Computer Science and Artificial Intelligence Laboratory

Technical Report

MIT-CSAIL-TR-2013-019
CBCL-313

August 6, 2013

Does invariant recognition predict tuning of neurons in sensory cortex?

Tomaso Poggio, Jim Mutch, Fabio Anselmi, Andrea Tacchetti, Lorenzo Rosasco, and Joel Z. Leibo

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

Does invariant recognition predict tuning of neurons in sensory cortex?

Tomaso Poggio, Jim Mutch, Fabio Anselmi,
Andrea Tacchetti, Lorenzo Rosasco, and Joel Z. Leibo
(tp@ai.mit.edu, jmutch@mit.edu, nfabio.anselmi@gmail.com,
atacchet@mit.edu, lrosasco@mit.edu, jzleibo@mit.edu)

Center for Biological and Computational Learning
Massachusetts Institute of Technology
Cambridge, MA 02139

August 6, 2013

Abstract

Tuning properties of simple cells in cortical V1 can be described in terms of a "universal shape" characterized by parameter values which hold across different species ([12], [33], [22]). This puzzling set of findings begs for a general explanation grounded on an evolutionarily important computational function of the visual cortex. We ask here whether these properties are predicted by the hypothesis that the goal of the ventral stream is to compute for each image a "signature" vector which is invariant to geometric transformations as postulated in [30] – with the the additional assumption that the mechanism for continuously learning and maintaining invariance consists of the memory storage of a sequence of neural images of a few objects undergoing transformations (such as translation, scale changes and rotation) via Hebbian synapses. For V1 simple cells the simplest version of this hypothesis is the online Oja rule which implies that the tuning of neurons converges to the eigenvectors of the covariance of their input. Starting with a set of dendritic fields spanning a range of sizes, simulations supported by a direct mathematical analysis show that the solution of the associated "cortical equation" provides a set of Gabor-like wavelets with parameter values that are in broad agreement with the physiology data. We show however that the simple version of the Hebbian assumption does not predict all the physiological properties. The same theoretical framework also provides predictions about the tuning of cells in V4 and in the face patch AL [17] which are in qualitative agreement with physiology data.

1 Computational goal and mechanism

The original work of Hubel and Wiesel, as well as subsequent research, left open the questions of (1) what is the function of the ventral stream in visual cortex and (2) how are the properties of its neurons related to it. Poggio et al.'s so-called "magic" theory [29, 30] – here "M-theory" – proposes

that the main goal of the ventral stream is to compute, at each level in the hierarchy of visual areas, a signature that is unique for the given image, invariant under geometric transformations and robust to small perturbations. M-theory suggests a mechanism for learning the relevant invariances during unsupervised visual experience: storing sequences of images (called "templates") of a few objects transforming, for instance translating, rotating and looming. It has been claimed that in this way invariant hierarchical architectures similar to models of the ventral stream such as Fukushima's Neocognitron [7] and HMAX [32, 35] – as well as deep neural network architectures called convolutional networks [15, 14] and related models—e.g. [28, 27, 34, 1]—can be learned from unsupervised visual experience. Here we focus on V1, making the assumption that the development of an array of initially untuned cells with spatially localized dendritic trees of different sizes is genetically determined, reflecting the organization of the retinal array of photoreceptors.

M-theory assumes that the templates and their transformations – corresponding to a set of "simple" cells – are memorized from unsupervised visual experience. In a second learning step, a "complex" cell is wired to simple cells that are activated in close temporal contiguity and thus are likely to correspond to the same patch of image undergoing a transformation in time [5]. However, the proposal of direct storage of sequences of images patches – seen through a Gaussian window – in a set of V1 cells is biologically implausible. Here we examine the biologically more plausible proposal that the neural memorization of frames (of transforming objects) is performed online via Hebbian synapses that change as an effect of visual experience. Specifically, we assume that the distribution of signals "seen" by a maturing simple cell is Gaussian in x, y reflecting the distribution on the dendritic tree of synapses from the lateral geniculate nucleus. We also assume that there is a range of Gaussian distributions with different σ (this range shifts towards larger σ as retinal eccentricity increases). As an effect of visual experience the weights of the synapses are modified by a Hebb rule [9]. Hebb's original rule, which states in conceptual terms that "neurons that fire together, wire together", can be written as $\dot{\mathbf{w}} = u(\mathbf{v})\mathbf{v}$, where $\mathbf{v}$ is the input vector $\mathbf{w}$ is the presynaptic weights vector, u is the postsynaptic response and $\dot{\mathbf{w}} = d\mathbf{w}/\mathbf{dt}$. In order for this dynamical system to actually converge, the weights have to be normalized. In fact, there is considerable experimental evidence that cortex employs normalization (cf. [38] and references therein).

2 The simplest Hebbian rule: the Oja flow

Mathematically, this requires a modified Hebbian rule. We consider here only the simplest among a large family of such rules, proposed by Oja [23]. Others (such as Independent Component Analysis, see later and [2, 10]), including biologically more realistic plasticity rules, would also be appropriate for our argument and possibly provide a better quantitative fit. Oja's equation $\dot{\mathbf{w}} = \gamma u(\mathbf{v})[\mathbf{v} - \mathbf{u}(\mathbf{v})\mathbf{w}]$ defines the change in presynaptic weights $\mathbf{w}$ where γ is the "learning rate" and the "output" u is assumed to depend on the "input" $\mathbf{v}$ as $u(\mathbf{v}) = \mathbf{w^T v}$. The equation follows from expanding to the first order the Hebb rule normalized to avoid divergence of the weights. Oja's version of Hebb's rule has been proven to converge to the top principal component of its input (technically to the eigenvector of the covariance of its inputs with the largest eigenvalue). Lateral inhibitory connections can enforce convergence of different neighboring neurons to several of the top eigenvectors ([23, 24] and see [30]). Our simulations with parameter values in the physiological range suggest that eigenvectors above the first three are almost always in the range of the noise. Because of this, we assume here that we can study the result of online Hebbian learning by studying

the properties of the top three eigenvectors of the covariance of the visual inputs to a cortical cell, seen through a Gaussian window.

In particular we consider the continuous version of the problem where images are transformed by the locally compact group of $2D$-translations. Notice that for small Gaussian apertures all motions are effectively translations (as confirmed by our simulations for natural images seen through a physiologically sized Gaussian). Thus, in this case, the tuning of each simple cell in V1 – given by the vector of its synaptic weights $\mathbf{w}$ – is predicted to converge to one of the top few eigenfunctions $\psi_n(x, y)$ of the following equation:

$$\int d\xi d\eta g(x, y) g(\xi, \eta) t^{\circledast}(\xi - x, \eta - y) \psi_n(\xi, \eta) = \nu_n \psi_n(x, y). \tag{1}$$

where the functions g are Gaussian distributions with the same, fixed width σ and $t^{\circledast}$ is the autocorrelation function of the input from the LGN. ψ_n is the eigenfunction and ν_n the associated eigenvalue. Equation (1), which depends on $t^{\circledast}$, defines a set of eigenfunctions parameterized by σ. We assume that the images generating the LGN signal $t(x, y)$ are natural images, with a power spectrum $\mathcal{F}t^{\circledast}(x) = 1/\omega^2$, where $\mathcal{F}$ is the Fourier transform [37]. In 1-D the solutions of equation 1 with this $t^{\circledast}$ are windowed Fourier transforms but for different σ they provide a very good approximation of Gabor wavelets for $n = 0, 1, 2$, since λ increase roughly proportionally to σ. An analytic solution for the specific input spectrum $\frac{1}{\omega^2}$ can be derived and will be given elsewhere. In 2D the known temporal high-pass properties of retinal processing (modeled as an imperfect high-pass, derivative-like operation in time) are compensated in the direction of motion by a local spatial average followed by a Difference of Gaussian (DOG) filter (see for instance [4]). Motion provides a selection mechanism that breaks the degeneracy of the 2D spectrum in the cortical equation (see Appendix).

3 Simulations

Our simulation pipeline consists of several filtering steps that mimic retinal processing, followed by a Gaussian mask which corresponds to the initial cortical cell receptive field, as shown in Figure 1. Values for the DoG filter were those suggested by [3]; the spatial lowpass filter has frequency response: $1/\sqrt{\omega_x^2 + \omega_y^2}$. The temporal derivative is performed using imbalanced weights $(-0.95, 1)$ so that the DC components is not zero. Each cells learns by extracting the principal components of a movie generated by a natural image patch undergoing a rigid translation. Each frame goes through the pipeline described here and is then fed to the unsupervised learning module (computing eigenvectors of the covariance). We used 40 natural images and 19 different Gaussian apertures for the simulations presented in this paper.

Simulations, suggested by a direct analysis of the equations show that, independently of the parameter values of the filtering, Gabor functions with modulation in the direction of motion (e.g. x), $G_n(x, y) \propto exp(-y^2/\sigma_{ny}^2 - x^2/\sigma_{nx}^2) \sin[(2\pi/\lambda_n)x]$ are approximate solutions of the equation. If for each σ only the first three eigenvectors are significant, then the set of solutions is well described well by a set of Gabor wavelets, that is a set of Gabor functions in which lambda is proportional to σ_x, which in turn is proportional to σ_y. These relations are captured in the ratio n_x/n_y (where $n_x = \sigma_x/\lambda$ and $n_y = \sigma_y/\lambda$) which was introduced to characterize tuning properties of simple cells

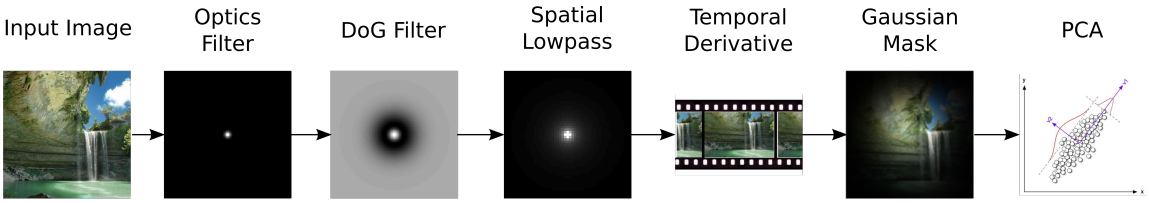

Figure 1: *Retinal processing pipeline used for V1 simulations. Though Gabor-like filters are obtained irrespectively of the presence or absence of any element of the pipeline the DoG filter is important in 1D and 2D for the emergence of actual Gabor wavelets with the correct dependence of λ on σ; the spatial low-pass filter together with the temporal derivative are necessary in our simulation to constrain λ to be proportional to σ.*

in V1 in the macaque [33]. It turns out that simple cells show Gabor-like tuning curves which are wavelet-like. Remarkably, the key parameter values are similar across three different species as shown by Figure 3 which includes, in addition to Ringach's, also Niell and Stryker's data on mouse V1 [22] and the original Palmer et al. experiments in cat cortex [12]. The theory of this paper, despite using the simplest Hebbian rule, seems to predicts the data in a satisfactory way. Equation (1) gives Gaussian eigenfunctions with no modulation, as well as with odd and even modulations, similar to data from simple cells. The general tuning is rather robust. In particular, we expect to find similar tuning if instead of natural images with a $\frac{1}{\omega^2}$ spectrum, the input from the retina is determined during the early stages of development by retinal waves [39].

Notice that our proposal does not necessarily require visual experience for the initial tuning to emerge during development: it is quite possible that a tuning originally discovered by evolution was eventually compiled into the genes. The theory however predicts that the *tuning is maintained and updated* by continuous visual experience (under the assumption of Hebbian plasticity). In particular, it predicts that tuning can be modified by disrupting normal visual experience. At the level of IT cortex, such a prediction is consistent with the rapid disruption of position and scale invariance induced by exposure to altered visual experience [19] as shown by simulations[11].

The original theory [30] posits that local invariance is obtained in complex cells by pooling the outputs of several simple cells in a way similar to "energy models". The wiring between a group of simple cells with the same orientation and a complex cell may develop according to a Hebbian trace rule[5]. Complex cells would thus inherit several of the properties of simple cells. Notice that a complex cell is invariant to translations in every direction even if its set of simple cells was "learned" while being exposed to motion in a specific direction. Thus the theory predicts the emergency of multiresolution analysis during development of V1 spanning a range of frequencies determined by a set of Gaussian distributions of synapses on dendritic trees with a range of σ which are assumed to be present at the beginning of visual development. More complex activity-dependent mechanisms than Oja's rule may automatically determine different sizes of receptive fields during development [40, 31]: the details of the rules operating during development are of course less important than the experimental confirmation of a key role of Hebbian rules in determining *and/or maintaining* the tuning of V1 cells.

A similar set of assumptions about invariance and Hebbian synapses leads to wavelets-of-wavelets at higher layers, representing local shifts in the 4-cube of x, y, scale, orientation learned at the level

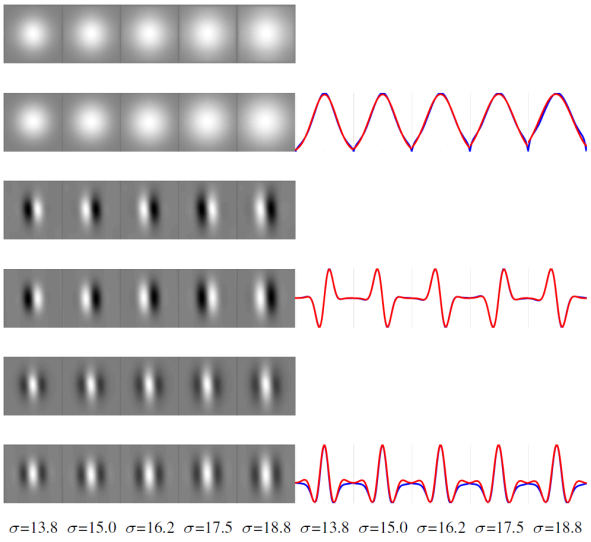

Figure 2: *Simulation results for V1 simple cells "learned" via PCA. Each "cell" receives as input all frames from* 40 *movies, each generated by a patch from a natural image undergoing a translation along the horizontal axis. A Gaussian filter with small sigma simulates the optics, a Difference of Gaussians filter and a spatial lowpass filter are applied to every frame to simulate retinal processing. Each frame is multiplied by a Gaussian mask to model a cell's initial distribution of input synapses on its dendritic tree. The weighted difference between subsequent frames is fed to the learning stage, to simulate an imperfect temporal derivative (the weights we used are (-0.95, 1.00)). Each cell "learns" its weight vector extracting the principal components of its input. On the left, for each row pair: the top row shows the best Gabor fit (least squares) and the bottom row shows the actual principal component vector; different columns represent different* σ *values for the Gaussian mask aperture. On the right we show 1D sections of the 2D tuning functions just described. The blue line is the learned function, red indicates the best least-squares fit to a Gabor wavelet, and green shows the difference (fitting error). The processing pipeline is described in the text. An orientation orthogonal to the direction of motion emerges.*

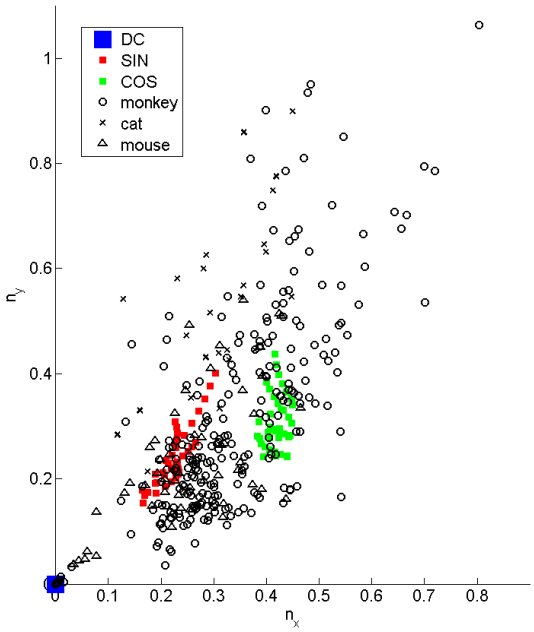

Figure 3: *This figure shows* $n_y = \frac{\sigma_y}{\lambda}$ *vs.* $n_x = \frac{\sigma_x}{\lambda}$ *for the modulated (x) and unmodulated (y) direction of the Gabor-like wavelet. Neurophysiology data from monkeys [33], cats [12], and mice [22] are reported together with our simulations (for* $n = 0(DC), 1(SIN), and 2(COS)$ *for which the eigenvectors had significant power). Notice that the range of* n_x *and* n_y *is between* 0.3 *and* 0.75 *across species. Usually* $n_y \geq n_x$ *– a robust finding in the theory. Simulated cells learn their weight vector according to the algorithm described in Figure 2. Note that* σ_x *and* σ_y *vary significantly across species and are not easy to obtain; Jones and Palmer [12] present two different methods to estimate them and report inconsistent results. Conversely* n_x *and* n_y*, as defined above, are dimensionless and consistent across different fitting methods. One of the predictions of the simple Oja rule assumed here is shown to be wrong by the data at a statistically significant level: the average* n_x *for the odd Gabor-like wavelets is smaller than for the even Gabor-like wavelets.*

of the simple cells in V1. Simulations show tuning that is qualitatively similar to physiology data in V2 and V4. A prediction that should be verifiable experimentally is that the tuning of cells in V2 corresponds to Gabor wavelets with a fixed relation between λ and σ in the four-dimensional cube of x, y, θ, s. Similar mechanisms, based on simple and complex cell modules can provide invariance to pose in face recognition; together with the Hebbian assumption, they may explain puzzling properties of neurons in one of the face patches recently found [6] in macaque IT [18].

4 Discussion

In summary, we study whether "universal" properties of simple cells in cortical V1 can be predicted from the hypothesis that the computational goal of the ventral stream is to learn via Hebbian synapses how objects transform – during and after development – in order to later compute for each image a "signature" vector which is invariant to geometric transformations. Taking into account the statistics of natural images, we derive that the solutions of an associated "cortical equation" are Gabor-like wavelets with parameter values that agree with the physiology data across different species. However, the data show that the prediction of a difference between odd and even wavelet-like tuning is incorrect. It is unclear at this point whether more realistic Hebbian rule than Oja's could overcome this disagreement with the data. Hebbian plasticity predicts the tuning of cells in V2, V4 and in the face patch AL, qualitatively in agreement with physiology data [6, 16]. It is important to notice that the emergence and maintenance of the tuning of simple cells is one of several predictions of the theory, whose goal is invariant recognition. The main result of the theory is the characterization of a class of systems for visual recognition that account for the architecture of the ventral stream and for several tuning and invariance properties of the neurons in different areas.

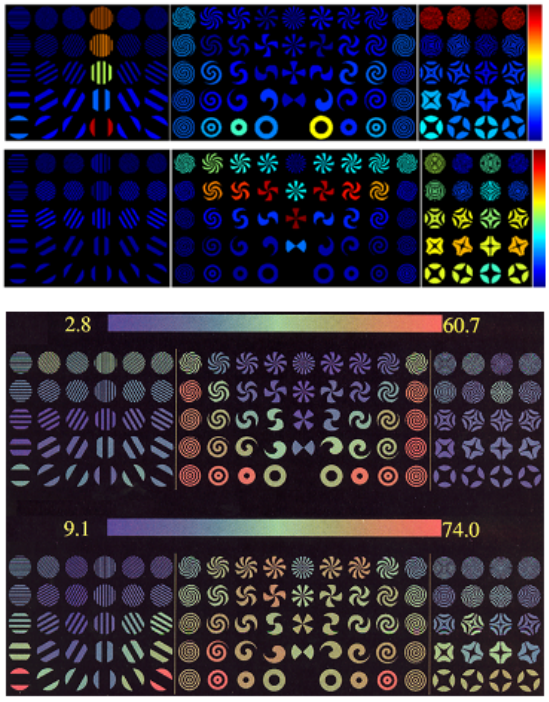

Figure 4: *At the next higher cortical level, similar Hebbian learning on the V1 output representation generates 4-dimensional wavelets (in x, y, θ, s, where s is scale). Here we computed the 3-dimensional wavelets – corresponding possibly to simple cells in V2 or V4 – that emerge if scale is kept constant. We show the responses of two model complex cells pooling such 3D wavelets (top) and of two real V4 cells [8] (bottom) to various stimuli used by Gallant [8]. Red/orange indicates a high response and blue/green indicates a low response. Note that we have not attempted to match particular model cells to real cells. We note that by varying only the orientation of a 3D higher-order wavelet, we are able to obtain a wide variety of selectivity patterns.*

Related architectures have been shown to perform well in computer vision recognition tasks and to mimic human performance in rapid categorization [35, 21, 13]. The results here are indirectly supported by Stevens' [36] symmetry argument showing that preserving shape invariance to rotation, translation and scale changes requires simple cells in V1 to perform a wavelet transform (Stevens also realized the significance of the Palmer and Ringach data and their "universality"). Similar indirect support can be found in Mallat's elegant mathematical theory of a scattering transform [20]. Independent Component Analysis(ICA) [2, 10], Sparse Coding (SC) [25] and similar unsupervised mechanisms [34, 40, 31, 26] describe plasticity rules similar to the basic Hebbian rule used in this paper. They can generate Gabor-like receptive fields and they do not need the assumption of different sizes of Gaussian distributions of LGN synapses; however, the required biophysical mechanisms and circuitry can be rather complex and, more importantly, their motivation depends on sparsity, whose computational and evolutionary significance is unclear – unlike our assumption of invariant recognition. It is interesting that in this theory a high level computational goal – invariant recognition – determines rather directly low-level properties of sensory cortical neurons.

5 Appendix: symmetry breaking by motion

In 2D the cortical equation has degenerate solutions if the sequence of images is a random sequence. Typically an orientation will emerge at random for the eigenvector with n=1 with the orthogonal orientation being associated with n=2. However if the sequence portrays an images continuously shifted in one direction, then the eigenvectors as in Figure 1 with the same orientation, orthogonal to

the direction of motion. The mechanism through which motion breaks symmetry can be explained as follows. Consider a $2D$ image moving through time t , $I(x(t), y(t)) = I(\mathbf{x}(\mathbf{t}))$ filtered, as in pipeline of Fig. 1, by a spatial low-pass filter and a band-pass filter and call the output $f(\mathbf{x}(\mathbf{t}))$. Suppose now a temporal filter is done by a high-pass impulse response $h(t)$. For example, let $h(t) \sim \frac{d}{dt}$. We consider the effect of the time derivative over the translated signal, $\mathbf{x}(t) = \mathbf{x} - \mathbf{v}t$ where $\mathbf{v} \in \mathbb{R}^{\mathbf{2}}$ is the velocity vector

$$\frac{\mathrm{d}\, f(\mathbf{x}(\mathbf{t}))}{\mathrm{d}\, t} = \nabla f(\mathbf{x}(\mathbf{t})) \cdot \mathbf{v}. \tag{2}$$

If, for instance, the direction of motion is along the x axis with constant velocity, $\mathbf{v} = (\mathbf{v_x}, \mathbf{0})$, then eq. (2) become

$$\frac{\mathrm{d}\, f(\mathbf{x}(\mathbf{t}))}{\mathrm{d}\, t} = \frac{\partial f(\mathbf{x}(\mathbf{t}))}{\partial x} v_x,$$

or, in Fourier domain of spatial and temporal frequencies:

$$\widehat{f}(i\omega_t) = i v_x \omega_x \hat{f}. \tag{3}$$

Consider now an image I with a symmetric spectrum $1/(\sqrt{\omega_x^2 + \omega_y^2})$. Equation (3) shows that the effect of the time derivative is to break the radial symmetry of the spectrum in the direction of motion (depending on the value of v_x). Intuitively, spatial frequencies in the x direction are enhanced. Thus motion effectively selects a specific orientation since it enhances the frequencies orthogonal to the direction of motion in Equation (1).

Thus the theory suggests that motion effectively "selects" the direction of the Gabor-like function (see previous section) during the emergence and maintenance of a simple cell tuning. It turns out that in addition to orientation other features of the eigenvectors are shaped by motion during learning. This is shown by an equivalent simulation but in which the order of frames was scrambled before the time derivative stage. The receptive fields are still Gabor-like functions but lack the important property of having $\sigma_x \propto \lambda$.

References

[1] O. Abdel-Hamid, A. Mohamed, H. Jiang, and G. Penn. Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 4277–4280. IEEE, 2012.

[2] A. Bell and T. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Research*, pages 3327–3338, 1997.

[3] L. Croner and E. Kaplan. Receptive fields of p and m ganglion cells across the primate retina. *Vision research*, 35(1):7–24, 1995.

[4] Y. Dan, A. J. J., and R. C. Reid. Effcient Coding of Natural Scenes in the Lateral Geniculate Nucleus: Experimental Test of a Computational Theory. *The Journal of Neuroscience*, (16):3351 – 3362, 1996.

[5] P. Földiák. Learning invariance from transformation sequences. *Neural Computation*, 3(2):194–200, 1991.

[6] W. Freiwald and D. Tsao. Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. *Science*, 330(6005):845, 2010.

[7] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, Apr. 1980.

[8] J. Gallant, C. Connor, S. Rakshit, J. Lewis, and D. Van Essen. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, 76:2718–2739, 1996.

[9] D. O. Hebb. The organization of behaviour: A neuropsychological theory. 1949.

[10] A. Hyvrinen and E. Oja. Independent component analysis by general non-linear hebbian-like learning rules. *Signal Processing*, 64:301–313, 1998.

[11] L. Isik, J. Z. Leibo, and T. Poggio. Learning and disrupting invariance in visual recognition with a temporal association rule. *Frontiers in Computational Neuroscience*, 6, 2012.

[12] J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.

[13] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 2012.

[14] Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, pages 255–258, 1995.

[15] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[16] J. Z. Leibo, F. Anselmi, J. Mutch, A. F. Ebihara, W. A. Freiwald, and T. Poggio. View-invariance and mirror-symmetric tuning in a model of the macaque face-processing system. In *Computational and Systems Neuroscience (COSYNE)*, 2013.

[17] J. Z. Leibo, J. Mutch, F. Anselmi, W. Freiwald, and T. Poggio. Part III:View-invariance and mirror-symmetric tuning in the macaque face-processing network. *in preparation*, 2013.

[18] J. Z. Leibo, J. Mutch, and T. Poggio. How can cells in the anterior medial face patch be viewpoint invariant? *MIT-CSAIL-TR-2010-057, CBCL-293; Presented at COSYNE 2011, Salt Lake City*, 2011.

[19] N. Li and J. J. DiCarlo. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, 321(5895):1502–7, Sept. 2008.

[20] S. Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[21] J. Mutch and D. Lowe. Object class recognition and localization using sparse features with limited receptive fields. *International Journal of Computer Vision*, 80(1):45–57, 2008.

[22] C. Niell and M. Stryker. Highly selective receptive fields in mouse visual cortex. *Journal of Neuroscience*, 28(30):7520–7536, 2008.

[23] E. Oja. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3):267–273, 1982.

[24] E. Oja. Principal components, minor components, and linear neural networks. *Neural Networks*, 5(6):927–935, 1992.

[25] B. Olshausen et al. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

[26] B. A. Olshausen, C. F. Cadieu, and D. Warland. Learning real and complex overcomplete representations from the statistics of natural images. *SPIE Proceedings, Vol. 7446: Wavelets XIII, (V.K. Goyal, M. Papadakis, D. van de Ville, Eds.)*, 2009.

[27] N. Pinto, J. J. DiCarlo, and D. Cox. How far can you get with a modern face recognition test set using only simple features? In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2591–2598. IEEE, 2009.

[28] T. Poggio and S. Edelman. A network that learns to recognize three-dimensional objects. *Nature*, 343(6255):263–266, 1990.

[29] T. Poggio, J. Mutch, F. Anselmi, J. Z. Leibo, L. Rosasco, and A. Tacchetti. The computational magic of the ventral stream: sketch of a theory (and why some deep architectures work). Technical Report MIT-CSAIL-TR-2012-035, MIT Computer Science and Artificial Intelligence Laboratory, 2012. Previously released in Nature Precedings, 2011.

[30] T. Poggio, J. Mutch, F. Anselmi, J. Z. Leibo, L. Rosasco, and A. Tacchetti. Invariances determine the hierarchical architecture and the tuning properties of the ventral stream. Technical Report available online, MIT CBCL, 2013. Previously released as MIT-CSAIL-TR-2012-035, 2012 and in Nature Precedings, 2011.

[31] M. Rehn, T., and F. T. A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *J. Comput. Neurosci.*, 22(2):135–146, 2007.

[32] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025, Nov. 1999.

[33] D. Ringach. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, 88(1):455–463, 2002.

[34] A. M. Saxe, M. Bhand, R. Mudur, B. Suresh, and A. Y. Ng. Unsupervised learning models of primary cortical receptive fields and receptive field plasticity. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 1971–1979. 2011.

[35] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust Object Recognition with Cortex-Like Mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(3):411–426, 2007.

[36] C. F. Stevens. Preserving properties of object shape by computations in primary visual cortex. *PNAS*, 101(11):15524–15529, 2004.

[37] A. Torralba and A. Oliva. Statistics of natural image categories. In *Network: Computation in Neural Systems*, pages 391–412, 2003.

[38] G. G. Turrigiano and S. B. Nelson. Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, 5(2):97–107, 2004.

[39] R. Wong, M. Meister, and C. Shatz. Transient period of correlated bursting activity during development of the mammalian retina. *Neuron*, 11(5):923–938, 1993.

[40] J. Zylberberg, J. T. Murphy, and M. R. DeWeese. A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of v1 simple cell receptive fields. *PLoS Comput Biol*, 7(10):135–146, 10 2011.