A HOPFIELD RECURRENT NEURAL NETWORK TRAINED ON NATURAL IMAGES PERFORMS STATE-OF-THE-ART IMAGE COMPRESSION

Christopher Hillar, Ram Mehta, Kilian Koepsell

Redwood Center for Theoretical Neuroscience
University of California, Berkeley

ABSTRACT

The Hopfield network is a well-known model of memory and collective processing in networks of abstract neurons, but it has been dismissed for use in signal processing because of its small pattern capacity, difficulty to train, and lack of practical applications. In the last few years, however, it has been demonstrated that exponential storage is possible for special classes of patterns and network connectivity structures. Over the same time period, advances in training large-scale networks have also appeared. Here, we train Hopfield networks on discretizations of grayscale digital photographs using a learning technique called minimum probability flow (MPF). After training, we demonstrate that these networks have exponential memory capacity, allowing them to perform state-of-the-art image compression in the high quality regime. Our findings suggest that the local structure of images is remarkably well-modeled by a binary recurrent neural network.

Index Terms— image compression, Hopfield network, Ising model, recurrent neural network, probability flow, JPEG

1. INTRODUCTION

Hopfield networks [1] are classical models of memory and collective processing in networks of abstract McCulloch-Pitts [2] neurons, but they have not been widely used in signal processing (although see [3]) as they usually have small memory capacity (scaling linearly in the number of neurons) and are challenging to train, especially on noisy data. Recently, however, it has been shown that exponential storage in Hopfield [4] (see also Fig. 2) and Hopfield-like [5, 6, 7, 8] networks is possible for special classes of patterns and connectivity structures. Additionally, training of large networks is now tractable [9], due to advances in statistical estimation [10].

Moreover, several studies in computer vision [11], retinal neuroscience [12], and even commercial quantum computation [13] have pointed to the importance and ubiquity of the underlying discrete probabilistic model in the Hopfield network: the Lenz–Ising model of statistical physics [14]. Additionally, "deep network" architectures, which have similar underlying models of data, have made a resurgence in the fields of machine learning [15] and image modeling [16].

We present a simple, efficient, high-quality compression scheme for digital images using discrete Hopfield networks trained on natural images. Our method performs $4\times$ compression (v.s. PNG originals) at high quality on two standard 512×512 grayscale images in computer vision (Figs. 4, 5), matching the corresponding coding cost of the JPEG algorithm [17]. Interestingly, our method has smaller coding cost compared to JPEG when compressing images with added noise. For instance, our scheme outperforms JPEG by 10% when compressing low additive Gaussian white noise ($\sigma \approx 6$; 2% of dynamic range) versions of these images. The model is also easy to train ($<$ 10 minutes on a standard desktop) and requires $<$ 17MBs of free space to store the $65,535$ (structure averages of) network memories that code discretized 4×4 grayscale digital image patches (Fig. 2).

In the next section, we review Hopfield networks, including training and capacity. The following section explains standard methods for image compression, and then our novel algorithm is outlined in detail in Section 3.3 (and Fig. 3).

2. BACKGROUND

2.1. Hopfield auto-associative pattern memory

We first define the underlying probabilistic model of data in the Hopfield network. This is the non-ferromagnetic[1] *Lenz–Ising model* [14] from statistical physics, more generally called a *Markov random field* in the machine learning literature, and the underlying probability distribution of a fully observable *Boltzmann machine* [18] in artificial intelligence. This discrete probability distribution has as *states* all length n column vectors of 0s and 1s, with the probability $p_{\mathbf{x}}$ of a particular state $\mathbf{x} = (x_1, \ldots, x_n) \in \{0,1\}^n$ given by:

$$p_{\mathbf{x}} = \frac{1}{Z} \exp\left(\sum_{i<j} J_{ij} x_i x_j - \sum_i \theta_i x_i\right) = \frac{1}{Z} \exp\left(-E_{\mathbf{x}}\right), \tag{1}$$

Research funded, in part, by NSF grant IIS-0917342 (CH and KK). Email: chillar@berkeley.edu, ram.mehta@gmail.com, kilian@berkeley.edu.

[1] In the literature, "non-ferromagnetic" (also "spin-glass") means that all-to-all and positive or negative connectivity is allowed in the network, unlike the classical "nearest-neighbor" connectivity of the Lenz–Ising model [14].

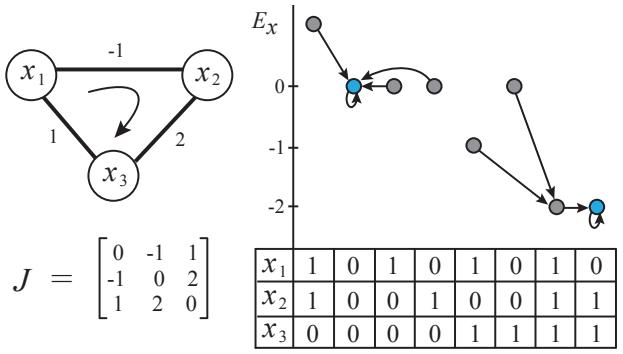

Fig. 1. **Small Hopfield network**. A 3-node Hopfield network with coupling matrix J and zero threshold vector θ. A state vector $\mathbf{x} = (x_1, x_2, x_3)^\top$ has energy $E_\mathbf{x}$ as labeled on the y-axis of the diagram. Arrows represent one iteration of the network dynamics; i.e., x_1, x_2, and x_3 are updated by Eq. (3) in the order of the clockwise arrow. Resulting memories / fixed-points $\mathbf{x}^*$ are indicated by blue circles.

in which $J \in \mathbb{R}^{n \times n}$ is a real symmetric matrix (the *coupling matrix*), the column vector $\theta \in \mathbb{R}^n$ is a bias or *threshold* term, and $Z = \sum_\mathbf{x} \exp(-E_\mathbf{x})$ is the *partition function*. The *energy* of a state $\mathbf{x}$ is given by the quadratic Hamiltonian:

$$E_\mathbf{x} = -\frac{1}{2}\mathbf{x}^\top J\mathbf{x} + \theta^\top \mathbf{x}. \tag{2}$$

Intuitively, we are to think of matrix entry J_{ij} as the weight of the "statistical coupling" between binary variables $\{x_i, x_j\}$.

A *Hopfield network* [1] is a recurrent network of binary nodes (representing spiking neurons) with deterministic dynamics that act to locally minimize an energy given by Eq. (2). Formally, the network on n nodes $\{1, \ldots, n\}$ consists of a symmetric coupling matrix $J \in \mathbb{R}^{n \times n}$ with zero diagonal and a threshold vector $\theta \in \mathbb{R}^n$. (See e.g. Fig. 1.) A *dynamics update* of state $\mathbf{x}$ consists of replacing each x_i in $\mathbf{x}$ with the value (in consecutive order starting with $i = 1$):

$$x_i = \begin{cases} 1 & \text{if } \sum_{j \neq i} J_{ij} x_j > \theta_i, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Update Eq. (3) is inspired by computations in neurons [19, 2].

A fundamental property of Hopfield networks is that asynchronous dynamics updates, Eq. (3), do not increase energy. Thus, after a finite (and usually small) number of updates, each initial state $\mathbf{x}$ converges to a *fixed-point* $\mathbf{x}^*$ (also called *stable-point* or *memory*) of the dynamics. Intuitively, we may interpret the dynamics as an inference technique, producing the most probable nearby memory given a noisy version.

2.2. Training Hopfield networks

A basic problem is to construct Hopfield networks with a given dataset $\mathcal{D}$ of binary patterns as memories. Such networks are useful for denoising and retrieval since corrupted versions of patterns in $\mathcal{D}$ will converge through the dynamics to the originals. In [1], Hopfield defined a learning rule that stores $n/(4 \log n)$ patterns without errors in an n-node network [20, 21], and since then improved methods to fit Hopfield networks have been developed (e.g., [22]).

To estimate Hopfield network parameters, we use the recently discovered *minimum probability flow* (MPF) technique [10] for fitting parameterized distributions that avoids computation with the partition function Z. Applied to the context of a Hopfield network / Lenz–Ising model, Eqn. (1), the *minimum probability flow* (MPF) objective function [10, 9] is:

$$K_\mathcal{D}(J, \theta) = \sum_{\mathbf{x} \in \mathcal{D}} \sum_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} \exp\left(\frac{E_\mathbf{x} - E_{\mathbf{x}'}}{2}\right). \tag{4}$$

Here, the *neighborhood* $\mathcal{N}(\mathbf{x})$ of $\mathbf{x}$ is defined as those binary vectors which are Hamming distance 1 away from $\mathbf{x}$ (i.e., those $\mathbf{x}'$ with exactly one bit different from $\mathbf{x}$).

When compared with classical techniques for Hopfield pattern storage, minimizing the MPF objective function, Eq. (4), provides superior efficiency and generalization; and, more surprisingly, allows for the storage of patterns from (unlabeled) highly corrupted / noisy training samples [9].

2.3. Exponential pattern capacity

Independent of the method to fit Hopfield networks, arguments of Cover [23] can be used to show that the number of generic (or "randomly generated") patterns robustly storable in a Hopfield network with n nodes is at most $2n$. Here, "robustly stored" means that the dynamics can recover the pattern even if a fixed, positive fraction of its bits are changed.

Nonetheless, theoretical and experimental evidence suggest that Hopfield networks usually have exponentially many memories (fixed-points of the dynamics). For instance, choosing couplings randomly from a normal distribution produces exponentially many fixed-points asymptotically [24]. Although a generic Hopfield network has exponential capacity, its basins of attraction are shallow and difficult to predetermine from the network, leading many researchers to speculate that such *spurious minima* are to be avoided.

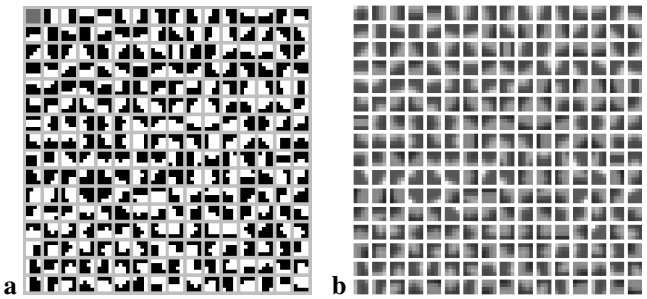

Fig. 2. **ON/OFF Hopfield network memories.** **a**) Top occurring $16 \times 16 = 256$ (of 65535 total) 4×4 ON/OFF 32-bit memories ordered more likely top-bottom, left-right. White pixels represent (ON,OFF) = (1,0); black, (0,1); gray, (0,0). **b**) Average grayscale normalized patch converging to the corresponding memory in **a**.

A surprising recent finding [4], however, is that special connectivity structures can create networks with robust memories in an exponential number of useful combinatorial configurations (e.g. cliques in graphs), opening up new possibilities. In fact, as demonstrated in Section 3.3, continuous natural images appear to have an exponential discrete structure (Fig. 2) that can be well-captured (Figs. 4, 5) with a Hopfield network that self-organizes its weights using MPF learning.

3. DIGITAL IMAGE CODING AND COMPRESSION

We explain standard strategies for image compression and then describe our method to use Hopfield networks.

3.1. Linear methods

Standard Fourier and wavelet-based methods, such as JPEG which uses the *discrete cosine transform* (DCT), first mean-zero an image patch (usually 8×8 pixels) and then code it with increasing numbers of (quantized) linear transform coefficients, much like *principal components analysis* (PCA) is used for dimensionality reduction in data analysis. There are also modern variants which do more complicated operations in the frequency domain [27]. Although these algorithms usually operate on non-overlapping patches of an image and are decades old, the high quality regimes of e.g. JPEG offer state-of-the-art compression. These schemes get expressive power from linear algebra – modeling a patch as a linear combination of columns of a real matrix (e.g. DCT matrix). We have not yet compared our work to wavelet-based JPEG 2000.

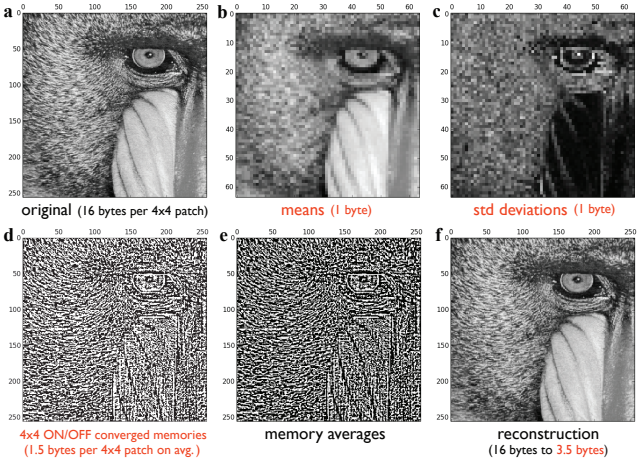

Fig. 3. **Hopfield neural network image compression algorithm**. **a**) 256×256 portion of 8-bit grayscale "baboon" PNG, **b**) means of each 4×4 non-overlapping continuous patch in **a**, **c**) standard deviations of these patches, **d**) replacement of each patch with its network converged ON/OFF discretization (Fig. 2a), **e**) as in **d** but with memory averages (Fig. 2b), **f**) reconstruction by adjusting the memory averages to have means **b** and standard deviations **c**.

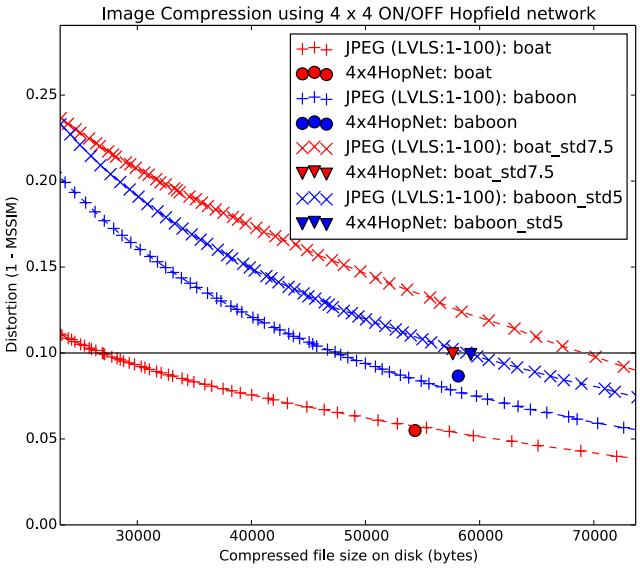

Fig. 4. **Rate-distortion performance** of image compression with a Hopfield network trained on discretized 4×4 natural image patches. The "+" / "x"s are JPEG codings of a standard 512×512 pixel, 8-bit image (*boat*, *baboon* / $\sigma = 7.5$, $\sigma = 5$ additive Gaussian white noisy versions). Circles / triangles indicate the file size (in bytes on disk) and reconstruction error (1 - MSSIM) of a 4×4 ON/OFF Hopfield network coding of these novel images. The (≤ 1) *mean structural similarity index* MSSIM [25] is thought to capture human perceptual image quality [26]. Black line is near-perceptual indistinguishability.

3.2. Unsupervised feature learning

There are several methods which try to leverage modeling the structure in natural images; e.g., *independent component analysis* [28, 29, 30], *sparse coding* [31]. Most of these techniques utilize "codebooks" of features, typically learned unsupervised over natural image datasets (usually by optimizing reconstruction error). These methods also code image patches as a linear combination of continuous structures. It is difficult to measure our performance against these methods since rarely is the rate expressed as bytes on disk. Instead, coding cost is in terms of the number of coefficients used in a coding.

3.3. High quality image coding with a Hopfield network

We map continuous image patches to binary vectors, inspired by the response properties of ON/OFF mammalian retinal ganglion cells [32]. Given a grayscale 4×4 patch, we remove the mean from each pixel and then normalize the patch's variance to be 1. We call such a patch *normalized*. Next, we partition the (mean-zero) intensity spectrum of the patch into three intervals and discretize pixel intensities accordingly.

A single pixel intensity is thus mapped onto two Hopfield neurons (one "ON" and one "OFF") as follows: if the pixel intensity is in the lowest interval, then only the OFF neuron fires; if the pixel intensity is in the middle interval, then neither neuron fires; and if the pixel intensity is in the highest interval, then only the ON neuron fires. In this way, we can

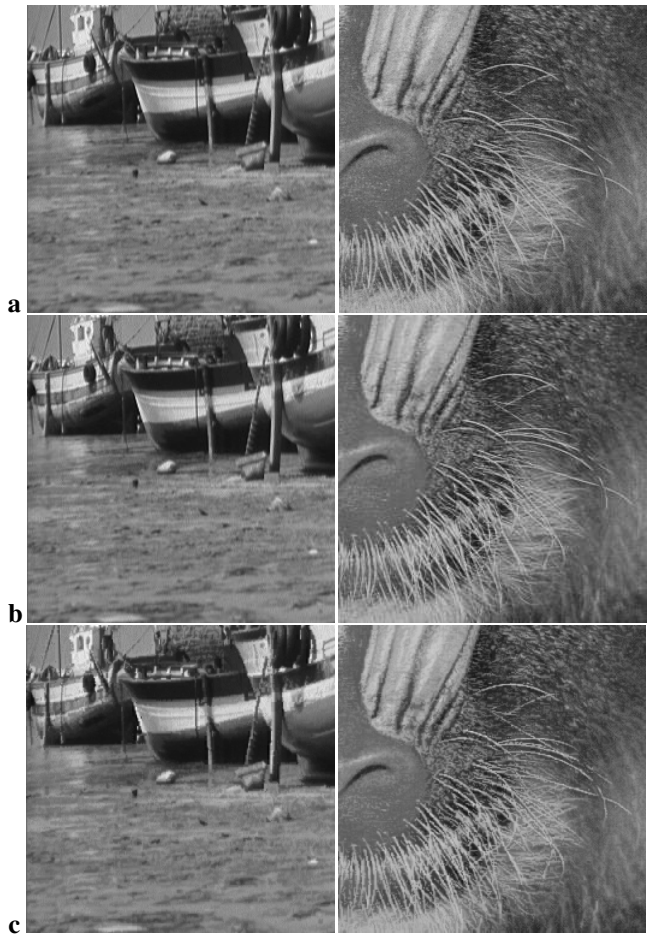

Fig. 5. 256×256 portions of images coded in Fig. 4. **a**) Originals; **b**) JPEG LVL (84) Boat: BYTES (57K), MSSIM (.95), PSNR (37), JPEG LVL (61) Baboon: BYTES (53K), MSSIM (.91), PSNR (29); **c**) ON/OFF Hopfield network Boat: BYTES (54K), MSSIM (.95), PSNR (33), Baboon: BYTES (58K), MSSIM (.91), PSNR (27).

convert any 4×4 grayscale image patch into a 32-bit binary vector of abstract ON and OFF neurons.

A collection of $3,000,000$ 4×4 natural image patches were chosen randomly from the van Hateren natural image database [30], and a Hopfield network with $n = 32$ nodes was trained using MPF parameter estimation on discretizations of these patches. Training time on a standard workstation computer running Mac OS X with 16GB RAM is < 10 minutes.

After training, we examined memories in the Hopfield network by collecting (over natural images) converged discretized ON/OFF patterns. We found that the dynamics collapses millions of ON/OFF binary activity patterns into one of $65,535$ memories[2], the most likely occurring 256 of which are displayed in Fig. 2a. For each of these 32-bit patterns, we also computed the average of normalized continuous patches converging to it; see Fig. 2b. The *entropy* H of 4×4 natural image patches after such a discretization is $H \approx 12.3$ bits (versus $H \approx 13.2$ when not applying the dynamics – although in this case we need at least 1GB on disk to store the more than 4 million binary patterns and continuous averages).

To compress a novel digital image, we partition it into non-overlapping 4×4 patches. We then normalize, discretize, and converge each patch to obtain an ON/OFF 32-bit codeword (which we Huffman encode), saving the means and variances as lossless PNG ("Portable Network Graphics") images. To reconstruct an image, we simply replace each binary patch code (Fig. 2a) with its corresponding continuous average (Fig. 2b) and then restore means and variances (see Fig. 3 for an example). Remarkably, this scheme performs state-of-the-art high quality compression (Figs. 4, 5), even though the model is not explicitly minimizing reconstruction error.

[2]Interestingly, these $2^{16} - 1$ converged patterns consist of all 4×4 ON/OFF patterns without (ON,OFF) $= (0, 0)$ pixels (except for the zero patch), but excluding all-ON and all-OFF configurations; see Fig. 2a.

4. SUMMARY OF RESULTS

On two standard images (see Fig. 4), we achieve an average coding cost on disk BYTES (56K), perceptual reconstruction quality MSSIM (.93), and peak signal-to-noise PSNR (30). Average bytes for the PNG originals is 210K. For each image, we determined the JPEG coding level with the same MSSIM score as the Hopfield reconstruction (see Fig. 5). These two JPEG codings averaged a cost BYTES (55K) and PSNR (33). With additive Gaussian white noise versions of these images (boat, baboon) having standard deviations $\sigma = (7.5, 5)$, our scheme achieves a coding cost BYTES (58K, 59K), MSSIM (.90, .90), and PSNR (30, 27); while the JPEG cost for this same MSSIM is BYTES (70K, 60K) with PSNR (33, 29).

5. REFERENCES

[1] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *PNAS*, vol. 79, no. 8, pp. 2554–2558, 1982.

[2] W.S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of mathematical biology*, vol. 5, no. 4, pp. 115–133, 1943.

[3] N. Martine and T. Jean-Bernard, "Neural approach for TV image compression using a Hopfield type network," in *Advances in Neural Information Processing Systems 1*, D.S. Touretzky, Ed., pp. 264–271. 1989.

[4] C. Hillar and N. M. Tran, "Robust exponential memory in hopfield networks," *ArXiv e-prints: nlin.AO 1411.4625*, 2014.

[5] V. Gripon and C. Berrou, "Sparse neural networks with large learning diversity," *IEEE Trans. Neural Networks*, vol. 22, no. 7, pp. 1087–1096, 2011.

[6] K.R. Kumar, A.H. Salavati, and A. Shokrollahi, "Exponential pattern retrieval capacity with non-binary associative memory," in *IEEE Information Theory Workshop (ITW)*, 2011, pp. 80–84.

[7] C. Curto, V. Itskov, K. Morrison, Z. Roth, and J.L. Walker, "Combinatorial neural codes from a mathematical coding theory perspective," *Neural computation*, vol. 25, no. 7, pp. 1891–1925, 2013.

[8] A Karbasi, A Salavati, and A Shokrollahi, "Iterative learning and denoising in convolutional neural associative memories," in *Proceedings of the 30th International Conference on Machine Learning (ICML)*, 2013.

[9] C. Hillar, J. Sohl-Dickstein, and K. Koepsell, "Efficient and optimal Little-Hopfield auto-associative memory storage using minimum probability flow," *NIPS (DISCML Workshop)*, 2012.

[10] J. Sohl-Dickstein, P.B. Battaglino, and M.R. DeWeese, "New method for parameter estimation in probabilistic models: minimum probability flow," *Physical review letters*, vol. 107, no. 22, pp. 220601, 2011.

[11] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[12] E. Granot-Atedgi, G. Tkačik, R. Segev, and E. Schneidman, "Stimulus-dependent maximum entropy models of neural population codes," *PLoS computational biology*, vol. 9, no. 3, pp. e1002922, 2013.

[13] R.H. Warren, "Numeric experiments on the commercial quantum computer," *Notices of the American Mathematical Society*, vol. 60, no. 11, 2013.

[14] E. Ising, "Beitrag zur Theorie des Ferromagnetismus," *Zeitschrift fur Physik*, vol. 31, pp. 253–258, 1925.

[15] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[16] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1106–1114.

[17] G.K. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 30–44, 1991.

[18] D.H. Ackley, G.E. Hinton, and T.J. Sejnowski, "A learning algorithm for boltzmann machines," *Cognitive science*, vol. 9, no. 1, pp. 147–169, 1985.

[19] D.O. Hebb, "The organization of behavior. 1949," *New York, Wiley*, 2002.

[20] G. Weisbuch and F. Fogelman-Soulié, "Scaling laws for the attractors of Hopfield networks," *Journal de Physique Lettres*, vol. 46, no. 14, pp. 623–630, 1985.

[21] R. McEliece, E. Posner, E. Rodemich, and S. Venkatesh, "The capacity of the Hopfield associative memory," *Information Theory, IEEE Trans. on*, vol. 33, no. 4, pp. 461–482, 1987.

[22] A.D. Bruce, A. Canning, B. Forrest, E. Gardner, and D.J. Wallace, "Learning and memory properties in fully connected networks," in *Neural Networks for Computing*. AIP Publishing, 1986, vol. 151, pp. 65–70.

[23] T.M. Cover, "Geometrical and statistical properties of systems of linear inequalities with application in pattern recognition," *IEEE Trans. Electronic Computers*, vol. 14, no. 3, pp. 326–334, 1965.

[24] R.J. McEliece and E.C. Posner, "The number of stable points of an infinite-range spin glass," *JPL Telecomm. and Data Acquisition Progress Report*, vol. 42-83, pp. 209–215, 1985.

[25] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[26] Zhou Wang and Alan C Bovik, "Mean squared error: love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[27] M.H. Asghari and B. Jalali, "Anamorphic transformation and its application to time–bandwidth compression," *Applied optics*, vol. 52, no. 27, pp. 6735–6743, 2013.

[28] J. Hurri, A. Hyvärinen, J. Karhunen, and E. Oja, "Image feature extraction using independent component analysis," in *Proc. NORSIG*, 1996.

[29] A.J. Bell and T.J. Sejnowski, "The independent components of natural scenes are edge filters," *Vision research*, vol. 37, no. 23, pp. 3327–3338, 1997.

[30] J.H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 265, no. 1394, pp. 359–366, 1998.

[31] B.A. Olshausen and D.J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[32] J.W. Pillow, J. Shlens, L. Paninski, A. Sher, A.M. Litke, E.J. Chichilnisky, and E.P. Simoncelli, "Spatio-temporal correlations and visual signalling in a complete neuronal population," *Nature*, vol. 454, no. 7207, pp. 995–999, 2008.