Translation-Invariant Neural Responses as Variational Messages in a Bayesian Network Model

Takashi Sano and Yuuji Ichisugi

The National Institute of Advanced Industrial Science and Technology (AIST)
{t.sano,y-ichisugi}@aist.go.jp

Abstract. In this paper, we propose the following interpretation: if a Bayesian network has acquired translation invariance in input images, its feedback messages from higher layers to lower layers can be interpreted as the response of complex cells in the visual system. To examine our proposal's validity, we trained a Bayesian network to acquire translation invariance using the standard belief propagation algorithm, and confirmed its feedback messages were translation invariant and thus they can be interpreted as the response of complex cells. Unlike previous studies, our model does not require specially prepared random variables. Furthermore, our model only uses the standard belief propagation algorithm. Therefore we believe that our model is more natural than the previous ones to integrate hierarchical Hubel-Wiesel architectures for the visual system, e.g. Hierarchical MAX models, and probabilistic graphical models.

Keywords: visual cortex, belief propagation, Bayesian network, complex cells, translation invariance

1 Introduction

Ranging from behavioral psychology to physiological experiments, probabilistic computation is strongly suggested as the fundamental computational principle of brain[1]. Especially for the visual cortex of mammals, probabilistic approaches are successfully employed to explain some extra-classical receptive field properties in the primary visual area (V1), for example, end-inhibition or context-dependent responses[2].

Most of the theories of vision are, however, based on non-probabilistic, feed-forward neural networks[3, 4]. Series of these studies originate from the finding of the simple and complex receptive fields[5]. In the primary visual area, both neurons that have the simple receptive fields and neurons that have the complex receptive fields are strongly tuned to their optimal orientations of stimuli (e.g. slit of light). For the position (or phase) of stimuli, however, simple cells have narrow, optimal positions, while complex cells respond relatively broad positions. This suggests that features extracted by simple cells are pooled by complex cells. The neocognitron architecture[6] mimics these findings to build an object recognition model with a hierarchical feedforward neural network, in which convolutional feature extraction and max-pooling operation are repeated alternatingly. This kind of architecture, called convolutional neural networks or HMAX (Hierarchical MAX) models[3, 4], is currently known to be the state-of-the-art visual recognition model.

These studies naturally lead us to the question; How can we integrate probabilistic models and the HMAX-like models? Although many probabilistic models are proposed for visual recognition tasks, most of them reproduce only responses of simple cells but not those of complex cells. For an exceptional example, convolutional deep belief networks (CDBNs)[7] consist of stacked restricted Boltzmann machines with special probabilistic pooling variables that correspond to complex cells. There are also models that are based on Bayesian networks in which variables for simple cells and complex cells are arranged in a way similar to the HMAX-model architecture[8]. In both cases, models of these types require special handling for complex-cell variables for model architecture building or for learning.

In this paper, we propose an alternative way to interpret simple cells and complex cells in probabilistic graphical models. Our approach is based on belief propagation, a bidirectional message propagation algorithm[9]. We point out that if a graphical model has successfully learned translation invariance in its higher levels, variational messages from those levels should have translation invariance. This means lower levels can receive translation invariant signals in feedback messages and therefore do not require special variables for complex-cells' behavior.

This paper is organized as follows; In the next section, we explain our approach to the integration of the graphical models and the HMAX models by comparing it to the previous attempts. To examine our idea, we performed an experiment by using a combination of a Bayesian network and the belief propagation algorithm. A preliminary of the experiment is given in Sec. 3, and results of the experiment appear in Sec. 4. Sec. 5 is devoted to describe the conclusion.

2 Related Work on Integration of Graphical Models and HMAX Models

A remarkable attempt to unify HMAX-like models and probabilistic graphical models is found in ref. [8]. The authors of ref. [8] propose the Bayesian network model that contains two kinds of random variables; one corresponds to simple cells, and the other to complex cells. Both cell types are trained so that each responds like the corresponding neuron type in HMAX models. The trained model successfully integrates bottom-up evidences and top-down predictions using belief propagation in a practical experiments.

However, their model has two disadvantages. The first one resides in the learning algorithm. In order to reproduce simple-cell like and complex-cell like responses, two different learning algorithms are used for the conditional probability tables. Furthermore, the learning algorithms are based on clustering, not on the maximal log-likelihood method.

The other one resides in the correspondence between the cells in model and the biological cells. In ref [8], the beliefs, or the marginal probabilities $\mathrm{BEL}(x) = P(x|e)$, are compared to the biological simple and complex cell activities. However, if the belief propagation algorithm is executed in the cerebral cortex, the activities corresponding to the intermediate variables, such as the messages, should also be observed experimentally. It is not sufficient to model the biological neural activities only with the beliefs.

In this paper, we present a new view for modeling simple and complex cells. In a hierarchical graphical model, simple representations in lower levels can be combined and stored in higher levels as complex representations. Hence, the receptive fields of higher-level variables are broader and less sensitive to the position of stimuli than that of lower-level variables. In a variational inference, the effect of such broader receptive fields are initially observed in beliefs of higher-level variables, then sent to lower-level variables as messages. This mechanism leads us to the view that simple cells can be modeled by feedforward messages, and complex cells can be modeled by feedback messages.

Note that, with this view, we do not need special handling for translation invariant variables when we construct the graphical model that responses like HMAX models. We only need to train the graphical model so that it acquires translation invariance using a standard method, such as supervised learning, or unsupervised, slow-feature-analysis learning. Complex-cell like responses can be computed by using messages coming from higher levels in a variational inference. This is a new way to "integrate" graphical models and HMAX models. This viewpoint is explained in detail in the experiment section by using a simple, specific model and the belief propagation algorithm as an inference algorithm.

Before proceeding, let us compare our idea with the other related studies. Many models for the cerebral cortex with probabilistic graphical models are presented so far[10–14], but the emergence of complex-cell like responses is hardly addressed, except in Ref.[7, 8].

The author of ref.[14] used hierarchical Bayesian network which has similar architecture to ours. The network was trained in an unsupervised manner with natural images. As a result, receptive fields similar to V1 and V2 cells were acquired by the model, and their details were carefully compared quantitatively to physiological experiments. The author concluded that the complex receptive fields could not be found.

From our point of view, we might give some explanations to this result. Firstly, in [14], cell activities were measured by the beliefs, or the marginal probabilities. As we mentioned above, if we execute a variational approximation in the model, we might find another type of responses in variational messages. Secondly, the author used unsupervised learning. If supervised learning had been used, the higher-level variables would have acquired broad receptive fields and would have shown complex-cell like responses.

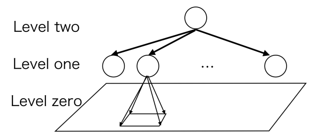

Fig. 1. Schematic representation of the Bayesian network in our experiment

In [11], the Bayesian network model was trained by a slow-feature-analysis like algorithm[15] to have the translation invariance. However, it was not clear whether there were responses corresponding to complex cells.

3 Bayesian Network Model

3.1 Model Architecture

To model the visual system in the cerebral cortex, we introduce a hierarchical Bayesian network that mimics hierarchical structure of the dorsal stream of the visual cortex. The Bayesian network model has three layers. The level zero consists of N_0 binary variables $L_0 = \{L_{01}, \ldots, L_{0N_0}\}$. The level one consists of N_1 variables $L_1 = \{L_{11}, \ldots, L_{1N_1}\}$. The level two consists of one variable L_2. The level zero is used as the visible, input layer, while the level one and the level two are hidden layers.

Each level has rough correspondence to a part of visual area: the level zero corresponds to lateral geniculate nucleus (LGN), the level one to V1, and the level two to V2 and higher areas. The hierarchical structure of the model is shown in Fig.(1).

The joint probability distribution of the Bayesian network is given by

$$P(L_0, L_1, L_2) = \prod_{i=1}^{N_0} P(L_{0i}|\mathrm{pa}_i) \prod_{j=1}^{N_1} P(L_{1j}|L_2)P(L_2), \tag{1}$$

where $\mathrm{pa}_i \subset L_1$ denotes the set of parent variables of L_{0i}. The conditional probability tables, $P(L_{0i}|\mathrm{pa}_i)$, $P(L_{1j}|L2)$, and $P(L_2)$ are the model parameters.

3.2 Belief Propagation

Belief propagation[9] is an efficient algorithm for inference in probabilistic graphical models, exploiting graph structure to reduce the computational costs. In this study, we employ the max-product belief propagation algorithm, which gives maximum a posteriori (MAP) combination of the states in the model:

$$X^{\mathrm{MAP}} = \operatorname*{argmax}_X P(X|e), \tag{2}$$

where X is the set of the hidden variables and e is the set of the visible variables with evidences given.

Generally, the max-product belief propagation algorithm gives an approximate MAP solution after sufficient iterative message passings. Since our model has a tree structure, however, the belief propagation gives the exact MAP solution in one cycle.

Because the set of message-update equations of the belief propagation algorithm is too lengthy, we no not describe its detail here. Instead, we give the key equation which connects the responses of beliefs in higher levels and the messages that are sent from a higher level to a lower level.

$$\pi_{L_{1j}}(L_2) = \prod_{k \neq j} \lambda_{L_{1k}}(L_2) P(L_2) \approx \mathrm{BEL}(L_2), \tag{3}$$

where $\pi_{L_{1j}}(L_2)$ is the feedback message sent from the higher variable L_2 to the lower variable L_{1j}, $\lambda_{L_{1k}}(L_2)$ is the feedforward message sent from L_{1j} to L_2, and $\mathrm{BEL}(L_2)$ is the belief of L_2. The approximation is justified if the number of the child variables of L_2 is large enough. This equation states how the feedback message $\pi_{L_{1j}}(L_2)$ is computed using the other messages and the conditional probability tables. As seen in the approximation, the feedback message $\pi_{L_{1j}}(L_2)$ behaves similar to the belief of L_2, $\mathrm{BEL}(L_2)$. This means that if $\mathrm{BEL}(L_2)$ responds translation invariantly, the feedback message $\pi_{L_{1j}}(L_2)$ also responds translation invariantly. This is the mathematical explanation of the translation invariant feedback messages in the belief propagation algorithm.

4 Simulation

4.1 Setup and Hyperparameters

We performed an experiment using a Bayesian network to concretely present our idea that translation invariant responses come from higher representations. The experiment consisted of two kinds of steps, i.e. the learning step and the recognition step.

The purpose of the learning step was to make the level-two variable learn translation invariant representations. For the sake of this, we used a supervised learning. We gave the translated images to the level-zero variables. Meanwhile, we gave the corresponding instruction signal to the level-two variable. Giving both the input and the instruction signal, the parameters in the model were updated so that the log-likelihood was maximized.

We used Gabor filters for the connections between the level-zero and the level-one variables instead of conditional probability tables. The Gabor filters were fixed throughout experiments, which means we only updated conditional probability table between level-one and two variables, $P(L_{1j}|L_2)$. This approximation is just for convenience and not essential. If one uses fully-probabilistic, maximal log-likelihood learning approach, one may obtain Gabor-filter like connection under an appropriate prior distribution[16]. With this approximation, the feedforward messages from the level-zero variables to the level-one variables were replaced by the convolutions of the image patches with the Gabor filters.

In the recognition step, we presented input signals in the level-zero variables. The hidden variables in the level one and level two were inferred with the belief propagation algorithm. We computed feedforward and feedback messages, as well as beliefs, as the responses of the model.

We used the following hyperparameters; The level zero had 16×16 binary variables, the level one had 4×4 variables each of which had 8 states corresponding to the 8 orientations of the Gabor filters (see below). The level two had one 4-state variable. Corresponding to the level-zero variables, an input image had 16×16 pixels. A level-one variable had 4×4 receptive field in level zero, and the level-two variable had also 4×4 receptive field in the level one. Throughout the network, there were no overlaps of receptive fields and therefore the model was tree-structured. Since the model was a tree, we used the belief propagation algorithm to get the exact MAP assignment in one cycle of iteration.

The Gabor filters employed to mimic the connections between the level zero and the level one were set by the parameters of $\gamma = 0.3, \sigma = 3.6$, and $\lambda = 4.6$ as in[4]. The angles were divided to eight directions by $\theta = 0, \pi/8, 2\pi/8, \ldots, 7\pi/8$, which consisted of the eight states of the level-one variables.

4.2 Results with Line Stimuli

We used horizontal line stimuli that covered the whole width of all level-one variables' receptive fields, namely, 16 pixels wide. In the learning step, we trained the model so that the sole level-two variable indicated the first state, out of the four possible states, whenever a horizontal line stimulus was presented regardless of its position. In the recognition step, we chose one level-one variable, L_1, and observed its two incoming messages $\lambda(L_1)$ and $\pi(L_1)$.

Fig. 2 shows the state of $\lambda(L_1)$ that corresponds to the horizontal angle of the Gabor filter, and its equivalent of $\pi(L_1)$. $\lambda(L_1)$ responded more strongly when a line stimulus was presented in the receptive field of L_1 than when a line stimulus was presented out of the receptive field. On the other hand, $\pi(L_1)$ responded more equally concerning the position of the presented stimulus. With these results, we can say that $\lambda(L_1)$ responds like simple cells and $\pi(L_1)$ responds like complex cells.

As a result of the supervised learning, the level-two variable predicted horizontal lines regardless of their positions. Since $\pi(L_1)$ was computed based on that prediction (Eqn.(3)), it responded more equally concerning the position of the stimulus.

Fig. 3 shows how the same state responded to the input stimuli that were roteted by 90 degrees, namely, vertical stimuli. The responses of both $\lambda(L_1)$ and $\pi(L_1)$ were suppressed. This indicates that both responses were selectively tuned to their optimal orientation, namely, horizontal orientation. Since the model had not been trained to recognize vertical stimuli, the L_2 belief for vertical stimuli was uniform. Therefore, the $\pi(L_1)$ message in the feedback was weak.

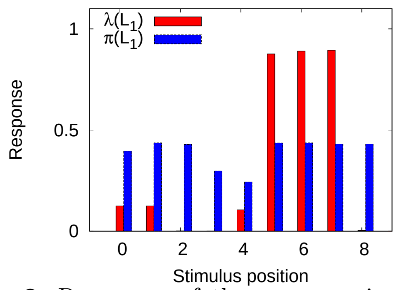

Fig. 2. Responses of the messages in the Bayesian network model. Horizontal axis indicates the position of presented horizontal stimuli, and vertical axis indicates the responses.

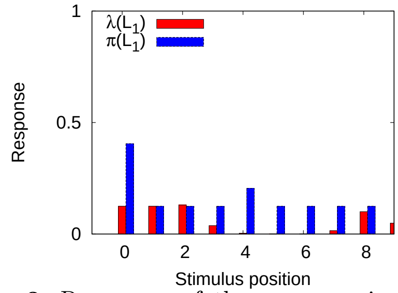

Fig. 3. Responses of the messages in the Bayesian network model. The same with Fig. 2, except that the presented stimuli are vertical.

5 Conclusion

We have shown that the biological neural responses can be interpreted as variational messages in a probabilistic graphical model. We have experimentally clarified this interpretation by using a Bayesian network and the belief propagation algorithm. We employed the standard supervised learning algorithm based on the maximal log-likelihood to obtain the translation invariance. The complex-cell like response in the level one was found in the computation of the feedback message from the higher level. Note that we do not need special architecture such as complex-cell random variables employed in the previous studies[7, 8] to reproduce the complex-cell like response. We therefore consider that our interpretation gives a new way to integrate the HMAX-like models and graphical models for visual system.

We remark that our interpretation is not restricted to Bayesian network models or the belief propagation algorithm. Many variational inference algorithms, such as the mean-field algorithm, share the computation in which a prediction of a higher-level variable is sent as a feedback message to lower-level variables. This fact suggests that the feedback messages in many variational inference algorithms convey translation invariance if the graphical model has acquired translation invariance appropriately.

Although our model is presented for the concreteness of the interpretation and too simple to model the real cortex, we can find interesting correspondence with the experiments. Complex cells in V1 are mainly distributed in superficial layers(II, III) and in deep layers (V, VI)[17]. For complex cells in the deep layers, it is suggested that they receive feedback signals from higher areas[18]. This is consistent with our interpretation that feedback signals can be used to realize a translation invariant response. On the other hand, complex cells in the superficial layers are suggested to receive direct projections from simple cells in the same area[19]. The role of the superficial complex cells in a variational inference model is left for a future work.

Acknowledgments. The authors thank to Naoto Takahashi for careful reading of the manuscript. This paper is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

This is a post-peer-review, pre-copyedit version of an article published in *Lecture Notes in Computer Science*. The final authenticated version is available online at: `https://doi.org/10.1007/978-3-319-68600-4_20`

References

1. Doya, K., Ishii, S., Pouget, A., Rao, R.P.N.: Bayesian brain: Probabilistic approaches to neural coding. MIT press (2007)
2. Rao, R. P. N., Ballard, D. H.: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci., 2(1), 79-87 (1999)
3. LeCun, Y., et al.: Handwritten digit recognition with a back-propagation network. In proc. NIPS 1989, 396404 (1990)
4. Serre, T., et al.: Robust object recognition with cortex-like mechanisms. IEEE Trans. Pattern Anal. Mach. Intell., 29(3) (2007)
5. Hubel, D. H., Wiesel, T. N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol., 160(1), 106-154 (1962)
6. Fukushima, K.: Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. Biol. Cybernetics 36, 193 202 (1980)
7. Lee, H., Grosse, R., Ranganath, R., Ng, A. Y.: Unsupervised learning of hierarchical representations with convolutional deep belief networks. Commun. ACM, 54(10), 95-103 (2011)
8. Dura-Bernal, S., Wennekers, T., Denham, S. L.: Top-down feedback in an HMAX-like cortical model of object perception based on hierarchical Bayesian networks and belief propagation. PLoS one, 7(11), e48216 (2012)
9. Pearl, J. : Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann. (1988)
10. Lee, T. S., Mumford, D.: Hierarchical Bayesian inference in the visual cortex. J. Opt. Soc. Am. A, 20(7), 1434-1448. (2003)
11. George, D., Hawkins, J. : A hierarchical Bayesian model of invariant pattern recognition in the visual cortex. In proc. IJCNN 2005, 1812-1817 (2005)
12. Ichisugi, Y.: A cerebral cortex model that self-organizes conditional probability tables and executes belief propagation. In proc. IJCNN 2007, 178-183 (2007)
13. Litvak, S., Ullman, S.: Cortical circuitry implementing graphical models. Neural Comput., 21(11), 3010-3056 (2009)
14. Hosoya, H.: Multinomial Bayesian learning for modeling classical and nonclassical receptive field properties. Neural Comput., 24(8), 2119-2150 (2012)
15. Wiskott, L., Sejnowski, T. J.: Slow feature analysis: Unsupervised learning of invariances. Neural Comput., 14(4), 715-770 (2002)
16. Olshausen, B. A., Field, D. J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381(6583), 607 (1996)
17. Gilbert, C. D.: Laminar differences in receptive field properties of cells in cat primary visual cortex. J. Physiol., 268(2), 391-421. (1977)

18. Sandell, J. H., Schiller, P. H.: Effect of cooling area 18 on striate cortex cells in the squirrel monkey. J. Neurophysiol., 48(1), 38-48 (1982)
19. Martinez, L. M., Alonso, J. M.: Construction of complex receptive fields in cat primary visual cortex. Neuron, 32(3), 515-525 (2001)