SCIENCE CHINA
Technological Sciences

•Article•

April 2019 Vol.62 No.4: 628–634
https://doi.org/10.1007/s11431-018-9327-9

Emergence of higher-level neuron properties using a hierarchical statistical distribution model

XIAN Ning, DENG YiMin & DUAN HaiBin*

School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China

Received February 5, 2018; accepted July 24, 2018; published online March 6, 2019

Essential to visual tasks such as object recognition is the formation of effective representations that generalize from specific instances of visual input. Neurons in primary visual cortex are typically hypothesized to efficiently encode image structures such as edge and textures from natural scenes. Here this paper proposed a novel hierarchical statistical distribution model to generalize higher-level neuron properties and encode distributed regularities that characterize local image regions. Two layers of our hierarchical model are presented to extract spiking activities of excitatory neurons decorrelated by inhibitory neurons and to construct the statistical patterns of input data, respectively. Trained on whitened natural images, parameters including neural connecting weights and distribution coding weights are estimated by their corresponding learning rules. To prove the feasibility and effectiveness of our model, several experiments on natural images are conducted. Adapting our model to natural scenes yields a distributed representation for higher-order statistical regularities. Comparison results provide insight into higher-level neurons which encode more abstract and invariant properties.

statistical distribution, spiking activity, object classification, localization

Citation: Xian N, Deng Y M, Duan H B. Emergence of higher-level neuron properties using a hierarchical statistical distribution model. Sci China Tech Sci, 2019, 62: 628–634, https://doi.org/10.1007/s11431-018-9327-9

1 Introduction

It has been hypothesized that primate visual system has evolved to accurately encode environmental signals with the minimal consumption of biological resources [1]. The retina transmits visual information to the brain with its anatomical and functional properties [2,3]. This pathway provides a natural choice for the study of coding efficiency [4]. Adapted to the statistic natural images, neurons in the primary visual cortex (V1) represent basic image features such as local orientation and edge using an efficient code [5]. However, a number of nonlinear effects, such as visual attention [6–8] and lateral inhibition [9,10], are difficult to be explained with those properties of simple cells. Neurons in the early visual areas extract simple features, which are transmitted to neurons in higher visual areas of primate visual cortex [11]. Thus, the perception of complex properties emerges from neuronal activity in higher visual areas, which carry necessary information for a completion of different tasks such as classification [12,13], object recognition [14] and pose estimation [15].

Even though individuals collected by the stimulus at the retina are inherently highly variables, they can be efficiently processed by the brain. Previous theoretical work has been done to explore the biological feasibility of explaining aspects of higher-level visual processing [16,17]. Early visual neurons such as the Gabor filter [18] and the Derivative of Gaussian (DoG) filter [19] are typically described as linear feature detectors. These filters have similar shapes to the receptive fields (RFs) of orientation-tuned cells in V1. Sparse coding has emerged as a useful principle for understanding neural representations in the cortex. Several quali-

*Corresponding author (email: hbduan@buaa.edu.cn) tative features of the RFs of neurons in V1 have been reproduced by establishing a sparse and independent local network (SAILnet) for natural images [20]. When retinal ganglion cells transmit the spatial information of natural images to the brain, a high level of efficiency with near-optimal redundancy appears in visual signaling by the retina [21]. An extended spiking model called E-I Net [22] is presented by adding a separate population of inhibitory neurons to provide conformance to Dale's Law. A hybrid method based on SIFT feature and sparse coding [23] is presented for image classification, and reached a competitive performance on public datasets.

To formulate the higher-level neuron properties, a standard model of visual cortex [24] is proposed for object recognition in a quantitative way. The standard model consists of four layers of alternate simple S units and complex C units. A series of extended methods [25,26] are developed based on the standard model. A multiscale convolutional network is presented to extract dense feature vectors by encoding regions of multiple sizes [27,28]. To capture higher-order nonlinear structure and represent nonstationary data distributions, a hierarchical Bayesian model [29] is presented through the generalization of independent component analysis (ICA). A distribution coding model [30] is proposed using the neural activity and probability distribution.

In this paper, a novel hierarchical statistical distribution model is proposed based on neural spiking activity and their internal distribution regularities. Being different from those traditional methods, two layers of hierarchical model are utilized to extract high-order structures from the nature images for target classification. The presented model can classify the abstract properties of input data. The dynamic spiking layer extracts image representations with spiking activities of excitatory neurons that decorrelated by separate inhibitory neurons. The distribution coding layer then constructs the statistical patterns of those spiking outputs. The contributions of this paper can be illustrated as: (1) the higher-order statistical regularities are considered for the presented hierarchical model; (2) adapting the presented hierarchical model to natural scenes can yield a distributed representation for higher-order statistical regularities, which can been further applied to target detection.

2 Hierarchical statistical distribution model

The hierarchical statistical distribution model is illustrated schematically in Figure 1.

The hierarchy stacks two neural activity extraction layers, thereby representing distributions of images patches. The statistical patterns are characterized by a Gaussian distribution with a fixed mean of zero and a covariance that is a function of the neural activity [30]. At the first layer, spiking activity of the retina ganglion cells is computed with some Gabor-like receptive fields. These responses are then encoded with a non-linear transformation layer to construct the statistical patterns at the second layer.

2.1 Dynamic spiking layer

The dynamic spiking layer consists of excitatory neurons (E) and inhibitory neurons (I) in separate populations. This layer obtains the dynamic spiking activity of these excitatory neurons. Instead of the activity of those simple cells that laterally inhibit each other directly [20], the excitatory neurons are decorrelated by those inhibitory ones. The inhibitory neurons provide feedback inhibition to excitatory ones to cancel out the redundant part and decorrelate the activity of excitatory neurons. Similar to LIF neurons [22], these neurons work together to learn a sparse representation of the input signal without violating Dale's Law.

Neurons at this layer compute a sparse representation of input data, which identifies visual features that match the Gabor-like receptive fields. The population of excitatory neurons receives retina input from whitened image patch, along with the feedback inhibition from the population of inhibitory neurons. The inhibitory neurons can receive input from the excitatory ones and send inhibition back. The inhibitory neurons also inhibit each other. Each connection can be viewed as the signal transfer process from pre-synaptic neuron to post-synaptic neuron. The neurons in this layer are fully connected and the connection diagram is shown as Figure 2. Three connection types between the excitatory and inhibitory neurons are defined by respective connection weights (i.e., $Q^{\mathrm{E}\to\mathrm{I}}$ is the weight from excitatory neurons to inhibitory neurons, $Q^{\mathrm{I}\to\mathrm{E}}$ is the weight from inhibitory neurons to excitatory neurons, and $Q^{\mathrm{I}\to\mathrm{I}}$ is the weight between inhibitory neurons). Connections between the input image patch and excitatory neurons are also represented as the connection weights $Q^{\mathrm{in}\to\mathrm{E}}$ (not shown in Figure 2).

Each E cell and I cell can be viewed as a pair of one feedforward excitatory connection and one feedback inhibitory connection. On the basis of those connecting weights, neural spiking activities of those excitatory cells constitute the sparse representation of the input stimuli.

The neuron state of these populations can be updated by the dynamic spiking function, which is similar to LIF neurons [22]:

$$y_i^C(t+1) = y_i^C(t)(1-\tau^C) + \tau^C \sum_{C_L} \beta^L \sum_j u_j^{C_L}(t) Q_{ij}^L, \tag{1}$$

where C is the neuron class in this layer (i.e., E or I), t is the simulation time step, $y_i^C(t)$ is the potential energy of neuron i of class C at time t, τ^C is the rate parameter governing the decay rate for neurons of class C, L is the connection type to

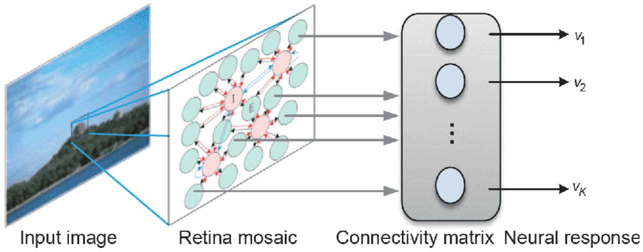

Figure 1 (Color online) Our proposed hierarchical statistical distribution model.

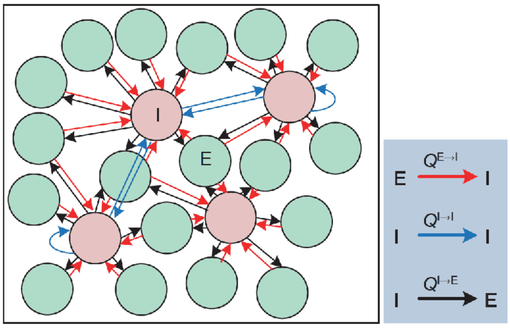

Figure 2 (Color online) Connection diagram at the dynamic spiking layer.

neuron i of class C, C_L represents the pre-synaptic neuron connecting to neuron i of class C, β^L is the impact sign of connection L (+1 for excitatory connections and −1 for inhibitory connections), $u_i^{C_L}(t)$ is the spike output (either 0 for no spike, or 1 for spike) of pre-synaptic neuron i at time t. The spike output of each neuron is computed by comparing the potential energy $y_i^C(t)$ with its corresponding spike threshold θ_i^C through eq. (2). If the neural potential energy crosses the corresponding spike threshold, the neuron spikes and then the potential energy is reset to zero.

$$u_i^C(t) = \begin{cases} 1, & y_i^C(t) \geq \theta_i^C, \\ 0, & \text{otherwise}. \end{cases} \tag{2}$$

In separate populations of excitatory and inhibitory neurons, the neuron state can be represented respectively as follows:

$$\begin{aligned} y_i^{\mathrm{E}}(t+1) &= y_i^{\mathrm{E}}(t)(1-\tau^{\mathrm{E}}) \\ &\quad +\tau^{\mathrm{E}}\left(\sum_j X_j Q_{ij}^{\mathrm{in}\to\mathrm{E}} - \sum_j u_j^{\mathrm{I}}(t) Q_{ij}^{\mathrm{I}\to\mathrm{E}}\right), \end{aligned} \tag{3}$$

$$\begin{aligned} y_i^{\mathrm{I}}(t+1) &= y_i^{\mathrm{I}}(t)(1-\tau^{\mathrm{I}}) \\ &\quad +\tau^{\mathrm{I}}\left(\sum_j u_j^{\mathrm{E}}(t) Q_{ij}^{\mathrm{E}\to\mathrm{I}} - \sum_j u_j^{\mathrm{I}}(t) Q_{ij}^{\mathrm{I}\to\mathrm{I}}\right), \end{aligned} \tag{4}$$

where X_i represents the value of the input image patch at pixel i. Following the precedent set by previous sparse coding studies (e.g., [21,30]), we used whitened input images to remove pairwise correlations in the input stimuli, which is similar to the process of visual signal passing through the retina to the visual cortex

The connection weights and spike thresholds can be obtained by training this dynamic spiking layer. The learning rules, similar to those in [20,22], are introduced to update the trainable parameters. Based on the spiking activities of the presynaptic and postsynaptic neurons, updates to the connection weights are computed locally. The weights from the input image patch to the excitatory neurons (i.e., $Q^{\mathrm{in}\to\mathrm{E}}$) and three connection weights from neurons of class C_1 to neurons of class C_2 (i.e., $Q^{\mathrm{E}\to\mathrm{I}}$, $Q^{\mathrm{I}\to\mathrm{E}}$, and $Q^{\mathrm{I}\to\mathrm{I}}$) are updated according to

$$\Delta Q_{ij}^{\mathrm{in}\to\mathrm{E}} \propto n_i^{\mathrm{E}}\left(X_j - n_i^{\mathrm{E}} Q_{ij}^{\mathrm{in}\to\mathrm{E}}\right), \tag{5}$$

$$\Delta Q_{ij}^{C_1\to C_2} \propto n_i^{C_1} n_j^{C_2} - p^{C_1} p^{C_2}\left(1 + Q_{ij}^{C_1\to C_2}\right), \tag{6}$$

where $n_i^C = \sum_t u_i^C(t)$ represents the number of the spikes emitted by neuron i of class C over time, and p^C is the mean spike rate of neurons of class C. The spike thresholds of neurons are also adjusted during the learning process, according to the adaptation rule $\Delta\theta_i^C \propto n_i^C - p^C$. Excitatory neurons become tuned to specific image features after exposure to thousands of input image patches. Their spiking activities come to resemble the Gabor-like response properties observed in simple cells of V1. At the end, the number of spikes generated by each excitatory neuron represents the sparse feature of input stimuli. This layer's representation of the image patch is the average spike rate of each neuron during the simulation time.

2.2 Distribution coding layer

The second layer, distribution coding layer, is to generate the high-order structure in input images on the basis of the high-lever neuron properties. With those sparsely distributed spiking outputs computed at the first layer, the distribution coding layer describes these spiking outputs n_i with multivariate Gaussian probability distributions as follows:

$$P(\mathbf{n} \mid \mathbf{v}) = N(0, \boldsymbol{\lambda}), \tag{7}$$

where the covariance $\boldsymbol{\lambda} = f(\mathbf{v})$ is a non-linear function of neural activities v_k. The transformation from the input image patch to the higher-order structure describes patterns in the variances of spiking outputs at the first layer, hence fundamentally non-linear. The logarithm of covariance is computed by neural activities weighted through W_{ik}:

$$\log\lambda_i = \sum_k W_{ik} v_k. \tag{8}$$

The covariance is a latent scale parameter in this layer. We set this parameter to connect the spiking outputs to neural activities and account for the high-order statistical distributions of the input image patches. The form of this connection in eq. (8) implies that absence of those neural activities ($v_k = 0$) corresponds to a canonical distribution ($\lambda_i = 1$). Figure 3 shows the schematic of the distribution coding layer. Nonzero activities of neurons in this layer describe changes in shaping the distributions of spiking outputs n_i at the first layer (dashed rectangle in Figure 3) using connecting weights W_{ik}. Each neuron at the second layer has a different set of weights (i.e., the J-dimensional vector $\mathbf{W}_k$ in Figure 3), corresponding to the role in modifying the encoded distribution.

The values of those neural activities $\mathbf{v}$ and connecting weights $\mathbf{W}$ for a given spiking outputs $\mathbf{n}$ are computed by evaluating the likelihood at the maximum a posteriori (MAP) estimate:

$$\hat{\mathbf{v}} = \underset{\mathbf{v}}{\operatorname{argmax}}\, p(\mathbf{v} \mid \mathbf{n}, \mathbf{W}), \tag{9}$$

$$\widehat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}}\, p(\mathbf{v} \mid \mathbf{n}, \mathbf{W})\, p(\mathbf{W}). \tag{10}$$

The log posterior distribution can be expressed as

$$\begin{aligned}\log p(\mathbf{v} \mid \mathbf{n}, \mathbf{W}) &\propto \log p(\mathbf{n} \mid \mathbf{W}, \mathbf{v}) p(\mathbf{v}) \\ &\propto \log \prod_{i=1}^{J} \frac{1}{\lambda_i} \exp(-n_i / \lambda_i) p(\mathbf{v}) \\ &\propto \sum_{i=1}^{J} (-\log\lambda_i - n_i / \lambda_i) + \log p(\mathbf{v}).\end{aligned} \tag{11}$$

The neural activities and weights are derived by gradient ascent:

$$\begin{aligned}&\frac{\partial \log p(\mathbf{v} \mid \mathbf{n}, \mathbf{W})}{\partial v_k} \\ &\propto \frac{\partial}{\partial v_k}\left(\sum_{i=1}^{J} (-\log\lambda_i - n_i / \lambda_i) + \log p(\mathbf{v})\right) \\ &\propto \sum_{i=1}^{J} (-W_{ik} + W_{ik} n_i / \lambda_i) + \frac{\partial}{\partial v_k} \log p(\mathbf{v}),\end{aligned} \tag{12}$$

$$\begin{aligned}&\frac{\partial \log p(\mathbf{v} \mid \mathbf{n}, \mathbf{W}) p(\mathbf{W})}{\partial w_{ik}} \\ &\propto \frac{\partial}{\partial w_{ik}}\left(\sum_{i=1}^{J} (-\log\lambda_i - n_i / \lambda_i) + \log p(\mathbf{v}) + \log p(\mathbf{W})\right) \\ &\propto -v_k + v_k n_i / \lambda_i + \frac{\partial}{\partial w_{ik}} \log p(\mathbf{W}).\end{aligned} \tag{13}$$

The gradients used for estimation are

$$\Delta \mathbf{v} \propto \mathbf{W}^{\mathrm{T}} (\mathbf{n} / \boldsymbol{\lambda} - 1) + \varphi(\mathbf{v}), \tag{14}$$

$$\Delta \mathbf{W} \propto (\mathbf{n} / \boldsymbol{\lambda} - 1) \mathbf{v}^{\mathrm{T}} + \varphi(\mathbf{W}), \tag{15}$$

where $\varphi(\mathbf{v}) = \mathrm{d}\log p(\mathbf{v}) / \mathrm{d}\mathbf{v}$ and $\varphi(\mathbf{W}) = \mathrm{d}\log p(\mathbf{W}) / \mathrm{d}\mathbf{W}$ are the sparse prior term placed on neural activities and weights, respectively. In this paper, we place a Laplacian prior on these parameters and infer their values for each input data. Neural activities are modeled with the Laplacian distribution such as $p(\mathbf{v}) = \prod p(v_k) \propto \prod \mathrm{e}^{-|v_k|}$. Then the sparse prior term of eq. (14) can be expressed as $\varphi(\mathbf{v}) \propto \sqrt{2}\,\mathrm{sign}(\mathbf{v})$, where sign is the symbolic function. We set the same prior on the weights and estimate their values. Using this parameterization, patterns of high-order statistical regularities are captured among the variances.

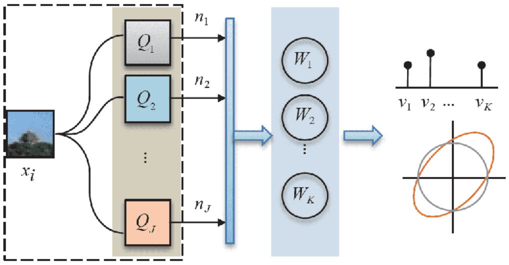

Figure 3 (Color online) Schematic of the distribution coding layer.

3 Parameter estimation and learning

Given the image patch $\mathbf{x}$, two sets of parameters (i.e., neural connecting weights $\mathbf{Q}$ and distribution coding weights $\mathbf{W}$) are adapted to the input data. Neural connecting weights $\mathbf{Q}$ will convert the input information into the spike activities of those excitatory cells, which constitute the sparse representation of the input image patch. Then distribution coding weights $\mathbf{W}$ describe the role of each neuron at the second layer in shaping the encoded image distribution. Two sets of neural activities (i.e., spiking outputs $\mathbf{n}$ at the first layer and high-level neural activities $\mathbf{v}$) are computed associated with the stimulus. For a given data sample (e.g., input image patch $\mathbf{x}$, which is expressed as the $N \times 1$ vector, see Figure 3), $\mathbf{n}$ is the $J \times 1$ vector of spiking outputs at the first layer while $\mathbf{v}$ is the $K \times 1$ vector of spiking outputs at the second layer. $\mathbf{Q}$ is the $J \times N$ matrix of neural connecting weights while $\mathbf{W}$ is the $J \times K$ matrix of distribution coding weights. Each row of matrix $\mathbf{Q}$ is fixed to unit-norm and regarded as one receptive field of excitatory neuron. The process of generating the spiking outputs can be viewed as a measure of the match between the input image patch and the receptive field of the neuron. The properties of components in distribution coding weights $\mathbf{W}$ are analogous to the neurobiological interpretation of complex cells, which pool squared output over specific first-order feature dimensions [2]. Thus, these parameters generate a hierarchical representation, in which the first layer encodes data precisely and the second layer describes more abstract properties associated with the shape of the distribution.

For computational efficiency, neural connections of two layers are assumed to be independent. Hence, neural connecting weights $\mathbf{Q}$ and distribution coding weights $\mathbf{W}$ are estimated separately in two layers in this paper. We first calculate the parameters at the first layer and adapt neural connecting weights **Q** to the input data using the learning rules discussed above. Then distribution coding weights **W** are optimized on the spiking outputs of the fixed neural connecting weights of excitatory neurons.

4 Experimental results

4.1 Image reconstruction

In order to validate the performance of the proposed method, experiments are performed on a standard set of natural images [5]. Image patches with 10 pixels×10 pixels (i.e., the dimension N is 100) are randomly extracted from standard grayscale images of natural scenes. 400 excitatory neurons (i.e., $J = 400$) and 49 inhibitory neurons are set to the dynamic spiking layer. Based on the input image patches and learning rules, parameters are estimated and a subset of the neural connecting weights (Gabor-like receptive fields) is learned by those neurons.

As shown in Figure 4, each square is an oriented and localized feature, which represents the Gabor-like receptive field of a single neuron. The gray value in each square represents zero. The lighter pixels correspond to positive stimuli and the darker ones correspond to negative stimuli. These Gabor-like receptive fields are consistent with those properties of simple cells in the primate visual cortex [20]. Instead of the convolution operator in previous models [28,31], the final neural spiking outputs at the first layer are calculated with the neural interaction of two separate neuron population. The connection weight between neurons is proportional to the degree of correlation between their tuning similarities.

To verify the coding performance of our dynamic spiking layer, reconstruction experiment is conducted and the result is shown in Figure 5. The input image in Figure 5(a) is whitened using the same preprocess as the training set. 5000 patches are randomly extracted from the input image. Thus, the spiking outputs are recorded from each excitatory neuron in response to each patch. Through multiplying each excitatory neuron's spiking output by the corresponding Gabor-like receptive field in Figure 4 and summing over all neurons, all decoded patches are tiled together at the previous positions. The reconstruction image is shown in Figure 5(b). The result indicates the dynamic spiking layer can successfully encode the input data precisely with remaining most features of the input scene. Owing to the sparse spiking rate, the reconstruction image is not identical to the original one. Some details are missing as these stimuli cannot produce enough spiking energy of those Gabor-like receptive fields. Better results can be obtained from more excitatory neurons with more completed orientation tuning.

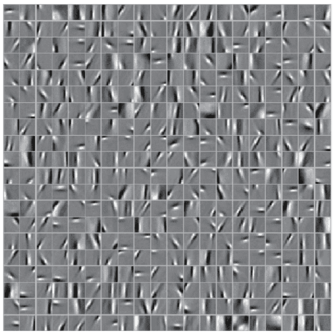

Figure 4 Neural connection weights learned by excitatory neurons.

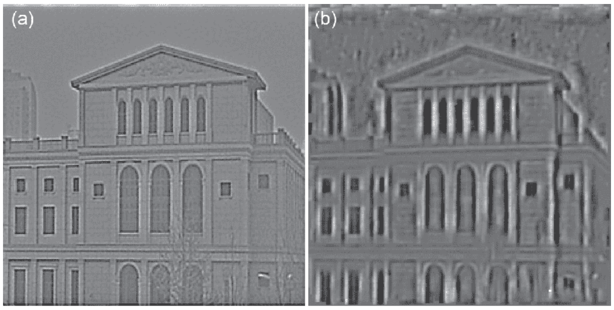

Figure 5 Image reconstruction with the neural activities at the dynamic spiking layer. (a) The original image; (b) the reconstruction image.

4.2 Region classification and detection

At the distribution coding layer, the number of high-lever neurons is set to 150 (i.e., $K = 150$). Sparse spiking outputs of training patches are utilized to derive weights by gradient ascent. The distribution coding layer captures the statistical regularities in the input data. To verify the validity of feature classification, patches in three regions of the test image in Figure 6 are used. Three regions contain distinct objects such as the tree (Region 1), shrub (Region 2) and wood plank (Region 3). The high-order features of different regions are calculated using the two-dimensional projection based on linear discriminant analysis (LDA) [32]. For comparison, the joint outputs of a pair of oriented Gabor filters are described with the scatter plot. The comparison results of capturing statistical regularities in the input data are given in Figures 6 and 7. Well-separated clusters reveals in Figure 6(b), while the joint output in Figure 6(c) are highly overlapping. Results in Figures 6 and 7 indicate that features of simple cells provide no means to distinguish between the regions. Being different from previous hierarchical models [24] that compute the similarity between primary features, image distributions are encoded in our presented method by analyzing the statistical regularities of outputs from primary Gabor receptive fields.

To predict the object region, the reconstruction process is utilized based on neural activities of those high-lever neurons. Then the difference map is computed by comparing the

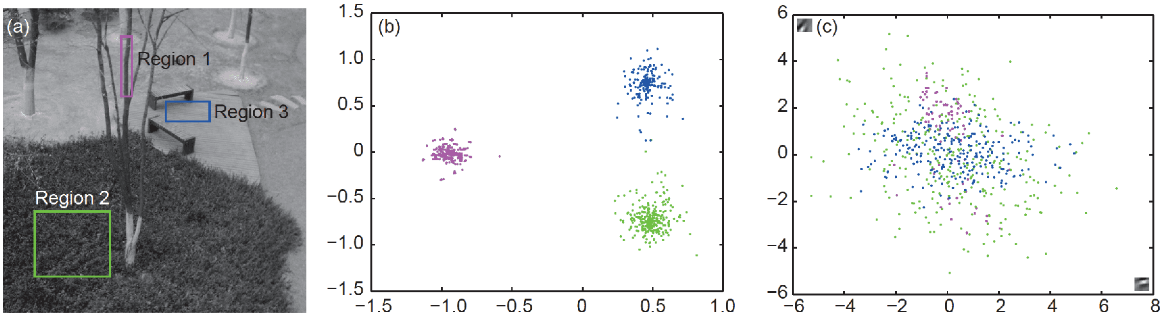

Figure 6 (Color online) Comparison results of capturing statistical regularities in the input data. (a) The original image and three regions (Region 1: tree, Region 2: shrub, Region 3: wood plank); (b) properties of patches using our distribution coding layer; (c) properties of patches using Gabor-like features.

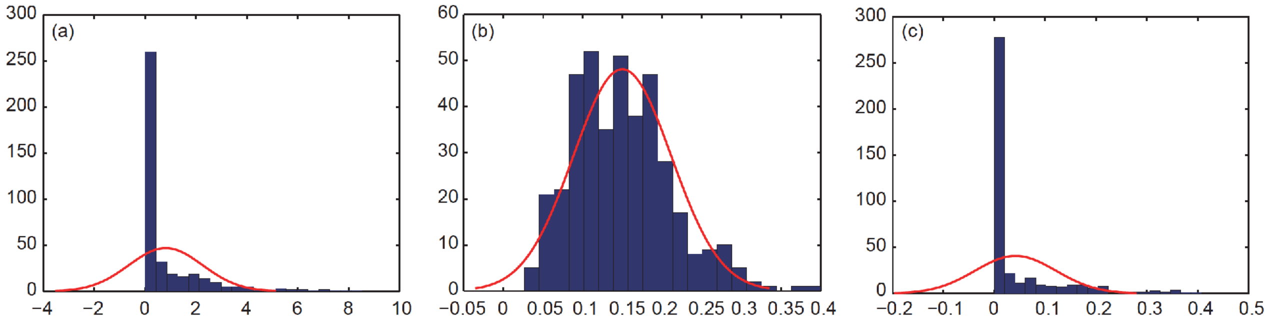

Figure 7 (Color online) Comparison distribution regularities of three regions in Figure 6 with histogram and Gaussian fitting analysis. (a) Region 1; (b) Region 2; (c) Region 3.

original input image and reconstruction image. The row-wise cumulative sum and column-wise cumulative sum of the difference map are calculated, respectively. Thus, the region containing (20, 80) percentiles of row-wise and column-wise total sum is selected as the object region. Experimental results are shown in Figure 8. The input image from Caltech-101 database [33] is sampled and non-overlapping patches are imported to our model.

From the reconstruction image and object region, it is obvious that the object (i.e., car) remain shows a greater difference. Note that our model is trained only on the nature images (background images), so that it can represent the patches from background more accurately than those from the object region. Compared to the results in Figure 8(d) using spectral residual approach [34], our method is superior to segment out the object region from the background.

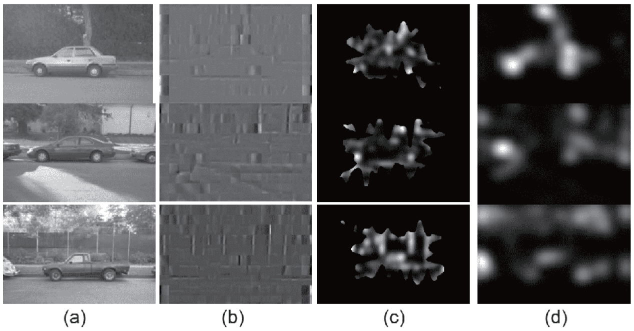

Figure 8 Results of object region detection. (a) The original image; (b) the reconstruction image; (c) our model; (d) spectral residual approach.

5 Conclusion

In this paper, a novel hierarchical statistical distribution model is presented to generalize the higher-level neuron properties and encode the distribution regularity, which is consistent with the input image. Two layers (i.e., dynamic spiking layer and distribution coding layer) are introduced to our hierarchical model to form distributed representations. The dynamic spiking layer learns a sparse code with Gabor-like receptive fields of excitatory neurons decorrelated by those inhibitory neurons. Then the distribution coding layer encodes the statistical distribution of the spiking outputs of those excitatory neurons. With those general set of representations, which are determined by the statistical structure, our model can classify the abstract properties of input data. To demonstrate the feasibility and effectiveness of our method, several experiments on natural scenes are conducted.

Adapted to patches sampled from natural images, parameters including connecting weights and distribution coding weights are estimated, respectively. The experimental results show that our hierarchical model is able to learn nonlinear statistical regularities and recognize similar images with their similar high-order intrinsic representations. Rather than coding the pixel intensities of a patch, our model yields a distributed representation, which includes higher-order spatial relationships for image data. Classification results indicate that our proposed model is able to extract effective representations of patches from different regions, which can be utilized as a reliable feature for the following recognition process.

Although the proposed hierarchical statistical distribution model is able to capture some nonlinear statistical representations, the encoded image structure is still quite low level. We would like to further improve our model and extend it to solve more specific problems such as perceptual invariance or scene segmentation.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 91648205 & 61333004).

1 van Hateren J H. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. J Comp Physiol A, 1992, 171: 157–170
2 Freeman J, Ziemba C M, Heeger D J, et al. A functional and perceptual signature of the second visual area in primates. Nat Neurosci, 2013, 16: 974–981
3 Field G D, Gauthier J L, Sher A, et al. Functional connectivity in the retina at the resolution of photoreceptors. Nature, 2010, 467: 673–677
4 Dan Y, Atick J J, Reid R C. Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. J Neurosci, 1996, 16: 3351–3362
5 Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 1996, 381: 607–609
6 Duan H B, Li P. Bio-inspired Computation in Unmanned Aerial Vehicles. Berlin, Heidelberg: Springer, 2014
7 Duan H, Deng Y, Wang X, et al. Biological eagle-eye: Based visual imaging guidance simulation platform for unmanned flying vehicles. IEEE Aerosp Electron Syst Mag, 2013, 28: 36–45
8 Itti L, Koch C. Computational modelling of visual attention. Nat Rev Neurosci, 2001, 2: 194–203
9 Deng Y M, Duan H B. Avian contrast sensitivity inspired contour detector for unmanned aerial vehicle landing. Sci China Tech Sci, 2017, 60: 1958–1965
10 Duan H, Deng Y, Wang X, et al. Small and dim target detection via lateral inhibition filtering and artificial bee colony based selective visual attention. PLoS ONE, 2013, 8: e72035
11 Krüger N, Janssen P, Kalkan S, et al. Deep hierarchies in the primate visual cortex: What can we learn for computer vision? IEEE Trans Pattern Anal Mach Intell, 2013, 35: 1847–1871
12 Pajares G, Guijarro M, Herrera P J, et al. Combining classifiers through fuzzy cognitive maps in natural images. IET Comput Vis, 2009, 3: 112–123
13 Wang K, Gu X F, Yu T, et al. Classification of hyperspectral remote sensing images using frequency spectrum similarity. Sci China Tech Sci, 2013, 56: 980–988
14 Sohn K, Zhou G, Lee C, et al. Learning and selecting features jointly with point-wise gated boltzmann machines. In: International Conference on Machine Learning. Atlanta, 2013. 217–225
15 Li H, Duan H B. Verification of monocular and binocular pose estimation algorithms in vision-based UAVs autonomous aerial refueling system. Sci China Tech Sci, 2016, 59: 1730–1738
16 Balakrishnan N, Hariharakrishnan K, Schonfeld D. A new image representation algorithm inspired by image submodality models, redundancy reduction, and learning in biological vision. IEEE Trans Pattern Anal Machine Intell, 2005, 27: 1367–1378
17 Spratling M W. Image segmentation using a sparse coding model of cortical area V1. IEEE Trans Image Process, 2013, 22: 1631–1643
18 Lee T S. Image representation using 2D Gabor wavelets. IEEE Trans Pattern Anal Machine Intell, 1996, 18: 959–971
19 Derpanis K G, Gryn J M. Three-dimensional *n*th derivative of Gaussian separable steerable filters. In: IEEE International Conference on Image Processing. Genoa: IEEE, 2005. 553–556
20 Zylberberg J, Murphy J T, DeWeese M R. 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, 2011, 7: e1002250
21 Doi E, Gauthier J L, Field G D, et al. Efficient coding of spatial information in the primate retina. J Neurosci, 2012, 32: 16256–16264
22 King P D, Zylberberg J, DeWeese M R. Inhibitory interneurons decorrelate excitatory cells to drive sparse code formation in a spiking model of V1. J Neurosci, 2013, 33: 5475–5485
23 Gu J, Han H, Li X, et al. Hierarchical spatial pyramid max pooling based on SIFT features and sparse coding for image classification. IET Comput Vision, 2013, 7: 144–150
24 Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nat Neurosci, 1999, 2: 1019–1025
25 Serre T, Wolf L, Bileschi S, et al. Robust object recognition with cortex-like mechanisms. IEEE Trans Pattern Anal Mach Intell, 2007, 29: 411–426
26 Deng Y, Duan H. Hybrid C2 features and spectral residual approach to object recognition. Optik-Int J Light Electron Opt, 2013, 124: 3590–3595
27 Farabet C, Couprie C, Najman L, et al. Learning hierarchical features for scene labeling. IEEE Trans Pattern Anal Mach Intell, 2013, 35: 1915–1929
28 Jarrett K, Kavukcuoglu K, Ranzato M A, et al. What is the best multi-stage architecture for object recognition? In: IEEE International Conference on Computer Vision. Kyoto: IEEE, 2009. 2146–2153
29 Karklin Y, Lewicki M S. A hierarchical Bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. Neural Comput, 2005, 17: 397–423
30 Karklin Y, Lewicki M S. Emergence of complex cell properties by learning to generalize in natural scenes. Nature, 2009, 457: 83–86
31 Faivre O, Juusola M. Visual coding in locust photoreceptors. PLoS ONE, 2008, 3: e2173
32 Mika S, Ratsch G, Weston J, et al. Fisher discriminant analysis with kernels. In: IEEE Conference on Neural Networks for Signal Processing. Madison: IEEE, 1999. 41–48
33 Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. Comput Vis Image Underst, 2007, 106: 59–70
34 Hou X D, Zhang L Q. Saliency detection: A spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis: IEEE, 2007. 1–8