Visual Cortex and Deep Networks

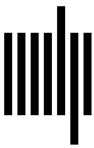
Learning Invariant Representations

By: Tomaso A. Poggio, Fabio Anselmi

Citation:

Visual Cortex and Deep Networks: Learning Invariant Representations
By: Tomaso A. Poggio, Fabio Anselmi
DOI: 10.7551/mitpress/10177.001.0001
ISBN (electronic): 9780262336710
Publisher: The MIT Press
Published: 2016

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.



The MIT Press

2 Biophysical Mechanisms of Invariance

Unsupervised Learning, Tuning, and Pooling

2.1 A Single-Cell Model of Simple and Complex Cells

At least two possible biophysical models for the Hubel-Wiesel module are implied by theory. The first is the original HW model of simple cells feeding into a complex cell. i-theory proposes the ideal computation of a cumulative distribution function (cdf), in which case the nonlinearity at the output of the simple cells is a threshold. A complex cell summing the outputs of a set of simple cells would then represent a bin of the histogram. A different complex cell in the same position pooling a set of similar simple cells with a different threshold would represent another bin of the histogram. Another possibility is that the nonlinearity at the output of the simple cells is approximately a square (or any power or combination of powers). In this case the complex cell pooling simple cells with the same nonlinearity would represent a moment of the distribution, including the linear average. Note that in this case some of the complex cells would be linear and would be classified by neurophysiologists using the standard criteria as simple. The nonlinear transformation at the output of the simple cells may correspond to the spiking mechanism in populations of cells (see references in [22]). The fact that a nonlinear biophysical mechanism can be approximated locally by a polynomial and thus can be used to compute combinations of moments has been noted several times [48, 49]. This is also the reason for the similar performance of models such as HMAX and the Freeman-Simoncelli model [48]. The nonlinearity in our model neurons (max or threshold) is well approximated by a few terms in a Taylor series approximation. As an aside, it should be emphasized that the distributions computed in our theory by complex cells do not have anything to do with the classical statistical interpretation of images and textures. The distributions do not come from the statistics of images but from the action of a group, and thus are independent of the nature of the images and whether or not they are textures.

The second biophysical model for the HW module that implements the computation required by i-theory consists of a single cell where dendritic branches play the role of simple cells (each branch containing a set of synapses with

weights providing, for instance, Gabor-like tuning of the dendritic branch) with inputs from the lateral geniculate nucleus (LGN). Active properties of the dendritic membrane distal to the soma provide separate thresholdlike nonlinearities for each branch separately, while the soma sums the contributions for all the branches. This model would solve the puzzle that so far there seems to be no morphological difference between pyramidal cells classified as simple versus complex by physiologists.

It is interesting that i-theory is robust with respect to the nonlinearity from simple to complex cells. We conjecture that almost any set of nontrivial nonlinearities will work. The argument rests on the fact that a set of different complex cells pooling from the same simple cells should compute the cumulative distribution or equivalently its moments or combinations of moments (each combination is a specific nonlinearity). Any nonlinearity will provide invariance if the nonlinearity does not change with time and is the same for all the simple cells pooled by the same complex cells. A sufficient number of different nonlinearities, each corresponding to a complex cell, can provide appropriate selectivity, assuming that each nonlinearity can be represented by a truncated power series and that the associated complex cells provide therefore a sufficient number of linearly independent combinations of moments.

2.2 Learning the Wiring in the Single-Cell Model

A simple model of how the wiring between a group of simple cells with the same tuning (e.g., representing the same eigenvector, having the same orientation) and a complex cell may develop is to invoke a Hebbian trace rule [49]. In a first phase, complex cells may have subunits with different selectivities (e.g., orientations), for instance, because natural images are rotation invariant and thus eigenvectors with different orientations are degenerate. In a second, plastic phase, subunits that are not active when the majority of subunits are active will be pruned out according to a Földiák-like rule [49]. Földiák's original trace rule says that the weight of a synapse between an input cell and an output cell is strengthened proportionately to the input activity and the trace or average of recent output activity at time t, where the dependence of the trace on previous activity decays over time with a certain decay constant.

2.3 Hebb Synapses and Principal Components

i-theory provides the following algorithm for learning the relevant invariances during unsupervised visual experience: storing sequences of images for each of a few objects (templates) while transforming (e.g., translating, rotating, and looming). We prove that in this way invariant hierarchical architectures

can be learned from unsupervised visual experience. Such architectures represent a significant extension, beyond simple translation invariance and beyond hardwired connectivity, of models of the ventral stream such as Fukushima's Neocognitron [2] and HMAX [47, 50] as well as deep neural networks called convolutional networks [34, 51] and related models [52–59].

In biological terms, the sequence of transformations of one template would correspond to a set of simple cells, each storing in its tuning a frame of the sequence. In a second learning step a complex cell would be wired to those simple cells. However, the idea of direct storage of sequences of images or image patches in the tuning of a set of V1 cells by exposure to a single object transformation is biologically rather implausible. Since Hebbian-like synapses are known to exist in the visual cortex, a biologically more plausible hypothesis is that synapses incrementally change over time as an effect of the visual inputs, that is, over many sequences of images resulting from transformations of objects (e.g., templates). The question is whether such a mechanism is compatible with i-theory, and how.

We explore this question for V1 in a simplified setup that can be extended to other areas. We assume the following:

1. The synapses between LGN inputs and (immature) simple cells are Hebbian, and in particular their dynamics follow Oja's flow. In this case, the synaptic weights will converge to the eigenvector with the largest eigenvalue of the covariance of the input images.
2. The position and size of the untuned simple cells are set during development according to an inverted pyramidal lattice (see figure 3.2 in the next chapter). The key point is that the size of the Gaussian spread of the synaptic inputs and the positions of the ensemble of simple cells are assumed to be set independently of visual experience.

In summary, we assume that the neural equivalent of the 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-like 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, which increase with retinal eccentricity. As an effect of visual experience, the weights of the synapses are modified by a Hebb rule [60]. Hebb's original rule can be written as

$$\delta \mathbf{w}_n = \alpha y(\mathbf{x}_n)\mathbf{x}_n, \tag{2.1}$$

where $\mathbf{x}_n$ is the input vector, $\mathbf{w}_n$ is the presynaptic weights vector, y is the postsynaptic response $y(\mathbf{x}_n) = y_n = \mathbf{w}_n^T \mathbf{x}_n$, and α is a positive constant called the learning rate. In order for this dynamical system to actually converge, the weights have to be normalized. In fact, there is considerable experimental evidence that the cortex employs normalization [61]. Hebb's rule, appropriately modified with a normalization factor, turns out to be an online algorithm to compute the eigenvectors of the covariance matrix of a set of input vectors. In this case it is called Oja's flow. Oja's rule [62, 63] defines the change in presynaptic weights $\mathbf{w}_n$ given the output response y of a neuron to its inputs to be

$$\Delta \mathbf{w}_n = \mathbf{w}_{n+1} - \mathbf{w}_n = \alpha y_n (\mathbf{x}_n - y_n \mathbf{w}_n). \tag{2.2}$$

The equation follows from expanding to the first-order Hebb rule normalized to avoid divergence of the weights.

Box 2.1
Principal Components

Principal component analysis (PCA) is a statistical procedure that allows identification of the principal directions in which the data vary. Those principal directions of variations are called principal components and are found by calculating the eigenvectors and eigenvalues of the data covariance matrix, defined as the matrix product between the data matrix and its transpose.

Since the Oja flow converges to the eigenvector of the covariance matrix of the $\mathbf{x}_n$ that has the largest eigenvalue, we are led to analyze the spectral properties of the inputs to simple cells and to study whether the eigenvectors (also called principal components (PCA); see box 2.1) can be used by the HW algorithm and in particular whether they satisfy the selectivity and invariance theorems.

Alternatives to Oja's rule can and should be considered [64, 65]. Note that a relatively small change in the Oja equation gives an online algorithm for computing independent components instead of PCs (see [66]). Which kind of plasticity is closer to biology remains an open question.

2.4 Spectral Theory and Pooling

Consider stage 1 of figure 1.5 in the previous chapter, which is retinotopic, and in particular the case of simple cells in V1. From assumption 1 in section 2.3, the lattice in space (x, y) and scale (s) of an immature simple cell is set during development of the organism (s is related to the variance of the Gaussian

envelope of the dendritic tree of the immature cell). Assume that all the simple cells are exposed while in a plastic state to a possibly large set of images $T = (t^1, \ldots, t^K)$. A specific cell at a certain position in x, y, s is exposed to the set of transformed templates $g_* T$, where g_* corresponds to the specific translation and scale that transforms the zero cell to the chosen neuron in the lattice. The associated covariance matrix is $C = g_* T T^T g_*^T$. Calling ϕ the first eigenvector of TT^T, we have that $g_* \phi$ is the first eigenvector of C. Thus it is possible to choose ϕ as a new template, and pooling over corresponding PCs across different cells (reached by different transformations from the zero cell) is equivalent to pooling over ϕ and its transformations. Both the invariance and selectivity theorems are valid. Empirically, we find [13] that the covariance of natural images yields eigenvectors that are Gabor-like wavelets with a random orientation for each size receptive field (second hypothesis in section 2.3). The random orientation follows from the argument together with the fact that the covariance of natural images is approximately rotation invariant. The Gabor-like shape can be qualitatively explained in terms of translation invariance of the correlation matrix associated with a set of natural images (and their approximate scale invariance, which corresponds to a $1/f$ spectrum; see also [67, 68]). Thus the Oja rule acting on natural images provides equivalent templates that are Gabor-like and are the optimal ones according to the theory of chapter 1 and appendix section A.2.5.

Consider now the nonretinotopic stage 2 of figure 1.5, in which transformations are not in scale or position, such as the transformation induced by a rotation of a face. Assume that a simple cell is exposed to all transformations g_i (g_i is a group element of the finite group G) of each set $T = (t_1, \cdots, t_K)$ of K templates. The cell is thus exposed to a set of images (columns of X) $X = (g_1 T, \cdots, g_{|G|} T)$. For the sake of this example, assume that G is a discrete group. Then the covariance matrix determining the Oja's flow is

$$C = XX^T = \sum_{i=1}^{|G|} g_i T T^T g_i^T. \tag{2.3}$$

It is evident that if ϕ is an eigenvector of C, then $g_i \phi$ is also an eigenvector with the same eigenvalue (for details on how receptive fields look in V1 and higher layers, see [11] and [69–71]). Consider, for example, G to be the discrete rotation group in the plane; then all the (discrete) rotations of an eigenvector are also eigenvectors. The Oja rule will converge to the eigenvectors with the top eigenvalue and thus to the subspace spanned by them. It can be shown that L^2 pooling over the PCs with the same eigenvalues represented by different simple cells is then equivalent to L^2 pooling over transformations (see appendix

section A.4). This argument can be formalized in the following variation of the pooling step in the HW algorithm.

2.4.1 Spectral Pooling Proposition

Suppose that Λ_k is the matrix corresponding to the group transformations of template t^k (each column is a transformation of the template). Consider the set of eigenvectors of the associated covariance matrix. Because of the preceding argument, $\langle g_m I, \phi_k \rangle = \langle I, \phi_p \rangle$, where $g_m^{-1}\phi_k = \phi_p$. In particular, it is easy to check that the first two moments can be computed either by pooling over eigenvectors or pooling over the transformations of the template (eq. (2.4)). The argument is still valid if the pooling is over part of the eigenvectors of $\Lambda = \cup_{i=1}^{K} \Lambda_i$. The first two moments, which can be thought of as part of an invariant signature, are

$$\mu_1 = \sum_i \langle I, \phi_i \rangle \,, \quad \mu_2(I) = \sum_i \langle I, \phi_i \rangle^2 \,; \tag{2.4}$$

to achieve invariance a complex cell can pool with a linear or quadratic nonlinearity over the eigenvectors of Λ_k instead of over the transformations of the template.

2.5 Tuning of Simple Cells

The theorems of chapter 1 on the HW module imply that the templates and therefore the tuning of the simple cells can be the image of any object. At higher levels in the hierarchy, the templates are neuroimages—patterns of neural activity—induced by actual images in the visual field. Section 2.4, however, offers a biologically more plausible way to learn the templates from unsupervised visual experience, via Hebbian plasticity. In chapter 3 we discuss predictions following from these assumptions for the tuning of neurons in the various areas of the ventral stream.

Background and Bibliography

This part of the book describes work in our group. Its main sources are the following technical reports: [9, 11].

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.

Library of Congress Cataloging-in-Publication Data

Names: Poggio, Tomaso, author. | Anselmi, Fabio, author.
Title: Visual cortex and deep networks : learning invariant representations / Tomaso A. Poggio and Fabio Anselmi.
Description: Cambridge, MA : MIT Press, [2016] | Series: Computational neuroscience | Includes bibliographical references and index.
Identifiers: LCCN 2016005774 | ISBN 9780262034722 (hardcover : alk. paper)
Subjects: LCSH: Visual cortex. | Vision. | Neural networks (Neurobiology) | Perceptual learning. | Computational neuroscience.
Classification: LCC QP383.15 .P64 2016 | DDC 612.8—dc23 LC record available at http://lccn.loc.gov/2016005774

10 9 8 7 6 5 4 3 2 1