Modelling object perception in cortex: hierarchical Bayesian networks and belief propagation

Salvador Dura-Bernal[1,2], Thomas Wennekers [2] and Susan L. Denham [2]

***Abstract*—Hierarchical generative models and Bayesian belief propagation have been shown to provide a theoretical framework that can account for perceptual processes, including feedback modulation. The framework explains both psychophysical and physiological experimental data and maps well onto the hierarchical distributed cortical anatomy. The complexity required to model cortical processes makes inference, even using approximate methods, very computationally expensive. Thus, existing models are typically limited to tree-structured networks with no loops, use small toy examples or fail to account for certain perceptual aspects such as invariance to transformations or feedback reconstruction.**

We propose a novel, rigorous methodology to 1) implement selectivity and invariance using belief propagation on Bayesian networks; 2) combine feedback information from multiple parents, significantly reducing the number of parameters and operations; and 3) deal with loops using loopy belief propagation and different sampling methods. To demonstrate these properties we implement a Bayesian network that reproduces the structure and approximates the operations of HMAX, a biologically-inspired large-scale hierarchical model of object recognition. Hence, the proposed model not only achieves successful feed forward recognition invariant to position and size, but also extends the original model by including high-level feedback connections that reproduce modulatory effects such as illusory contour completion, attention and mental imagery.

Overall, the proposed methodology, based on state-of-the-art probabilistic approaches, can be used to build biologically plausible models of hierarchical perceptual organization that include top-down and bottom-up interactions. Furthermore, the proposed framework is suitable for large-scale parallel distributed hardware implementations.

Keywords: Bayesian belief propagation, hierarchical perception

I. Introduction

A. The Bayesian brain hypothesis

Experimental evidence shows that feedback originating in higher-level areas, such as V4, inferotemporal (IT) cortex, lateral occipital complex (LOC) or middle temporal (MT) cortex with bigger and more complex receptive fields, can modify and shape V1 responses, accounting for contextual or extra-classical receptive field effects [1]–[3].

While there is relative agreement that feedback connections play a role in integrating global and local information from different cortical regions to generate an integrated percept [4], [5], several differing approaches have attempted to explain the underlying mechanisms. Generative models and the Bayesian brain hypothesis [6] provide a framework that can quantitatively model the interaction between prior knowledge and sensory evidence, in order to represent the physical and statistical properties of the environment.

Overall, increasing evidence supports the proposal that Bayesian inference provides a theoretical framework that maps well onto cortical connectivity, explains both psychophysical and neurophysiological results, and can be used to build biologically plausible models of brain function [6]–[9]. Within this framework, Bayesian networks and belief propagation provide a rigorous mathematical implementation of these principles. Belief propagation has been found to be particularly well-suited for neural implementation, due to its hierarchical distributed organization and homogeneous internal structure and operations [5], [10]–[13].

B. Current limitations

However, modelling cortical perceptual processes using this framework poses a number of problems. First of all, there is no evident way of modelling invariance to object transformations using exclusively Bayesian networks with belief propagation. Some have proposed adding a two-layer associative network at the top of the Bayesian network [14], whereas others combine simple and complex features within the same node [10], to avoid much of the complexity, and benefits, inherent in a rigorous implementation of belief propagation.

Secondly, the complexity that emerges from the large-scale and intricate cortical connectivity means that exact inference methods are intractable, making it necessary to use approximate solutions such as loopy belief propagation [10], sampling methods [5], [14], [15] or variational methods [6], [16], [17]. Exact inference is only possible when the generative model avoids physiological constraints, for example by using simple tree-structured networks [18]; or models exclusively higher level phenomena such as attention, relying on non-Bayesian object recognition models [19].

However, even approximate methods struggle to deal with networks where nodes have multiple parents and loops. The size of the connectiviy matrix and the number of operations to perform inference is exponential to the number of parents. Additionally, networks with loops require several iterations of these operations to converge to an approximate solution. Consequently, the results of model simulations on real-world data are still limited. Some models remain purely theoretical [5], or provide simple toy examples [6], [14], [15]. Those that use bigger and more complex input images fail to account for certain aspects of object perception, such as position and

[1]Email: salvador.durabernal@plymouth.ac.uk

[2]Centre for Robotics and Neural Systems, University of Plymouth, Plymouth, UK.

This work was supported by the European Research Area Specific Targeted Project SCANDLE (IST-FP7-231168).

scale invariance [16], [17], [19], or feedback reconstruction (required to perform illusory contour completion) [14], [18]; or are not implementing rigorous, theoretically-grounded generative models [10].

C. Proposal

Firstly, we propose a method to model selectivity and invariance in alternating layers, a popular structure present in many object recognition models [20], [21], using exclusively Bayesian networks and belief propagation. Secondly, to approximate the connectivity matrices such that the number of parameters grows linearly and not exponentially with the number of parents, we implement the weighted sum approach proposed by [22]. Furthermore, in order to reduce the exponential number of operations, we propose sampling the incoming messages to keep only the highest values of the distributions with the highest variance. Finally, to deal with loops in the network we propose implementing loopy belief propagation and explore different belief updating methods to compare the effects these have on the resulting beliefs.

We illustrate the methods proposed by implementing a Bayesian network that captures the structure of the HMAX model [20], [21], a biologically-inspired hiearchical model of object recognition. The network approximates the selectivity and invariance operations of HMAX using the belief propagation algorithm, thus accounting for feedforward categorization invariant to position and size. Crucially, the inherent properties of Bayesian networks allow us to naturally extend the original feedforward model to include recursive feedback connectivity and account for high-level modulatory effects, such as illusory contour completion. This example demonstrates how the proposed methods facilitate the implementation of large-scale Bayesian networks that model hierarchical object perception.

II. Methods

A. Bayesian networks and belief propagation

A Bayesian network is a specific type of graphical model called a *directed acyclic graph*, where each node in the network represents a random variable, and arrows establish a causal dependency between nodes. Therefore, each arrow represents a conditional probability distribution $P(X|\Pi_X)$ which relates node X with its parents Π_X. Crucially, the network is defined such that the probability of a node X being in a particular state depends only on the state of its parents, Π_X. Consequently, a Bayesian network of N random variables X_i defines a joint probability distribution which can be factorized as follows,

$$P(X_1, ..., X_N) = \prod_i P(X_i|\Pi_{X_i}) \tag{1}$$

Belief propagation is a message-passing algorithm that manages to perform inference in a Bayesian network in a way that grows only linearly with the number of nodes, as it exploits the common intermediate terms that appear in the calculations. In belief propagation the effects of the observation are propagated throughout the network by passing messages between nodes.

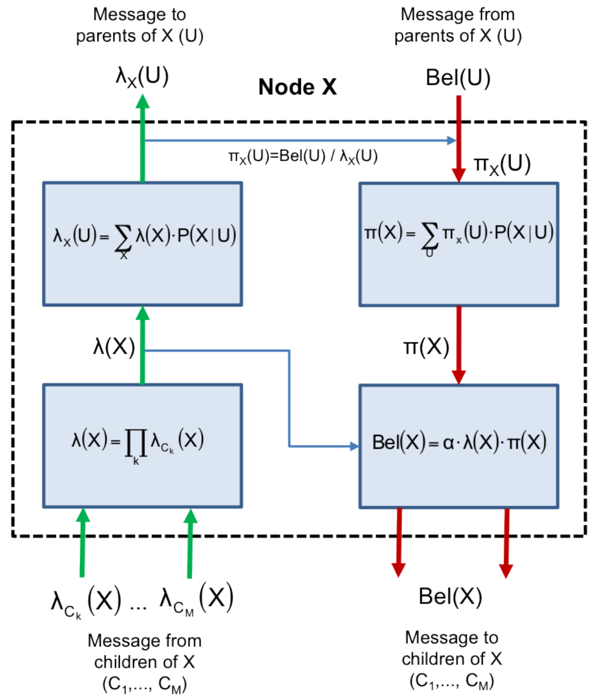

Fig. 1. Internal structure of a Bayesian node implementing belief propagation

The final belief, or posterior probability, is computed locally at each node by combining all incoming messages, i.e. evidence from higher and lower levels. The internal structure of a Bayesian node implementing belief propagation is shown in Figure 1. For a detailed description of Bayesian networks and the belief propagation algorithm refer to [23].

B. Approximating selectivity and invariance operations

We start by defining a generic scenario composed of a three layer hierarchical network with alternating simple and complex layers (S1, C1 and S2). Our aim is to approximate the invariance operation, typically implemented by *max-pooling* a subset of S1 nodes onto a C1 node, and the selectivity operation, typically implemented in S2 by performing a convolutional or distance operation between a subset of C1 nodes and a learned set of prototypes.

The Bayesian network in Figure 2 represents our proposed solution using belief propagation. Each square or rectangle represents a Bayesian node (3x9 nodes in layer S1, 3 nodes in layer C1 and 1 node in layer S2) and the dotted lines represent the convergent connectivity between the layers (e.g. 3x3 S1 nodes feed onto 1 C1 node). An example of the probability distribution over the states or features of the node is shown for each layer.

The C1 layer becomes an intermediate step that converts different patterns of afferent S1 nodes into the states of a single C1 node through the conditional probability table (CPT) $P(S1|C1)$ (Equations 2 and 3). The invariance or *max-pooling* operation only occurs during the generation of the $\lambda_{C1}(S2)$ output messages to the S2 layers (Equation

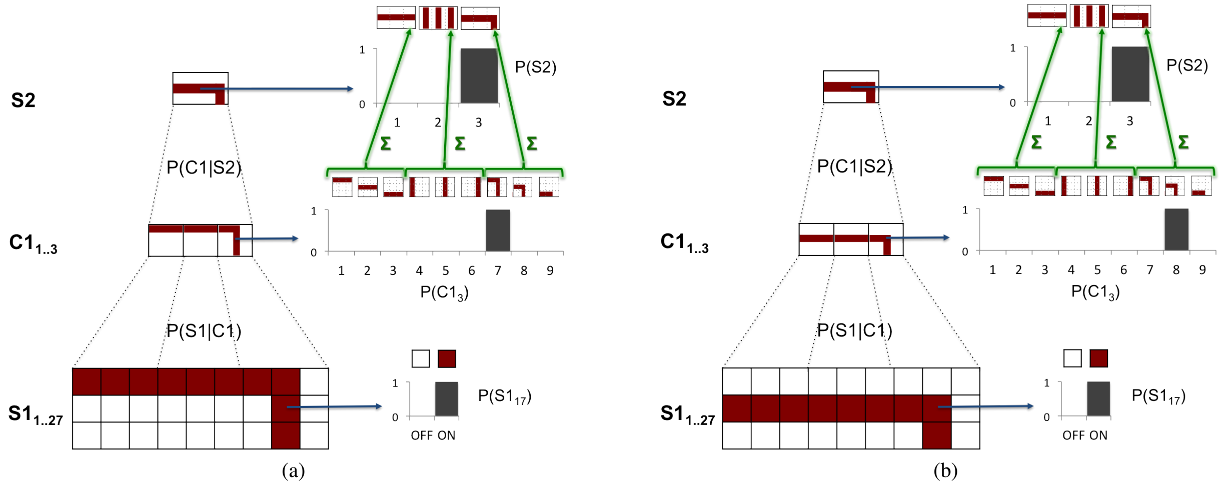

Fig. 2. Bayesian network that approximates the selectivity and invariance operations using belief propagation. This example illustrates how grouping C1 states achives translation invariance. The S1 and C1 representation in a) and b) differ in the location of their pattern, but the S2 representation is invariant to position and, thus, equivalent in both cases.

4), which groups these states according the learned wight matrix, or CPT, $P(C1|S2)$. Each group contains $K_{C1group}$ states corresponding to the different patterns of S1 nodes that encode each complex S2 feature (eg. horizontal line or corner). The learning method used was *k-means* clustering, which aims at obtaining a single activation per group of states, thus approximating the *max-pooling* operation. Figure 2 illustrates how the grouping of features in C1 leads to translation invariance.

Selectivity is approximated as the simultaneous co-ocurrence of the features in the afferent nodes (Equations 4 and 5), where the weight matrix for each prototype is stored in the CPT $P(C1|S2)$. In this sense it can be argued that both the selectivity operation and the invariance operation are implemented using the weights in $P(C1|S2)$, whereas the weights in $P(S1|C1)$ implement a necessary pre-processing step. The Equations are shown below:

$$\lambda_{S1_i}(c1_j) = \sum_{s1_i=1..K_{S1}} \lambda(s1_i) \cdot P(s1_i|c1_j) \quad (2)$$

$$\lambda(c1_j) = \prod_{i=1..N_{S1}} \lambda_{S1_i}(c1_j) \quad (3)$$

$$\lambda_{C1_j}(s2) = \sum_{c1_j=1..K_{C1}} \lambda(c1_j) \cdot P(c1_j|s2) \quad (4)$$

$$\lambda(s2) = \prod_{j=1..N_{C1}} \lambda_{C1_j}(s2) \quad (5)$$

where low-case letters denote specific states of a node, and N_{layer} and K_{layer} denote the number of afferent nodes and states in each layer, respectively. In the example in Figure 2, the corresponding values are $N_{S1} = 9, N_{C1} = 3, K_{S1} = 2, K_{C1} = 9, K_{S2} = 3, K_{C1group} = 3,$.

C. Multiple parents and loops

Bayesian networks that try to model the visual cortex will inevitably require multiple parent interactions as this arise as a consequence of overlapping receptive fields. The number of elements of the CPT $P(X|U_1, \cdots, U_N)$ grows exponentially with the number of parents, N, as it includes entries for all possible combinations of the states in node X and its parent nodes, e.g. given $k_X = k_U = 4, N = 8$, the number of parameters in the CPT is $4 \cdot 4^8 = 262,144$, where k_X and k_U represent the number of states in node X and its parent nodes, respectively.

We propose implementing the weighted sum of simpler CPTs based on the concept of *compatible parental configurations* [22]. This method obtains a $k_X \times k_U$ CPT, $Comp\{P(X|U_i)\}$, between node X and each of its N parent nodes, and assumes the rest of the parents, U_j, where $j \neq i$, are in *compatible* states. The final CPT $P(X|U_1, \cdots, U_N)$ is obtained as a weighted sum of the N $Comp\{P(X|U_i)\}$ CPTs. Therefore, the total number of parameters required to be learned grows linearly with the number of parents, more precisely, is equal to $k_X \cdot k_N \cdot N$. Using the values of the previous example, the number of elements now becomes $4 \cdot 4 \cdot 8 = 128$, several orders of magnitude smaller than using conventional methods.

To reduce the excessive number of operations required to calculate the belief, only the k_{max} states, with the highest values, from the N_{max} π messages, with the highest variance, are used in the calculation, where $k_{max} \leq k_u$ and $N_{max} \leq N$. The states with the strongest responses of the messages with highest variance are likely to carry most of the information content. Note this applies only to discrete probability distributions. To ensure the belief calculations are still valid it is necessary to select the appropriate columns of the CPTs, i.e. those that correspond to the sampled states of the π messages. This reduces the number of operations to $k_{max}^{N_{max}}$ sums and

$N_{max} \cdot k_{max}^{N_{max}}$ product operations. The sampling method is analogously applied to the incoming λ messages.

For Bayesian networks with loops, such as those that arise from modelling the visual cortex, the original belief propagation algorithm is no longer valid and approximate methods have to be implemented. The method we choose is loopy belief propagation, which has been empirically demonstrated to obtain good approximations to the exact beliefs in pyramidal networks (similar to that of the model) once the approximate beliefs have converged after several iterations [24]. The fact that belief propagation now requires several iterations means that a temporal dimension must be added to the original formulation. The resulting dynamical model is captured by the set of Equations 6 and these replace the standard belief propagation equations shown in Figure 1. The equations also implement the weighted sum method used to approximate the CPT of multiple parents, and the sampling methods used to reduce the number of operations.

$$
\begin{aligned}
Bel^{t+1}(x) &= \alpha \cdot \lambda^{t+1}(x) \cdot \pi^{t+1}(x) \\
\lambda^{t+1}(x) &= \prod_{j \in M_s} \lambda^t_{C_j}(x) \\
\pi^{t+1}(x) &= \sum_{u_s} \sum_{g} w_g \cdot P(x|u_g) \cdot \prod_{i \in N_s} \pi^t_X(u_i) \\
\lambda^{t+1}_X(u_i) &= \beta \sum_x [\lambda^{t+1}(x) \cdot \sum_{u_s \setminus u_i} \\
&\quad \left(\sum_g w_g \cdot P(x|Comp\{u_g\}) \right) \cdot \prod_{k \in N_s \setminus i} \pi^t_X(u_k)] \\
\pi^{t+1}_{C_j}(x) &= \alpha \prod_{k \in M_s \setminus j} \lambda^t_{C_k}(x) \cdot \pi(x) = \alpha \cdot \frac{Bel^{t+1}(x)}{\lambda_{C^t_j}(x)} \\
&\approx Bel^{t+1}(x)
\end{aligned}
\tag{6}
$$

where M_s represents the indices of the M_{max} incoming λ messages with the highest variance; N_s represents the indices of the N_{max} incoming π message with highest variance; u_s represents the indices of the k_{max} states with highest values out of each of the N_s incoming π messages.

III. Results

A. Example: HMAX as a Bayesian network

The HMAX model reproduces physiological and anatomical data of the activity and functionality observed along the ventral visual pathway, comprising areas V1, V2, V4 and IT [20], [21]. The model is based upon widely accepted basic principles such as the hierarchical arrangement of these areas, with a progressive increase in receptive field size and complexity of preferred stimuli, as well as a gradual build-up of invariance to position and scale as we move further up the hierarchy.

We develop a Bayesian network that reproduces the structure of a specific HMAX version with five layers per level, which alternatively implements the selectivity and invariance operations. The exact parameters of the model can be found in [21]. To illustrate the large scale of the resulting Bayesian network we should note that the S1 layer has over 200,000 Bayesian nodes, and the nodes in the S2 layer have 1000 states. Notably, every node in the network has identical internal structure (Figure 1) implementing the set of Equations 6. The methods previously described allowed us to approximate the HMAX operations using belief propagation despite the size and complexity of the network.

The steps required to interpret the HMAX model as a Bayesian network are are as follows:

1) Each node of the Bayesian network represents a specific location, scale-band and layer of the HMAX model.
2) The discrete states of each node of the Bayesian network represent the different features coded at that location, scale-band and layer of the HMAX model. For example each Bayesian node at layer S1 will have $K_{S1}(= 4)$ states, representing the four different Gabor filter orientations of HMAX.
3) The discrete probability distribution over the states of each Bayesian node comprises the sum-normalized responses of the K HMAX units coding the different features at that location, scale-band and layer.
4) The conditional probability tables (CPTs) that link each node of the Bayesian network with its parent nodes in the layer above, represent the prototype weights used to implement selectivity in the HMAX model. Additionally, the CPTs are used to approximate the *max-pooling* (invariance) operation between simple and complex layers of the HMAX model. Learning the appropriate CPT parameters allows the model to approximate the HMAX functionality using belief propagation.

TABLE I
Comparison of the feedforward categorization performance.

Model	Normal	Occluded	Scaled 10%	Scaled 20%	Translated
HMAX	100%	86%	100%	93%	87%
HTM	74%	48%	60%	56%	40%
BN+BP	100%	92%	83%	67%	58%

B. Feedforward categorization

The network was trained using 60 object silhouette images of size 160x160 pixels, from which the S2 and S3 prototypes were learned and encoded in the model CPTs. The trained network was then tested on different transformations of the same images including occluded, translated and scaled versions. The model's performance is measured as a percentage of correctly categorized images for each dataset of 60 images.

Table I compares the categorization performance of 1) the HMAX model [21], 2) a Hierarchical Temporal Memory network, which employs similar methods to that described in this paper, and 3) the Bayesian network and belief propagation (BN+BP) proposed model, with the same structure as the HMAX model.

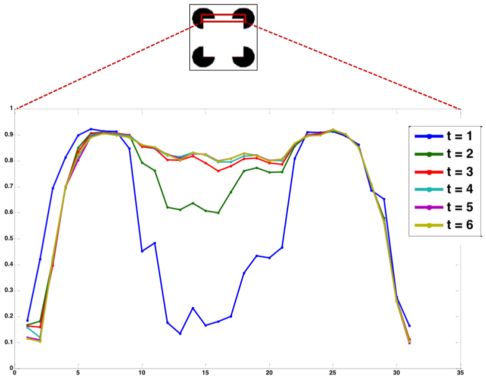

Fig. 3. Simulation results reproducing the illusory contour formation phenomenon in a Kanizsa figure. The graph shows the C1 belief response over time for the nodes coding the illusory contour region.

C. *Feedback modulation*

To study illusory contour completion in the model, we use a Kanizsa square input image while the top-down feedback from C2 is *clamped* to a square representation. Figure 3 shows the temporal response of the C1 belief for the region corresponding to the top horizontal illusory contour of the Kanizsa figure. This illustrates how the bottom-up evidence from the input image, $\lambda(S1)$, is recursively combined with top-down information from the C2 layer, $\pi(S2)$, and how the representation at the C1 layer evolves over time.

Figure 4 shows the model S3 belief response to an input image with a lamp occluding a dog, given two different conditions of the S3 prior, $\pi(S3)$: 1) an equiprobable or flat distribution and 2) a biased distribution where the prior probability of objects that are animals has been doubled. This example, despite depicting a very trivial problem, serves to illustrate the capacity of the model to simulate feedback effects, such as priming or feature attention. The model could similarly simulate spatial attention by defining a prior distribution that favours certain locations, specially when processing larger images with several objects.

IV. Discussion

As has been extensively argued in the literature, the parallel, distributed and hierarchical architecture of the cortex presents significant similarities with the structure of Bayesian networks [5], [10], [23]. Furthermore, the homogeneous internal structure of cortical columns (the canonical microcircuit) is comparable to the homogeneous internal operations (belief propagation) of each Bayesian node, which has lead to the proposal of possible cortical mappings and biologically plausible implementations of belief propagation [6], [10], [12], [13].

Here we propose methods to deal with some of the problems that arise when using Bayesian networks and belief propagation to model hierarchical object perception, which includes invariance to transformations and recursive feedback modulation. A recent feedforward model, which employs alternating simple and complex layers, has achieved the best pub-

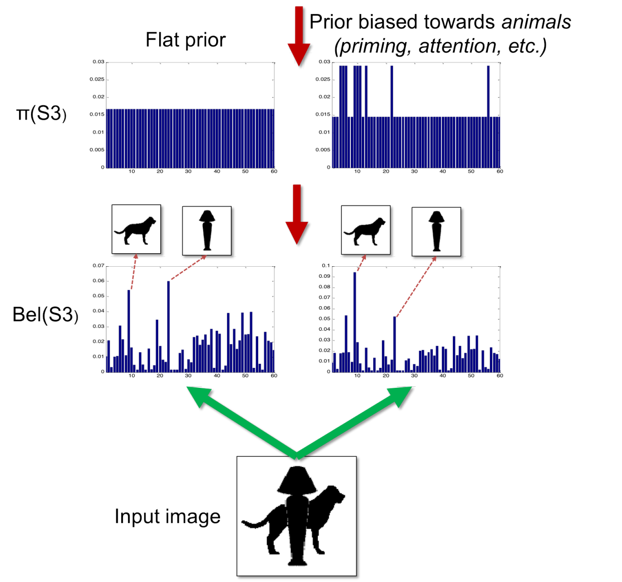

Fig. 4. Simulation results reproducing priming or attentional effects. The S3 belief distributions are modulated by the different prior distributions, $\pi(S3)$.

lished results on benchmarks for object classification (NORB, CIFAR10) and handwritten digit recognition (MNIST) [28]. Thus, first we show how to approximate the selectivity (distance to prototypes) and invariance (*max-pooling*) operations in alternating layers, usually referred to as simple and complex layers.

The proposed method also provides a way to feed back information from complex to simple layers, where each complex feature corresponds to a specific subset of simple nodes. It is is equivalent to that employed by Hierarchical Temporal Networks [10], where features in each node are combined into temporal groups or Markov chains. The method used here, however, preserves the Bayesian network structure by implementing the grouping of features in the weights of the CPTs.

Secondly, we show how to facilitate the implementation of belief propagation in large-scale Bayesian networks where nodes have multiple parents. To reduce the size of the connectivity matrices we particularize the weighted sum method proposed by [22] to the hierarchical object recognition domain. This method has a strong advantage over the Noisy-OR gate [23], a widely used method that is limited to non-graded variables. The proposed method can be applied to categoric variables with no given order, such as the visual features coded at a given location. Additionally, to reduce the exponential growth of the number of operations we propose sampling the probability distributions of the incoming messages. We demonstrate empirically that this method provides a relatively good fit to the exact distributions given a moderate number of samples (unpublished results).

The final set of dynamical equations implementing loopy belief propagation (Equations 6), propagates messages recursively and allows us to observe the temporal response across the network. We obtained preliminary results (unpublished) suggesting that updating only the layers that receive new evidence at each time step (similar to Gibbs sampling [14]), as opposed to updating all layers, yields similar results while strongly reducing the computational cost.

To illustrate the benefits of these methods we implement a Bayesian network that captures the structure of the HMAX model [20], [21] and approximates its feedforward operations using the proposed loopy belief propagation algorithm. Additionally, the network naturally extends the original HMAX model to include dynamic and recursive feedback, which had been pinpointed as its main limitation by the model authors [21].

Simulation results (Table I) demonstrate that the proposed Bayesian network can achieve similar feedforward categorization results to the original HMAX model. The still superior performance of HMAX is, however, not surprising as it was specifically designed to perform feedforward categorization. In fact, it is remarkable that the proposed model can achieve comparable categorization results using the homogeneous belief propagation operations implemented locally by every Bayesian node of the network. The HTM model [10] categorization results were obtained with the Numenta Vision Toolkit, but alternative HTM network implementations are likely to improve the results. Nonetheless, the results are intended to illustrate that, for belief propagation models with feedback modulation, such as HTM networks and our model, it is not trivial to succesfully perform feedforward categorization tasks.

Evidence shows the Kanizsa figure is represented as a complete figure in high cortical levels and, as time progresses, the illusory contour representation can be measured in lower levels [25]. Simulation results, some of which are shown in Figure 3, are consistent with the qualitative response pattern and temporal sequence of events observed across cortical layers, suggesting the model is able to reproduce the phenomena of illusory contour formation.

An additional proof-of-concept example (Figure 4) also demonstrates the model can account for higher-level feedback effects such as priming or attentional effects, which arise from areas outside the ventral pathway such as the prefrontal cortex, fusiform gyrus, posterior parietal cortex or the amygdala [26], [27]. The Bayesian implementation of attention resembles that proposed by [19]. Importantly, these effects are accommodated as part of the Bayesian network parameters (S3 prior distribution), without the need to include any external artifacts.

The proposed framework can potentially be applied to model other scenarios with similar hierarchical perceptual properties, such as cortical auditory processing. Furthermore, the resulting models are well-suited for real-time, parallel and distributed hardware implementations.

REFERENCES

[1] S. O. Murray, P. Schrater, and D. Kersten, "Perceptual grouping and the interactions between visual cortical areas," *Neural Networks*, vol. 17, no. 5-6, pp. 695–705, 2004.

[2] J. Y. Huang, C. Wang, and B. Dreher, "The effects of reversible inactivation of postero-temporal visual cortex on neuronal activities in cat's area 17," *Brain Res*, vol. 1138, pp. 111–128, 2007.

[3] M. A. Williams, C. I. Baker, H. P. Op de Beeck, W. Mok Shim, S. Dang, C. Triantafyllou, and N. Kanwisher, "Feedback of visual object information to foveal retinotopic cortex," *Nat Neurosci*, vol. 11, no. 12, pp. 1439–1445, 2008.

[4] J. Bullier, "Integrated model of visual processing," *Brain Res Rev*, vol. 36, no. 2-3, pp. 96–107, 2001.

[5] T. S. Lee and D. Mumford, "Hierarchical bayesian inference in the visual cortex," *J Opt Soc Am A*, vol. 20, no. 7, pp. 1434–1448, 2003.

[6] K. Friston and S. Kiebel, "Cortical circuits for perceptual inference," *Neural Networks*, vol. 22, no. 8, pp. 1093–1104, 2009.

[7] D. C. Knill and W. Richards, Eds., *Perception as Bayesian inference*. Cambridge University Press, 1996.

[8] A. Yuille and D. Kersten, "Vision as bayesian inference: analysis by synthesis?" *Trends in Cogn Sci*, vol. 10, no. 7, pp. 301–308, 2006.

[9] S. Deneve, "Bayesian spiking neurons i: Inference," *Neural Comput*, vol. 20, no. 1, pp. 91–117, 2008.

[10] D. George and J. Hawkins, "Towards a mathematical theory of cortical micro-circuits," *PLoS Comput Biol*, vol. 5, no. 10, p. e1000532, 2009.

[11] R. P. N. Rao, "Neural models of bayesian belief propagation," in *Bayesian brain: Probabilistic approaches to neural coding*, K. Doya, Ed. MIT Press, 2006, pp. 239–268.

[12] S. Litvak and S. Ullman, "Cortical circuitry implementing graphical models," *Neural Comput*, p. In press, 2009.

[13] A. Steimer, W. Maass, and R. Douglas, "Belief propagation in networks of spiking neurons," *Neural Comput*, vol. 21, no. 9, pp. 2502–2523, 2009.

[14] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput*, vol. 18, no. 7, pp. 1527–54, 2006.

[15] M. S. Lewicki and T. J. Sejnowski, "Bayesian unsupervised learning of higher order structure," in *Adv Neur In*, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds., vol. 9. The MIT Press, 1997, p. 529.

[16] J. Murray and K. Kreutz-Delgado, "Visual recognition and inference using dynamic overcomplete sparse learning," *Neural Comput*, vol. 19, no. 9, pp. 2301–2352, 2007.

[17] R. P. N. Rao and D. H. Ballard, "Dynamic model of visual recognition predicts neural response properties in the visual cortex," *Neural Comput*, vol. 9, no. 4, pp. 721–763, 1997.

[18] B. Epshtein, I. Lifshitz, and S. Ullman, "Image interpretation by a single bottom-up top-down cycle," *P Natl Acad Sci*, vol. 105, no. 38, pp. 14 298–14 303, 2008.

[19] S. Chikkerur, T. Serre, C. Tan, and T. Poggio, "What and where: A bayesian inference theory of attention," *Vision Res*, vol. In Press, Corrected Proof, 2010.

[20] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nat Neurosci*, vol. 2, no. 11, pp. 1019–25, 1999.

[21] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE T Pattern Anal*, vol. 29, no. 3, pp. 411–426, 2007.

[22] B. Das, "Generating conditional probabilities for bayesian networks: Easing the knowledge acquisition problem," *CoRR*, vol. cs.AI/0411034, 2004.

[23] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan, 1988.

[24] Y. Weiss, "Correctness of local probability propagation in graphical models with loops," *Neural Comput*, vol. 12, no. 1, pp. 1–41, 2000.

[25] M. Maertens, S. Pollmann, M. Hanke, T. Mildner, and H. E. Mller, "Retinotopic activation in response to subjective contours in primary visual cortex," *Front Hum Neurosci*, vol. 2, no. 2, p. doi:10.3389/neuro.09.002.2008, 2008.

[26] M. Bar, K. S. Kassam, A. S. Ghuman, J. Boshyan, A. M. Schmid, A. M. Dale, M. S. Hamalainen, K. Marinkovic, D. L. Schacter, B. R. Rosen, and E. Halgren, "Top-down facilitation of visual recognition," *P Natl Acad Sci USA*, vol. 103, no. 2, pp. 449–454, 2006.

[27] S. Grossberg, P. Cisek, T. Drew, and J. F. Kalaska, "Towards a unified theory of neocortex: laminar cortical circuits for vision and cognition," in *Prog Brain Res*. Elsevier, 2007, vol. Volume 165, pp. 79–104.

[28] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, J. Schmidhuber. "High-Performance Neural Networks for Visual Object Classification." *ArXiv Preprint* arXiv:1102.0183v1 [cs.AI], 1 Feb 2011.