Modeling the Object Recognition Pathway: A Deep Hierarchical Model Using Gnostic Fields

Panqu Wang (pawang@ucsd.edu)
Department of Electrical and Computer Engineering, University of California San Diego
9500 Gilman Dr 0407, La Jolla, CA 92093 USA

Garrison Cottrell (gary@ucsd.edu)
Department of Computer Science and Engineering, University of California San Diego
9500 Gilman Dr 0404, La Jolla, CA 92093 USA

Christopher Kanan (ckanan@caltech.edu)
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91101, USA

Abstract

To recognize objects, the human visual system processes information through a network of hierarchically organized brain regions. Many neurocomputational models have modeled this hierarchical structure, but they have often used hand-crafted features to model early visual areas. According to the linear efficient coding hypothesis, the goal of the early visual pathway is to capture the statistical structure of sensory stimuli, removing redundancy, and factoring the input into independent features. In this work, we use a hierarchical Independent Components Analysis (ICA) algorithm to automatically learn the visual features that account for early visual cortex. We then continue modeling the object recognition pathway using Gnostic Fields, a theory for how the brain does object categorization, in which brain regions devoted to classifying mutually-exclusive categories exist near the top of sensory processing hierarchies. The whole biologically-inspired model not only allows us to develop representations similar to those in primary visual cortex, but also to perform well on standard computer vision object recognition benchmarks.

Keywords: object recognition; Independent Component Analysis; Gnostic Fields; deep model

Introduction

Over the years, researchers have built many models of object recognition that are informed by findings in neuroscience. Four of the best known models are the O'Reilly and Munakata model (OReilly & Munakata, 2000), HMAX (Riesenhuber & Poggio, 1999), VisNet (Wallis & Rolls, 1996), and The Model ("TM", Dailey and Cottrell (1999)). In O'Reilly and Munakata, the visual input is first transformed by a center-surround transformation, followed by the processing of V1, V2, V4 and PFC, with realistic neural constraints and increasingly large receptive fields. The HMAX model uses a hierarchical structure with alternating layers of units that are selective for complex features (S units) and units that have increasing tolerance to position and scale (C units), with the top layer of the hierarchy containing view-tuned and task-related units that correspond to processing done by IT and PFC. VisNet shares a similar structure with HMAX, but the input is filtered by Differential of Gaussian (DoG) filters before feeding into the 4 hierarchical competitive network that correspond to V2, V4, PIT and AIT, and the synapse weights are learned using Hebbian learning rule. TM is aimed to model cognitive phenomena such as the development of hemispheric lateralization (Wang & Cottrell, 2013) and experience moderation effect of face and object recognition (Wang, Gauthier, & Cottrell, 2014), and the input is processed by a Gabor-PCA system, which is used as input to a multi-layer perceptron neural network.

One limitation of these models is that they all use hand-crafted features to simulate visual processing done by the retina, LGN, or primary visual cortex: DoG filters in VisNet, Gabor filters in TM and S1 units of HMAX, and the preset parameter of S2 units in HMAX. In mammals, this is not how the early visual system develops its representations. It develops visual features as visual experience is acquired, and these visual representations are likely learned in a mostly unsupervised manner, since these features are universal across visual tasks. In this paper, we describe a model that uses hierarchical Independent Components Analysis (ICA) to learn a hierarchy of visual filters that can extract diagnostic visual features from images. To recognize objects, we combine the learned ICA filters with gnostic fields, which simulates IT.

The ICA algorithm is an implementation of Barlow's linear efficient encoding hypothesis (Barlow, 1961), which hypothesizes that the goal of early vision is to reduce the redundancy of the input, from which the statistical structure can be captured. One-layer ICA has been used to explain the very first layer of information processing in the visual cortex (Bell & Sejnowski, 1997; Olshausen & Field, 1996). More recently, Shan, Zhang, and Cottrell (2006) proposed a recursive implementation of ICA, which captures the higher order structure. Our hierarchical model adopts the basic idea of recursive ICA, but with different implementation.

ICA is not sufficient for the system to output labels for each category; supervised learning is needed. To do this, we use Gnostic Fields, a state-of-the-art algorithm for object recognition (Kanan, 2013a, 2014). Gnostic Fields are based on based on Konorski's theory (Konorski, 1967) for how the brain recognizes objects. In this theory, Gnostic Fields are brain regions that exist near the top of the sensory information processing hierarchy, and they are responsible for object recognition. A gnostic field is composed of competing gnostic sets, with each set representing one category. Each gnostic set contains multiple gnostic neurons, and they encode particular properties of an object while maintaining a degree of invariance to scale, location, or appearance. fMRI studies suggest that the brain does develop regions that are especially active during object recognition tasks, such as the fusiform face area (FFA) for face recognition (Kanwisher, McDermott, & Chun, 1997) and the visual word form area for recognizing words (McCandliss, Cohen, & Dehaene, 2003).

In the next sections, we give implementation details for our model, an analysis of visual filters learned by the model, results on benchmark computer vision datasets, and we discuss future directions for the model.

Methods

Figure 1 depicts the structure of our model, which consists of image pre-processing inspired by the retina, visual filters learned using hierarchical ICA, and then a gnostic field that performs object classification.

Image Preprocessing

While most object recognition systems start from modeling with V1, we begin with pre-cortical processing done by the retina. We first resize the input image so that its smallest dimension is 128 pixels, with the other dimension resized to preserve the aspect ratio. Subsequently, we convert the image from standard RGB (sRGB) color space to LMS colorspace (Fairchild, 2013), which simulates the retina's long, medium, and short wavelength cone photoreceptors' responses. We then apply a cone-like nonlinearity to the LMS pixels, which helps the model deal with changes in brightness (Caywood, Willmore, & Tolhurst, 2004). The formula we use is

$$I_C^{'}(z) = \max\left(\frac{\log(\theta+1) - \log(I_C(z)+\theta)}{(\log(\theta+1) - \log(\theta))(\theta-1)} + 1, 0\right), \quad (1)$$

where $I_C(z)$ is the image for LMS channel C at location z. θ controls the normalization strength and $\theta = 0.01$ in all of our experiments.

Hierarchical ICA

The two central assumptions of hierarchical ICA are: 1) different brain regions share similar anatomical structures and work under similar computational principles; 2) the input should follow generalized Gaussian distribution in order to let the statistical structure of the system be captured. In the formulation of ICA, the observed data X is assumed to be generated by underlying neural signal source S:

$$X = AS + \varepsilon, \quad (2)$$

where A is the ICA basis matrix and ε is the Gaussian noise term.

In the first layer of ICA, in order to form a Gaussian-like input, the stimulus X has to be whitened. In the visual system, the input is whitened in retina and LGN before transmission to V1. Here we use whitened PCA (WPCA) (Bell & Sejnowski, 1997) to decorrelate the stimulus and normalize the variance. This transformation can be written as:

$$W_C = (D_c + \delta I)^{-\frac{1}{2}} \Phi_c^T, \quad (3)$$

where Φ_c contains the eigenvectors of the covariance matrix of the input stimulus, D_c is the diagonal eigenvalue matrix, and I is the identity matrix. The regularization parameter δ is set to be 0.01 in all experiments.

As the neural signal S is assumed to be sparse and independent, the filter response, which is the input for next layer's ICA, is not Gaussian. As suggested by Shan et al. (2006) and Kanan (2013a), we take the absolute value of the filter response and apply the cumulative distribution function (CDF) of the exponential distribution to the response to efficiently increase the discriminative power of ICA filters. We then whiten the filter response again to form the input for the second layer ICA.

The above process is repeated again for the second layer ICA. This two-layer hierarchical structure can efficiently capture the statistical property of the input stimulus. As the features are learned gradually from the images, this process can simulate the formation of the early visual areas in our brain, which is believed to develop and mature in our early life and are shared components in the entire visual pathway.

Gnostic Fields

In our model, the unsupervised hierarchical ICA algorithm simulates the early visual pathway. In order to model the full object recognition pathway, supervision is needed as we need information to distinguish between different object categories. We use Gnostic Fields to model the higher visual pathway. Below is a brief review of the necessary information to implement a gnostic field. Please see Kanan (2013b) and Kanan (2014) for additional details.

The feature response vector for a given ICA layer (channel) c and location t is $g_{c,t}$. We then augment this descriptor by adding a vector that contains the feature's spatial information: $l_{c,t} = [x_t, y_t, x_t^2, y_t^2, 1]^T$, where (x_t, y_t) is the location of $g_{c,t}$ normalized by the dimension of the input image. $l_{c,t}$ is then also normalized, and the appended new vector $\hat{g}_{c,t}$ is whitened by W_c, which is learned by a collection of training images of the task. This whitening step can also be served as dimensionality reduction, and the whitened feature vectors are made unit length. The final whitened and normalized feature vector $f_{c,t}$ is given by

$$f_{c,t} = \frac{W_c \hat{g}_{c,t}}{\parallel W_c \hat{g}_{c,t} \parallel}. \quad (4)$$

In our model, a gnostic field for channel c is composed of K gnostic sets, and each set represents one object category. Each gnostic set contains a different number of gnostic units $m_{k,c}$, which is given by

$$m(k,c) = \min(\lceil b(\log(n_{k,c}) + 1)^2 \rceil, n_{k,c}), \quad (5)$$

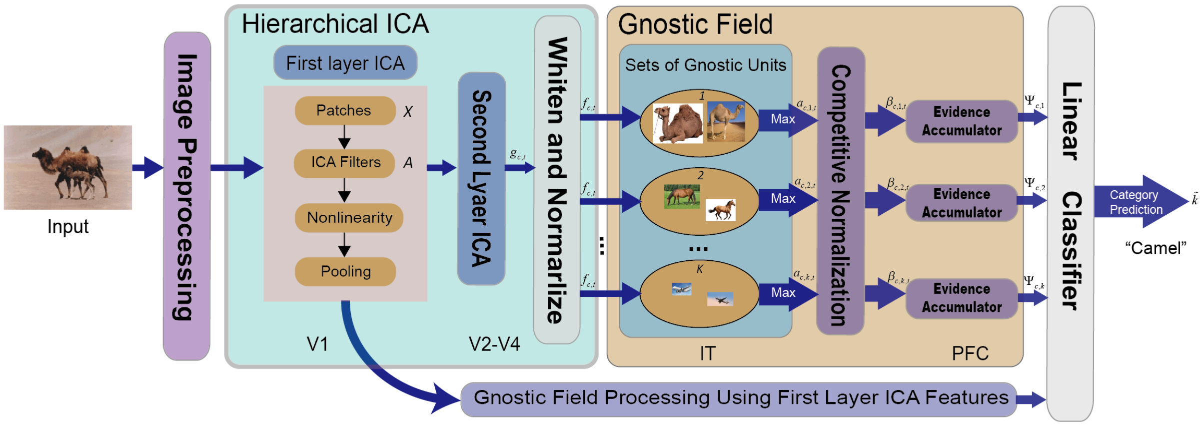

Figure 1: A high-level description of our model. Hierarchical ICA learns features of lower visual areas, and the visual information projects to gnostic sets, with units in the target gnostic set responding strongest. The competitive normalization step suppresses non-relevant set responses, and the information for each feature adds to the model's beliefs. The linear classifier makes the final prediction using information from all categories and layers.

where $n_{k,c}$ is the total number of feature vectors from category k and channel c, and b regulates how many units learned in the category (here $b = 10$ in all experiments). Since the number of feature vectors is directly proportional to the number of examples, the number of gnostic units increases logarithmically in the number of training examples.

A gnostic set measures how similar the input feature vector $f_{c,1}, \ldots, f_{c,T}$ is to the previous examples (memory) of that category. The output vector of the given category is generated by:

$$a_{c,k,t} = \max_j (v_{c,k,j} \cdot f_{c,t}), \tag{6}$$

where $v_{c,k,j}$ denotes the weight vector for each gnostic unit j in category k, and is learned by spherical k-means unsupervised clustering algorithm for unit length data (Dhillon & Modha, 2001). The max operation is taken across all units in the category, so we can treat it like a max pooling step, which enables the gnostic set to activate strongly to any stimuli that matches previous observations of that category.

The above "max pooling" step is performed on every gnostic set in the whole gnostic field. Subsequently, inhibitive competition is used to suppress the response of least active gnostic sets. To implement this on all K sets of channel c, their outputs are first attenuated by half-way rectification, i.e.,

$$q_{c,k,t} = \max(a_{c,k,t} - \theta_{c,t}, 0), \tag{7}$$

where $\theta_{c,t} = \frac{1}{K}\sum_{k'} a_{c,k',t}$. The non-zero responses are then normalized using

$$\beta_{c,k,t} = \nu_{c,t} \cdot q_{c,k,t}, \tag{8}$$

where

$$\nu_{c,t} = \frac{\sum_{k'} q_{c,k',t}}{(K^{-1} + \sum_{k'} q^2_{c,k',t})^{3/2}}, \tag{9}$$

which acts as a form of divisive normalization and has been reported crucial in obtaining good object recognition accuracy in Kanan (2013b).

The next step is categorical evidence accumulation, which simply sums the activation of $\beta_{c,k,t}$ across all time steps or locations of the input:

$$\psi_{c,k} = \sum_{t=1}^{T} \beta_{c,k,t}, \tag{10}$$

and the evidence accumulated from all category and channels of the gnostic field forms the vector Ψ, which is made mean zero and unit length.

Finally, a linear multi-category classifier is used to decode the activity of all units and deal with confused categories. The model's predicted category is given by $\tilde{k} = \operatorname{argmax}_k w_k \cdot \Psi$, where w_k is the weighting vector of category k. In our experiments, the weights were learned with the LIBLINEAR toolbox (Fan, Chang, Hsieh, Wang, & Lin, 2008) using multiclass Support Vector Machine (SVM) formulation by Crammer and Singer (Crammer & Singer, 2001), and the SVM cost parameter was set to be 0.0001.

In our model, the gnostic field sits on top of the low-level visual information processing produced by hierarchical ICA, simulating the fact that the gnostic sets sit near top of a sensory processing hierarchy (vision in our model), as hypothesized in Konorski's Gnostic Fields theory. In general, our biological-inspired hierarchical model not only develops the low-level visual features, but also possesses the capability of learning categorical information necessary to perform well in the high-level object recognition task. We show the results in the next section.

Experiments

In this section, we will first analyze the ICA filters learned by hierarchical ICA, and then show the object recognition experiment results on computer vision datasets using the whole model.

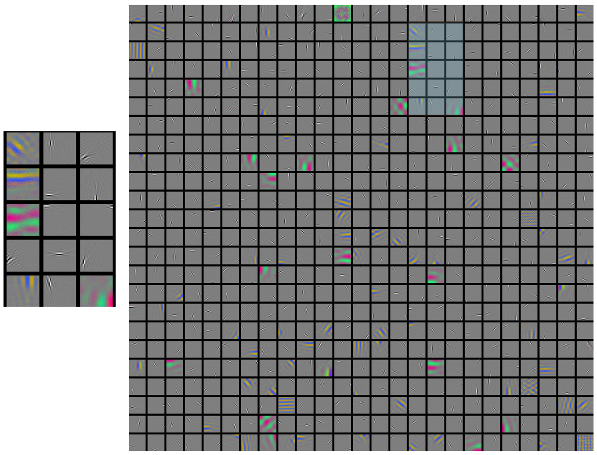

Figure 2: The 600 basis function learned from first layer ICA. The learned Gabor-like filters share the three color opponency characteristics (dark-light, yellow-blue, and red-green) of V1 neurons.

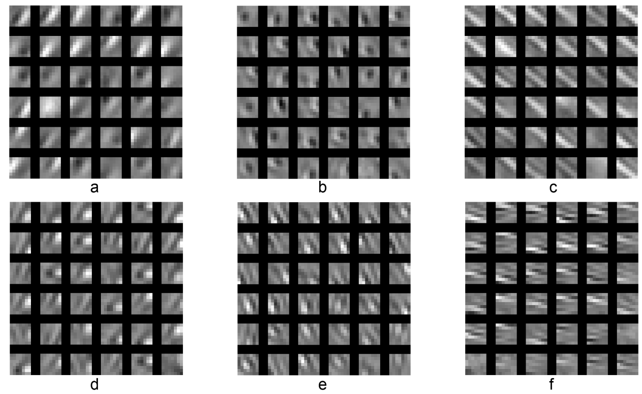

Figure 3: Visualization of 6 randomly selected examples of second-layer ICA filters. Each sub-figure shows a filter with the dimensions of top 6×6 entropy values. Dimension for each patch is 7×7.

Feature Learning using Hierarchical ICA

Just like the early visual cortex matures by gradually perceiving the surrounding environment, our model develops its representation for unsupervised feature learning by using the natural image dataset. We learned the ICA filters from 625 images from the Mcgill color image dataset (Olmos et al., 2003). Each image was preprocessed using the method described in the previous section. For each image, 300 patches were randomly selected for 25×25 patch size. Prior to ICA, all patches were whitened using WPCA and the dimensionality was reduced to 600, which preserves about 98.86% of the total variance. Next, we learned the filters using the Efficient Fast ICA algorithm (Koldovsky, Tichavsky, & Oja, 2006). The learned matrix $A = [a_1, ..., a_n]^T$ consists of ICA basis row vector a_i, and the learned filters are shown in Figure 2.

From Figure 2, we can see a population of filters that respond to both chromatic and achromatic features. All features are Gabor-like, and they share all three color opponency characteristics of V1 neurons: dark-light, blue-yellow and red-green (Caywood et al., 2004; Lee, Wachtler, & Sejnowski, 2002).

In the second layer of ICA, the filter responses were first processed by a 2×2 max pooling operation, which enables the system to gain invariance for small translation. The patch size at this layer is 7×7, which make the visual feature learned in this layer account for a larger receptive field. Next, the filter responses were processed through the non-linearity and the dimensionality was reduced to 300 by WPCA. To visualize this layer's features, we obtained the row vector for each ICA basis that can be reshaped to a 49×600 matrix, then ranked each column of the matrix based on the entropy of the learned filters. Examples of the learned second layer ICA filters are shown in Figure 3.

From Figure 3, we can see that the second layer filters primarily respond to edges with different frequencies and orientations (a,c and f), and they show some degree of spatial invariance (same filter pattern appears at different location of the patch), like V1 complex cells. In addition, there are also filters that correspond to contours (b and d), like the cells in V2 do. In general, the second-layer ICA features are good representation for higher layers of early visual cortex.

Object Recognition Using Gnostic Fields

Given the feature learned by Hierarchical ICA, we assessed the performance of our object recognition system using two major computer vision datasets: Caltech-101 (Li, Fergus, & Perona, 2007) and Caltech-256 (Griffin, Holub, & Perona, 2007). Caltech-101 dataset contains 9146 images of 101 distinct object categories (and one background). Caltech-256 dataset is the successor of Caltech-101, and it contains a broader number (256) of object categories and more varied images, so the task is more difficult. All results below are reported as the mean-class accuracy over five cross-validation runs.

On Caltech-101 dataset, we analyzed the contribution of each component in our model to the overall performance of object recognition task. Specifically, we aim to explore whether we will benefit from the hierarchical structure, and how much performance will boost using the gnostic field. The results are shown in Figure 4.

As can be seen from Figure 4, the recognition accuracy benefits a lot from adding the gnostic field. This is because if learning only involves unsupervised approach, ICA features themselves are not very discriminative, just like we cannot recognize objects relying only on V1 and V2. On the other hand, the performance of using two-layer ICA is better than using only one-layer ICA. This indicates that a hierarchical system not only better models the early visual cortex, but also generates the feature that is more discriminative than that a single-layer model, although both are unsupervised. Considering the "deep" models, no matter unsupervised (Le, 2013)

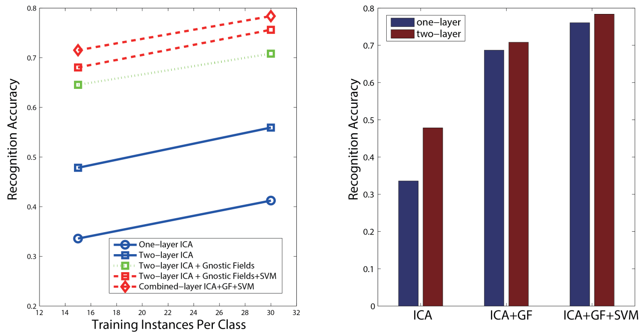

Figure 4: Mean per-class accuracy on Caltech-101 dataset using different model structures. Left: Recognition accuracy as a function of number of training instances. Right: Comparison of using only one-layer ICA and two-layer ICA, with different model structure: 1) ICA: gather the filter reponse for each image, downsample to uniform size, vectorize and apply SVM as classifier; 2) ICA+GF: Apply gnostic field on top of ICA, use the naive argmax classifier; 3) ICA+GF+SVM: replace argmax by SVM.

or fully supervised (Krizhevsky, Sutskever, & Hinton, 2012), they all benefit from the rich representation that the intermediate layer could develop, which helps the model reach very good object recognition performance. Finally, we can observe that combining information of both ICA layers and adding linear classifier generates the best performance, which predicts that the categorizer in IT or PFC may need the information from all previous visual layers to make the more reliable decision.

On Caltech-256 dataset, using the same settings on Caltech-101 dataset, we measured the mean per-class accuracy of up to 50 test images using 15, 30, 45, 60 training images, respectively. The results, as well as the comparison with other methods, are shown in Figure 5.

From Figure 5, we can see that our model's performance is very competitive when compared with other methods from computer vision. Our model achieves the recognition accuracy of 51.8% when using 60 training images, and outperforms all other methods mentioned in the figure. The closest performance is achieved by Kanan (2013a), but they used manually-designed CSIFT feature, as opposed to our biologically-inspired learned ICA features. One thing to note is that for deep convolutional neural network, using a pre-trained network on big dataset usually yields a better performance than training on Caltech-256 alone. This suggests that our hierarchical model is more competent on learning from a medium-sized dataset.

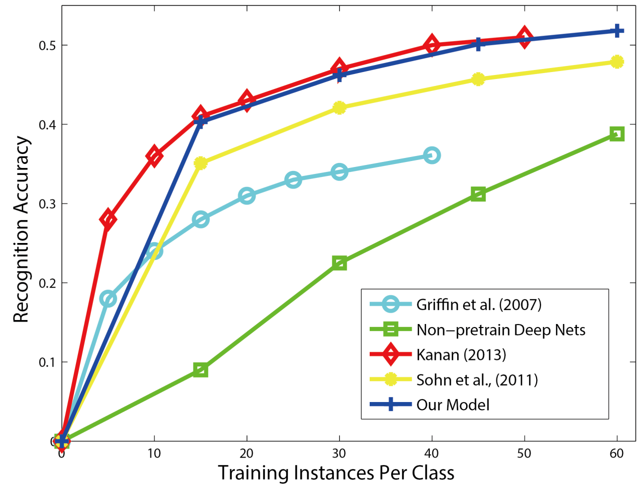

Figure 5: Mean per-class accuracy of Caltech-256 dataset as a function of number of training images. Griffin et al. (2007) provides baseline results. Deep convolutional neural networks do not perform well when trained solely on Caltech-256, because they overfit without additional training on significantly larger datasets, e.g., ImageNet (Zeiler & Fergus, 2014). Sohn et al. (2011) used features learned from a convolutional RBM. Kanan (2013a) used CSIFT features and multichannel gnostic fields, and we achieved almost the same accuracy. Chance is 0.0039.

Discussion

In this paper, we proposed a biologically-inspired deep hierarchical model for visual object recognition. We combined unsupervised feature learning (hierarchical ICA) and a supervised learning algorithm (Gnostic Fields) to account for the processing of early visual cortex and higher visual pathways, respectively. We learned V1 and V2-like filters automatically from natural images using a hierarchical ICA algorithm, that simulates the development and maturation of early visual cortex. With Gnostic fields, we achieve good performance on two object recognition tasks. We suggest that the Gnostic field models the categorization process in areas IT and PFC. Overall, our biologically-inspired model provides an end-to-end model of the human object recognition pathway.

Recently, deep convolutional neural networks (CNNs) have emerged as a powerful machine learning tool for object classification, particularly when millions of labeled images are available (Krizhevsky et al., 2012). A deep CNN has multiple convolutional layers followed by multiple fully connected layers. In our model, ICA layers are stacked twice, and the Gnostic Fields serve as the classification layer. One notable difference between our model and deep networks is that our model is composed of both unsupervised and supervised learning, while state-of-the-art CNNs are fully supervised.

One question is whether more layers of ICA will help or not. The learning method is highly computationally intensive, so we were unable to add a third layer of ICA processing. It is possible that more invariance would arise in a deeper unsupervised network. The question is whether the loss of information through dropping the sign and spatial pooling will lead to beneficial invariants being learned. Also, it is possibly the case that deeper layers, which are closer to the temporal pole, receive more category information, so that using strictly unsupervised learning might not be appropriate. In future work, we would like to apply this model to video to investigate whether it can learn useful spatiotemporal features and achieve good performance in tracking and activity recognition.

Acknowledgments

This work was supported in part by NSF Science of Learning Center grants SBE-0542013 and SMA-1041755 to the Temporal Dynamics of Learning Center, NSF grant IIS-1219252 to GWC. PW was supported by a fellowship from Hewlett-Packard. This work was initiated while CK was affiliated with UCSD.

References

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*, 217–234.

Bell, A. J., & Sejnowski, T. J. (1997). The independent components of natural scenes are edge filters. *Vision Research*, *37*(23), 3327–3338.

Caywood, M. S., Willmore, B., & Tolhurst, D. J. (2004). Independent components of color natural scenes resemble v1 neurons in their spatial and color tuning. *Journal of Neurophysiology*, *91*(6), 2859–2873.

Crammer, K., & Singer, Y. (2001). On the algorithmic implementation of multiclass kernel-based vector machines. *J Machine arning Research*, *2*, 265-292.

Dailey, M. N., & Cottrell, G. W. (1999). Organization of face and object recognition in modular neural network models. *Neural Networks*, *12*, 1053–1073.

Dhillon, I. S., & Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. *Machine learning*, *42*(1-2), 143–175.

Fairchild, M. D. (2013). *Color appearance models*. John Wiley & Sons.

Fan, R., Chang, K., Hsieh, C., Wang, X., & Lin, C. (2008). LIBLINEAR: A library for large linear classification. *J Machine Learning Research*, *9*, 1871-1874.

Griffin, G., Holub, A., & Perona, P. (2007). Caltech-256 object category dataset.

Kanan, C. (2013a). Active object recognition with a space-variant retina. *ISRN Machine Vision*, *2013*, 138057.

Kanan, C. (2013b). Recognizing sights, smells, and sounds with gnostic fields. *PLoS ONE*, *e54088*. doi: 10.1371/journal.pone.0054088

Kanan, C. (2014). Fine-grained object recognition with gnostic fields. In *Proceedings of the IEEE Winter Applications of Computer Vision Conference (WACV-2014).*

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.

Koldovsky, Z., Tichavsky, P., & Oja, E. (2006). Efficient variant of algorithm fastica for independent component analysis attaining the cramer-rao lower bound. *Neural Networks, IEEE Transactions on*, *17*(5), 1265–1277.

Konorski, J. (1967). *Integrative activity of the brain.* Chicago.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097–1105).

Le, Q. V. (2013). Building high-level features using large scale unsupervised learning. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 8595–8598).

Lee, T.-W., Wachtler, T., & Sejnowski, T. J. (2002). Color opponency is an efficient representation of spectral properties in natural scenes. *Vision Research*, *42*(17), 2095–2103.

Li, F., Fergus, R., & Perona, P. (2007). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, *106*(1), 59–70.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*(7), 293–299.

Olmos, A., et al. (2003). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, *33*(12), 1463–1473.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

OReilly, R., & Munakata, Y. (2000). Neurocomputational models of face processing. In *Computational Explorations in Cognitive Neuroscience.* Cambridge, MA: MIT Press.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.

Shan, H., Zhang, L., & Cottrell, G. (2006). Recursive ICA. In *Advances in Neural Information Processing Systems* (pp. 1273–1280).

Sohn, K., Jung, D. Y., Lee, H., & Hero, A. O. (2011). Efficient learning of sparse, distributed, convolutional feature representations for object recognition. In *Computer Vision, 2011 IEEE International Conference on* (pp. 2643–2650).

Wallis, G., & Rolls, E. T. (1996). A model of invariant object recognition in the visual system. *Prog. Neurobiol*, *51*, 167–194.

Wang, P., & Cottrell, G. W. (2013). A computational model of the development of hemispheric asymmetry of face processing. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Wang, P., Gauthier, I., & Cottrell, G. (2014). Experience matters: Modeling the relationship between face and object recognition. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014* (pp. 818–833). Springer.