general conditions, a symmetric template $t = t \circ s$ has the same response to an image or its symmetric version*

$$N_n(f)(t) = N_n(f \circ s)(t),$$

assuming that the template sets at lower layers are balanced.

The proof of the theorem, as well as the details of the assumptions, is given in (Bouvier et al., 2010). The intuition behind the proof is that replacing the image with its mirror-symmetric version is equivalent to looking at the original image with the mirror-symmetric version of the architecture. Under the assumptions of the theorem it is easy to see that the architecture and its mirror-symmetric version are identical, up to a reshuffling of the units that does not affect the top units pooling signals over the whole visual field.

Acknowledgments

This report describes research done at the Center for Biological & Computational Learning, which is in the McGovern Institute for Brain Research at MIT, as well as in the Department of Brain & Cognitive Sciences, and which is affiliated with the Computer Sciences & Artificial Intelligence Laboratory (CSAIL). This research was sponsored by grants from DARPA (IPTO and DSO), AFOSR and the National Science Foundation.

Additional support was provided by: Adobe, Honda Research Institute USA, King Abdullah University Science and Technology grant to B. DeVore, NEC, Sony and especially by the Eugene McDermott Foundation.

References

- Bouvier, J., Poggio, T., Rosasco, L., Smale, S., and Wibisono, A. (2010). Generalization and Properties of the Neural Response. *MIT-CSAIL-TR-2010-051/CBCL-292*.
- Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3(2):194–200.
- Freiwald, W. and Tsao, D. (2010). Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. *Science*, 330(6005):845.
- Leibo, J., Mutch, J., Ullman, S., and Poggio, T. (2010). From primal templates to invariant recognition. *MIT-CSAIL-TR-2010-057, CBCL-293*.
- Logothetis, N., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5(5):552–563.

- Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025.
- Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., and Poggio, T. (2005). A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. *CBCL Paper #259/AI Memo #2005-036*.
- Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15):6424–6429.
- Smale, S., Rosasco, L., Bouvrie, J., Caponnetto, A., and Poggio, T. (2010). Mathematics of the neural response. *Foundations of Computational Mathematics*, 10(1):67–91.

