

Temporal Information as Top-down Context in Binocular Disparity Detection

Mojtaba Solgi and Juyang Weng
 Department of Computer Science and Engineering
 Michigan State University
 East Lansing, Michigan 48824
 Email: {solgi, weng}@cse.msu.edu

Abstract—Recently, it has been shown that motor initiated context through top-down connections boosts the performance of network models in object recognition applications. Moreover, models of the 6-layer architecture of the laminar cortex have been shown to have computational advantage over single-layer models of the cortex. In this study, we present a temporal model of the laminar cortex that applies expectation feedback signals as top-down temporal context in a binocular network supervised to learn disparities. The work reported here shows that the 6-layer architecture drastically reduced the disparity detection error by as much as 7 times with context enabled. Top-down context reduced the error by a factor of 2 in the same 6-layer architecture. For the first time, an end-to-end model inspired by the 6-layer architecture with emergent binocular representation has reached a sub-pixel accuracy in the challenging problem of binocular disparity detection from natural images. In addition, our model demonstrates biologically-plausible gradually changing topographic maps; the representation of disparity sensitivity changes smoothly along the cortex.

I. INTRODUCTION

The importance of temporal signals for the acquisition of stereo vision capabilities in visual cortices (i.e. how time contributes to binocular cognitive abilities to emerge) has been underestimated in the past. Temporal context information from the previous time step(s) guide the human visual system to develop stereo receptive fields in childhood, and to improve disparity detection later on. We study the role of context signals, as context information, in the development of disparity tuned cells, and their contribution to help the laminar cortex to detect disparity after development.

In the real world, objects do not come into and disappear from the field of view randomly, but rather, they move continuously across the field of view, given their motion is not too fast for the brain to respond. At the pixel level, views are very discontinuous as image patches sweep across the field of view. Motivated by cerebral cortex, our model explores the temporal context at the later cortical areas including the intermediate areas and motor output area. These later areas are more “*abstract*” than the pixel level, and thus should be more useful as temporal context. However, how to use such information is a great challenge.

Physiological studies (e.g. [3] and [8]) have shown that the primary visual cortex in macaque monkeys and cats have a

laminar structure with a local circuitry similar to our model in Fig. 2. However, a computational model that explains how this laminar architecture contributes to the perception and cognition abilities is still unknown. LAMINART [11] presented a schematic model of the 6-layer circuitry, accompanied with simulation results that explained how top-down attentional enhancement in V1 can laterally propagate along a traced curve, and also how contrast-sensitive perceptual grouping is carried out in V1. Franz & Triesch 2007 [7] studied the development of disparity tuning in toy objects data using an artificial neural network based on back-propagation and reinforcement learning. They reported 90% correct recognition rate for 11 classes of disparity. In Solgi & Weng 2008 [13], a multilayer in-place learning network was used to detect binocular disparities which are discretized into classes of 4 pixels apart from image rows of 20 pixels wide. This classification scheme does not fit well for higher accuracy needs, as a misclassification between disparity class -1 and class 0 is very different from that between a class -1 and class 4 . The work presented here, investigates the more challenging problem of regression with subpixel precision, in contrast with the prior scheme of classification in Solgi & Weng 2008 [13].

For the first time, we present a spatio-temporal regression model of the laminar architecture of the cortex that is able to perform competitively on the difficult task of stereo disparity detection in natural images with subpixel precision. The model of the inter-cortical connections we present here was informed by Felleman & Van Essen [6] and that for the intra-cortical connections was informed by Callaway [2] and Wiser & Callaway [17].

Luciw & Weng 2008 [10] presented a model for top-down context signals in spatio-temporal object recognition problems. Similar to their work, in this paper the emergent recursive top-down context is provided from the response pattern of the motor cortex at the previous time to the feature detection cortex at the current time. Biologically plausible networks (Hebbian learning instead of error back-propagation) that use both bottom-up and top-down inputs with engineering-grade performance evaluation have not been studied until recently [13], [14].

It has been known that orientation preference usually changes smoothly along the cortex [1]. Chen et. al. [4] has recently discovered that the same pattern applies to the disparity sensitivity maps in monkey V2. Our model shows

that defining disparity detection as a regression problem (as apposed to classification) helps to form similar patterns in topographic maps; disparity sensitivity of neurons changes smoothly along the neural plane.

In the remainder of the paper, we first introduce the architecture of the network in Section II. Section III provides analysis. Next, the implementation and results are presented in Section IV. Finally, we provide some concluding remarks in Section V.

II. NETWORK ARCHITECTURE AND OPERATION

The network applied in this paper is an extension of the previous models of Multilayer In-place Learning Network (MILN) [14]. To comply with the principles of Autonomous Mental Development (AMD) [15], these networks autonomously develop features of the presented input, and no hand-designed feature detection is needed.

An overall sketch of the network is illustrated in Fig. 1. In this particular study, we deal with a network consisting of a sensory array (marked as *Input* in Fig. 1), a stereo feature-detection cortex which has a 6-layer architecture inspired by the laminar architecture of human cortex, and a motor cortex that functions as a regressor.

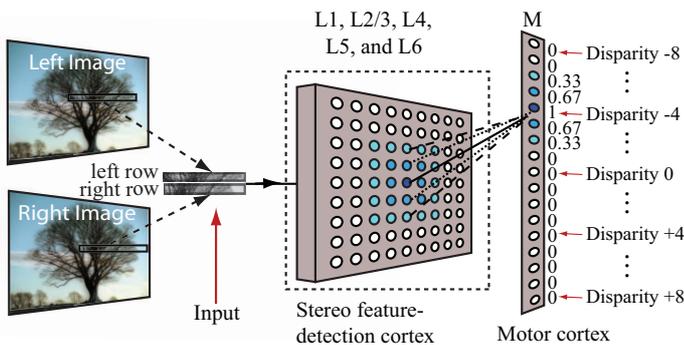


Fig. 1. A schematic of the architecture of the network studied here. Input to the network (on the left) consists of a pair of rows taken from slightly different positions (depending on the degree of disparity) of a set of natural images. The stereo feature-detection cortex has a 6-layer laminar architecture (see Fig. 2). Each circle is a neuron. Activation level of the neurons is shown by the darkness of the circles. The diagram shows an instance of the network during training phase when the disparity of the presented input is -4 . A triangular kernel, centered at the neuron of Disparity -4 , imposes the activation level of Disparity -4 neuron and four of its neighbors. The lines between neurons in the motor cortex and feature detection cortex represent two-way synaptic connections. The denser the line, the stronger the connection.

The architecture of the feature-detection cortex is sketched in Fig. 2. Layer $L1$ is connected to the sensory input in a *one-to-one* fashion; there is one neuron matched with each pixel, and the activation level of each neuron is equal to the intensity of the corresponding pixel (i.e. $\mathbf{z}^{(L1)}(t) = \mathbf{I}(t)$). We use no hand-designed feature detector (e.g. Laplacian of Gaussian, Gabor filters, etc.), as it would be against the paradigm of AMD [15]. The other four layers¹ are matched in functional-assistant pairs (referred as feedforward-feedback pairs in [3]).

¹ $L2/3$ is counted as one layer

$L6$ assists $L4$ (called assistant layer for $L4$) and $L5$ assists $L2/3$.

Layer $L4$ is globally connected to $L1$, meaning that each neuron in $L4$ has a connection to every neuron in $L1$. All the two-way connections between $L4$ and $L6$, and between $L2/3$ and $L5$, and also all the one-way connections from $L4$ to $L2/3$ are *one-to-one* and constant. In other words, each neuron in one layer is connected to only one neuron in the other layer at the same position in neural plane coordinates, and the weight of the connections is fixed to 1. Finally, neurons in the motor cortex are globally and bidirectionally connected to those in $L4$. There is no connections from $L2/3$ to $L4$.

The stereo feature-detection cortex takes a pair of stereo rows from the sensory input array. Then it runs the following developmental algorithm.

1. Fetching input in $L1$ and imposing supervision signals (if any) in motor cortex – $L1$ is a retina-like grid of neurons which captures the input and sends signals to $L4$ proportional to pixel intensities, without any further processing. During developmental training phase, an external teacher mechanism sets the activation levels of the motor cortex according to the input. If n_i is the neuron representative for the disparity of the currently presented input, then the activation level of n_i and its neighbors are set according to a triangular kernel centered on n_i . The activation level of all the other neurons is set to zero:

$$z_j^{(M)}(t) = \begin{cases} 1 - \frac{d(i,j)}{\kappa} & \text{if } d(i,j) < \kappa \\ 0 & \text{if } d(i,j) \geq \kappa \end{cases} \quad (1)$$

where $d(i,j)$ is the distance between neuron n_i and neuron n_j in the neural plane, and κ is the radius of the triangular kernel.

Then the activation level of motor neurons from the previous time step, $z_j^{(M)}(t-1)$, is projected onto $L2/3$ neurons via top-down connections.

2. Pre-response in $L4$ and $L2/3$ – Neurons in $L4(L2/3)$ compute their *pre-response* (response prior to competition) solely based on their bottom-up(top-down) input. The pre-response of the i 'th neuron, n_i , at time step t is computed as

$$\hat{z}_i^{(L4)}(t) = \frac{\mathbf{b}_i^{(L4)}(t) \cdot \mathbf{w}_{b,i}^{(L4)}(t)}{\|\mathbf{b}_i^{(L4)}(t)\| \|\mathbf{w}_{b,i}^{(L4)}(t)\|} \quad (2)$$

and

$$\hat{z}_i^{(L2/3)}(t) = \frac{\mathbf{e}^{(L2/3)}(t) \cdot \mathbf{w}_{e,i}^{(L2/3)}(t)}{\|\mathbf{e}^{(L2/3)}(t)\| \|\mathbf{w}_{e,i}^{(L2/3)}(t)\|} \quad (3)$$

where $\mathbf{b}_i^{(L4)}(t) = \mathbf{z}^{(L1)}(t)$ is the bottom-up input to neuron n_i in $L4$ (It is equal to the activation level vector of $L1$ for all the neurons in $L4$, since they are all globally-connected to $L1$). Similarly, $\mathbf{e}^{(L2/3)}(t) = \mathbf{z}^{(M)}(t-1)$ is the top-down input to neurons in $L2/3$. Also, $\mathbf{w}_{b,i}^{(L4)}(t)$ and $\mathbf{w}_{e,i}^{(L2/3)}(t)$ are the bottom-up(top-down) weight vectors of the i 'th neuron in $L4$ ($L2/3$).

3. $L6$ and $L5$ provide modulatory signals to $L4$ and $L2/3$ – $L6$ and $L5$ receive the firing pattern of $L4$ and $L2/3$, respectively, via their one-to-one connections. Then they send

modulatory signals back to their paired layers, which will enable the functional layers to do long-range lateral inhibition in the next step.

4. Response in $L4$ and second pre-response in $L2/3$ – Provided by feedback signals from $L6$, the neurons in $L4$ internally compete via lateral inhibition; k neurons with the highest pre-response win, and the others get suppressed. If $r_i = \text{rank}(\hat{z}_i^{(L4)}(t))$ is the ranking of the pre-response of the i 'th neuron (with the highest active neuron ranked as 0), we have $z_i^{(L4)}(t) = s(r_i)\hat{z}_i^{(L4)}(t)$, where

$$s(r_i) = \begin{cases} \frac{k-r_i}{k} & \text{if } 0 \leq r_i < k \\ 0 & \text{if } r_i \geq k \end{cases} \quad (4)$$

The same mechanism concurrently happens in $L2/3$ assisted by $L5$, except the output of $L2/3$ is called the second pre-response (denoted by $\hat{z}_i^{(L2/3)}(t)$).

5. Response in $L2/3$ – Each neuron, n_i in $L2/3$ receives its bottom-up input from one-to-one connection with the corresponding neuron in $L4$ (i.e. $b_i^{(L2/3)}(t) = z_i^{(L4)}(t)$). Then it applies the following formula to merge bottom-up and top-down information and compute its response.

$$z_i^{(L2/3)}(t) = (1 - \alpha) \cdot b_i^{(L2/3)}(t) + \alpha \cdot \hat{z}_i^{(L2/3)}(t) \quad (5)$$

where α is the relative top-down coefficient. We will discuss the effect of this parameter in detail in Section IV-A.

6. Response of motor Neurons in Testing – The activation level of the motor neurons is not imposed during testing, rather it is computed utilizing the output of feature-detection cortex, and used as context information in the next time step. The neurons take their input from $L2/3$ (i.e. $\mathbf{b}_i^{(M)}(t) = \mathbf{z}^{(L2/3)}(t)$). Then, they compute their response using the same equation as in Eq. 2, and laterally compete. The response of the winner neurons is scaled using the same algorithm as in Eq. 4 (with a different k for the motor layer), and the response of the rest of the neurons will be suppressed to zero. The output of the motor layer is the response weighted average of the disparity of the winner neurons:

$$disparity = \frac{\sum_{n_i \text{ is winner}} d_i \times z_i^{(M)}(t)}{\sum_{n_i \text{ is winner}} z_i^{(M)}(t)} \quad (6)$$

where d_i is the disparity level that the winner neuron n_i is representative for.

7. Hebian Updating with LCA in Training – The top winner neurons in $L4$ and motor cortex and also their neighbors in neural plane (excited by 3×3 short-range lateral excitatory connections) update their bottom-up connection weights. Lobe component analysis (LCA) [16] is used as the updating rule. Neurons in $L4$ and motor cortex update their bottom-up weights using their own internal temporally scheduled plasticity as $\mathbf{w}_{b,i}(t) = \beta_1 \mathbf{w}_{b,i}(t-1) + \beta_2 z_i \mathbf{b}(t)$ where the scheduled plasticity is determined by its two age-dependent weights:

$$\beta_1 = \frac{m_i - 1 - \mu(m_i)}{m_i}, \beta_2 = \frac{1 + \mu(m_i)}{m_i}, \quad (7)$$

with $\beta_1 + \beta_2 \equiv 1$. Finally, the cell age (maturity) m_i for the winner neurons increments: $m_i \leftarrow m_i + 1$. Afterwards, the motor cortex bottom-up weights are directly copied to $L4$ top-down weights.

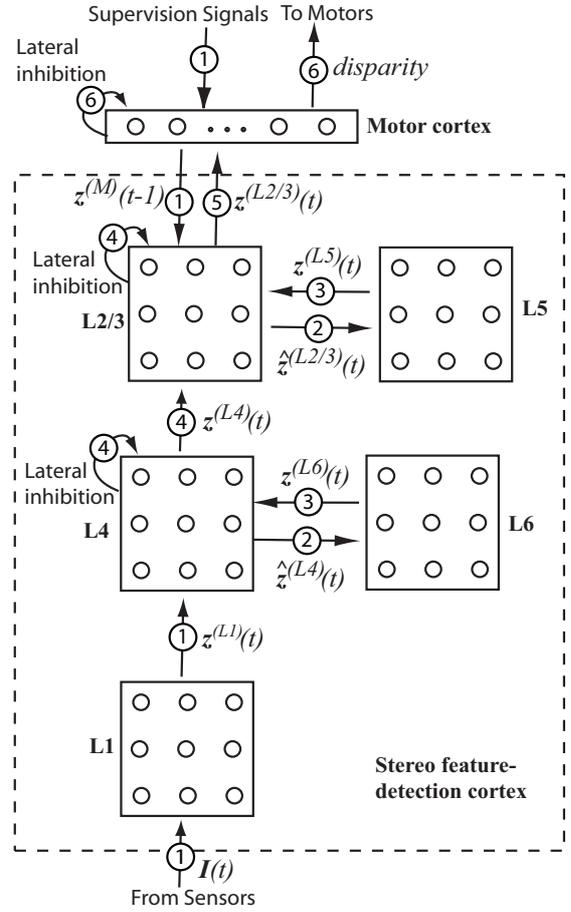


Fig. 2. Architecture diagram of the 6-layer laminar cortex studied in this paper, which also introduces some notation. The numbers in circles are the steps of the algorithm described in Section II. See the text for notations.

III. ANALYSIS

Analytical reasoning as to why top-down context signals help the networks to perform better is presented in [10]. In this section, we analytically investigate why and how the 6-layer laminar architecture outperforms the single-layer architecture model. Fig. 3 compares the algorithms by which the activation level of the neurons in single-layer and 6-layers architectures is computed. In single-layer architecture (the top row in Fig. 3), the top-down and bottom-up energies are first computed and proportionally added according to Eq. 8.

$$z_i = (1 - \alpha) \cdot E_{b,i} + \alpha \cdot E_{e,i} \quad (8)$$

$$E_{b,i} = \frac{\mathbf{b}_i \cdot \mathbf{w}_{b,i}}{\|\mathbf{b}_i\| \|\mathbf{w}_{b,i}\|}, E_{e,i} = \frac{\mathbf{e}_i \cdot \mathbf{w}_{e,i}}{\|\mathbf{e}_i\| \|\mathbf{w}_{e,i}\|} \quad (9)$$

The notation here is consistent with those in Equations 2, 3 and 5². In most real world sensory data, such as stereo pairs in our

²Except we dropped the time and layer ID components, for the sake of simplicity.

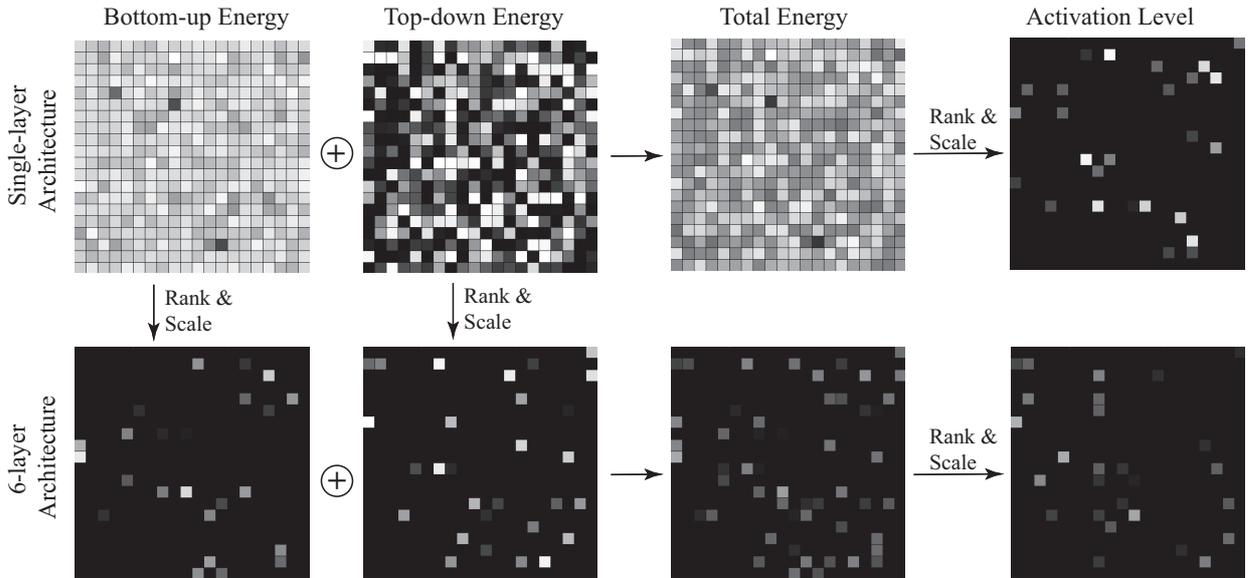


Fig. 3. The mechanisms of neuron winner selection (via lateral inhibition) in single-layer and 6-layer architectures. The maps are taken from a snap-shot of the 20×20 neurons in the networks performing on real data. Each small square projects the value for a neuron in that particular position (black/white): minimum(maximum) values). The top row shows the steps in the single-layer architecture, and the bottom row shows the steps for the 6-layer architecture (which shares some steps with the single-layer architecture). \oplus represents the operation of taking weighted average of two vectors (similar to Eq. 5).

case, the bottom-up sensory vector (\mathbf{b}_i in Eq. 9) is significantly more uniform than the top-down supervision/context vector \mathbf{e}_i .³ In the case of binocular disparity detection, the input pair of images is often featureless with similar intensities for the majority of pixels, while the top-down context/supervision vector is relatively more variant. As a result we have

$$\text{var}(E_b) \ll \text{var}(E_e) \quad (10)$$

where E_b and E_e are two random variables that can get any of the values $E_{b,i}$ and $E_{e,i}$, respectively. Here, we show that as a result of the lack of variation in bottom-up stimuli in such a single-layer architecture, activation level of the feature detection neurons is mostly determined by only top-down energy and the bottom-up energy is almost discarded. Obviously, this greatly reduces the performance of the network, as the top-down context signals are misleading when the input to the network at time t is considerably different from the input at time $t-1$. We call this effect “hallucination”.

Let us define $\tilde{\mathbf{E}}_b = \mathbf{E}_b - \bar{E}_b \mathbf{I}$ where \bar{E}_b is the mean value of the elements in \mathbf{E}_b (scalar value) and \mathbf{I} is the unit matrix of the same size as \mathbf{E}_b . Also, $\tilde{\mathbf{E}}_e = \mathbf{E}_e - \bar{E}_e \mathbf{I}$ in the same manner, and $\tilde{\mathbf{z}} = (1-\alpha) \cdot \tilde{\mathbf{E}}_b + \alpha \cdot \tilde{\mathbf{E}}_e$. Since $\tilde{\mathbf{z}}$ is only a constant term different from \mathbf{z} , we have

$$\text{rank}(z_i) = \text{rank}(\tilde{z}_i) \quad (11)$$

which is, the rank of each element z_i in \mathbf{z} is the same as the rank of the corresponding element \tilde{z}_i in $\tilde{\mathbf{z}}$. In addition, the rank of each element $\tilde{z}_i = (1-\alpha) \cdot \tilde{E}_{b,i} + \alpha \cdot \tilde{E}_{t,i}$ is mostly determined by its top-down component, $\tilde{E}_{t,i}$. The reason is because Eq.

10 induces the absolute value of the top-down component for most of the neurons is much greater than the absolute value of the bottom-up component, i.e. $|\tilde{E}_{t,i}| \gg |\tilde{E}_{b,i}|$. Hence, the ranking of neurons’ activation is largely effected only by their top-down component, and the reasoning is complete.

On the other hand, in the case of 6-layer architecture (the bottom row in Fig. 3), the bottom-up and top-down energies are ranked separately in $L4$ and $L2/3$, respectively, before they get mixed and compete again to decide the winner neurons in $L2/3$. Therefore, as a result of separation of bottom-up and top-down energies in different laminar layers, the 6-layer architecture manages to out-perform the single-layer architecture, specially when the imperfect context top-down signals are active (as apposed to supervision top-down signals which are always perfect).

IV. EXPERIMENTS AND RESULTS

From a set of natural images (available from <http://www.cis.hut.fi/projects/ica/imageica/>), 7 images were randomly selected, 5 of them were randomly chosen for training and 2 for testing. A pair of rows, each 20 pixels wide, were extracted from slightly different positions in the images. The right-view row was shifted by $-8, -7, -6, \dots, 0, \dots, +6, +7, +8$ pixels from the left-view row, resulting in 17 disparity degrees.

In each training epoch, for each degree of disparity, 50 spatially continuous samples were taken from each of the 5 training images. Therefore, there was $5 \times 50 \times 17 = 4250$ training samples in each epoch. For testing, 100 spatially continuous samples were taken from each of the 2 testing images (disjoint test), resulting in $2 \times 100 \times 17 = 3400$ testing samples in each epoch.

³Variance of the elements of the bottom-up sensory vector (\mathbf{b}_i in Eq. 9) is significantly lower than variance of the elements of the top-down supervision/context vector (\mathbf{e}_i in Eq. 9)

We trained networks with 40×40 neurons in each of $L2/3$, $L4$, $L5$ and $L6$ layers of stereo feature-detection cortex (of course there were 2×20 neurons in $L1$, as there is a one-to-one correspondence between input and $L1$ neurons). The k parameter (the number of neurons allowed to fire in each layer) was set to 100 for the stereo feature-detection cortex, and 5 for the motor cortex. We set $\kappa = 5$ in Eq. 1 and $\alpha = 0.4$ in Eq. 5 for all of the experiments, unless otherwise is stated.

A. The Advantage of Spatio-temporal 6-layer Architecture

Fig. 4 shows that applying top-down context signals in single-layer architecture (traditional MILN networks [14]), increases the error rate up to over 5 pixels (we intentionally set the relative top-down coefficient, α , as low as 0.15 in this case, otherwise the error rate would be around chance level). As discussed in Section III, this observation is due the absolute dominance of misleading top-down context signals provided complex input (natural images in this study). On the other hand, context signals reduce the error rate of the network to a sub-pixel level in 6-layer architecture networks. This result shows the important role of assistant layers (i.e. $L5$ and $L6$) in the laminar cortex to modulate the top-down and bottom-up energies received at the cortex before mixing them.

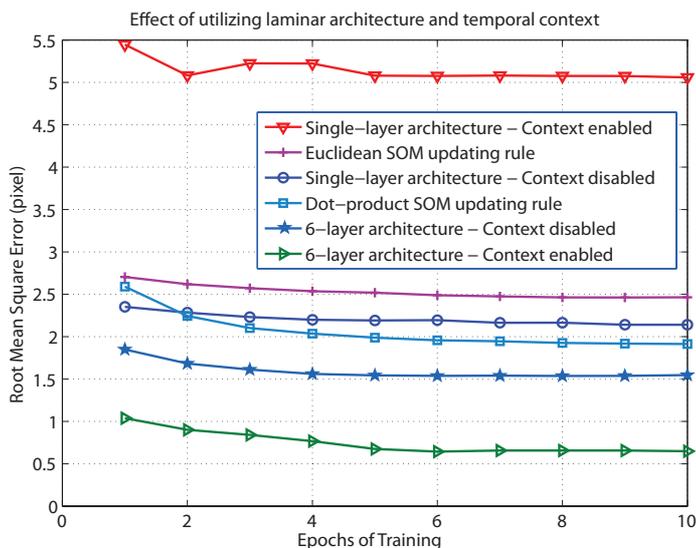


Fig. 4. How temporal context signals and 6-layer architecture improve the performance.

For comparison, we implemented two versions of Self-Organizing Maps updating rules, Euclidean SOM and dot-product SOM [9]. With the same amount of resources, the 6-layer architecture outperformed both versions of SOM by as much as at least 3 times lower error rate.

B. Smoothly Changing Receptive Fields

In two separate experiments, we studied the topographic maps formed in $L2/3$.

1) *Experiment A* – $\kappa = 5$: As depicted in Fig. 5a, the disparity-probability vectors for neurons tuned to close-by disparities are similar; neurons tuned to close-by disparities are

more likely to fire together. Equivalently, a neuron in the stereo feature-detection cortex is not tuned to only one exact disparity, but to a disparity range with a Gaussian-like probability for different disparities (e.g. neuron n_i could fire for disparities $+1, +2, +3, +4, +5$ with probabilities 0.1, 0.3, 0.7, 0.3, 0.1, respectively). This fuzziness in neuron's disparity sensitivity is caused by smoothly changing motor initiated top-down signals ($\kappa > 1$ in Eq. 1) during training. Fig. 5b shows this effect on topographic maps; having $\kappa = 5$ causes the regions sensitive to close-by disparities quite often reside next to each other and change gradually in neural plane (in many areas in Fig. 5b the colors change smoothly from dark blue to red).

2) *Experiment B* – $\kappa = 1$: However, if we define disparity detection as a classification problem, and set $\kappa = 1$ in Eq. 1 (only one neuron active in motor layer), then there is no smoothness in the change of the disparity sensitivity of neurons in the neural plane.

These observations are consistent with recent physiological discoveries about the smooth change of stimuli preference in topographic maps in the brain [5] and disparity maps in particular [4], [12].

V. CONCLUSION

Presented is the first spatio-temporal model of the 6-layer architecture of the cortex which incorporated temporal aspects of the stimuli in the form of top-down context signals. It outperformed simpler single-layer models of the cortex by a significant amount. Furthermore, defining the problem of binocular disparity detection as a regression problem by training a few nearby neurons to relate to the presented stimuli (as apposed to only one neuron in the case of classification), resulted in biologically-observed smoothly changing disparity sensitivity along the neural layers.

Since the brain generates actions through numerical signals (spikes) that drive muscles and other internal body effectors (e.g. glands), regression (output signals) seems closer to what the brain does, compared to many classification models that have been published in the literature. The regression extension of the MILN [14] has potentially a wide scope of application, from autonomous robots to machines that can learn to talk. A major open challenge is the complexity of the motor actions to be learned and autonomously generated.

As presented here, an emergent-representation based binocular system has shown disparity detection abilities with sub-pixel accuracy. In contrast with engineering methods that used explicit matching between the left and right search windows, a remarkable computational advantage of our work is the potential for integrated use of a variety of image information for tasks that require disparity as well as other visual cues.

Our model suggests a computational reason as to why there is no top-down connection from $L2/3$ to $L4$ in laminar cortex; to prevent the top-down and bottom-up energies received at the cortex from mixing before they internally compete to sort out winners.

Utilization of more complex temporal aspects of the stimuli and using real-time stereo movies will be a part of our future work.

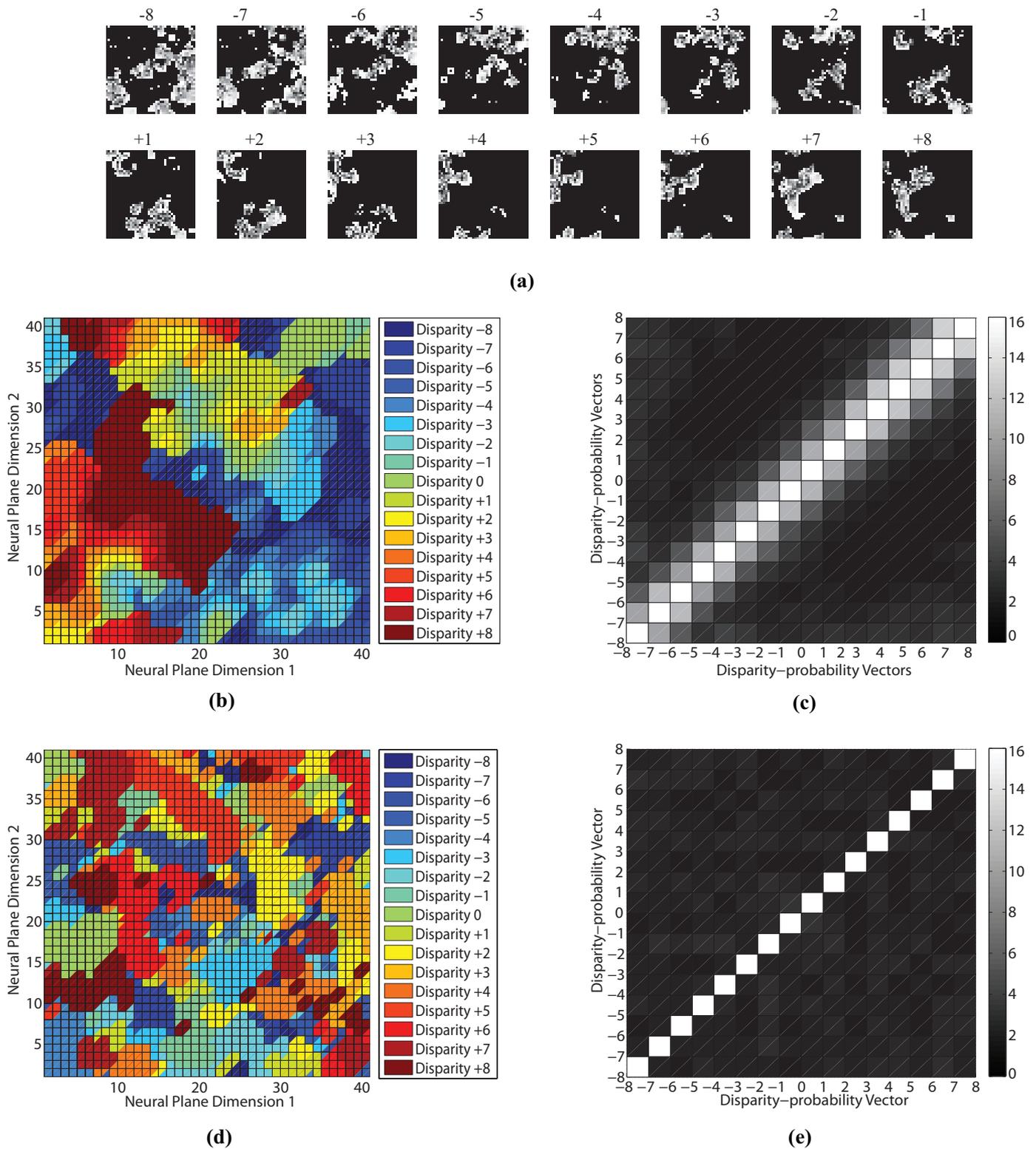


Fig. 5. (a) Disparity-probability vectors of $L2/3$ neurons for different disparities when $\kappa = 5$. Disparity-probability vector for each disparity is a $40 \times 40 = 1600$ dimensional vector containing the probability of neurons to fire for that particular disparity (black/white): minimum(maximum) probability. (b,d). Disparity-probability maps in $L2/3$ when $\kappa = 5$ and $\kappa = 1$. For each neuron, the largest disparity-probability (the disparity for which the neuron is most probable to fire) is shown by the color corresponding to that particular disparity. (c,e). Cross-correlation of disparity-probability where $\kappa = 5$ in (c) and $\kappa = 1$ in (e). Higher value of cross-correlation means higher similarity between two vectors, and hence more probable that neurons fire together for the corresponding classes.

REFERENCES

- [1] W. H. Bosking, Y. Zhang, B. Shoefeld, and D. Fitzpatrick. Orientation selectivity and arrangement of horizontal connections in tree shrew

striate cortex. *Journal of neuroscience*, 17:2112–2127, 1997.

- [2] E. M. Callaway. Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, 21:47–74, 1998.
- [3] Edward M. Callaway. Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Netw.*, 17(5-6):625–632, 2004.
- [4] G. Chen, H. D. Lu, and A. W. Roe. A map for horizontal disparity in monkey v2. *Neuron*, 58(3):442–450, May 2008.
- [5] D. B. Chklovskii and A. A. Koulakov. Maps in the brain: What can we learn from them? *Annual Review of Neuroscience*, 27:369–392, 2004.
- [6] D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1–47, 1991.
- [7] A. Franz and J. Triesch. Emergence of disparity tuning during the development of vergence eye movements. In *International Conference on Development and Learning*, pages 31–36, 2007.
- [8] C.D. Gilbert and T.N. Wiesel. Microcircuitry of the visual cortex. *Annu. Rev. Neurosci.*, 6:217–247, 1983.
- [9] T. Kohonen. *Self-Organizing Maps*. 1997.
- [10] M. D. Luciw and J. Weng. Motor initiated expectation through top-down connections as abstract context in a physical world. In *Proc. 7th International Conference on Development and Learning (ICDL'08)*, 2008.
- [11] R. D. Raizada and S. Grossberg. Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. *Cereb Cortex*, 13(1):100–113, January 2003.
- [12] A. W. Roe, A. J. Parker, R. T. Born, and G. C. DeAngelis. Disparity channels in early vision. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(44):11820–11831, October 2007.
- [13] M. Solgi and J. Weng. Developmental stereo: Topographic iconic-abstract map from top-down connection. In *Proc. the First of the Symposia Series New developments in Neural Networks (NNN'08)*, 2008.
- [14] J. Weng, T. Luwang, H. Lu, and X. Xue. Multilayer in-place learning networks for modeling functional layers in the laminar cortex. *Neural Networks*, 21:150–159, 2008.
- [15] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. *Science*, 291(5504):599–600, 2001.
- [16] J. Weng and N. Zhang. Optimal in-place learning and the lobe component analysis. In *Proc. World Congress on Computational Intelligence*, Vancouver, Canada, July 16–21 2006.
- [17] A. K. Wisner and E. M. Callaway. Contributions of individual layer 6 pyramidal neurons to local circuitry in macaque primary visual cortex. *Journal of neuroscience*, 16:2724–2739, 1996.