Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2009-047
CBCL-280

October 3, 2009

A Bayesian inference theory of attention: neuroscience and algorithms

Sharat Chikkerur, Thomas Serre, and Tomaso Poggio

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

A Bayesian inference theory of attention: neuroscience and algorithms

Sharat Chikkerur, Thomas Serre and Tomaso Poggio
Center for Biological and Computational Learning, MIT

Abstract

The past four decades of research in visual neuroscience has generated a large and disparate body of literature on the role of attention [Itti et al., 2005]. Although several models have been developed to describe specific properties of attention, a theoretical framework that explains the computational role of attention *and* is consistent with all known effects is still needed. Recently, several authors have suggested that visual perception can be interpreted as a Bayesian inference process [Rao et al., 2002, Knill and Richards, 1996, Lee and Mumford, 2003]. Within this framework, top-down priors via cortical feedback help disambiguate noisy bottom-up sensory input signals. Building on earlier work by Rao [2005], we show that this Bayesian inference proposal can be extended to explain the role and predict the main properties of attention: namely to facilitate the recognition of objects in clutter. Visual recognition proceeds by estimating the posterior probabilities for objects and their locations within an image via an exchange of messages between ventral and parietal areas of the visual cortex. Within this framework, spatial attention is used to reduce the uncertainty in feature information; feature-based attention is used to reduce the uncertainty in location information. In conjunction, they are used to recognize objects in clutter. Here, we find that several key attentional phenomena such such as pop-out, multiplicative modulation and change in contrast response emerge naturally as a property of the network. We explain the idea in three stages. We start with developing a simplified model of attention in the brain identifying the primary areas involved and their interconnections. Secondly, we propose a Bayesian network where each node has direct neural correlates within our simplified biological model. Finally, we elucidate the properties of the resulting model, showing that the predictions are consistent with physiological and behavioral evidence.

1 Introduction

The recognition and localization of multiple objects in complex cluttered visual scenes is a difficult problem for both biological and machine vision systems. Most modern computer vision algorithms (*e.g.*, [Schneiderman and Kanade, 2000, Viola and Jones, 2001]) scan the image over a range of positions and scales. Instead the primate visual system proceeds in discrete shifts of attention and eye movements. The past four decades of behavioral research in attention has generated a large body of experimental data (see [Wolfe, 2007, Itti et al., 2005] for a recent review). A theoretical framework, however, that explains the computational role of attention *and* predicts its effects is still missing.

Several theoretical proposals and computational models have been described to try to explain the main functional and computational role of visual attention. One influential proposal by Tsotsos [1997] is that attention reflects evolution's attempt to fix the processing bottleneck in the visual system [Broadbent, 1958] by directing the finite computational capacity of the visual system preferentially to relevant stimuli within the visual field while ignoring everything else. Treisman and Gelade [1980] suggested that attention is used to *bind* different features (e.g. color and form) of an object during visual perception. Duncan [1995] suggested that the goal of attention is to bias the choice between competing stimuli within the visual field. These proposals however remain agnostic about how attention should be implemented in the visual cortex and do not yield any prediction about the various behavioral and physiological effects of attention. On the other hand, computational models attempt to model specific behavioral and physiological effects of attention. Behavioral effects include pop-out of salient objects [Itti et al., 1998, Zhang et al., 2008, Rosenholtz and Mansfield, 2005], top-down bias of target features [Wolfe, 2007, Navalpakkam and Itti, 2006], influence from scene context [Torralba, 2003], serial vs. parallel-search effect [Wolfe, 2007] etc. Physiological effects include multiplicative modulation of neuron response under spatial attention [Rao, 2005] and feature based attention [Bichot et al., 2005]. A unifying computational framework that can account for these disparate effects is missing.

Recently, several authors have suggested that visual perception can be interpreted as a Bayesian inference process [Dayan and Zemel, 1999, Lee and Mumford, 2003, Rao, 2005, Yu and Dayan, 2005, Epshtein et al., 2008] where top-down signals are used to disambiguate noisy bottom-up sensory input signals. Following this idea, attention can be regarded as an inference process that disambiguates form and location information [Yu and Dayan, 2005]. In particular, Rao [2005] proposed a probabilistic Bayesian model of spatial attention whereby perception corresponds to estimating posterior probabilities of features and their locations in an input image.

Here we extend Rao's model of spatial attention to incorporate top-down feature-based attention. We show that the resulting model is capable of performing visual searches in cluttered visual scenes. We show that the model is consistent with neurophysiology experiments about the role of feature-based and spatial attention [Bichot et al., 2005, Reynolds J. H., 2009]. We also find that ,surprisingly, several key attentional phenomena such such as pop-out, multiplicative modulation and change in contrast response also emerge naturally as a property of the model.

1.1 Background: Recognition in clutter

Our central hypothesis is that attention is a mechanism that has evolved to improve recognition in clutter. The ventral stream in the primate visual cortex [Ungerleider and Haxby, 1994], which plays a key role in object recognition, achieves a difficult tradeoff between selectivity and invariance – in particular tolerance to position and scale [1]. Studies examining the tuning properties of these neurons have shown that the complexity of the preferred stimulus increases as we go further along the ventral stream: from simple oriented bars in area V1 [Hubel and Wiesel, 1959], curvature in V2[Hegde and Van Essen, 2000, Ito and Komatsu, 2004], simple shapes in V4 [Pasupathy and Connor, 2001, Desimone and Schein, 1987, Gallant et al., 1996] and finally to object selective cell in area IT [Tanaka et al., 1991, Tanaka, 1996]. This gradual increase in selectivity provides a convincing explanation to how the brain recognizes complex objects, but cannot explain how it can do so in a location invariant manner. In their seminal study, Hubel and Wiesel observed the existence of simple and complex cells in the striate cortex [Hubel and Wiesel, 1959]. Simple cells were found to be sensitive to both position and orientation of the stimulus while complex cells exhibited the same tuning with greater invariance to position. It was suggested that a complex cell achieves translation invariance by pooling responses from simple cells. Oram and Perret [Perrett and Oram, 1993] suggested that pooling mechanism between simple and complex cell may be extended to higher regions of the ventral stream to achieve invariant recognition.

Based on existing neurobiological models [Wallis and Rolls, 1997, Mel, 1997, Riesenhuber and Poggio, 1999, Ullman et al., 2002, Thorpe, 2002, Amit and Mascaro, 2003, Wersing and Koerner, 2003], conceptual proposals [Hubel and Wiesel, 1968, Perrett and Oram, 1993, Hochstein and Ahissar, 2002, Biederman, 1987] and computer vision systems [Fukushima, 1980, LeCun et al., 1998] Serre et al. [Riesenhuber and Poggio, 1999, Serre et al., 2002, 2005a, 2007b] showed that a computational model of the ventral stream that gradually builds up specificity through composition and invariance through max-pooling and can account for responses of V4 [Cadieu et al., 2007], IT neurons [Hung et al., 2005] in the ventral stream and also behavioral aspect of human vision, namely ultra-rapid complex object recognition [Fabre-Thorpe et al., 2001]. The main limitation of the model is that it does not take into account the massive amount of backprojections that convey information from higher cortical areas back to the lower areas [Felleman and Essen, 1991].

It was also shown that recognition performance of these "tolerant" networks is significantly affected by clutter and image "crowding". The effect of crowding and clutter has been observed not only in human psychophysics [Serre et al., 2007b] but also at all stages of the ventral stream [Reynolds et al., 1999, Zoccolan et al., 2007]. We propose that this problem is symptomatic of feed forward systems that achieve invariance through pooling. A natural solution to the clutter problem of architectures such as the ventral stream is a mechanism to suppress regions of the image that are unlikely to contain objects to be recognized. We propose that this mechanism – and the algorithms to support its function, in particular the choice of

[1] The tolerance property may also play a role in helping maintain a stable percept of visual objects in spite of ego-motion, eye movements or motion of the objects themselves.

the regions to be suppressed – is in fact visual attention. This argument is different from the usual “bottleneck” argument [Tsotsos, 1997], which assumes limited computational resources to analyze the visual scene. Instead the model suggests that attention is needed to eliminate interference by the clutter.

1.2 Background: Visual attention in the brain

In this section, we identify the primary regions of the brain involved in attention and their interconnections. This simplified model will serve as the basis for our Bayesian formulation. Koch and Ullman [1985] postulated the existence of a *saliency map* in the visual system. The *saliency map* combines information from several abstract feature maps (*e.g.*, local contrast, orientations, color) into a global saliency measure that indicates the relevance of each position in the image. Consistent with this hypothesis, models of attention have assumed that there exist two stages of visual processing. In a pre-attentive parallel processing mode, the entire visual field is processed at once to generate a saliency map which is then used to guide a slow serial attentive processing stage, in which a region of interest (attentional spotlight) is selected for “specialized” analysis. However, the neural correlate of the saliency map remains to be found.

Prior studies [Colby and Goldberg, 1999] have shown that the parietal cortex maintains a spatial map of the visual environment and in fact maintains several frames of reference (eye-centered, head-centered etc) making it a likely candidate for the saliency computation. Studies show that response of LIP neurons within the parietal are correlated with likelihood ratio of the target object [Bisley and Goldberg, 2003]. In this paper, our hypothesis – which is not critical for the theory and is mainly dictated by simplicity – is that the saliency map is represented in LIP.

In addition to computing saliency, circuits are also needed to plan the shifts of attention, that is, to plan and serialize the search by prioritizing candidate shifts of attention and holding them in working-memory until the saccade has been initiated. Because of its overlap with the prefrontal cortex (PFC), the frontal eye field (FEF) is a good candidate for shifting the focus of attention. Recent evidence [Buschman and Miller, 2007] further supports the role of FEF in spatial and feature based attention. We speculate that the FEF stimulation effect reported by Moore and Armstrong [2003] (i.e., an enhancement observed in V4 receptive field locations that match the region of the visual field represented at the FEF stimulation site) is indirect, mediated through LIP.

In addition to the parietal region, the ventral stream is also intimately involved in attention. Zhaoping and Snowden [2006], Itti et al. [1998] have proposed computational models based on V1-like features showing that they are sufficient to reproduce attentional effects such as pop-out and search asymmetries. However, recent evidence shows V1 to be relatively unaffected by top-down attentional modulation [Hegde and Felleman, 2003], thus moving the locus of attention away from V1 and towards higher regions such as V4. Experiments on spatial attention [McAdams and Maunsell, 1999] and feature-based attention [Bichot et al., 2005] have shown attentional modulation in V4. In particular, feature-based attention is found to modulate the response of V4 neurons at all locations–the activities are increased if the preferred stimulus of the neurons is the same as the target stimulus and suppressed otherwise. Under spatial attention, V4 neurons that have receptive fields overlapping with the locus of attention are found to be enhanced. Thus V4 neurons are involved in feature-based attention as well as spatial attention suggesting that V4 serves as the area of interaction between ventral and parietal cortices. Several computational models have attempted to model the interaction between the ventral and parietal/dorsal regions [Rao, 2005, Van Der Velde and De Kamps, 2001].

2 Our approach

In this work, we explicitly model the interaction between the ventral and parietal cortical regions and integrate these interactions within a feed forward model of the ventral stream [Serre et al., 2007a]. The main addition to the feed forward model is (i) the inclusion of cortical feedback within the ventral stream (providing feature-based attention) and (2) from areas of the parietal cortex onto areas of the ventral stream (providing spatial attention) and, (3) feed forward connections to the parietal cortex that serve as a ’saliency map’ encoding the visual relevance of individual locations [Koch and Ullman,

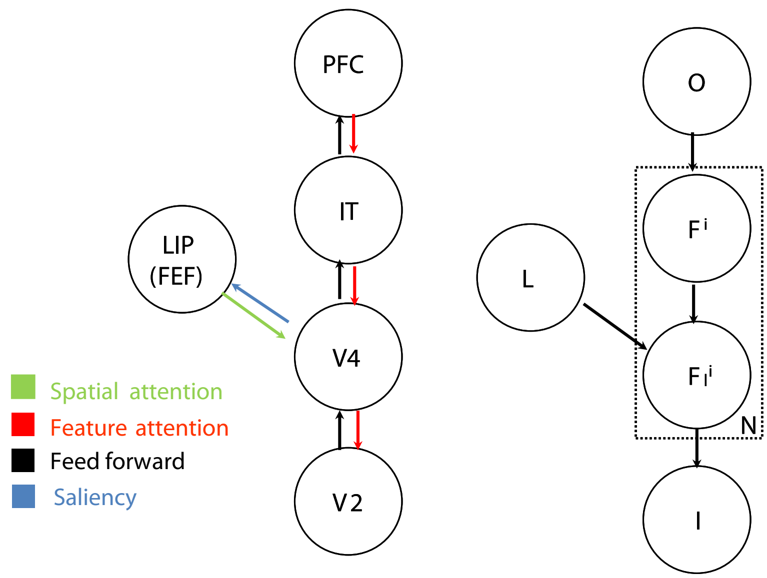


Figure 1: Left: A model illustrating the interaction between the parietal and ventral streams mediated by feed forward and feedback connections. Right: The proposed Bayes network that models these interactions.

1985]. The model is directly inspired by the physiology of attention and extends a Bayesian model of spatial attention proposed by Rao [2005].

2.1 Computational model

The model (see Fig. 1) consists of a location encoding unit (L), object encoding units O, non-retinotopic feature encoding units F^i and *combination* units F^i_l, that encode position-feature combinations. These units receive input I from lower areas in the ventral stream. L models LIP area in the parietal cortex and encodes position and scale independently of features. F^i units model non-retinotopic, spatial and scale invariant cells found in higher layers of the ventral stream [2].

The location variable L is modeled as a multinomial random variable with $|L|$ distinct locations. Attentional spotlight due to spatial attention is modeled by concentrating the prior at and around the location of interest. The prior on the location variable can be specified explicitly by spatial cues or task-dependent constraints (such as location constraints imposed by the scene [Torralba, 2003]). The spatial prior biases the saliency map toward task-relevant locations.

The object variable, O is a multinomial random variable with $|\mathrm{O}|$ values corresponding to objects known by the model. The prior $\mathrm{P(O)}$ is set based on the search task. Each feature-encoding unit $\mathrm{F^i}$ is a binary random variable that represents the presence or absence of a feature irrespective of location and scale. The arrays of feature detectors $\mathrm{F^i_l}$ are assumed to be modulated according to how diagnostic they are for the specific categorization task at hand. For instance, if the task is to find a pedestrian, the pedestrian-selective O units at the top will modulate units in lower areas that are important to discriminate between pedestrians vs. the background. The main effect of this feature-based modulation is a suppression of the response of task-irrelevant but otherwise salient units so as to bias the saliency map towards task-relevant locations. In practice, top-down bias on features is implemented by learning the conditional probability $P(F^i|O)$ over a set of training data. The variable F^i_l represents the feature map and is modeled as a multinomial variable with $|L|+1$ values $(0, 1 \cdots L)$. The conditional probability $P(F^i_l|F^i, L)$ is specified such that when feature F^i is present ($F^i = 1$), at location l^* ($L = l^*$),

[2]It is worth nothing that neurons in the brain have limited invariance properties–extending to few degrees of visual angle. In this work, we assume that the invariance extends across the entire image for simplicity.

the value $F_l^i = l^*$ and nearby locations are likely to be activated. However, when the feature F^i is absent ($F^i = 0$), only the value $F_l^i = 0$ is likely to be activated. In addition to this top-down prior, each value $\mathrm{F_l^i}$ receives bottom-up evidence from the input $\mathrm{P(I|F_l^i)}$, that is proportional to the activity of the feature F^i at location $L = l$. For instance, if features F^i correspond to orientations, the feature map is computed using oriented Gabor filters [Daugman, 1980]. Given the image I, for each orientation and location, $P(I|F_l^i)$ is set proportional to the output of the filter. The response can be passed through a sigmoid or even discretized without affecting the model. In our case, the features F^i correspond to shape-based features of intermediate complexity [Serre et al., 2007a]. Furthermore, we discretized the feature response for simplicity. (For detailed description about the units and their conditional probability, see the appendix). The relative dynamics between these three main components is governed by a series of messages passed within the ventral stream and between the ventral stream and the parietal cortex, which we describe below.

2.2 Message passing

Within the Bayesian framework, feed forward signals are interpreted as bottom-up evidence and top-down feedback is modeled as priors. Given the input image, the posterior probability over location corresponds to the saliency map. In order to understand the model, we examine the messages passed between units in the system under a single feature . We adopt the notation proposed in [Rao, 2005], where the top-down messages, $\pi()$ and bottom-up messages $\lambda()$ are replaced by a uniform $m()$ term.

$$
\begin{aligned}
m_{O\rightarrow F^i} &= P(O) && (1)\\
m_{F^i\rightarrow F_l^i} &= \sum_O P(F^i|O)P(O) && (2)\\
m_{L\rightarrow F_l^i} &= P(L) && (3)\\
m_{I\rightarrow F_l^i} &= P(I|F_l^i) && (4)\\
m_{F_l^i\rightarrow F^i} &= \sum_L \sum_{F_l^i} P(F_l^i|F^i,L)(m_{L\rightarrow F_l^i})(m_{I\rightarrow F_l^i}) && (5)\\
m_{F_l^i\rightarrow L} &= \sum_{F^i} \sum_{F_l^i} P(F_l^i|F^i,L)(m_{F^i\rightarrow F_l^i})(m_{I\rightarrow F_l^i}) && (6)
\end{aligned}
$$

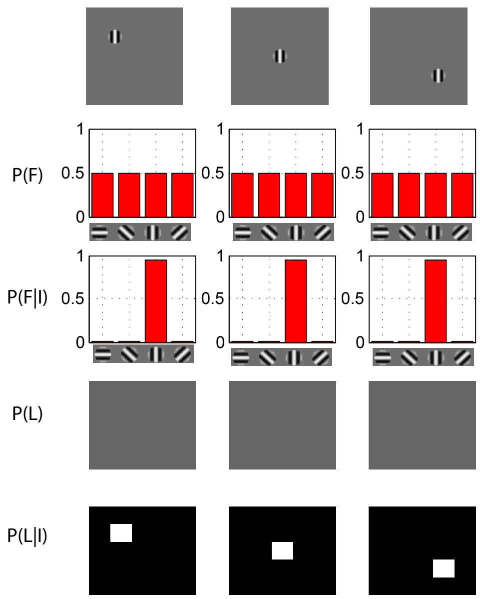
The first three messages correspond to the priors imposed by the task. The rest correspond to bottom-up evidence propagated upwards within the model. The posterior probability of location (saliency map) is given by

$$P(L|I) \propto (m_{L\rightarrow F_l^i})(m_{F_l^i\rightarrow L}) \qquad (7)$$

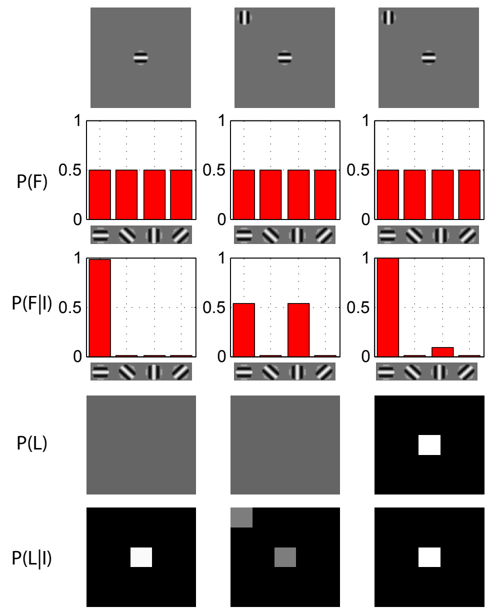
The constant of proportionality can be resolved after computing marginals over all values of the random variable. Thus, the saliency map is influenced by task dependent prior on location $P(L)$, prior on features $P(F^i|O)$ as well as the evidence from the ventral stream $m_{F_l^i\rightarrow L}$. As a side note, it is worth mentioning that the summations in the message passing equations are done over all the discrete states of the variable. Thus, L is summed over its states, $\{1, 2 \cdots |L|\}$, F^i is summed over $\{0, 1\}$ and F_l^i, over states $\{0, 1, \cdots |L|\}$.

Multiple features: Under multiple features, the Bayesian inference proceeds as in a general polytree [Pearl, 1988]. Most messages remain identical. However, the bottom-up evidence for location is influenced by the presence of other features and is now given by

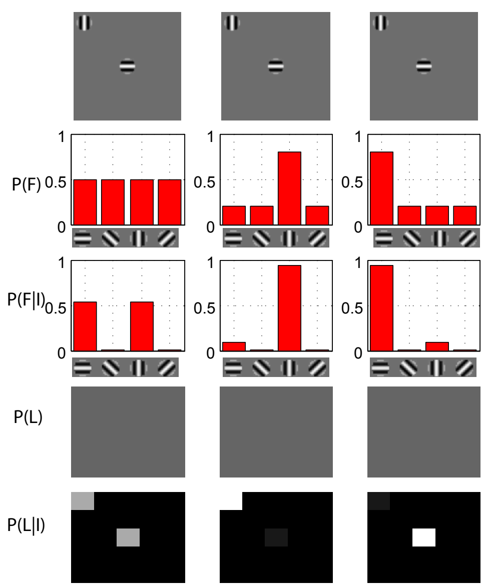
$$
\begin{aligned}
m_{L\rightarrow F_l^i} &= P(L)\prod_{j\neq i} m_{F_l^j\rightarrow L} && (8)\\
P(L|I) &\propto P(L)\prod_i m_{F_l^i\rightarrow L} && (9)
\end{aligned}
$$



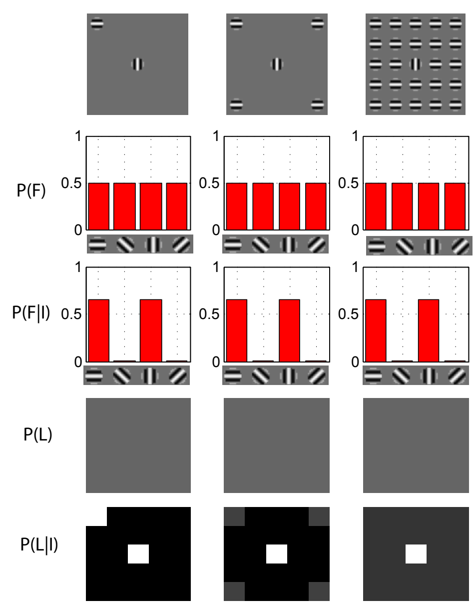
(a) Spatial invariance.



(b) Spatial attention.

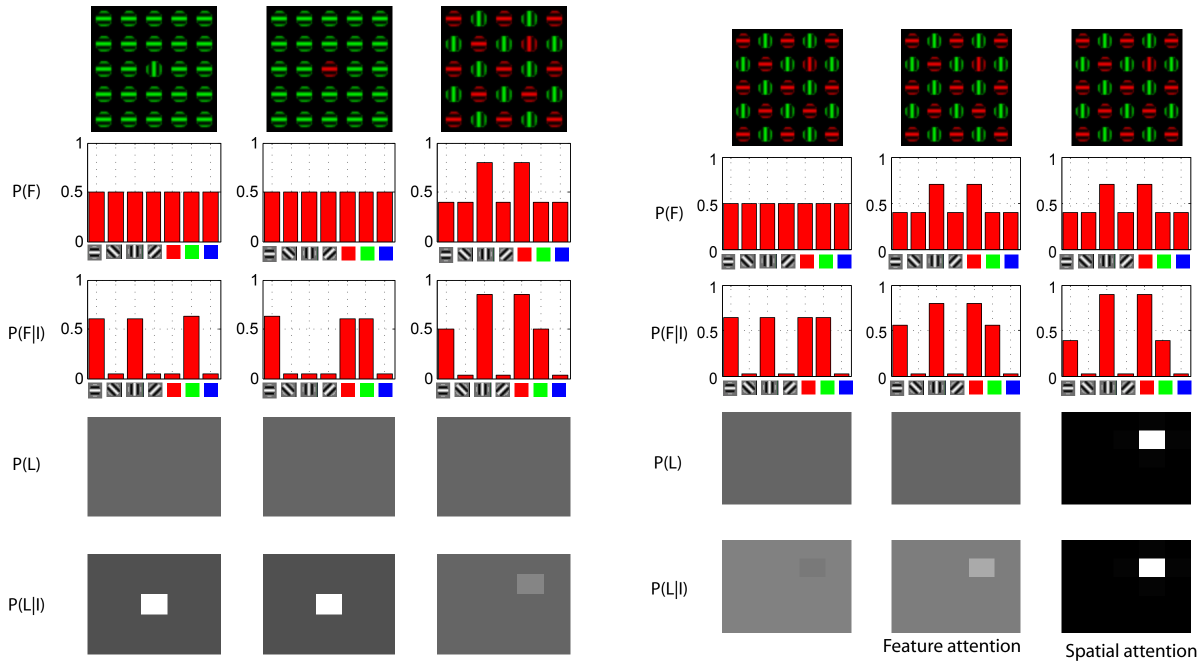


(c) Feature attention.



(d) Feature popout.

Figure 2: (a) The response of the feature encoding units is independent of the location of the stimuli. This is achieved through marginalization across locations.(b) With isolated stimuli (left column), the uncertainty in features and locations are minimal. The uncertainty increases when under the presence of distracting stimuli (middle column). However, when spatial attention is deployed (adjusting $P(L)$), the response returns to its initial value as if the distractor was not present. (c) In the absence of attention, the saliency map is agnostic about the identity of the stimulus (left column). During a search task, the prior on the desired on the desired feature is increased (vertical in the middle column, horizontal in the right column). The saliency map is now biased towards the location of the target stimulus. (d) The saliency map is biased towards locations containing stimuli that differ from stimuli at other locations (across each feature dimension). Left column shows equally salient stimuli. In the middle column, the vertical stimuli is more salient. The pop-out effect is more pronounced in the right column.



(a) Parallel vs. serial search.

(b) Object recognition under clutter.

Figure 3: (a) When a stimulus differs from the background in a single dimension (first and second column) search is easy (indicated by a high contrast saliency map) and independent of the number of distractors. However, when both features are shared between the target and distractors (third column), search is more difficult (indicated by a low contrast saliency map). (b) Object recognition under clutter consists of feature-based attention followed by spatial attention. The most likely location of the target is found by deploying feature-based attention (the feature priors are changed middle column). The hypothesis is then verified by deploying spatial attention (the spatial priors are changed in the right column). The value of the feature units $P(F^i|I)$ indicate the presence or absence of the object feature.

3 Results

3.1 Translation invariance

The F^i units encode the presence or absence of individual features in a translation/scale invariant manner. The invariance is achieved by pooling responses from all locations. The posterior probability of the feature F^i is given by

$$P(F^i|I) \propto (m_{F^i \to F^i_l})(m_{F^i_l \to F^i}) \quad (10)$$

$$\text{where, } m_{F^i_l \to F^i} = \sum_L \sum_{F^i_l} P(F^i_l|F^i, L)P(L)P(I|F^i_l) \quad (11)$$

Spatial invariance is achieved by marginalizing (summing over) the L variables. Thus, F^i behaves similarly to non-retinotopic units (tuned to specific features) found in the ventral stream [Serre et al., 2005b].

3.2 Spatial attention

Spatial attention corresponds toconcentrating the prior $\mathrm{P(L)}$ around the location/scale of interest. Such a change in the prior is propagated from L to $\mathrm{F^i_l}$ (through messages in the Bayesian network). This results in a selective enhancement of all feature maps $\mathrm{F^i_l}$ for $\mathrm{i = 1 \dots n}$ at locations $\mathrm{l_1 \dots l_m}$ where the attentional spotlight $\mathrm{P(L)}$ is placed and in suppression at other locations. This shrinks the effective receptive field of the non-retinotopic $\mathrm{F^i}$ units at the next stage.

The message passing isinitiated in the L units (assumed to be in parietal cortex) and manifests itself after a short delay in the $\mathrm{F^i}$ units (assumed to be found in the ventral stream), in agreement with physiological data [Buschman and Miller, 2007]. Thus, spatial attention results in a sequence of messages $\mathrm{L \rightarrow F^i_l \rightarrow F^i \rightarrow O}$ (see Fig.2).

3.3 Feature-based attention

During feature-based search task, the exact opposite sequence of messages is initiated. Priors on the object are changed based on the specific task so as to be concentrated on the target of interest (*e.g.*, cars vs. pedestrians). Spatial priors can still be imposed (based on task demands,*e.g.*,, spatial cueing) independently of the object priors. The change in object prior is propagated to the feature units, through the message $\mathrm{O \rightarrow F^i}$. This results in a selective enhancement of the features that are present in the target object and suppression of others. This preference propagates to all feature-map locations through message $\mathrm{m_{F^i \rightarrow F^i_l} = \sum_O P(F^i|O)P(O)}$. The L unit pools across all features $\mathrm{F^j_l}$ for $\mathrm{j = 1 \dots n}$ at a specific location l. However, because of the feature-based modulation, locations that have features associated with the object are selectively enhanced. Thus, priors of the objects in the ventral stream generate an attentional spotlight in the parietal cortex that corresponds to locations most likely to contain the object of interest. The message passing is thus initiated in the ventral stream first and is manifested in the parietal cortex (L units) only after a delay, in agreement with the recent data by Buschman & Miller [Buschman and Miller, 2007]. In summary, feature based attention results in a sequence of messages $\mathrm{O \rightarrow F^i \rightarrow F^i_l \rightarrow L}$ (see Fig.2).

To train the feature-based attention (determine $P(F^i|O)$), we compute feature maps for each of the training image. The feature maps are discretized to maximize classification accuracy [Fleuret, 2004]. The feature F^i is said to be present if its detected at any location in the feature map. $P(F^i|OZ)$ is determined by simply counting the frequency of occurrence of each features.

3.4 Feature pop-out

Since the states of F^i_l are mutually exclusive ($\forall i, \sum_{F^i_l} P(F^i_l|F^i, L) = 1$), increasing the activity at one location (through $m_{I \rightarrow F^i_l}$), has the effect of inhibiting the likelihood of the stimuli being present at other locations. This reproduces the well known effect of lateral inhibition observed in real neurons [Carandini et al., 1997]. Further, these changes are propagated to the location unit via the messages $m_{F^i_l \rightarrow L}$. As a result, feature dimensions that have fewer active stimuli induce a higher likelihood for individual locations (through $m_{F^i_l \rightarrow L}$) than feature dimensions with more active stimuli. This results in a feature 'pop-out' effect, where the saliency map is biased towards locations of 'surprising' stimuli (see Fig. 2).

3.5 Object recognition under clutter

During a visual search for a specific feature or object, top-down feature-based attention is first used to bias the saliency map towards locations that share features with the target. The sequence of messages are identical with feature-based attention ($\mathrm{O \rightarrow F^i \rightarrow F^i_l \rightarrow L}$). The saliency map ($P(L|I)$), provides the most likely location containing the target. The search now proceeds with the deployment of the spatial attention around the region of interest. The direct effect of this spatial attention is a shrinking of the receptive fields in the ventral stream around the attended region. The sequence of messages are identical with that of spatial attention ($\mathrm{L \rightarrow F^i_l \rightarrow F^i \rightarrow O}$).

Thus, object recognition under clutter involves the sequential application of feature-based attention and spatial attention (see Fig 3). To locate subsequent objects, the attentional spotlight is then shifted (possibly via the PFC and/or FEF onto LIP) to the next location [Posner and Cohen, 1984]. Mechanisms for selecting the size of the attentional spotlight are not well known. Computational models of attention have relied upon fixed size [Itti and Koch, 2001] or adapted it to the image, extending the spotlight to include proto-object boundaries [Walther and Koch, 2007]. Furthermore, the matter or visual acuity further confounds this choice. In this work, we assume the attentional spotlight to be of a fixed size.

3.6 Multiplicative modulation

Spatial attention: In [McAdams and Maunsell, 1999], it was observed that the tuning curve of a V4 neuron is enhanced (multiplicatively) when attention is directed towards its receptive field. We observe that this effect is also reproduced within the computational model. The response of a simulated neuron encoding feature i and at location x, is given by

$$P(F_l^i = x|I) \quad \propto \quad \sum_{F^i,L} P(F_l^i = x|F^i, L)P(I|F_l^i)P(F^i)P(L) \tag{12}$$

Under normal conditions, $P(L)$ and $P(F^i)$ have a uniform distribution and thus the response of the neuron is largely determined by the underlying stimulus ($P(I|F_l^i)$). Under spatial attention, the location priors are concentrated around $P(L = x)$. This leads a multiplicative change (from $P(L = x) = 1/|L|$ to $P(L = x) \approx 1$) enhancing the response, even under the same stimulus condition (see Fig. 4).

Feature based attention: We assume that classification units in the PFC and higher areas modulate arrays of feature detectors in intermediate areas of the ventral stream (PIT and V4) according to how diagnostic they are for the specific categorization task at hand. This is consistent with recent findings in physiology [Bichot et al., 2005] that that show multiplicative modulation of neuronal response under attention. It is also suggested that this modulation is effective at all locations within the receptive field. This effect is also observed within the computational model. Under normal conditions, $P(L)$ and $P(F^i)$ have a uniform distribution and thus the response of the neuron is largely determined by the underlying stimulus ($P(I|F_l^i)$). Under feature-based attention, the feature priors are modified to $P(F^i = 1) \approx 1$. This leads to a multiplicative change (from $P(F^i = 1) = 1/2$ to $P(F^i = 1) \approx 1$) enhancing the response at all locations. The response is more pronounced when the stimulus is preferred ($P(I|F_l^i)$ is high).

3.7 Contrast response

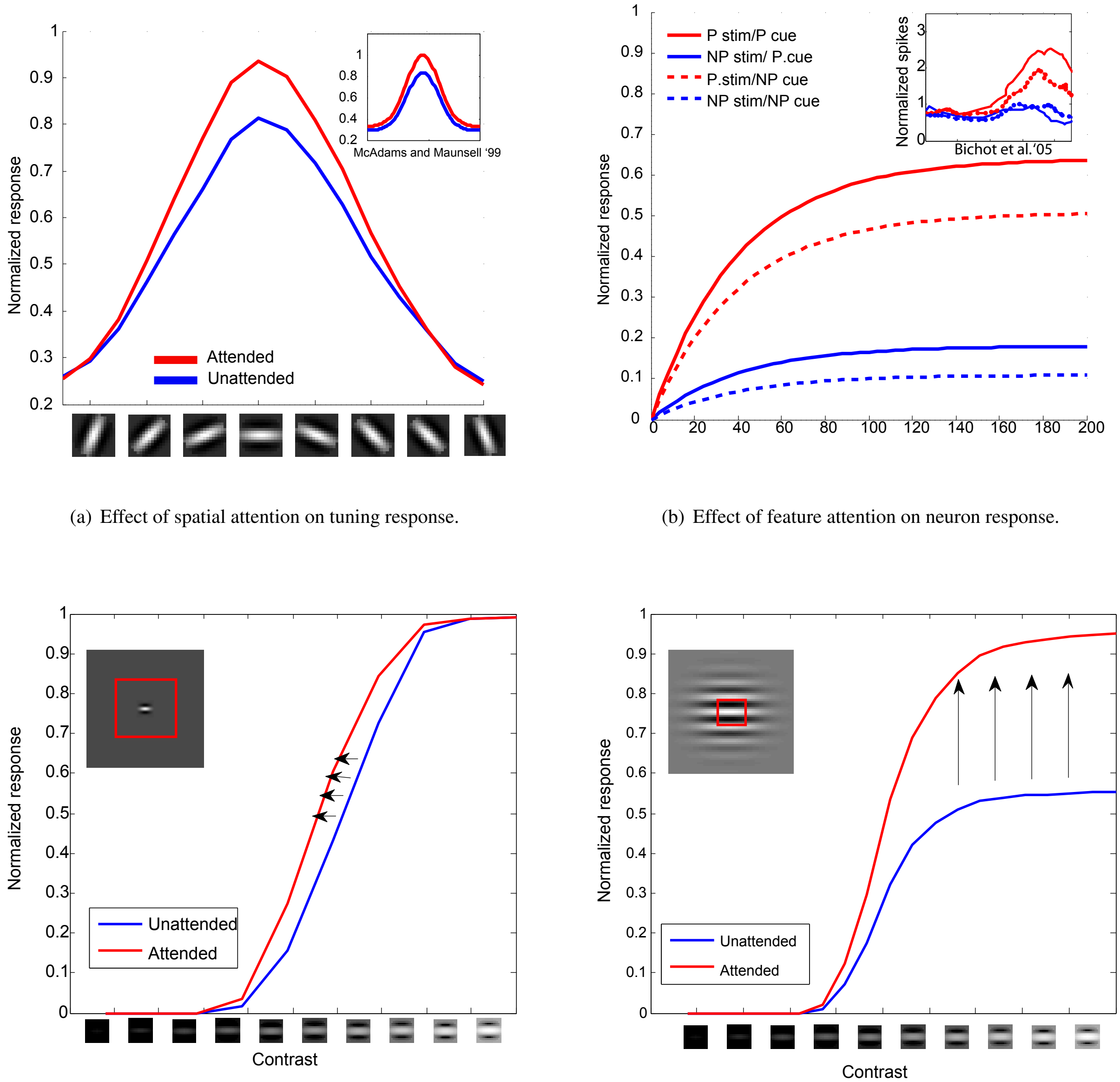
The influence of spatial attention on the contrast response of V4 neurons have been studied extensively. However, prior work have shown two major effects: In [Martınez-Trujillo and Treue, 2002, Reynolds et al., 2000], it was shown that attention serves to enhances the contrast gain of neurons. At the same time, in [McAdams and Maunsell, 1999, Treue and Trujillo, 1999], it was shown that attention creates a multiplicative gain in the contrast response of neurons. Reynolds and Heeger reconciled these differences through a normalization model of attention [Reynolds J. H., 2009]. In this model, the response of a neuron at location x and tuned to orientation θ is given by

$$R(x,\theta) \quad = \quad \frac{A(x,\theta)E(x,\theta)}{S(x,\theta)+\sigma} \tag{13}$$

$$\tag{14}$$

Here, $E(x,\theta)$ represents the excitatory component of neuron response. $S(x,\theta)$ represents the suppressive component of the neuron response derived by pooling activity over a larger area. $A(x,\theta)$ represents the attentional modulation that enhances specific orientation and locations based on the search task. They showed that this model produce contrast gain behavior when the area attended to is larger than the stimulus and can produce response gain behavior when the area attended to is smaller than the stimulus. We show that the Bayesian model can – surprisingly – reproduce the same behavior. Consider the response of the model neuron conditional on a stimulus. It is given by

$$P(F_l^i|I) \quad = \quad \frac{P(I|F_l^i)\sum_{F^i,L} P(F_l^i|F^i, L)P(L)P(F^i)}{\sum_{F_l^i}\left\{P(I|F_l^i)\sum_{F^i,L} P(F_l^i|F^i, L)P(L)P(F^i)\right\}} \tag{15}$$



(a) Effect of spatial attention on tuning response.

(b) Effect of feature attention on neuron response.

(c) Contrast gain under attention.

(d) Response gain under attention.

Figure 4: (a) Multiplicative modulation of neuron responses due to spatial attention: The tuning curve undergoes a multiplicative enhancement under attention. (b) Contrast response: modulation of neuron responses due to feature-based attention. The response of the neuron is enhanced (multiplicatively) when the stimulus feature matches the target feature. The absolute value of the response depends on whether the neuron prefers the stimulus or not. (c) The model exhibits contrast gain when the attentional spotlight is larger than the stimulus. This is equivalent to a shift in the contrast tuning curve. (d) Units in the model exhibit response gain modulations when the spotlight of attention is smaller than the stimulus.

Here, the term $P(I|F_l^i)$ represents the excitatory component, $P(L)$, $P(F^i)$ serve as the attentional modulation component and the sum over all F_l^i (used to generate normalized probabilities) serves as the divisive normalization factor. Thus, the model may be seen as a special case of the normalization model of attention (the normalization model has a constant term in the denominator). Notice that normalization in our model emerges directly from the Bayesian formulation. In Fig.4c and Fig.4d we show that the proposed model is consistent with both contrast gain and response gain effects observed in earlier studies.

4 Prior related work

Studies have shown that image-based *bottom-up* cues can capture attention, particularly during free viewing conditions. Locations where stimulus differs significantly from rest of the image is said to 'pop-out'. In [Itti et al., 1998], center-surround difference across color, intensity and orientation dimensions is used as measure of saliency. In [Gao and Vasconcelos, 2007], self information of the stimuli $(-\log(P(I)))$ is used as measure of distinctiveness [Zhang et al., 2008]. In [Rosenholtz, 1985], the normalized deviation from mean response is used instead. Spectral methods for computing bottom-up saliency have also been proposed [Hou and Zhang, 2007]. These models, however, cannot account for the task-dependency of eye movements [Yarbus, 1967].

A seminal proposal to explain how top-down visual search may operate is the *Guided Search* model proposed by Wolfe [Wolfe, 2007] according to which the various feature maps are weighted according to their relevance for the task at hand to compute a solitary saliency map. Building on Wolfe's model, several approaches have been suggested [Navalpakkam and Itti, 2006, Gao and Vasconcelos, 2005, Zhang et al., 2008]. However, these models ignore the role of spatial attention. In situations where the location of the target is explicitly cued, the role of spatial attention cannot be ignored. In [Desimone, 1998], it was shown that activity of V4 neurons are reduced when multiple stimuli are present within its receptive field. However, when a specific location is cued and subsequently attended, the neurons at the attended locations are selectively enhanced. The neuron responds as if there is a single stimulus within the receptive field. In [Rao, 2005], a Bayesian model of spatial attention is proposed that reproduces this effect. Our work can be viewed as an extension of this approach. In addition to direct cueing, spatial cues may also be derived indirectly, by context, in natural scenes. Spatial relations between objects and their locations within a scene have been shown to play a significant role in visual search and object recognition [Biederman et al., 1982]. In [Oliva et al., 2003], Oliva, Torralba and colleagues showed that a combination of spatial context and bottom-up attention could predict a large fraction of human eye movements during real-world visual search tasks in complex natural images. With the exception of Ehinger et al. [2009], computational models have not considered the interaction between spatial, bottom-up and top-down attentional effects. Table 3 provides a succinct comparison of our approach with existing work in literature.

5 Conclusion

The past four decades of behavioral research in attention has generated a large body of experimental data. A theoretical framework, however, that explains the computational role of attention *and* predicts its effects is still missing. Our theory is based on the hypothesis that attention is a natural mechanism to fix the sensitivity to clutter of hierarchical, transformation-tolerant recognition architectures such as the ventral stream. We conjecture that for a very general class of hierarchical networks, tolerance to image transformations (without a large number of specific training examples) *implies* sensitivity to clutter. This limitation of the feed forward architecture in the presence of clutter suggests a key role for cortical feedback and top-down attention. The hypothesis assumes that attention suppresses the effect of clutter by focusing processing in regions of the image that do not contains objects to be recognized. The hypothesis is novel (see [Serre et al., 2005b]). It suggests an implementation – a computational model based on Bayesian inference. We describe the properties of this extended model and its parallel to biology. The overall model, developed to simulate feature and spatial attention, also predicts effects such such as pop-out, multiplicative modulation and change in contrast response that are consistent with physiological studies.

References

Y. Amit and M. Mascaro. An integrated network for invariant visual detection and recognition. *Vision Research*, 43(19): 2073–2088, 2003.

N.P. Bichot, A.F. Rossi, and R. Desimone. Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4. *Science*, 308(5721):529–534, 2005.

I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147, 1987.

I. Biederman, RJ Mezzanotte, and JC Rabinowitz. Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14(2):143, 1982.

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

D. E. Broadbent. *Perception and communication*. 1958.

N. Bruce and J. Tsotsos. Saliency based on information maximization. *Advances in neural information processing systems*, 18:155, 2006.

T.J. Buschman and E.K. Miller. Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science*, 315(5820):1860, 2007.

C. Cadieu, M. Kouh, A. Pasupathy, C.E. Connor, M. Riesenhuber, and T. Poggio. A model of V4 shape selectivity and invariance. *Journal of Neurophysiology*, 98(3):1733, 2007.

M. Carandini, D.J. Heeger, and J.A. Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17(21):8621–8644, 1997.

C.L. Colby and M.E. Goldberg. Space and attention in parietal cortex. *Annual review of Neuroscience*, 22(1):319–349, 1999.

J.G. Daugman. Two-dimensional spectral analysis of cortical receptive field profile. *Vision Research*, 20:847–856, 1980.

P. Dayan and R. Zemel. Statistical models and sensory attention. *Proceedings of the International Conference on Artificial Neural Networks*, page 2, 1999.

G. Deco and E.T. Rolls. A Neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, 44(6):621–642, 2004.

R. Desimone. Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society*, 1998.

R. Desimone and SJ Schein. Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *Journal of Neurophysiology*, 57(3):835–868, 1987.

J. Duncan. Target and nontarget grouping in visual search [comment]. *Percept. Psychophys.*, 57(1):117–20, January 1995. Comment on: Percept Psychophys 1992 Jan;51(1):79-85.

K. Ehinger, B. Hidalgo-Sotelo, A. Torralba, and A. Oliva. odeling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, 2009.

B. Epshtein, I. Lifshitz, and S. Ullman. Image interpretation by a single bottom-up top-down cycle. *Proceedings of the National Academy of Sciences*, 105(38):14298, 2008.

M. Fabre-Thorpe, A. Delorme, C. Marlot, and S. Thorpe. A Limit to the Speed of Processing in Ultra-Rapid Visual Categorization of Novel Natural Scenes. *Journal of Cognitive Neuroscience*, 13(2), 2001.

D.J. Felleman and D.C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, 1: 1–47, 1991.

F. Fleuret. Fast binary feature selection with conditional mutual information. *The Journal of Machine Learning Research*, 5:1531–1555, 2004.

K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cyb.*, 36:193–202, 1980.

K. Fukushima. A neural network model for selective attention in visual pattern recognition. *Biological Cybernetics*, 55(1): 5–15, 1986.

JL Gallant, CE Connor, S. Rakshit, JW Lewis, and DC Van Essen. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, 76(4):2718–2739, 1996.

D. Gao and N. Vasconcelos. Integrated learning of saliency, complex features, and object detectors from cluttered scenes. In *Computer Vision and Pattern Recognition*, 2005.

D. Gao and N. Vasconcelos. Bottom-up saliency is a discriminant process. In *International Conference on Computer Vision*, 2007.

J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *Advances in neural information processing systems*, 19: 545, 2007.

J. Hegde and D. J. Felleman. How selective are V1 cells for pop-out stimuli? *J Neurosci*, 23(31), 2003.

J. Hegde and D.C. Van Essen. Selectivity for complex shapes in primate visual area V2. *Journal of Neuroscience*, 20(5): 61–61, 2000.

S. Hochstein and M. Ahissar. View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36: 791–804, 2002.

X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition*, 2007.

D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *J Physiol*, 148:574–91, 1959.

D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *J. Phys.*, 195:215–243, 1968.

C.P. Hung, G. Kreiman, T. Poggio, and J.J. DiCarlo. Fast Readout of Object Identity from Macaque Inferior Temporal Cortex. *Science*, 310(5749):863–866, 2005.

M. Ito and H. Komatsu. Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *Journal of Neuroscience*, 24(13):3313–3324, 2004.

L. Itti and C. Koch. Computational modelling of visual attention. *Nat Rev Neurosci*, 2(3):194–203., 2001.

L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *PAMI*, 20(11), 1998.

L. Itti, G. Rees, and J.K. Tsotsos. Neurobiology of attention. 2005.

D.C. Knill and W. Richards. *Perception as Bayesian Inference*. Cambridge: Cambridge University Press, 1996.

C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4):219–27, 1985.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proc. of the IEEE*, 86(11):2278–2324, November 1998.

T. S. Lee and D. Mumford. Hierarchical bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 2003.

J.C. Martınez-Trujillo and S. Treue. Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron*, 35(2):365–370, 2002.

C.J. McAdams and J.H.R. Maunsell. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience*, 19(1):431–441, 1999.

B. W. Mel. SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Comp.*, 9:777–804, 1997.

T. Moore and K. M. Armstrong. Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, 421 (6921):370–3, 2003.

V. Navalpakkam and L. Itti. An integrated model of top-down and bottom-up attention for optimizing detection speed. In *Computer Vision and Pattern Recognition*, 2006.

A. Oliva, A. Torralba, M. S. Castelhano, and J. M. Henderson. Top-down control of visual attention in object detection. In *International Conference on Image Processing*, 2003.

A. Pasupathy and C.E. Connor. Shape representation in area V4: position-specific tuning for boundary conformation. *Journal of Neurophysiology*, 86(5):2505–2519, 2001.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, 1988.

DI Perrett and MW Oram. Neurophysiology of shape processing: Understanding shape: perspectives from natural and machine vision. *Image and Vision Computing*, 11(6):317–333, 1993.

M.I. Posner and Y. Cohen. Components of visual orienting. *Attention and performance X*, pages 531–556, 1984.

R.P.N. Rao. Bayesian inference and attentional modulation in the visual cortex. *Neuroreport*, 16(16):1843–1848, 2005.

R.P.N. Rao, B.A. Olshausen, and M.S. Lewicki. *Probabilistic models of the brain: Perception and neural function*. The MIT Press, 2002.

J.H. Reynolds, L. Chelazzi, and R. Desimone. Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4. *Journal of Neuroscience*, 19(5):1736, 1999.

J.H. Reynolds, T. Pasternak, and R. Desimone. Attention increases sensitivity of V4 neurons. *Neuron*, 26(3):703–714, 2000.

Heeger D. J. Reynolds J. H. The normalization model of attention. *Neuron Review*, 61:168–184, 2009.

M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nat. Neurosci.*, 2:1019–1025, 1999.

R. Rosenholtz. A simple saliency model predicts a number of motion poput phenomena. *Human Neurobiology*, 39(19): 3157–3163, 1985.

R. Rosenholtz and J. Mansfield. Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 761–770. ACM New York, NY, USA, 2005.

H. Schneiderman and T. Kanade. A statistical method for 3D object detection applied to faces and cars. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 746–751, 2000.

T. Serre, M. Riesenhuber, J. Louie, and T. Poggio. On the role of object-specific features for real world object recognition. *Biologically Motivated Computer Vision, Second International Workshop (BMCV 2002)*, pages 387–397, 2002.

T. Serre, M. Kouh., C. Cadieu, U. Knoblich, G. Kreiman, and T. Poggio. A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. *MIT AI Memo 2005-036 / CBCL Memo 259*, 2005a. URL `ftp://publications.ai.mit.edu/ai-publications/2005/AIM-2005-036.pdf`.

T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman, and T. Poggio. A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex, CBCL MIT paper, November 2005, 2005b.

T. Serre, Wolf L., S. Bileschi, M. Reisenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007a.

T. Serre, A. Oliva, and T. Poggio. A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, 104(15):6424, 2007b.

K. Tanaka. Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19(1):109–139, 1996.

K. Tanaka, H. Saito, Y. Fukada, and M. Moriya. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of neurophysiology*, 66(1):170–189, 1991.

S.J. Thorpe. Ultra-rapid scene categorisation with a wave of spikes. *Biologically Motivated Computer Vision, Second International Workshop (BMCV 2002)*, pages 1–15, 2002.

A. Torralba. Modeling global scene factors in attention. *Journal of Optical Society of America*, 20(7):1407–1418, 2003.

A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

S. Treue and J.C.M. Trujillo. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399:575–579, 1999.

J.K. Tsotsos. Limited capacity of any realizable perceptual system is a sufficient reason for attentive behavior. *Consciousness and cognition*, 6(2-3):429–436, 1997.

S. Ullman, M. Vidal-Naquet, and E. Sali. Visual features of intermdediate complexity and their use in classification. *Nat. Neurosci.*, 5(7):682–687, 2002.

L.G. Ungerleider and J.V. Haxby. 'What'and'where'in the human brain. *Current Opinion in Neurobiology*, 4(2):157–165, 1994.

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.

P. Viola and M. Jones. Robust real-time face detection. In *ICCV*, volume 20(11), pages 1254–1259, 2001.

G. Wallis and E. T. Rolls. A model of invariant object recognition in the visual system. *Prog. Neurobiol.*, 51:167–194, 1997.

D.B Walther and C. Koch. *Computational Neuroscience: Theoretical insights into brain function, Progress in Brain Research*, chapter Attention in Hierarchical Models of Object Recognition. 2007.

H. Wersing and E. Koerner. Learning optimized features for hierarchical models of invariant recognition. *Neural Comp.*, 15(7):1559–1588, 2003.

Jeremy M. Wolfe. Guided search 4.0: Current progress with a model of visual search. *Integrated Models of Cognitive System*, pages 99–119, 2007.

AL Yarbus. *Eye movements and vision*. Plenum press, 1967.

A.J. Yu and P. Dayan. Inference, attention, and decision in a Bayesian neural architecture. *Advances in Neural Information Processing Systems*, 17:1577–1584, 2005.

L. Zhang, M. H Tong, T. K Marks, H. Shan, and G. W Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of Vision*, 8(7):1–20, 2008.

L. Zhaoping and R.J. Snowden. A theory of a saliency map in primary visual cortex (V1) tested by psychophysics of colour–orientation interference in texture segmentation. *Visual cognition*, 14(4):911–933, 2006.

D. Zoccolan, M. Kouh, T. Poggio, and J.J. DiCarlo. Trade-Off between Object Selectivity and Tolerance in Monkey Inferotemporal Cortex. *Journal of Neuroscience*, 27(45):12292, 2007.

APPENDIX

A Computational Model

We implemented the Bayesian network using Kevin Murphy's Bayesian toolbox available at http://bnt.sourceforge.net. The Bayesian model consists of a location encoding unit (L), object encoding units O, non-retinotopic feature encoding units F^i and *combination* units F^i_l, that encode position-feature combinations. These units receive input I from lower areas in the ventral stream. L models LIP area in the parietal cortex and encodes position and scale independently of features. F^i units model non-retinotopic, spatial and scale invariant cells found in higher layers of the ventral stream. More details about the model units are provided in Table 1.

Model unit	Brain area	Representation/Model
L	LIP/FEF	This variable encodes the location and scale of the target object. It is modeled as a discrete multinomial variable with $\|L\|$ distinct values.
O	PFC	This variable encodes the identity of the object. It is modeled as a discrete multinomial variable that can take $\|O\|$ distinct values.
F^i	IT	Each feature variable F^i encodes the presence of a simple shape feature. Each such unit is modeled as a discrete binary variable that can be either on or off. It is to be noted that presence or absence is indicated in a position/scale invariant manner. In practice $10 \sim\ 100$ such features are used.
F_l^i	V4	This variable can be thought of as a feature map that encodes the joint occurrence of the feature (F^i) at location $L = l$. It is modeled as a discrete multinomial variable with $\|L\|$+1 distinct values $(0, 1 \cdots L)$. Values $(1 \cdots L)$ correspond to valid locations. Value $F_l^i = 0$ indicates that the feature is completely absent from the input.
I	V2	This is the feed-forward evidence obtained from the lower areas of ventral stream model.

Table 1: Bayesian model units and tentative mapping to brain areas.

<table>
<tr><th>Conditional Probability</th><th>Modeling</th></tr>
<tr><td>$P(L)$</td><td>Each scene or view-point places constraints on the location and sizes of objects that can be encountered in the image. Such constraints can be specified explicitly (e.g. during spatial attention) or learned using a set of training examples [Torralba, 2003].</td></tr>
<tr><td>$P(F^i|O)$</td><td>The probability of each feature being present or absent given the object and is directly learned from the training data.</td></tr>
<tr><td>$P(F_l^i|F^i, L)$</td><td>When the feature F^i is present and location $L = l^*$ is active, the F_l^i units that are nearby unit $L = l*$ are most likely to be activated. When the feature F^i is absent, only the $F_l^i = 0$ location in the feature map is activated. This conditional probability can be captured succinctly by the following table
<table>
<tr><th></th><th>$F^i = 1, L = l$</th><th>$F^i = 0, L = l$</th></tr>
<tr><td>$F_l^i = 0$</td><td>$P(F_l^i|F^i, L) = \delta_1$</td><td>$P(F_l^i|F^i, L) = 1 - \delta_2$</td></tr>
<tr><td>$F_l^i \neq 0$</td><td>$P(F_l^i|F^i, L)$ $\sim$ Gaussian centered around $L = l$</td><td>$P(F_l^i|F^i, L) = \delta_2$</td></tr>
</table>
δ_1 and δ_2 are small values. They are chosen to ensure that $\sum P(F_l^i|F^i, L) = 1$.</td></tr>
<tr><td>$P(I|F_l^i)$</td><td>For each location within the feature map, $P(I|F_l^i)$ provides the likelihood that F_l^i is active. In the model, this bottom-up evidence or likelihood is set proportional to the activations of the shape-based units (see [Serre et al., 2007a]).</td></tr>
</table>

Table 2: Conditional probabilities.

	Proposed	[Bruce and Tsotsos, 2006]	[Zhang et al., 2008]	[Deco and Rolls, 2004]	[Ehinger et al., 2009]	[Fukushima, 1986]	[Hou and Zhang, 2007]	[Harel et al., 2007]	[Itti and Koch, 2001]	[Rao, 2005]	[Torralba, 2003]	[Walther and Koch, 2007]	[Wolfe, 2007]	[Yu and Dayan, 2005]
Biologically plausible	✓	✓	✓	✓	×	✓	×	×	✓	✓	✓	✓	✓	✓
Real world stimuli	✓	✓	✓	×	✓	×	✓	✓	✓	×	✓	✓	×	×
Pop-out	✓	✓	✓	×	✓	×	✓	✓	✓	×	✓	✓	×	×
Feature-based attention	✓	×	×	✓	✓	✓	×	×	×	×	✓	×	✓	✓
Spatial attention	✓	×	×	×	✓	×	×	×	×	✓	✓	✓	×	✓
Parallel vs. serial search	✓	×	×	×	×	×	×	×	×	×	×	×	✓	×
Explicitly models ventral/parietal	✓	×	×	✓	×	×	×	×	×	✓	×	×	×	×

Table 3: The matrix compares the features of prior computational models