Multinomial Bayesian model of Early Visual Cortex

Haruo Hosoya\textsuperscript{1,2}
\begin{itemize}
\item \textsuperscript{1} Brain Science Institute, RIKEN
\item \textsuperscript{2} Presto, Japan Science and Technology Agency
\end{itemize}
E-mail: hahosoya@brain.riken.jp

\textbf{Abstract}—Using a multi-layer multinomial Bayesian network, we study the interplay between Bayesian inference and natural image learning in relation to receptive field properties of early visual cortex.

\textbf{Keywords}—Bayesian inference, natural image learning

\section{Introduction}
In this work, we study a possible interplay between Bayesian inference \cite{3,7} and natural image learning \cite{6} in relation to receptive field properties of early visual cortex. We particularly focus on a multi-layer multinomial Bayesian network \cite{7} and show that maximal-likelihood learning using sampling-based Bayesian inference can yield a hierarchy of representations similar to the receptive field properties of V1 simple cells and V2 cells, in which multinomial variables are important in enforcing sparsity for learning efficient representations. Furthermore, we show that, in the network with such learned representations, Bayesian inference can reproduce a non-classical receptive field property known as filling-in \cite{5}, in which feedback from the V2-like layer is crucial. Much previous work in the same direction has focused on predictive coding \cite{8}. However, the important difference is that feedforward signals represent prediction matches in our case while these represent prediction errors in their case; in the latter case, filling-in appears difficult to explain.

\section{Computational Model}
We consider a Bayesian network consisting of a set $X$ of variables (nodes), divided into visible continuous nodes $V$ and hidden multinomial nodes $H$. Each multinomial variable represents a discrete-finite feature space much like hypercolumns gathering dozens of minicolumns. As usual, we postulate the factorizable joint distribution $P(X|w) = \prod_{X \in X} P(X|\text{pa}(X), w)$, where each conditional probability distribution $P(X|\text{pa}(X), w)$ (where $\text{pa}(X)$ denotes the set of $X$’s parent nodes) is defined in terms of the parameters $w$: for a hidden variable $X$,

$$P(x|u_1, \ldots, u_p, w) = \frac{\exp(\sum_{j=1}^p w_{x,u,j}x_j)}{\sum_{x' \in \text{st}(X)} \exp(\sum_{j=1}^p w_{x',u,j}x'_j)}$$

where we assume a weight matrix $\{w_{x,u}\}_{x \in \text{st}(X), u \in \text{st}(U)}$ of real values for each parent variable $U$ (where $\text{st}(X)$ denotes the set of $X$’s states); for a visible variable $X$,

$$P(x|u_1, \ldots, u_p, w) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x - \sum_{j=1}^p w_{x,u,j}u_j)^2}{2\sigma^2} \right)$$

where we assume a weight vector $\{w_{x,u}\}_{u \in \text{st}(U)}$ of real values for each parent variable $U$ with a fixed hyperparameter $\sigma$; the prior $P(X|w)$ of each root node is uniform.

Our learning scheme can be derived as a stochastic gradient method on the expected log likelihood $E^*[\log P(V|w)]$, where $E^*[\cdot]$ denotes the expectation taken under the external data distribution $P^*(V)$. The actual algorithm iterates the following two steps alternately. In the inference step, we take a random sample $\hat{y}$ from the distribution $P^*(V)$ and then take a random sample $\hat{h}$ from the posterior distribution $P(H|\hat{y}, w)$ by using Gibbs-sampling \cite{2}; we repeat this process several times. In the update step, for each pair of a hidden node $X$ and a parent $U \in \text{pa}(X)$, we update each weight $w_{x,u}$ by

$$\Delta w_{x,u} \propto (\delta_{u,\hat{u}}(\delta_{x,\hat{x}} - P(x|\text{pa}(x)))$$

Also, for each pair of a visible node $X$ and a parent $U \in \text{pa}(X)$, we update each weight $w_{X,u}$ by

$$\Delta w_{X,u} \propto \left\langle \frac{1}{\sqrt{2\pi}\sigma} \delta_{u,\hat{u}}(\hat{x} - \sum_{u_j \in \text{pa}(x)} w_{X,u_j}) \right\rangle$$

In both, $\delta_{\cdot,\cdot}$ denotes Kronecker’s delta, $\langle \cdot \rangle$ is the average over the samples taken in the inference step, $\hat{u}$ and $\hat{x}$ each denote the states of $U$ and $X$ in each sample, and $\text{pa}(x)$ denotes the set of the states of the parent nodes $\text{pa}(X)$ in each sample. A formal derivation is given in the full version \cite{4}.}

\section{Simulation}
We employed a three-layer Bayesian network consisting of layer-0 with $24 \times 24$ visible nodes, layer-1 with $3 \times 3$ hidden nodes each having 100 discrete states, and layer-2 with 4 nodes each having 100 states. Each
layer-2 node is connected to all layer 1 nodes, while each layer-1 node is connected to a set of $12 \times 12$ layer-0 nodes such that two adjacent layer-1 nodes have overlapping sets of subordinates.

After training the network in a layer-by-layer manner using the Olshausen’s dataset of gray-scale natural images (preprocessed with the whitening and low-pass filter) with $1/\sigma^2 = 4.5$, we obtained a hierarchy of representations as illustrated in Figure 1. The basis images in layer-1 resembled oriented Gabor filters qualitatively similar to typical receptive field shapes of V1 simple cells. Quantitative comparison was also made and some similarities were found [4]. In layer-2, some units combined similar orientations and thus looked like collinear or parallel lines, but other units combined orientations that were slightly or largely different and thus looked like contours or angles. Since such properties were qualitatively similar to the receptive field shapes of V2 cells [1], we conducted quantitative analysis on the tendencies of differences between the combined orientations and found that near-zero differences were the most typical and larger differences were progressively less frequent [4], which is consistently with the physiological data (Figure 2).

Further, Bayesian inference (based on Gibbs-sampling) performed on the network with such learned representations exhibited complex non-classical receptive field properties, namely, filling-in. Filling-in is the phenomenon that one perceives within the blind spot visual attributes similar to the surrounds even though it receives no retinal input. The particular task we study here is inspired by the neural correlates discovered in V1 [5], illustrated in Figure 3(a). In this, when a horizontal bar stimulus is short and the right end stays within the blind spot (A), the target unit (right-most unit of node 1b) only weakly detects the orientation based on the bottom-up signals from the partial retinal input; however, when the bar is long enough to exceed the blind spot boundary, the unit clearly detects the orientation since layer-2 nodes now recognize the orientation in the surround and thus exert a strong top-down influence on the target node. Indeed, the simulation reproduced the predicted behavior (Figure 3(b)). Note that layer-2 units representing combinations of similar orientations provide the key ground that the orientation inside the blind spot is likely to be similar to the one outside.

References