Modeling correlations in spontaneous activity of visual cortex with centered Gaussian-binary deep Boltzmann machines

Nan Wang
Institut für Neuroinformatik
Ruhr-Universität Bochum
Bochum, 44780, Germany
nan.wang@ini.rub.de

Dirk Jancke
Institut für Neuroinformatik
Ruhr-Universität Bochum
Bochum, 44780, Germany
jancke@neurobiologie.rub.de

Laurenz Wiskott
Institut für Neuroinformatik
Ruhr-Universität Bochum
Bochum, 44780, Germany
laurenz.wiskott@rub.de

Abstract

Spontaneous cortical activity – the ongoing cortical activities in absence of intentional sensory input – is considered to play a vital role in many aspects of both normal brain functions [1] and mental dysfunctions [2]. We present a centered Gaussian-binary Deep Boltzmann Machine (GDBM) for modeling the spontaneous activity in early cortical visual area and relate the random sampling in GDBMs to the spontaneous cortical activity. After training the proposed model on natural image patches, we show that the samples collected from the model's probability distribution encompass similar activity patterns as found in the spontaneous activity. Specifically, filters having the same orientation preference tend to be active together during random sampling. Our work demonstrates GDBM is a meaningful model approach for basic receptive field properties and the emergence of spontaneous activity patterns in early cortical visual areas. Besides, we show empirically that centered GDBMs do not suffer from the difficulties during training as GDBMs do and can be properly trained without the layer-wise pretraining as described in [3].

1 Introduction

Spontaneous cortical activity has been studied in various contexts ranging from somato-sensory to visual and auditory perception [4]. These ongoing cortical activities in absence of intentional sensory input are considered to play a vital role in many aspects of both normal brain functions [1], and mental dysfunctions [2]. In [1], the spontaneous activity in the early cortical visual area is reported to have a set of states, several of which resemble cortical representation of orientation, i.e. neurons in the visual cortex having similar orientation preference tend to be active together. One hypothesis about this finding is that these states reflect expectations about the sensory input, which reflects an aspect of approaches such as Bayesian models[5] and generative models. Concretely, a brain learn to synthesize (or generate) representations of sensory inputs.

Previous studies have considered Deep Boltzmann Machines (DBMs) as a potential model framework for modeling the relevant aspects of this generative process and have related the inference in

DBM to the mechanisms of cortical perception[6, 7]. The authors have trained a DBM on binary images and have shown that trained DBMs can qualitatively reproduce several aspects of hallucinations in Charles Bonnet Syndrome. In this work, we have chosen a variant of DBM, a Gaussian-Binary DBM (GDBM), as a model for early cortical visual areas, which extends DBMs to modeling continuous data. In particular, we are interested in modeling the spontaneous activity found in [1] and demonstrate how GDBM reproduces the findings of spontaneous activity in early visual cortex.

As for training GDBMs on the visual inputs, we adapt the centering proposed in [8] to GDBMs by rewriting the energy function as centered states. We show empirically the proposed centered GDBMs can learn features from the natural image patches without the layer-wise pretraining procedure and do not suffer from the difficulties during training of GDBMs as reported in [3].

In Section 2, the proposed centered GDBM is introduced. Then we describe the training procedure in Section 3.1, and show that the centered GDBM can not only learn Gabor-like filters, but also meaningful features in the higher layer. Finally, by considering the samples of the hidden units from the model distribution as spontaneous activity, we demonstrate in Section 3.2 that these random samples present similar activity patterns as the spontaneous visual cortical activity reported in [1].

2 Algorithm

Like Deep Belief Networks (DBN) [9] and deep autoencoders [10], DBMs have been proposed for learning multi-layer representations that are increasingly complex. In particular, DBM incorporates both the top-down messages and the bottom-up passes during the inference for each layer, which gives DBM an advantage in propagating the input uncertainties. DBMs have been applied to many problems and show promising results[11, 12, 13].

To model the natural image patches that have continuous values, a variant of DBMs is required since the original DBM is designed for binary data. There are two common ways to extend DBMs to modeling continuous values. The most common way is to train a Gaussian-binary restricted Boltzmann machine (GRBM) as a preprocessing model and use the output of the trained model as the input data for training a DBM [14]. However, this practice loses the ability to train the model as a whole, i.e. the preprocessing part needs to be trained beforehand. An alternative is a Gaussian-binary deep Boltzmann machine (also known as Gaussian-Bernoulli DBM [3]), in which the binary units in the bottom layer are replaced by the real-value ones as in GRBMs. Moreover, GDBMs have been proved to be a universal approximator. The pitfall of GDBMs is the difficulty in training it [3].

The centering has been proposed in [8] for DBMs and has been shown to improve learning. The idea is to have the output of each unit to all the other units to be centered around zero and take effects only when its activation deviate from its mean activation. In [8], the authors show that centering produces a better conditioned optimization problem. Therefore, we adapted the centering to the GDBM and refer to the new model as centered GDBM. Compared with the training recipe in [3], we find empirically that a centered GDBM can easily be trained even without pre-training phase and is insensitive to the choice of hyper-parameters.

2.1 Centered GDBM

To avoid cluttering, we present a centered GDBM of two hidden layers as an example, although the model can be extended to an arbitrary number of layers. A two-layer centered GDBM as illustrated in Figure 1, consisted of an input layer $\mathbf{X}$ and two hidden layers, has an energy defined as:

$$
\begin{aligned}
E\left(\mathbf{X}, \mathbf{Y}, \mathbf{Z}; \Theta, \mathcal{C}\right): \quad &= \quad \sum_i^M \frac{(X_i - c_{x_i})^2}{2\sigma_i^2} - \sum_{i,j}^{L,M} \frac{(X_i - c_{x_i})w_{ij}(Y_j - c_{y_j})}{\sigma_i^2} - \sum_j^M b_{y_j}(Y_j - c_{y_j}) \\
&\quad - \sum_k^N b_{z_k}(Z_k - c_{z_k}) - \sum_{j,k}^{M,N} (Y_j - c_{y_j})u_{jk}(Z_k - c_{z_k}) \quad (1) \\
&= \quad \frac{(\mathbf{X} - \mathbf{c_x})^T \mathbf{\Lambda}^{-1}(\mathbf{X} - \mathbf{c_x})}{2} - (\mathbf{X} - \mathbf{c_x})^T \mathbf{\Lambda}^{-1} \mathbf{W}(\mathbf{Y} - \mathbf{c_y}) \\
&\quad -(\mathbf{Y} - \mathbf{c_y})^T \mathbf{b_y} - (\mathbf{Z} - \mathbf{c_z})^T \mathbf{b_z} - (\mathbf{Y} - \mathbf{c_y})^T \mathbf{U}(\mathbf{Z} - \mathbf{c_z}), \quad (2)
\end{aligned}
$$

where $\mathbf{Y}$ represents the first hidden layer, and $\mathbf{Z}$ denotes the second hidden layer. L, M, N are the dimensionality of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ respectively. $\Theta := \{\mathbf{W}, \mathbf{U}, \mathbf{b_y}, \mathbf{b_z}\}$ denotes the parameters trained for maximizing the loglikelihood. $\mathcal{C} := \{\mathbf{c_x}, \mathbf{c_y}, \mathbf{c_z}\}$ represents the offsets for centering the units and is set to the mean activation of each unit. $\mathbf{\Lambda}$ is a diagonal matrix with the elements σ_i^2. The probability of any given state $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ in the DBM is

$$P(\mathbf{x}, \mathbf{y}, \mathbf{z}; \Theta, \mathcal{C}) : \quad = \quad \frac{1}{\mathcal{Z}(\Theta)} \exp\left(-E(\mathbf{x}, \mathbf{y}, \mathbf{z}; \Theta, \mathcal{C})\right). \tag{3}$$

Here $\mathcal{Z}(\Theta)$ is the partition function depended on the parameters of the model. Inference in centered GDBM is simple, because the states of the units in each layer are independent of the other ones in the same layer given the adjacent upper and lower layer.

$$P(X_i|\mathbf{y}) \quad = \quad \mathcal{N}\left(X_i; \mathbf{w}_{i*}(\mathbf{y} - \mathbf{c_y}) + c_{x_i}, \sigma_i^2\right), \tag{4}$$

$$P(Y_j = 1|\mathbf{x}, \mathbf{z}) \quad = \quad f\left((\mathbf{x} - \mathbf{c_x})^T \mathbf{w}_{*j} + \mathbf{u}_{j*}(\mathbf{z} - \mathbf{c_z}) + b_{y_j}\right), \tag{5}$$

$$P(Z_k = 1|\mathbf{y}) \quad = \quad f\left((\mathbf{y} - \mathbf{c_y})^T \mathbf{u}_{*k} + b_{z_k}\right), \tag{6}$$

where $f(\cdot)$ is a sigmoid function and $\mathcal{N}(\cdot; \mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. $\mathbf{w}_{i*}$ and $\mathbf{w}_{*j}$ denote the ith row and the jth column of matrix $\mathbf{W}$. $\mathbf{u}_{j*}$ and $\mathbf{u}_{*k}$ are defined correspondingly.
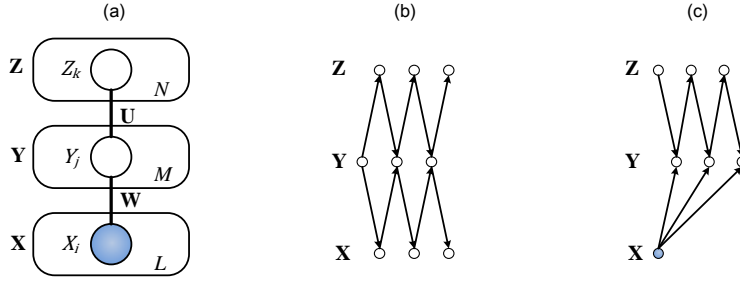

Figure 1: (a) A graphical description of a two-layer GDBM. (b) Illustration of the sampling procedure for estimating model dependent expectation. (c) Illustration of the sampling procedure for estimating data dependent expectation.

For training a centered GDBM, the objective is to maximize the loglikelihood $\hat{\ell}$, of which the partial derivative for each parameter $\theta \in \Theta$ is

$$\frac{\partial \hat{\ell}}{\partial \theta} \quad = \quad \left\langle \frac{\partial(-E(\mathbf{x}, \mathbf{y}, \mathbf{z}))}{\partial \theta} \right\rangle_{data} - \left\langle \frac{\partial(-E(\mathbf{x}, \mathbf{y}, \mathbf{z}))}{\partial \theta} \right\rangle_{model}, \tag{7}$$

where $\langle \cdot \rangle_{data}$ and $\langle \cdot \rangle_{model}$ represent the expectation with respect to the data and the model distribution, respectively. By using the mean-field approximation, we can estimate the data-dependent expectation. And the model-dependent expectation is usually approximated by using persistent Markov Chains with Gibbs sampling. See [14] for details.

As for the offsets, we adjust them along the training procedure with a moving average. To ensure the units are centered from the beginning of training, $\mathbf{c_y}$ and $\mathbf{c_z}$ are set to $f(\mathbf{b_y})$ and $f(\mathbf{b_z})$, respectively. And $\mathbf{c_X}$ is set to the data mean $\langle \mathbf{x} \rangle_{data}$. In general, we follow the learning algorithm in [8] with several modifications as shown in Algorithm 1.

3 Experiments and results

3.1 Learning from natural image patches

We applied the centered GDBM to image patches of 32×32 pixels[1], taken randomly from the van Hateren natural image dataset [15]. The patches were firstly whitened by principal component analysis and reduced to the dimensionality of 256 in order to avoid aliasing [16].

---

[1]We generated 60,000 image patches by randomly taken patches of 32×32 pixels from 2,000 natural images. A subset of size 50,000 was used for training. The rest was used for testing.

Algorithm 1 Training algorithm for centered GDBMs

1: $Initialize\ \mathbf{W}, \mathbf{U}$ $\quad\big(i.e.\ w_{ij} \sim \mathcal{U}\big[-\sqrt{\frac{6}{L+M}}, \sqrt{\frac{6}{L+M}}\big],\ u_{jk} \sim \mathcal{U}\big[-\sqrt{\frac{6}{M+N}}, \sqrt{\frac{6}{M+N}}\big]\big)$

2: $Initialize\ \mathbf{\Lambda}$ $\quad\big(i.e.\ \sigma_i^2 \sim \mathcal{N}(0.5, 0.01)\big)$

3: $Initialize\ \mathbf{b_y}, \mathbf{b_z}$ $\quad\big(i.e.\ b_{y_j}, b_{z_k} \sim \mathcal{N}(-4.0, 0.01)\big)$

4: $Initialize\ \mathbf{c_x}, \mathbf{c_y}, \mathbf{c_z}$ $\quad\big(i.e.\ \mathbf{c_x} \leftarrow <\mathbf{x}>_{data}, \mathbf{c_y} \leftarrow f(\mathbf{b_y}), \mathbf{c_z} \leftarrow f(\mathbf{b_z})\big)$

5: **loop**

6: $\quad \mathbf{y}_{model} \leftarrow \mathbf{c_y}$

7: $\quad$ **for all** batches $\mathbf{x}_{data}$ **do**

8: $\quad\quad \mathbf{z}_{data} \leftarrow \mathbf{c_z}$

9: $\quad\quad$ **loop**

10: $\quad\quad\quad \mathbf{y}_{data} \leftarrow P(\mathbf{Y}|\mathbf{x}_{data}, \mathbf{z}_{data})$

11: $\quad\quad\quad \mathbf{z}_{data} \leftarrow P(\mathbf{Z}|\mathbf{y}_{data})$

12: $\quad\quad$ **end loop** until stop criteria is met

13: $\quad\quad$ **loop**

14: $\quad\quad\quad \mathbf{z}_{model} \sim P(\mathbf{Z}|\mathbf{y}_{model})$

15: $\quad\quad\quad \mathbf{x}_{model} \sim P(\mathbf{X}|\mathbf{y}_{model})$

16: $\quad\quad\quad \mathbf{y}_{model} \sim P(\mathbf{Y}|\mathbf{x}_{model}, \mathbf{z}_{model})$

17: $\quad\quad$ **end loop** until stop criteria is met

18: $\quad\quad \mathbf{c_y} \leftarrow (1-\nu) \cdot \mathbf{c_y} + \nu \cdot \langle \mathbf{y}_{data} \rangle$

19: $\quad\quad \mathbf{c_z} \leftarrow (1-\nu) \cdot \mathbf{c_z} + \nu \cdot \langle \mathbf{z}_{data} \rangle$

20: $\quad\quad \mathbf{W} \leftarrow \mathbf{W} + \eta\big(\langle(\mathbf{x}_{data} - \mathbf{c_x})\mathbf{\Lambda}^{-1}(\mathbf{y}_{data} - \mathbf{c_y})^T\rangle - \langle(\mathbf{x}_{model} - \mathbf{c_x})\mathbf{\Lambda}^{-1}(\mathbf{y}_{model} - \mathbf{c_y})^T\rangle\big)$

21: $\quad\quad \mathbf{U} \leftarrow \mathbf{U} + \eta\big(\langle(\mathbf{y}_{data} - \mathbf{c_y})(\mathbf{z}_{data} - \mathbf{c_z})^T\rangle - \langle(\mathbf{y}_{model} - \mathbf{c_y})(\mathbf{z}_{model} - \mathbf{c_z})^T\rangle\big)$

22: $\quad\quad \sigma_i \leftarrow \sigma_i + \eta_\sigma\big(\langle\frac{(x_{i,data} - c_{x_i})^2}{\sigma^3} - \frac{2(x_{i,data} - c_{x_i})\mathbf{w}_{i*}(\mathbf{y}_{data} - \mathbf{c_y})}{\sigma^3}\rangle - \langle\frac{(x_{i,model} - c_{x_i})^2}{\sigma^3} - \frac{2(x_{i,model} - c_{x_i})\mathbf{w}_{i*}(\mathbf{y}_{model} - \mathbf{c_y})}{\sigma^3}\rangle\big)$

23: $\quad\quad \mathbf{b_y} \leftarrow \mathbf{b_y} + \eta\big(\langle\mathbf{y}_{model}\rangle - \langle\mathbf{y}_{model}\rangle\big) + \nu\mathbf{W}^T\mathbf{\Lambda}^{-1}(\mathbf{x}_{data} - \mathbf{c_x}) + \nu\mathbf{U}\big(\mathbf{z}_{data} - \mathbf{c_z}\big)$

24: $\quad\quad \mathbf{b_z} \leftarrow \mathbf{b_z} + \eta\big(\langle\mathbf{z}_{data}\rangle - \langle\mathbf{z}_{model}\rangle\big) + \nu\mathbf{U}^T\big(\mathbf{y}_{data} - \mathbf{c_y}\big)$

25: $\quad\quad \mathbf{y}_{model} \leftarrow P(\mathbf{Y}|\mathbf{x}_{model}, \mathbf{z}_{model})$

26: $\quad$ **end for**

27: **end loop**

Afterwards, we trained a centered GDBM with 256 visible units and two hidden layers. There were 900 units in the first hidden layer and 100 units in the second hidden layer[2]. Despite the difficulties during training of GDBMs, we found empirically the centered GDBM can be trained much easier. Without the layer-wise pretraining, the centered GDBM did not suffer from the issue that the higher layer units are either always inactive or always active as reported in [3]. Since any centered GDBM can be reparameterized as a normal GDBM [18], this may imply that the centering in GDBM plays an important role in the optimization procedure.

After training, the centered GDBM had oriented, Gabor-like filters in the first hidden layer (Figure 2a). Most of the units in the second hidden layer had either strong positive or negative connections to the filters in the first layer that have similar patterns. As shown in Figure 2b, the filters having strong connections to the same second-layer units either have the similar orientation or the same location. The results further suggest that the model learned to encode more complex features such as contours, angles, and junctions of edges. These results resemble the properties of the neurons in V1 of visual cortex and imply that centered GDBM is a meaningful model for basic receptive field properties in early cortical visual areas. Despite the resemblance of these results to those from sparse DBNs [19], sparse DBNs show worse match to the biological findings in [1]. A quantitative comparison between centered GDBMs and sparse DBNs is still open for future studies.

---

[2]The size of the hidden layers were chosen to get a good model of the spontaneous cortical activity as described in Section 3.2. The training procedure started with a learning rate of 0.03 and a momentum of 0.9, which is annealed to 0.001 and 0.0, respectively. But the standard deviation σ_i have a different learning rate, which is only one-tenth of the other's learning rate [17]. Neither weight decay nor sparse penalty is used during training. Mini-batch learning is used with a batch size of 100. The updating rate for centering parameters was 0.001. The training procedure was stopped when the reconstruction error stopped decreasing.
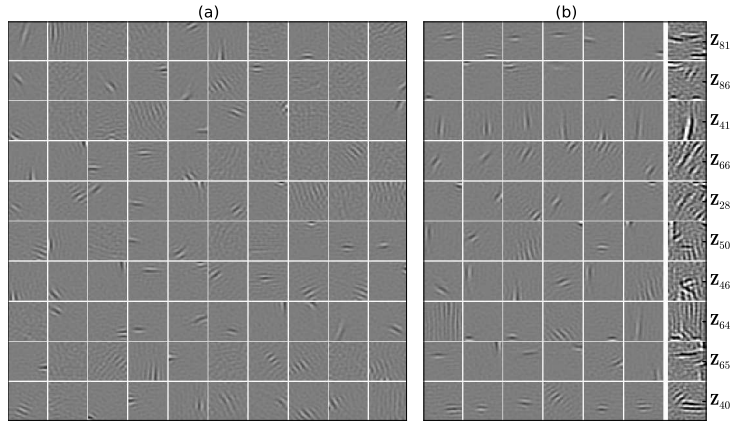
Figure 2: (a) 100 randomly selected first-layer filters learned from the natural image. (b) The leading six columns visualize the first-layer filters that have strongest connections to one selected second-layer unit. The filters are arranged from left to right in descending order by the absolute value of their weight to the selected unit. The last column depicts the weighted sum of the six strongest-connected filters, which can be considered as an approximation of the receptive fields of the selected second-layer units.

3.2 Comparing with biological experiments

After successful training of centered GDBM, we quantitatively analyzed the results of the trained model[3] with the methods used in [1]. For conducting the measurements, we made two basic assumptions. Firstly, since only the first-layer filters of centered GDBM present strong orientation preferences, we assumed these filters correspond to the visual cortical neurons recorded in the literature. Secondly, the samples of the hidden units collected according to the model distribution was assumed to be the counterpart of the neuron's activity.

3.2.1 Generating orientation maps

To compare with the original experiments, we firstly generated full-field gratings as input data and measured the response of the centered GDBM model. We collected the responses of the first-layer hidden units to eight orientations from $0°$ to $157.5°$ with a difference of $22.5°$.

The amplitude of the gratings was chosen such that the average norm of the input stimuli is the same as that of the natural image patches before whitening. For each orientation, the grating stimuli of various frequencies and phases was fed to the model. To be consist with the learning procedure, we used the mean-field variational approximation to approximate the responses of the model to each stimulus, which can also be estimated by using the Gibbs sampling with the visible units clamped to the input. An illustration of the sampling procedure in given in Figure 1(c). After collecting the responses, the average response over all the stimuli of each orientation were calculated and considered to be the model's response to the corresponding orientation. These activity patterns correspond to the *single-condition orientation maps* in [1]. Figure 3 (top) visualizes the most active filters in the single-condition orientation maps of four selected orientations.

3.2.2 Generating spontaneous frames

To simulate spontaneously emerging cortical activity, we sampled the states of the trained model starting from a random initialization. Neither the visible nor the hidden unit was clamped to fixed value during sampling. Figure 1(b) shows an illustration of the sampling procedure. This sampling procedure generated samples from the model's distribution $P(\mathbf{Y}, \mathbf{X}, \mathbf{Z})$, which were considered as

---

[3]Considering the authors of the original paper [1] only presented the results from one hemisphere of a selected cat, here we only present the results of one centered GDBM. However, all the centered GDBMs trained in our experience showed consistent results.
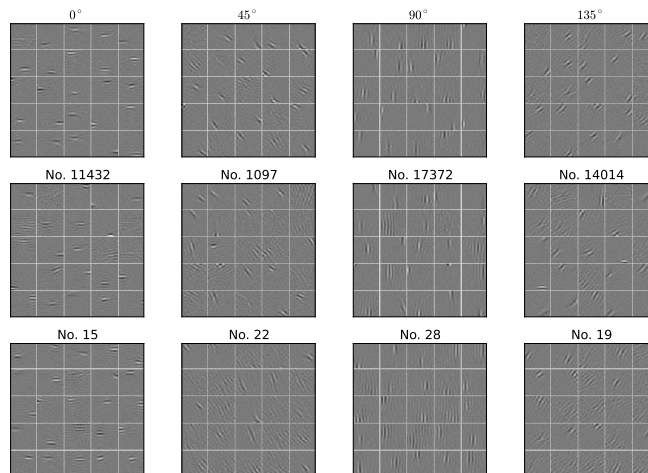
Figure 3: (top) The 25 most active filters in each of the single-condition orientation maps. For each map the filters are arranged in a descending order of activity level. (middle) The same but using the filters in the spontaneous frames that are best correlated to the corresponding single-condition maps. (bottom) The same but using the filters of the best-correlated nodes in the self-organizing map.

the expected states of the units in the model. This differs from the sampling procedure in [6], where the visible units were clamped to zeroes and the hidden biases were adapted during sampling. The difference is discussed at length in Section 4.

In total, there were 100 Markov chains running Gibbs sampling for 2,000 iterations. For each Markov chain, the initial states of the first-hidden-layer units were set to be active with a probability that is equal to the average $P(\mathbf{Y}|\mathbf{x}_{data}, \mathbf{z}_{data})$ over the training data. By recording the samples every 10 sampling steps, we collected 20,000 samples of $\mathbf{Y}$ and considered the corresponding probabilities $P(\mathbf{Y}|\mathbf{x}, \mathbf{z})$ as *spontaneous frames*. Such a sampling procedure is referred to as a session. In the experiments we repeated the session for many times with every trained model.

3.2.3 Correlation between spontaneous frames and orientation maps

To establish the similarity between the single-condition orientation maps and the spontaneous frames, we calculated the spatial correlation coefficients between them. Figure 4(a, red) presents an example of the distribution of these correlation coefficients for four selected orientations[4]. In order to show that these correlations are stronger than expected by chances, we generated random activity patterns of the first hidden layer with the same probability as the one used for initializing the Markov chains, and the same correlation coefficients were calculated with these random activity patterns as shown in Figure 4(a, blue). Specifically, the maximal correlation coefficients is 0.50 ± 0.06 whereas the correlation coefficients between the spontaneous frames and the random generated patterns seldom reach 0.2. The same observation were made in Figure 2 in [1].

In [1], the authors observed that there were more spontaneous activity patterns corresponding to the cardinal orientations than to the oblique ones. Therefore we further calculated the orientation preference of these spontaneous frames. By this point, only the spontaneous frames that are significantly correlated were chosen. As in the biological experiment [1], we chose a significance level of $P < 0.01$, resulting in a threshold of $|0.182|$ and a selection of $18 \pm 9\%$ spontaneous frames. We then calculated the orientation preference of these frames by searching the orientation that is maximally correlated for each frame. Figure 4(b) plots the relative occurrences of the different orientation preferences together with the maximal correlation coefficients. The results match those from the cats' visual cortex in [1] fairly well, i.e. the spontaneous frames corresponding to the car-

---

[4]The following results were collected in a single session, but the results are consistent across simulation runs. In all sessions, the similar observation as shown in Figure 2–4 can be observed. The only difference is the shape of the curves in Figure 4b might vary in different sessions. However, the dominance of the cardinal orientation is always present.

dinal orientations emerged more often than those corresponding to the oblique ones and the former also have larger correlation coefficients.
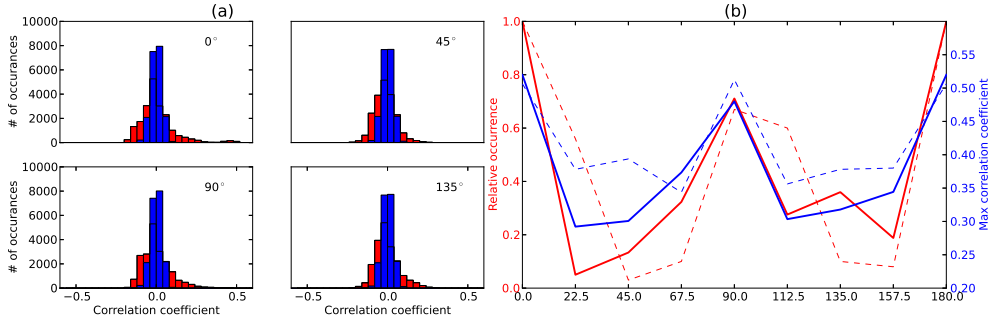

Figure 4: (a) Red, example distribution of the correlation coefficients between the spontaneous frames and the four selected single-condition orientation maps. Blue, the same, using random generated activity patterns with the same average probability. The threshold for significant correlation with the significance level of $P < 0.01$ is $|0.182|$. (b) The relationship between the relative occurrence of the different orientation preferences (red) and the maximal correlation coefficient (blue). The results from the centered GDBM are plotted by solid lines in comparison with the results in [1] plotted by dotted lines. The relative occurrence is calculated relative to the occurrence of the horizontal orientation. The figure is adapted to Figure 3b in [1].

Next we compared the most active filters in the spontaneous frames with those in the single-condition orientation map. Figure 3 (middle) visualizes these filters for four spontaneous frames that are best correlated with the selected single-condition orientation maps. These filters demonstrate similar features as those in the corresponding orientation maps shown in Figure 3 (top). Thus the result supports the similarity between the spontaneous frames and the orientation maps in centered GDBMs.

3.2.4 Learning SOM from the spontaneous frames

We followed the methods in [1] and applied Self-Organizing Map (SOM) algorithm [20] to the spontaneous frames in order to study the intrinsic structures of these spontaneous activities. We trained a SOM on the 20,000 spontaneous frames collected from a single session. The SOM projects the spontaneous frames onto 40 nodes that were arranged on a 1-D circle. See [1] for details of training. After training the SOM, we examined the correlation between the weight vectors of the 40 nodes and the single-condition orientation maps. Figure 3 (bottom) illustrates the most active filters in the weight vectors of four nodes that are best correlated with the selected single-condition orientation maps. The remarkable resemblance between these filters and those in the single-condition orientation maps suggests that the spontaneous frames encompass several states of the hidden variables in the first layer, which resemble the model's representation of orientation.

4 Discussion

In this work, we present a variant of DBMs, centered Gaussian-binary deep Boltzmann machines (GDBM) for modeling spontaneous activity in visual cortex. We show empirically that the proposed centered GDBM does not require the layer-wise pretraining procedure by virtue of centering the units' activation. An intuitive explanation for the success of centering is that centering prevents the units from being either always active or inactive and thus forming a bias for the other units. Because the centering offsets keep the output of each unit to zero unless its activation differs from its mean activation. Formally, the authors in [8] show that the centering produces improved learning conditions. Nevertheless, the authors show that a centered DBM can be reparameterized as a normal DBM with the same probability distribution in [18]. Therefore, centered GDBM can be considered as a GDBM with better optimization conditions. In other word, the results presented in this work can also be expected in a GDBM, despite its optimization difficulties.

The results of modeling natural image patches with centered GDBM in this work suggest centering helps to overcome the commonly observed difficulties during training. In addition, we also trained a centered GDBM on the Olivetti face dataset [21] following the same setting as in [3]. We achieved a reconstruction error on the test set of 41.6 ± 0.40 compared to the result of about 40 in [3]. Although the results are not sufficient to claim the superiority of centered GDBMs over the proposed training algorithm in [3], we argue that a centered GDBM is an alternative to GDBM.

Our main contribution is to consider centered GDBMs as a model for the early cortical visual areas and to reproduce the findings of the spontaneous cortical activity in[1]. Compared to previous work [6] of modeling cortical activity with DBMs, we extend the model from binary data to continuous data with the proposed centered GDBMs. This extension makes it possible to use centered GDBMs to model other cortical activities besides vision.

For modeling spontaneous activity, we've also tested other models, i.e. GRBMs and DBN. None of them can match the results in [1] as well as centered GDBMs. The correlation between the spontaneous frames and the orientation maps are significantly less than in centered GDBMs, where $18 \pm 9\%$ of the frames are significantly correlated to the maps compared to 2% or less in other models. A possible explanation is that centered GDBMs is the only model using both top-down and bottom-up interactions during inference. In comparison, GRBMs and DBN only use either the bottom-up or the top-down information. On one hand, this suggests that the observed spontaneous activity is a result of interactions between incoming stimulation and feedback from higher areas. On the other hand, our results also predicts the states of spontaneous activity found in [1] are the result of the interactions within the early cortical visual areas. Because a centered GDBM with two hidden layers is enough to faithfully reproduce the reported results.

In [6], the authors clamped the visible units to zero in order to model profound visual impairment or blindness. This is equivalent to clamping the visible units of a centered GDBM to the centering offsets, which leads to zero inputs to the first hidden layer from bottom-up. However, in this work we sampled the visible units freely during generating spontaneous frames. The main reason is that spontaneous cortical activity is the ongoing activity in the absence of intentional sensory input, which does not exclude incoming stimuli [22]. The authors also reported no difference in their findings when a uniform grey screen is used instead of having the room darkened [1]. From the model's persepctive, our sampling procedure is supposed to approximate samples from the model's prior distribution $P(\mathbf{Y})$, which are the expected states of the first-hidden-layer units without any knowledge of bottom-up or top-down information. As a result, our results suggest that the spontaneous activity patterns appear to be an expectation of internal states in brain and indicate that a brain might learn to generate such expectations as a generative model. Moreover, we observed that the correlation between the spontaneous frames and the orientation maps disappeared when we clamped either the visible or the second hidden layer units to the centering offsets during generating the spontaneous frames. This supports our prediction that generating the observed spontaneous activity needs both incoming stimulation and feedback from higher areas.

5   Conclusion

We present centered GDBMs by applying the centering to GDBMs and used the centered GDBM as a model of spontaneous activity in early cortical visual areas. Our work extend the previous work of using DBM for modeling cortical activities to continuous data. The results demonstrate that a centered GDBM is a meaningful model approach for basic receptive field properties and the emergence of spontaneous activity patterns in early cortical visual area and has the potential to give further insights to spontaneous activity in brain. Our work also show empirically centered GDBMs can be properly trained without layer-wise pretraining.

References

[1] Tal Kenet, Dmitri Bibitchkov, Misha Tsodyks, Amiram Grinvald, and Amos Arieli. Spontaneously emerging cortical representations of visual attributes. *Nature*, 424:954–956, 2003.

[2] W Burke. The neural basis of Charles Bonnet hallucinations: a hypothesis. *Journal of Neurology, Neurosurgery & Psychiatry*, 73(5):535–541, 2002.

[3] KyungHyun Cho, Tapani Raiko, and Alexander Ilin. Gaussian-Bernoulli deep Boltzmann machines. In *the proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2013.

[4] Rémy Lestienne. Spike timing, synchronization and information processing on the sensory side of the central nervous system. *Progress in Neurobiology*, 65(6):545 – 591, 2001.

[5] Iris Vilares and Konrad Kording. Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, 1224(1):22–39, 2011.

[6] David Reichert, Peggy Series, and Amos Storkey. Hallucinations in Charles Bonnet syndrome induced by homeostasis: a deep Boltzmann machine model. In *the proceedings of the Conference on Neural Information Processing Systems (NIPS)*, pages 2020–2028, 2010.

[7] David Reichert, Peggy Series, and Amos Storkey. Charles Bonnet syndrome: Evidence for a generative model in the cortex? *PLoS Comput Biol*, 9(7):e1003134, 07 2013.

[8] Grégoire Montavon and Klaus-Robert Müller. Deep Boltzmann machines and the centering trick. In *Neural Networks: Tricks of the Trade*, volume 7700 of *Lecture Notes in Computer Science*, pages 621–637. Springer Berlin Heidelberg, 2012.

[9] Geoffrey Hinton and Ruslan Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[10] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *the proceedings of the International Conference on Machine Learning (ICML)*, pages 1096–1103. ACM, 2008.

[11] Nitish Srivastava and Ruslan Salakhutdinov. Multimodal learning with deep Boltzmann machines. In *the proceedings of the Conference on Neural Information Processing Systems (NIPS)*, pages 2231–2239, 2012.

[12] Nitish Srivastava, Ruslan Salakhutdinov, and Geoffrey Hinton. Modeling documents with deep Boltzmann machines. In *the proceedings of the Internation Conference on Uncertainty in Artificial Intelligence (UAI)*, 2013.

[13] Ruslan Salakhutdinov, J Tenenbaum, and Antonio Torralba. Learning with hierarchical-deep models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1958–1971, 2013.

[14] Ruslan Salakhutdinov and Geoffrey Hinton. Deep Boltzmann machines. In *the proceedings of the internation conference on Artificial Intelligence and Statistics (AISTATS)*, volume 5, pages 448–455, 2009.

[15] A. van der Schaaf and J.H. van Hateren. Modelling the power spectra of natural images: Statistics and information. *Vision Research*, 36(17):2759 – 2770, 1996.

[16] Aapo. Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. Adaptive and Learning Systems for Signal Processing, Communications and Control Series. Wiley, 2004.

[17] Alex Krizhevsky. Learning multiple layers of features from tiny images. Master's thesis, University of Toronto, Toronto, 4 2009.

[18] Jan Melchior, Asja Fischer, Nan Wang, and Laurenz Wiskott. How to center binary restricted Boltzmann machines. *arXiv preprint arXiv:1311.1354*, 2013.

[19] Honglak Lee, Chaitanya Ekanadham, and Andrew Y.Ng. Sparse deep belief net model for visual area v2. In *the proceedings of the Conference on Neural Information Processing Systems (NIPS)*. MIT Press, 2007.

[20] Teuvo Kohonen. *Self-Organizing Maps*. Physics and astronomy online library. Springer Berlin Heidelberg, 2001.

[21] F. S. Samaria and A. C. Harter. Parameterisation of a stochastic model for human face identification. In *the proceedings of IEEE Workshop on Applications of Computer Vision*, pages 138–142, 1994.

[22] Amos Arieli, Alexander Sterkin, Amiram Grinvald, and Ad Aertsen. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273:1868–1871, 1996.