

Extensions of the Von Der Malsburg Self-Organized Visual Cortex Model

Joshua Goh

Psych 591 Neural Network Modeling

Project #3

Spring 2007

Introduction

The primary visual cortex has columnar functional organization according to neuronal sensitivity to specific line stimuli orientations (Hubel et al., 1962). Von Der Malsburg (VDM; 1973) modeled this using a self-organizing network. This paper further evaluates the VDM network to add plausibility to this model as an account of human visual cortical function.

Specifically, two main issues are addressed. 1) The VDM network was relatively unresponsive to untrained stimuli, predicting the insensitivity of cats to horizontal lines when raised in an environment with only vertical lines (Blakemore et al., 1970). However, an unexplored point in VDM's paper is how neural plasticity might still develop sensitivity to untrained stimuli if given enough time. 2) The VDM network dealt with a relatively primitive visual feature: line orientation at one spatial location. Unexplored is how different line orientations at different spatial location can be bound together into higher representations (e.g. a group of lines forming a square object).

This study uses a modified VDM network (VDM-mod) to explore whether a self-organizing network is capable of modeling issues 1 and 2 above.

Model Operation

Visual neurons are realized as vectors of unit activation values. There are three layers of units: input, middle, and output (Fig. 1). The units are organized hexagonally to facilitate simulating surrounding connectivity between neighboring units. The input layer simulates retinal neurons, with 9 visual locations projected onto groups of 19 input units (note: simulations 1 and 2 below only use one location and no output layer). At each location, units have activation 1 or 0, jointly coding for the stimulus line orientation at that location. Each input unit has synaptic connections to every excitatory unit in the

corresponding middle layer location group, that are initialized as a weight matrix, s_{lkij} , of random numbers ranging from 0 to s (where the weight subscripts refer to the l^{th} layer (1=middle, 2=output), k^{th} location group, i^{th} input unit, j^{th} output unit; refer to Table 1 for constants). To avoid unrestricted increasing weight values (VDM, 1973), weights are normalized to their mean value as:

$$s'_{lkij} = s_{lkij} \cdot N \cdot s / 2 / s_{lk \cdot j}$$

Where s'_{lkij} is the normalize weight matrix used for computing subsequent net input, N is the number units in the input layer, and $s_{lk \cdot j}$ is a vector consisting of the sum of all weights leading to each unit in the layer above.

There are 9 middle layer location groups, each with 169 excitatory and 169 inhibitory. Each excitatory unit receives input from all 19 input units in the corresponding location group below, projects to excitatory units immediately surrounding it with connection strength p , and to the inhibitory units immediately beneath and around it with strength r . The inhibitory units project to surrounding excitatory units two steps away with strength, q . There are no connections modeled between location groups. For each location group, at time t , the activation of excitatory units in the middle layer, $E_{lkj(t)}$, is defined by the sum of the net positive signal from the input layer, excitatory units and negative signal from inhibitory units:

$$E_{lkj(t)} = \sum s_{lkij(t)} a_{ki(t)} + \sum_{m=1, m \neq j}^M p_{lkmj(t)} E_{lkm(t)}^* - \sum_{m=1}^M q_{lkmj(t)} I_{lkm(t)}^*$$

Where a_{ki} is the input vector at location k , E^* is the output signal from surrounding excitatory units computed as $E - \theta$ if $E > \theta$, zero otherwise, I^* is the output signal from surrounding inhibitory units computed as $I - \theta$ if $I > \theta$, zero otherwise, M is the number of middle layer excitatory or inhibitory units per location group. At time t_l with a fixed

input, neighboring units will excite each other but inhibit those two steps away. If the input remains constant, a current high excitation will lead to an increased inhibition at t_2 , which in turn leads to lower excitation at t_3 . The result is oscillatory change towards a stable state based on a fixed input. Once stable, the final excitatory signals, E_l^* , are used to update s_{lkij} with the learning function defined as:

$$s_{lij(t+1)} = s_{lij(t)} + h \cdot a_{i(t)} \cdot E_{lj(t)}^*$$

Where h is the growth rate. This function additively increases weights for units that are jointly active utilizing the Hebbian rule (Hebb, 1970). The resulting weight matrix is then normalized as described above for the next run.

At the output layer, there are 19 units that each receives inputs from all excitatory units in the middle layer through weight matrix $s_{2,ij}$. There are no location groups in this layer. In order to obtain binary inputs as in the input layer, the excitatory signals from the middle to output layer, $a_{2,i}$, was first made binary such that $a_{2,i(t)} = 1$ if $E_{lki(t)} > \theta$, zero otherwise. All other unit activation and weight computations are arrived at in exactly the same manner as for the middle layer.

Simulations

For all simulations, the excitatory unit activation states were allowed to evolve for 20 iterations to stable state (as in Von Der Malsburg's paper) for each run before weights were updated.

Simulation 1. To replicate the VDM paper, a one-location network was trained on 9 orientation inputs (Fig. 2a) for 100 runs. The results were similar to the VDM paper. At run 1, middle layer units were already showing some orientation specific activation clustering (Fig. 2b). At run 100, the clustering was even more restricted (Fig. 2c). The number of uni-orientation sensitive units increased from 75 at run 1 to 104 at run 100

(Table 2). Thus, the VDM basic network was replicated.

Simulation 2. To test the effect of untrained stimuli, a new one-location network was trained on vertical lines only (inputs 2 and 9; Fig. 2a) for 20 runs. The middle layer units became selective and had activation clusters (Fig. 3). The network was further trained on horizontal lines (inputs 5 and 6; Fig. 2a) for 50 runs. The first run response to horizontal inputs was similar to run 1 of vertical inputs, but clustering occurred with training (Fig. 3). Similarly, the mean activation state of the layer units (entropy) was high when the horizontal input was first introduced to the vertically trained network (Table 3) but decreased with further training on horizontal lines. In relation to issue 1, this simulation predicts that visual neurons can acquire untrained stimuli through learning.

Simulation 3. To test the 2-layer network on some form of higher-level feature representation, the network was trained on “object” stimuli at various retinotopic locations (Fig. 4) for 50 time steps. Note that retinotopic information is represented as the net input from middle layer location groups to every output layer unit. Thus, like in higher visual cortices, direct retinotopy is lost at the output layer. After 50 steps, the output layer showed a cluster organization that related to the location of the object regardless of object type (Fig. 6). As seen in the figure, the output response to a cross and square at the bottom location is in fact identical. Although retinotopy was lost at this layer, the network still represented that information in some form. However, the output layer was not able to distinguish between different objects. Indeed, the tuning curves of the units showed that most units were sensitive to both types of objects (Fig. 7).

Discussion

This paper replicated VDM’s model of the primary visual cortex and showed that though it was initially unresponsive to untrained stimuli, it could learn them with time

(issue 1). The use of an additional layer that integrates lower-level features was unable to distinguish between two objects (issue 2). However, it still maintained location information despite loss of retinotopy. This lack of object identification may be because more units are required. In any case, simple integration of low-level features with a small network is insufficient to form object-level representations. The problem may also arise from independent retinotopic groups in this network. Inhibitory and excitatory connections probably extend throughout the primary visual cortex even across retinotopic locations. Object representation in the human visual cortex is far more complex than just squares and crosses. Further studies into this interesting field would reveal more about how feature combinations give rise to a relatively stable perception of objects. Presumably, this would also involve top-down feedback based on prior knowledge

Importantly, this network was able to learn untrained stimuli given enough time. Since no limits of neural plasticity are imposed here, it would be interesting to see if this were the case in real life, whether we can acquire new “orientations” if we were raised in a limited environment, or if there are critical periods for such neural development and learning.

Understanding how our visual system works provides a key to knowing how we organize the information that constantly bombards our retinas. Perhaps even more important, is to show that the complex perception we have can arise from simple principles of neurobiology interacting with the environment. This further raises the question of how this interaction might operate at even higher levels of cognitive processing such as decision-making and intentions.

References

Blakemore, C. and Cooper, G.F. (1970). Development of the brain depends on the visual environment. *Nature*, 228, 477-478.

Hebb, D.O. (1949). *Organization of behavior*. New York: Wiley.

Hubel, D.H. and Wiesel, T.N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160, 106-154.

Von Der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14, 85-100.

p	0.4
q	0.3
r	0.286
s	0.25
h	0.05
θ	1.0

Table 1. Constants used in the model. Taken from Von Der Malsburg (1973).

Type of unit	Number of runs		
	1	20	100
Non-modal	49	32	27
Unimodal	75	97	104
Polymodal	45	40	38

Table 2. Number of units in simulation 1 that were maximally sensitive to more than one input (polymodal), one input (unimodal), or having no selectivity (non-modal), after learning.

Run	Entropy for vertical input	
	Input 2	Input 9
1	1.74	1.74
20	0.09	0.09

Run	Entropy for horizontal input	
	Input 5	Input 6
1	1.41	1.41
20	0.02	0.02
50	0.02	0.02

Table 3. The entropy of the middle layer in simulation 2. The entropy for the network decreased with greater learning of vertical inputs (top panel). At the first run of the horizontal inputs on the vertically trained network (bottom panel), the entropy increased again, but decreased with more learning of the horizontal inputs.

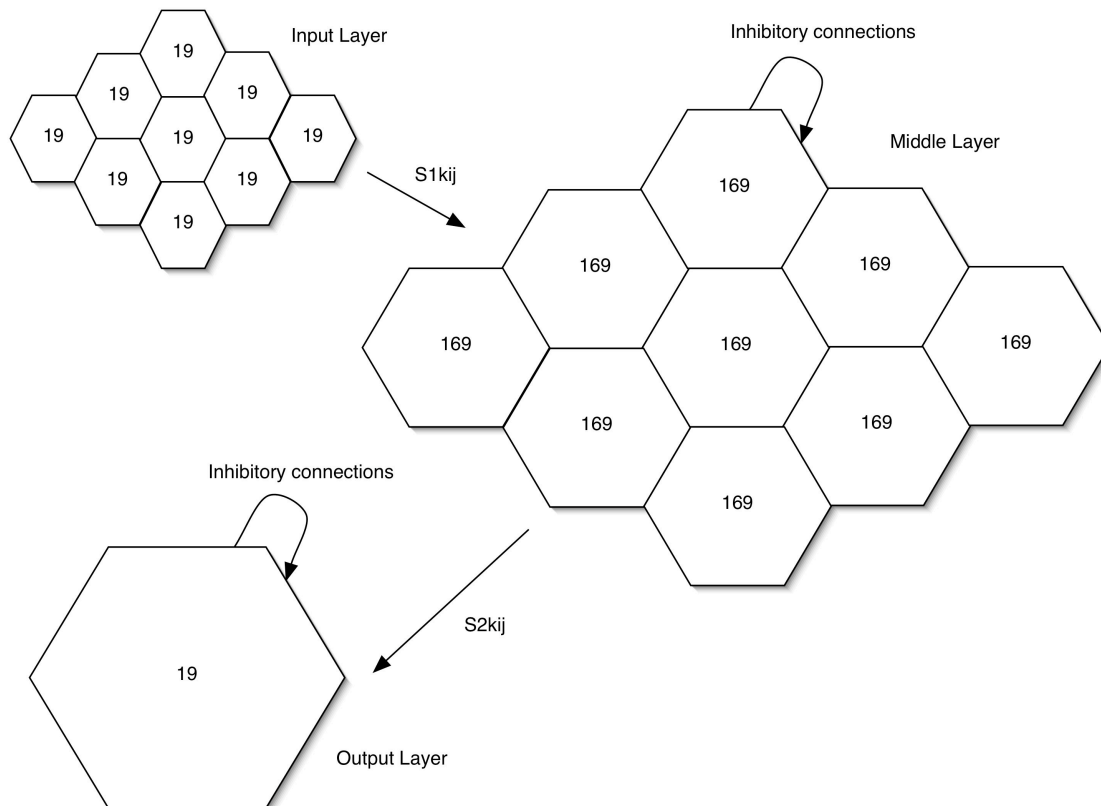


Fig. 1. A schematic of the modified VDM-mod network. The input layer simulates the retinal cells from 9 spatial locations. Each retinal location group contain 19 units that send inputs to all 169 excitatory units in the middle layer, in the corresponding location. In the middle layer, there are also 169 inhibitory units. Excitatory-Inhibitory connections are specified in the text. All units in the middle layer project to the output layer which has no retinotopy. The output layer has 19 units with similar excitatory-inhibitory organization as the middle layer.

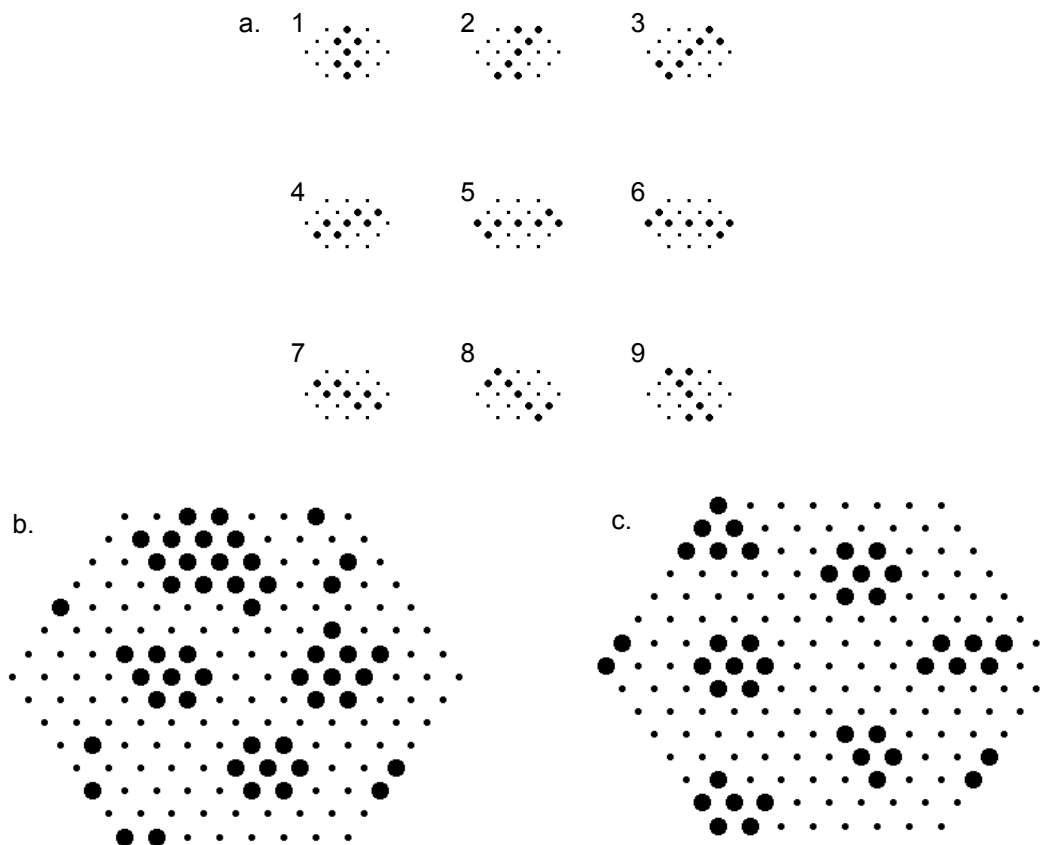


Fig. 2. a) 9 different activation states of input units in simulation 1 representing rotating line orientations. b) Sample activation state, in response to input 5, of the middle layer without learning at run 1. c) Sample activation state, in response to input 5, of the middle layer after 100 learning runs.

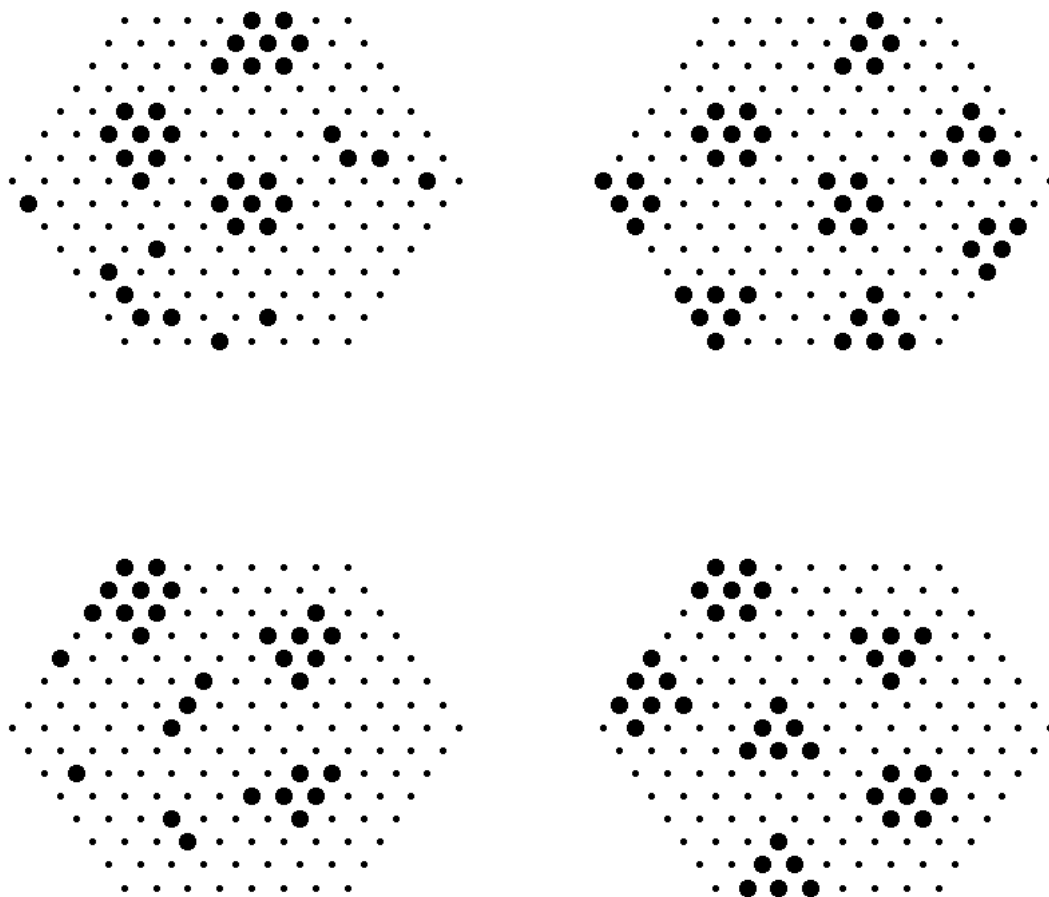


Fig. 3. a) Activation states of middle layer units in simulation 2. The top row shows the network response to vertical inputs after 1 run (top left), and 20 runs (top right). The bottom row shows the vertically-trained network's response to horizontal inputs at run 1 (bottom left) and run 20 (bottom right). The network response was at first similar to the 1st run of the vertical inputs, but became more self-organized with training.



Fig. 4. Sample “object” stimuli created for simulation 3. The top row shows the cross stimulus at two sample retinotopic locations, and the bottom row shows the square stimulus at two locations. There were four possible locations for each object type.

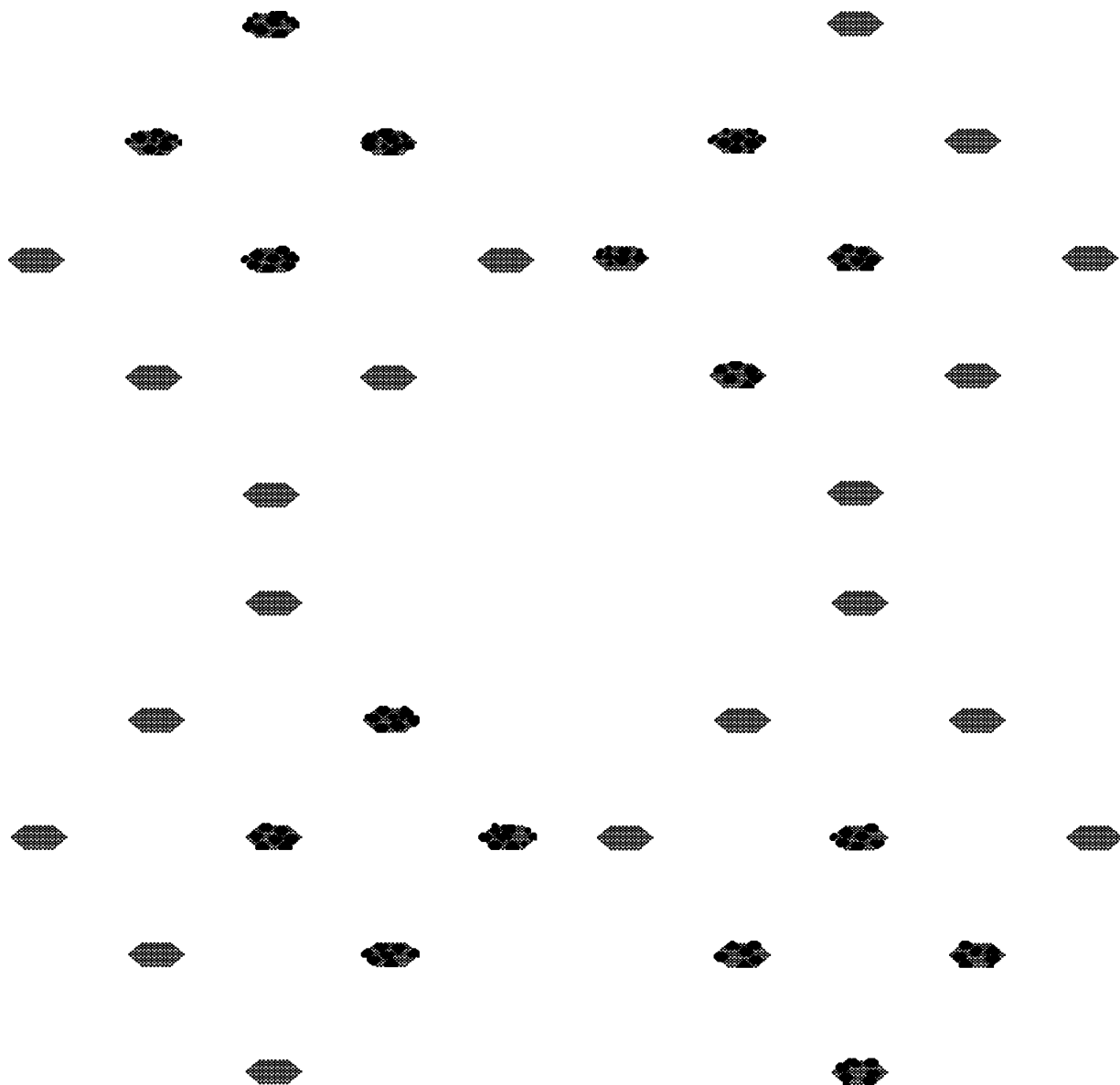


Fig. 5. Middle layer unit responses to the square input at four locations for simulation 3. Locations with no input did not respond, locations with input responded with the corresponding cluster of activity related to the line orientation at that location.

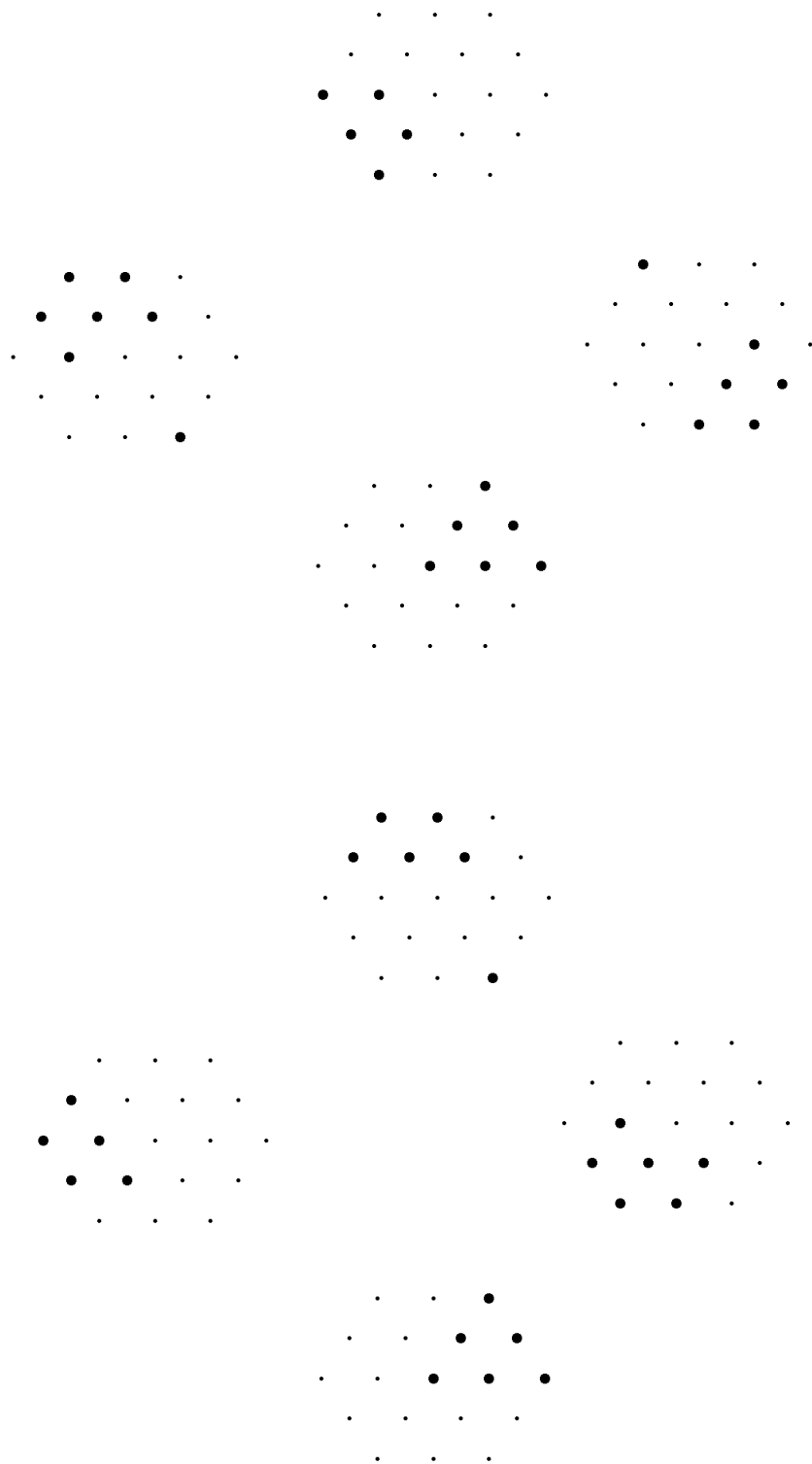


Fig. 6. Output layer units' responses to square inputs at 4 locations (top panel) and cross inputs (bottom panel).



Fig. 7. Sensitivity response curves for all 19 units in the output layer. Most units are bimodal, being sensitive to both cross and square objects at a specific retinotopic location.