
Paper:
Harrison, DG and De Kamps, M (2011) *A dynamical model of feature-based attention with strong lateral inhibition to resolve competition among candidate feature locations.* In: AISB 2011 Convention - Proceedings of the AISB 2011 Symposium on Architectures for Active Vision. AISB 2011 Convention , 4-7 April 2011, University of York, UK. Society for the Study of Artificial Intelligence and Simulation of Behaviour , 43 - 48.

Lateral Inhibition and Competitive Mechanisms for resolving cued features to locations

David G. Harrison[1] and **Marc De Kamps**

Abstract. Feature-based attention acts across the visual field and so presents a problem with resolving the location of features when their presence is detected in the visual array. We present here a computational model of feature-based attention using strong inhibitory interactions to reduce activity of neurons without attended features in their receptive fields to baseline levels. This mechanism allows the locations of attended features to be determined by interactions of an attentional template and stimulus driven activities in the ventral stream.

1 INTRODUCTION

The amount of visual information entering the eye would overwhelm the brain's visual processing capability if the input is processed to the same clarity at all locations[16]. As such, the fovea, a small central portion of the retina with has a greater resolution than peripheral retinal cells and a correspondingly greater cortical processing capacity. In order to perceive visual objects in detail, the object of interest is brought into focus on the fovea. The mechanism by which we designate what is of interest and to fixate the fovea upon it is described as visual attention. The ability to visually attend to an object is so crucial to visual perception it has been argued that "to see is to attend"[19].

Three types of visual attention have been described: attention may be deployed to a location (spatial attention), an individual object (object-based attention), or to a collection of features (feature-based attention). Object-based attention may be described as an example of feature based attention, as evidence[15] suggests that object representations are composed of distributed sub-object, feature building blocks bound together as the neural object representation as needed, or determined, by the visual stimulus.

1.1 SPATIAL ATTENTION

Motter [13] describes spatial attention as shrinking the receptive field around the attended to object. We achieve this effect through lateral inhibition of cortical feature-binding circuits with receptive fields containing non-matching features to an endogenously initiated attentional template in the ventral stream. The lateral inhibition from non-matching populations subdues activity in neighbouring populations, effectively removing external stimuli in their receptive fields from further visual processing. Regions with few mismatches between the attentional template and stimulus driven activity project excitatory activity to a separate cortical area in the dorsal stream, where the saccade necessary to foveate the object may be generated.

Bisley [1] describes priority maps as like saliency maps [9], but includes top down influences in addition to bottom up. This interaction is modeled with DI circuits. Influence of top down flow is necessary to sustain output to LIP for captured attention, or to find location of cued feature through lateral inhibition or mismatches to attentional template.

Inhibition of neighbours is not enough for them to become mismatched and in turn inhibit their matching neighbours. Therefore our result is actually shrinking the receptive field to slightly within the border of the attended to object, but this is future work to resolve.

1.2 FEATURE-BASED ATTENTION

Feature similarity gain principle [11] Moran and Desimone - biased competition

Lots of people have found that feature-based attention acts across the visual field [16, 3, 4, 20] in the way the model works.

Addition of LI and ILI populations to DI circuit. Why do we collapse the separate positive and negative circuits? Because we're more interested in non-matches than matches. But the strength of activity in the top-down network shows the degree of confidence we have of that location being positive or negative, so we use this information to inhibit neighbours to an appropriate level. For example, if we are more confident that a receptive field should be positive Pr will have a high activation rate, so if the matching Pf population is negative, we want to strongly inhibit its neighbours, but if we are less confident, Pr will have a low activation rate and on mismatches, will inhibit its neighbours much less. See 1.

1.3 ATTENTIONAL CAPTURE

Bottom-up flow causes Pf/Nf to activate when stimuli fall within their receptive fields. This causes excitation of the Gp/Gn and Ep/En populations. The Ep/En populations output a brief period of activity to LIP when the corresponding Pf/Nf population is activated, prior to the inhibition from Gp/Gn. This allows attention to be captured, as a sudden region of activity in LIP could cause spatial attention to activate Pr/Nr, allowing sustained attention to that location. In instances of attentional capture the lateral inhibition mechanism is not activated as LI needs some excitatory input from the top-down network.The Ep/En population excites the inhibitory lateral inhibition population (ILI) briefly when stimuli first appear, but without activity in the top-down layer, LI has no output, allowing changes in the visual array to breakthrough to LIP when the external stimulus is large enough, as neighbouring populations are not inhibited.

The neural correlates of visual attention have been much studied (see [1] for a recent review). We build upon a model of feature-based

---

[1] University of Leeds, UK, email: pab2dgh@leeds.ac.uk

attention to resolve the binding problem which occurs from a distributed object representation [17], by adding a mechanism of lateral inhibition to resolve a collection of features to a spatial location.

2 THE MODEL

The model consists of two artificial neural networks, one modelling the bottom-up flow of stimulus activities and the other modelling the top-down flow. The bottom-up network is a feed-forward network of five layers corresponding to V1, V2, V4, posterior inferotemporal (PIT) and anterior inferotemporal (AIT) visual areas. A widening receptive field in higher layers allows AIT neurons to project across the entire V1 layer, allowing objects to be recognised in all locations. This network is trained using backpropagation to associate objects presented at the V1 layer, to individual neurons in AIT. V1 consists of 4 feature layers which detect lines of 45° orientations, and objects are presented by direct stimulation of neurons in the appropriate feature layer to simulate neural inputs from the lateral geniculate nucleus (LGN), which we do not model.

Once the forward network has been trained to recognise the objects, it is used to train the top-down network through Hebbian learning: each training pattern is evolved through the forward network and conditions reciprocal connection weights in the reverse network. This mechanism creates the attentional template.

With the forward and reverse artificial neural networks trained, they are then converted into a dynamical model as neural populations of Wilson-Cowan differential equations [18]. During this conversion the neurons in the forward and reverse networks are converted to a neural circuit separating out positive and negative activities in the ANN's to two neural populations of positive spike rates, with one of the pair's populations implicitly coding for the negative activities. The architecture and conversion of the model into dynamical populations is detailed in [6].

Layers of neural circuits are created between layers V2, V4 and PIT of the converted forward and reverse networks to detect correlated neural activations in paired populations of the forward and reverse networks. Correlating activations in the forward and reverse networks is achieved through implementing the disinhibition mechanism described in [17]. We extended this disinhibition circuit with two inhibitory populations, to create a mechanism to inhibit the activity of neighbouring circuits when there is a mismatch of activities in the disinhibition circuit.

Figure 1 shows the populations of the disinhibition circuit in grey, and matching populations in the forward and reverse network as positive-forward (Pf) and positive-reverse (Pr), likewise Nf and Nr for the negative populations. Open triangles represent excitatory connections, and black triangles inhibitory. If we consider a stimulus-driven activation in Pf, the excitatory-positive (Ep) and gating positive (Gp) populations of the disinhibition circuit receive equal rates of excitatory spikes. The Ep population is inhibited by Gp, but there is a small delay in inhibition as the driving activity in Pf passes through Gp. This delay allows the attentional capture mechanism, as there is a brief output of Ep to LIP from increases in the spike rate of Pf. However, the inhibitory output of Gp is itself inhibited by the inhibitory-positive (Ip) population if there is matching positive activity in the reverse population, Pr.

The Ep population sends excitatory projections to the LIP layer, but also excites the inhibitory-lateral-inhibition population (ILI), which in turn inhibits the lateral-inhibition population (LI) from reducing the output to LIP of neighbouring circuits. Due to the mutually exclusive activities in the positive and negative populations, it is guaranteed that the disinhibition circuit can only receive strong excitatory spikes from one reverse and one forward population at a time[2].

In the situation of matching positive activity in the forward and reverse network just described (and analogously for matching negative activity), the circuit will output excitatory spikes to LIP from the Ep population, while the inhibitory spikes from LI to neighbouring circuits is prevented by inhibition from ILI. Now we consider a mismatch of activities in the forward and reverse networks. If, for example, Pf and Nr have high activity, Gp is not inhibited by Ip, shutting off the output to LIP and ILI from Ep in a few tens of milliseconds. With no inhibitory activity from ILI, the excitatory input to LI from Nr is unchecked and neighbouring populations of Ep and En neurons are inhibited.

It is worth stressing that the circuit only produces lateral inhibition on mismatches in the (implied) sign of the forward and reverse networks. We can consider the spike rate of populations in the forward network to be the weight of evidence from the visual array, and the spike rate of the reverse populations as the expectation of a positive or negative value in the receptive field of the circuit if that receptive field was to contain the searched for feature. Thus strong evidence and strong expectation output a strong excitatory activity to LIP, whereas some evidence and high expectation is inconclusive. Similarly, strong evidence which contradicts a strong expectation suggests the sought for feature is probably not in the receptive field, and this confidence is passed on to neighbouring circuits via lateral inhibition. If neighbouring circuits have strong evidence of a match, they can compete with the inhibition, and the most confident circuit (i.e. that with the highest initial activity) wins the competition.

The benefits of mismatches being determined as opposite signed activities allow a non-linear selection of matches and mismatches and allows for slight differences between instances of objects belonging to the same category. This non-linearity is desirable, so the use of lateral inhibition to inhibit neighbouring neurons on mismatches allows only neurons to survive the attentional template that are in regions with high correlation between stimulus and attentional template, without the mechanism running away and completely extinguishing matches throughout the visual pathway.

The model is implemented using MIIND [5], a computational neuroscience framework, for modelling the artificial neural networks and the simulation of the neural dynamics.

3 RESULTS

The forward network was created as a 16×16 grid of neurons, consisting of four feature detectors for lines at 45° orientations, and four AIT neurons to code for each of the square, diamond, horizontal cross and diagonal cross as seen in the top left layer of the following figures. Each shape was presented to V1 by direct stimulation of the appropriate feature neurons. Each shape was presented at every location in V1 which allowed the shape to be contained wholly, to avoiding border effects. As the shapes are simple, training continued until the global network error, measured as the sum of differences between expected and actual AIT activity for every training exemplar, was below 10^{-5}.

Simulations were then run by converting the trained ANN's into dynamical networks as previously described. During the conversion the inputs patterns presented to V1 were varied, as was the attentional

---

[2] Baseline activity, such as thermal noise, may cause occasional low level rates, but these are not significant.

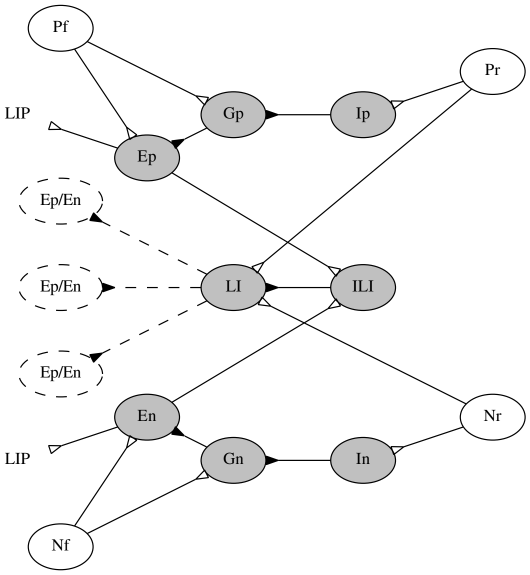

Figure 1. Disinhibition circuit with lateral inhibition. Grey nodes form the disinhibition circuit: Ep: excitatory population (positive), Gp: gating population (positive), Ip: Inhibitory population (positive). En, Gn, In denote negative populations. ILI: Inhibitory lateral inhibition population, LI: Lateral inhibition population inhibits positive/negative neighbouring populations of the forward network. 'f' suffix denotes populations from forward network, 'r' suffix denotes populations from reverse network. Black triangles denote inhibitory synapses, white shows excitatory synapse.

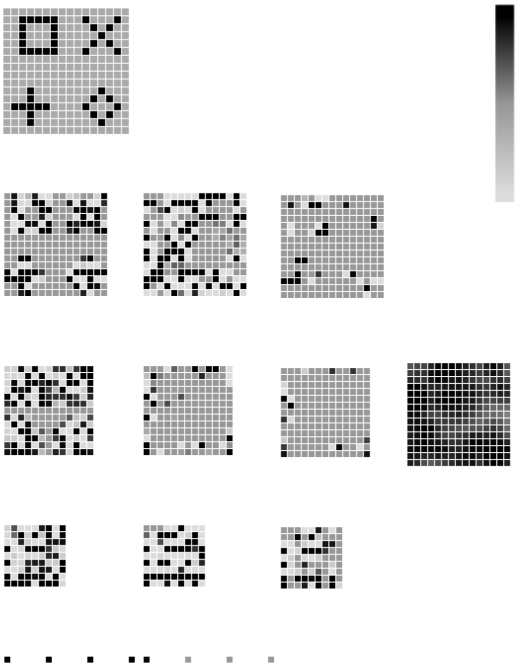

Figure 2. Final activity of network without lateral inhibition. The spread of activity across LIP shows no clear locus for spatial attention. From the left, the columns show the stimulus driven pathway of the ventral stream, the top-down stream, the disinhibition layers, then an LIP layer on the right. The rows of the first 3 columns show V1 (top), V2, V4, PIT and AIT (bottom).

template, selected by activating an AIT population in the reverse network. The connectivity of the disinhibition and lateral inhibition circuit could also be modified at this stage to create simulations from identical inputs with and without lateral inhibition, by removing the projections from LI to neighbouring populations.

Figure 2 shows an image of a simulation without lateral inhibition at the end of the two second simulation. Inputs to V1 and AIT were constant, so the image shows the steady state of the network. The bottom row of the image shows four AIT populations of the forward network, all showing high activity, demonstrating the presence of the four types of objects in the input array. Next to these are the four AIT populations of the reverse network, showing high activity in the left most population which codes for squares. The three layers above, are PIT, V4 and V2. The third column shows matching activity in the forward and reverse layers, and shows the activity of the Ep and En populations of the disinhibition circuits. Output of these populations is mutually exclusive, allowing activity of both populations to be visualised as a single element. Strong positive activity (Ep) is coloured black, strong negative activity (En) the lightest grey, and zero activity the middle grey in the depicted colour bar.

The interesting image in figure 2 is the activity in the LIP layer. Without lateral inhibition, the high levels of activity in the forward network causes a high level of activity across the LIP layer, and this activity is poorly correlated with the location of activity in V1. This shows that busy visual scenes with stimuli across the visual array prevents a winning location being clearly resolved, and therefore the location to direct attention or a saccade cannot be determined. This image shows the need for large scale inhibition to reduce the location noise in LIP.

Enabling lateral inhibition raises a number of interesting events in the neural dynamics. Visualisation software shows shifting patterns of activity in the disinhibition network and the LIP layer through the time course of the simulation. All simulations were run for 1.5 seconds and activation rates for each 5 millisecond period were recorded. Due to space constraints only three images are given in this paper, showing the more interesting features of the simulation.

Figure 3 shows a snapshot of the network 30 milliseconds after stimulus presentation. At this early stage the forward network causes activity in the Ep/En and Gp/Gn populations, but Ep/En is not yet inhibited by Gp/Gn so LIP receives excitatory spikes from all locations containing the presented objects. The image of the LIP layer shows high activity in the four quadrants corresponding to the activity in V1. This demonstrates the capture of attention by new visual stimuli. Spatial or feature-based attention would maintain activity in some (or all) of these active regions. However, in this simulation these activities are ignored.

After 500 milliseconds, the AIT population coding for squares becomes active and the attentional template is propagated to lower layers, gating the stimulus initiated activity to LIP at matching locations. The temporal nature of the dynamic simulation allows the time course of this disinhibition to be captured. While not shown in the images, location activity in LIP is first generated from activity in higher layers, and supplemented with activity from lower layers as the spread of activations from the attentional template descends the reverse network.

Figure 4 shows the network activity 350 milliseconds after activation of the AIT population in the reverse network. Competition in the disinhibition layers is has finished, leaving a single population of high spike rates being output to LIP. Neighbouring populations are also elevated above baseline, as their influence on LIP can be seen. However, this activity is barely visible on the image. Similarly, the areas of LIP activity seen in figure 3 are slightly elevated above baseline, as the inhibition of the Ef/En populations is not total.

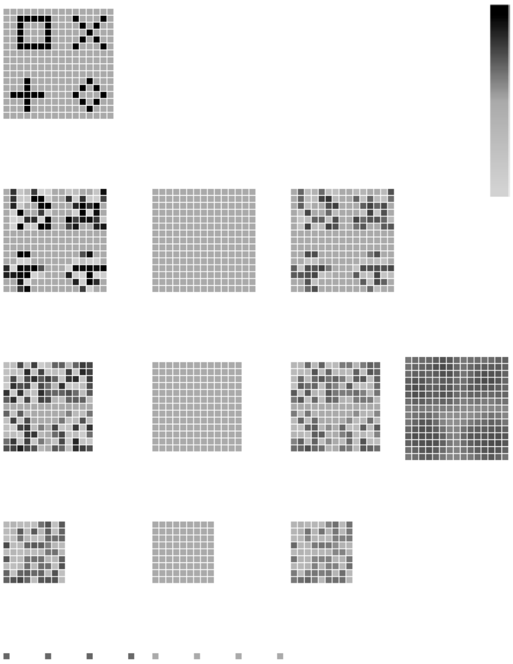

Figure 3. Network activity with lateral inhibition, trained on all locations, 30ms after stimulus onset. Prior to top-down activation, the stimulus activations cause location activity in LIP, demonstrating attentional capture. The layout is as described in figure 2.

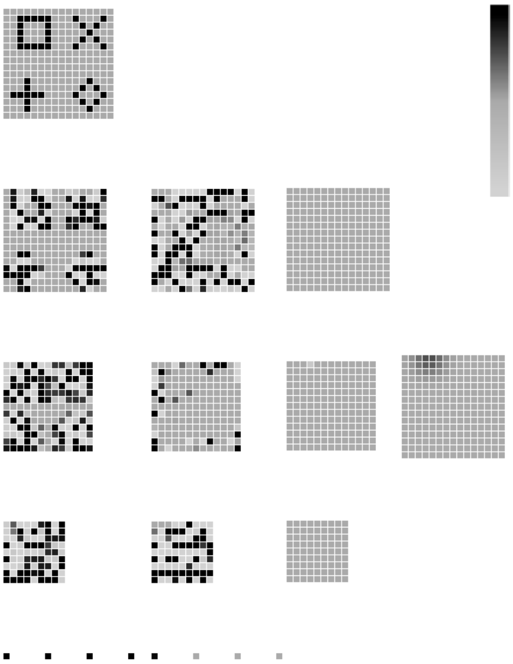

Figure 4. Network activity with lateral inhibition, trained on all locations, 350ms after simulation of the AIT population coding for the square. The layout is as described in figure 2.

Figure 5 shows the state of the simulation 1 second after activation of the cue population. This image demonstrates the location of the searched for feature persists with the application of attention. This contrasts with the LIP activity shown in figure 3, which is rapidly subdued.

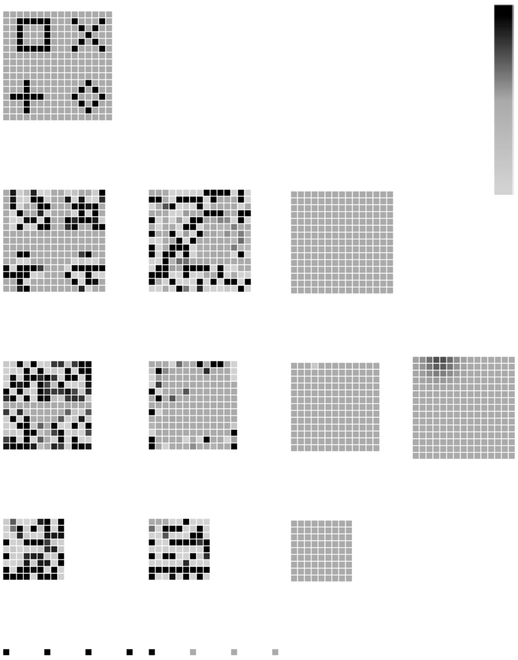

Figure 5. Network activity with lateral inhibition, trained on all locations, 500ms after stimulation of the square AIT population in the reverse network. The image shows maintained activity in LIP for locations of attended features. The layout is as described in figure 2.

In order to test the robustness of the model, simulations were generated with one to five objects in various positions of the visual array. In all cases the location of the resulting LIP activity covered at least part of the target object in V1. The model was able to resolve the location of objects even when distractor objects were overlaid on the target. Figure 6 shows an example of this, with the target square being partially obscured by an overlaid diamond. Comparison of figures 5 and 6 shows the activity in LIP to be slightly reduced. This reduction in activity occurs due to neurons with both the target and distractor in the receptive field experiencing more inhibition, and less excitatory stimulation as their receptive fields are effectively reduced in size.

4 DISCUSSION

Motter[13] showed that the shrinking of the receptive field around an attended object is accompanied by an enhancement of stimulus activity (in the forward network) in neurons within that receptive field in the ventral pathway. Currently the model shrinks the receptive field to inside the boundary of the attended object. Increasing activity in the forward network will help borderline circuits neighbouring the receptive field to win their competitions with their inhibitory neighbours, and expand the receptive field. This could be achieved in the ventral stream by augmenting the disinhibition circuit to project excitatory connections from Ep/En to neighbouring Pf/Pn populations, or by reciprocal excitatory connections from the dorsal stream. As lateral inhibition affects the Ep/En populations of the disinhibition circuit, modulation of the Pf/Pn populations through such mechanisms

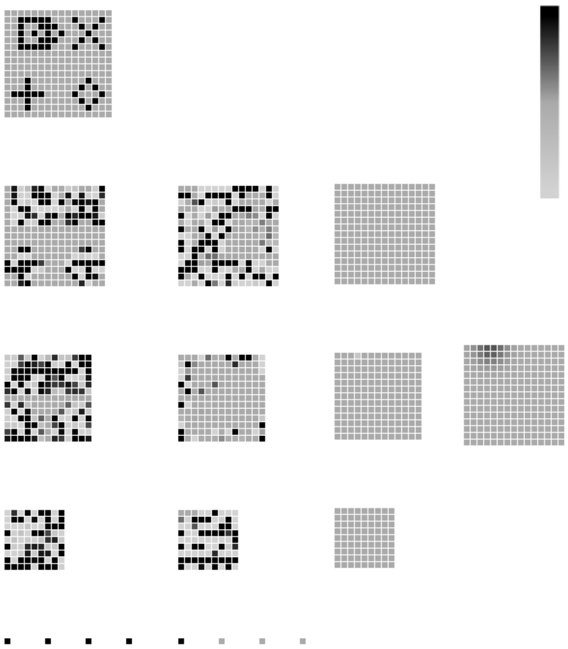

Figure 6. Network activity with lateral inhibition, trained on all locations, 500ms after stimulation of the square AIT population, with the target square (top left) partially obscured. The layout is as described in figure 2.

is feasible with the current architecture. Future work will investigate this process.

The attentional capture mechanism was not demonstrated in the simulations depicted here. However, we have run trials with activity in the reverse network prior to presentation of stimuli to V1. This models a Posner-like paradigm of a cue being presented and held in memory briefly before presentation of an array of targets and distractors. In simulations of this type, the location of the matching objects are not extinguished in LIP after the stimulus driven pulse of activity. While we do not provide direct evidence here, this can be seen from the architecture of the disinhibition circuit: stimuli matching the attentional template never receive the inhibition from Gp/Gn populations as they are already inhibited by activity from the top-down (reverse) network. Evidence for the modulation of activity in the the ventral stream through the action of attention without bottom-up visual stimuli has been demonstrated to act throughout the visual field[16], as exhibited by this model.

When the input array contains multiple objects to be attended, the objects' locations are found by appropriate activations in LIP. A form of inhibition of return could be implemented to resolve the multiple candidate locations to an ordered list of eye movements, based on criteria such as largest spike rate, or largest area of activity. The model as presented does not use the activity in LIP to generate actions or saccades, but the required neural information for such actions is available for use by these mechanisms.

5 CONCLUSION

We have shown that the use of an attentional template coupled with lateral inhibition of circuits neighbouring mismatched bottom-up and top-down populations can provide a resolution of targets from distractors in a mixed visual array. The lateral inhibition and disinhibition mechanism employed by the model allows for lateral inhibition for feature-based attention, while still supporting attentional capture: excitatory output to the dorsal stream (LIP) only occurs when the stimulus-driven activity is large enough (low pass filter), or when the attentional template matches.

We interpret our use of a top-down attentional template in combination with the bottom-up activity as a simple form of a priority map [2]. Attention effectively gates the output of the ventral stream [12] to visual areas in the dorsal stream, where planning of actions can be initiated. The inhibitory mechanism in effect causes the receptive field of neurons to shrink around the attended to object, as described by Motter[13, 14] and others [8, 1], while implementing a biased competition[7] between neurons coding for different features within a receptive field, and inhibiting activity of neighbouring neurons with distractor objects with their receptive field[10]. It should be possible to model other neural correlates associated with the deployment of attention, such as elevated bottom-up activity from attended stimuli, within the presented framework.

REFERENCES

[1] J. W. Bisley, 'The neural basis of visual attention', *The Journal of Physiology*, **589**(1), 49–57, (2011).

[2] J. W. Bisley and M. E. Goldberg, 'Neuronal activity in the lateral intraparietal area and spatial attention', *Science*, **299**(5603), 81 – 86, (2003).

[3] G. M. Boynton, 'Attention and visual perception', *Current Opinion in Neurobiology*, **15**(4), 465–469, (2005).

[4] G. M. Boynton, 'A framework for describing the effects of attention on visual responses', *Vision Research*, **49**(10), 1129–1143, (2009).

[5] M. de Kamps and V. Baier, 'Multiple Interacting Instantiations of Neuronal Dynamics (MIIND): aLibrary for Rapid Prototyping of Models in Cognitive Neuroscience', in *International Joint Conference on Neural Networks*, pp. 2829 – 2834, (2007).

[6] M. de Kamps and F. van der Velde, 'From artificial neural networks to spiking neuron populations and back again', *Neural Networks*, **14**(6-7), 941–953, (2001).

[7] R. Desimone and J. Duncan, 'Neural mechanisms of selective visual attention.', *Annual Review Neuroscience*, **18**, 193–222, (1995).

[8] F. H Hamker and M. Zirnsak, 'V4 receptive field dynamics as predicted by a systems-level model of visual attention using feedback from the frontal eye field', *Neural Networks*, **19**(9), 1371 – 1382, (11 2006).

[9] L. Itti and C. Koch, 'A saliency-based search mechanism for overt and covert shifts of visual attention', *Vision Research*, **40**, 1489–1506, (2000).

[10] S. Kastner, M. Pinsk, P. De Weerd, R. Desimone, and L. Ungerleider, 'Increased activity in human visual cortex during directed attention in the absence of visual stimulation', *Neuron*, **22**, 751–761, (1999).

[11] J. C. Martinez-Trujillo and S. Treue, 'Feature-based attention increases the selectivity of population responses in primate visual cortex', *Current Biology*, **14**(9), 744 – 751, (2004).

[12] J. Moran and R. Desimone, 'Selective attention gates visual processing in the extrastriate cortex', *Science*, **229**(4715), 782 – 784, (1985).

[13] B. C. Motter, 'Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli', *Journal of Neurophysiology*, **70**(3), 909 – 119, (1993).

[14] B. C. Motter, 'Neural correlates of feature selective memory and pop-out in extrastriate area V4', *The Journal of Neuroscience*, **14**(4), 2190 – 2199, (1994).

[15] R. A. Rensink, 'Seeing, sensing and scrutinizing', *Vision Research*, **40**, 1469–1487, (2000).

[16] J. T. Serences and G. M. Boynton, 'Feature-based attentional modulations in the absence of direct visual stimulation', *Neuron*, **55**(2), 301 – 312, (2007).

[17] F. Van Der Velde and M. De Kamps, 'From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation', *Journal of Cognitive Neuroscience*, **13**(4), 479–491, (2001).

[18] H. R. Wilson and J. D. Cowan, 'Excitatory and inhibitory interactions in localized populations of model neurons', *Biophysical Journal*, **12**, 1 – 24, (1972).

[19] S. Yantis, 'To see is to attend', *Science*, **299**(5603), 54 – 56, (2003).

[20] W.W. Zhang and S.J. Luck, 'Feature-based attention modulates feedforward visual processing', *Nature Neuroscience*, **12**(1), 24–25, (2009).