UNIVERSITY OF BIRMINGHAM

University of Birmingham

Modelling Visual Search with the Selective Attention for Identification Model (VS-SAIM): A Novel Explanation for Visual Search Asymmetries

Heinke, Dietmar; Backhaus, A

Link to publication on Research at Birmingham portal

Modelling Visual Search with the Selective Attention for Identification Model (VS-SAIM): A Novel Explanation for Visual Search Asymmetries

Dietmar Heinke · Andreas Backhaus

Abstract In earlier work, we developed the Selective Attention for Identification Model (SAIM [16]). SAIM models the human ability to perform translation-invariant object identification in multiple object scenes. SAIM suggests that central for this ability is an interaction between parallel competitive processes in a selection stage and a object identification stage. In this paper, we applied the model to visual search experiments involving simple lines and letters. We presented successful simulation results for asymmetric and symmetric searches and for the influence of background line orientations. Search asymmetry refers to changes in search performance when the roles of target item and non-target item (distractor) are swapped. In line with other models of visual search, the results suggest that a large part of the empirical evidence can be explained by competitive processes in the brain, which are modulated by the similarity between target and distractor. The simulations also suggest that another important factor is the feature properties of distractors. Finally, the simulations indicate that search asymmetries can be the outcome of interactions between top-down (knowledge about search items) and bottom-up (feature of search items) processing. This interaction in VS-SAIM is dominated by a novel mechanism, the knowledge-based on-centre-off-surround receptive field. This receptive field is reminiscent of the classical receptive fields but the exact shape is modulated by both, top-down and bottom-up processes. The paper discusses supporting evidence for the existence of this novel concept.

Keywords Visual attention · Visual search · Computational modelling · Search asymmetry

D. Heinke (✉)
School of Psychology, University of Birmingham, Birmingham B15 2TT, UK
e-mail: d.g.heinke@bham.ac.uk

A. Backhaus
Fraunhofer IFF, Biosystems Engineering, 39106 Magdeburg, Germany
e-mail: Andreas.Backhaus@iff.fraunhofer.de

Introduction

The visual search task is a commonly used experimental procedure to study human processing of multiple object scenes. In a standard visual search task, participants are asked to determine whether a pre-defined target item among non-targets (distractors) is present or absent. During the course of the experiments the number of distractors (display size) is varied. Typically, the time it takes participants to make this decision (reaction time) is measured as a function of the display size (search function). The slope of the search function is interpreted as indicator for the search efficiency for particular target-distractor pairings. For instance, search for a diagonal line among vertical lines is highly efficient with a slope close to 0ms/item whereas search for a 'T' among 'L's is inefficient with a slope of around 25 ms/item. Over 40 years or so, visual search tasks have produced a plethora of experimental evidence (see [31, 41] for reviews). There have been numerous attempts to develop qualitative theories of visual search, e.g. most prominently the Feature Integration Theory (FIT) by Treisman et al. [37] or the Attentional Engagement Theory (AET [12]). This article presents a connectionist model of visual search. This model is an extension of the Selective Attention for Identification Model (SAIM; [16, 19, 20]) adopted to simulate visual search and therefor is termed VS-SAIM.

SAIM was developed in a connectionist framework and aims to explain human behaviour in terms of the underlying neurophysiological processes in the brain. However, SAIM avoids the full complexity of neurophysiological processes, e.g. the dynamics of different neurotransmitters and employs rate-coded neuron models. On the other hand, this simplification is balanced with SAIM's objective to unify a broad range of behavioural data in one model (see [17]; for extensive discussions on the relationship between models of the neural substrate and modelling behavioural data). SAIM's starting point is the human ability to identify objects in multiple object scenes. SAIM suggests that central for this ability is an interaction between parallel competitive processes in a selection stage and a object identification stage. Based on this assumption, SAIM was able to simulate a broad range of experimental evidence usually associated with normal operation of attention and with dysfunctional attention [16]. The simulations of normal attention covered two-object costs on selection, global precedence, spatial cueing both within and between objects, and inhibition of return. The effects of disordered attention included view-centred and object-centred visual neglect. In Heinke et al. [19], SAIM was successfully applied to simulate a few visual search experiments. These studies showed that the search functions in visual search can be an emerged property of the competitive processes in the brain. The slopes of the search functions were influenced by the similarity between distractors and target. However, when we attempted to simulate a broader range of visual search experiments, it became clear that this initial version of VS-SAIM was not able to mimic this additional data. Consequently, we modified some operations within VS-SAIM. Especially, we replaced the original similarity measure, the scalar product, with the Euclidian distance. The present article reports on a first set of results of this extension.

For the first set of results we chose experimental evidence that, on the face of it, is particularly challenging to VS-SAIM's similarity-based approach, the search asymmetry (see [43]; for a review). In search asymmetries search slopes differ when the roles of target item and distractor item are swapped. For instance, it is easier to find a tilted line among vertical lines then vice versa [37]; a diagonal line among vertical lines than the reverse [3]. Other examples are: orange item (easier) versus red item [36], moving item (easier) versus static item [11, 34]. For a similarity-based approach these data are particular challenging, as the target-distractor similarity simply does not change when target and distractor are swaped around. A theoretical account needs to introduce an additional factor to explain these findings.

On a wider note, there is no satisfactory theoretical account for the occurrence of search asymmetry at present.

Initially, Treisman and Gormican [37] suggested that search asymmetries are indicative for the existence of feature maps assuming that detection of the presence of a feature is better than the detection of its absence [37]. However, subsequent evidence has not supported their theory. For instance, their assumption does not fit with the findings on diagonal line versus vertical line [3], as there are well-known feature maps for diagonal lines in the brain. Moreover, recent evidence showed that search for an "inverted elephant" among upright elephants is more efficient than the other way around [43] pointing towards the involvement of object knowledge in search asymmetries. The current paper aims to develop a first coherent account of search asymmetries. It focuses on the search asymmetries with line orientations.

The reminder of the paper is organized as follows. After introducing VS-SAIM in detail, we discuss how VS-SAIM relates to other important models and theories of visual search. Then we illustrate how the search process in VS-SAIM plays out in detail (Study 1). Study 2 demonstrates that VS-SAIM mimic the experimental findings of asymmetries of line orientation for both diagonal versus vertical line and titled versus vertical line. We also present detailed explanations for this success. The explanation also suggests that VS-SAIM's search efficiency depends not only on target-distractor similarity but also on the orientations of the distractors. Study 3 confirms this point through simulating findings by Foster and Westland [14]. To complete the picture, Study 4 shows that VS-SAIM can also simulate a visual search task with symmetric results [13]. The general discussion discusses the theoretical implications and present supporting evidence for VS-SAIM's explanation of search asymmetries.

VS-SAIM

Overview

This short description gives an overview of the interactions between VS-SAIM's modules. Afterwards the model will be explained in more details. The mathematics behind VS-SAIM are documented in the "Appendix". Figure 1 illustrates VS-SAIM's architecture. Overall VS-SAIM implements a translation-invariant object identification in multiple object scenes. VS-SAIM's first stage, the early visual processing stage (EVPS), extracts simple features, e.g. orientations, from the visual field. In the bottom-up path, the contents network maps a spatial selection of these feature through to the 'Focus of Attention' (FOA). This mapping is translation-invariant, meaning that the contents of any location in the input image can be mapped through to the FOA. The mapping is controlled by the selection
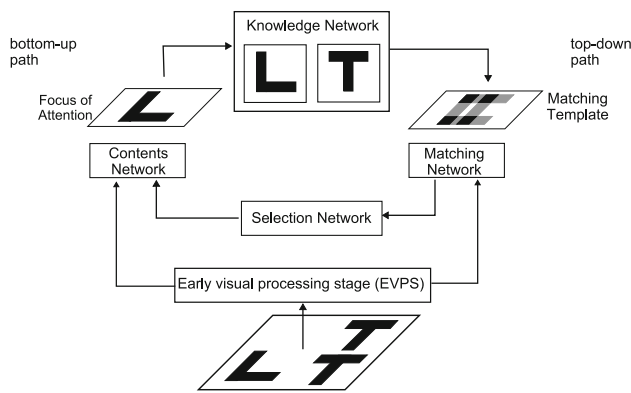
Fig. 1 Architecture of VS-SAIM (see text for details)

network. The selection network, on one hand, chooses the location from which the contents network takes its input and, on the other hand, ensures that the mapping does not distort the original input. These functions are implemented through competitive and co-operative interactions between units in the selection network. VS-SAIM also contains object knowledge stored in the knowledge network with template units. The knowledge network identifies the content of the FOA by matching it with the templates. The template matching utilizes a Euclidian distance as similarity measure. In addition, the knowledge network biases the selection process in VS-SAIM towards 'known/relevant' over 'unknown/irrelevant' objects. This is done via the top-down pathway from the knowledge network to the selection network mediated via the matching network. The role of the matching network is to compare the template information (matching template) with the output from the EVPS. Consistent with the bottom-up path, this comparison utilizes a Euclidian distance and is translation-invariant.

It is important to note that, as in the previous versions of SAIM, VS-SAIM was designed with the help of the principle of minimization of energy function. This idea was first introduced into connectionism by Hopfield and Tank [21] and implements a soft-constraint satisfaction. The design principle follows the following steps: First the problem is formulated as constraint satisfaction problem which defines the constraints a solution has to fulfil. These solutions are translated into activation patterns in a connectionist network. Then an energy function is designed in which these activation patterns are minimal energy values. Finally, to find these energy minima starting from pre-

defined activation pattern, a gradient decent procedure is applied to the energy function. The gradient decent procedure results in nonlinear differential equations which, in turn, define a biologically plausible network topology, including the weights between connections. The advantage of this approach is that the energy minima defines a stable state or attractor state for the nonlinear differential equations. This property makes this approach appealing to the design of connectionist models. However, while designing the model in such a way, we found that some of the terms in the equations did not lead to a successful object selection and identification. Subsequently, we relaxed the minimization approach. The details of this relaxation are discussed in the "Appendix". Nevertheless, the topology of the model is still directly motivated by the energy minimization approach.

Early Visual Processing Stage (EVPS)

VS-SAIM's early visual processing stage consists of Gabor-filters tuned to four orientations, 0°, 90°, 45° and 135°. Gabor-filters have been widely used to model receptive fields of orientation-selective simple cell in the primary visual cortex V1 (e.g. [9]). Details about the implementation of the filters and the parameters can be found in the "Appendix".

The output of the EVPS consists of five feature maps, intensity feature map and four orientation feature maps. Figure 2 shows an example of the feature maps for a search display with a 'L' among 'T's. In order to take into account random noise in the brain, a quasi-stochastic behaviour is added to each feature map (see "Appendix" for details). Finally, before the feature maps are fed into the remainder of the model, the activation of the maps is weighted. Horizontal and vertical orientations are weighted higher than diagonal features, as suggested by physiological evidence (e.g. [7]). As we will show later, these different weightings are important for simulating the line search asymmetries.

Contents Network

The contents network aims to enable a translation-invariant mapping from the output of the EVPS to the smaller focus of attention (FOA). The core mechanism of this mapping is

Fig. 2 Feature maps in the early visual processing stage (EVPS). The display on the *left* shows an example of an input image. The four images on the *right* show the resulting feature maps
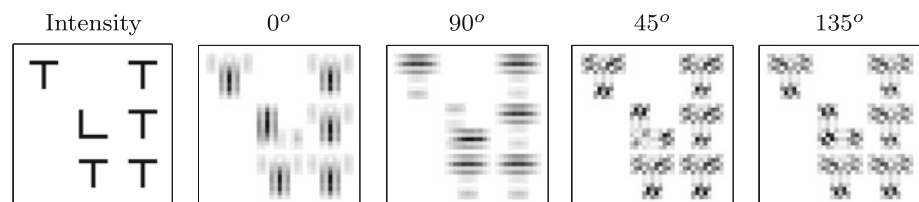
a gating mechanism implemented with sigma-pi units [26] which are controlled by layers in the selection network (see Fig. 3 for an illustration and "Appendix" for details on the operations of the contents network). Sigma-pi units produce an output activation by combining input activations with two operations, multiplications and additions. In Fig. 3, these two operations are illustrated separately. The circles illustrate the multiplications and the squares depict the additions. The multiplication operation combines the output activation from a layer in the selection network with the output of the feature maps at spatially corresponding locations. The addition produces the output activation of a sigma-pi unit by summing up the result of the multiplication. The output of each sigma-pi unit represents a pixel in the FOA. Hence, each layer in the selection network controls the activation of one pixel in the FOA. Figure 3 illustrates this gating mechanism for three locations in the visual field and three pixels in the FOA. For instance, in the central layer the unit corresponding to a pixel of the vertical T-stroke is switched on (filled circle). This activation gates this pixel through to the FOA, as indicated by the open circle in the contents network.

It is important to note that the content network can implement an arbitrary mapping which depends on the activation pattern in the selection network. For instance, if the unit in the centre of each layer in selection network had a high activation and all other units in the selection
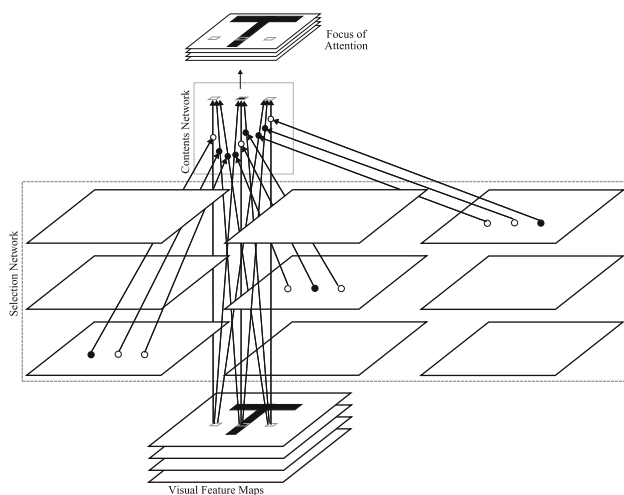

Fig. 3 Mapping in the contents network. The mapping is illustrated for three input locations and three FOA-locations. The contents network consists of sigma-pi units which combine two operations: multiplication (*filled circles*) and addition (*squares*). The multiplications combine the output of the selection network with the output of the feature maps at spatially corresponding locations. The addition produces the output activation of the sigma-pi units by summing up the results of the multiplications. The output of each sigma-pi unit represents a pixel in the FOA. The *filled circles* in the selection network indicate activated units. These activated units map the content of the corresponding location of the feature maps to the FOA via the sigma-pi units (see text for more details)

network were set to zero, the content of the centre of the input image would be represented in all FOA pixels. Hence, translation-invariant mapping is a special case that is achieved, if two constraints on the activation pattern in the selection network are fulfilled: First, only one unit in the each layer should be activated. With this restriction only the content of one image location is routed into the FOA, because the multiplication allows only one location to be passed into the FOA. Second, only units across the selection network that map neighbouring locations in input image onto neighbouring locations in FOA are allowed to be active. The constraint ensures that the FOA forms veridical representation of the selected object in the input image and is implemented through a "diagonal" activation pattern in the selection network. The necessity of "diagonality" arises from the following rational: If one unit in one layer is activated, the layer that controls the adjacent FOA-pixel has to activated the unit adjacent to the first unit. In this way, two locations adjacent in the input image are mapped into adjacent pixels in the FOA. The connections in the selection network implement the corresponding constraint satisfaction process.

Selection Network

The selection network aims to select a stimulus in the input image by producing an appropriate activation pattern. Since the selection network controls the mapping in the contents network this activation pattern has to ensure a veridical representation of this stimulus in the FOA. The selection network is structured in layers whereby every layer controls the routing for one of the FOA pixels in the contents network (see Fig. 4 for an illustration). To ensure a veridical stimulus representation in the FOA, the activation pattern in these layer has to fulfil two constraints (see also section on contents network): (1) Only one unit in each layer is allowed to become active. (2) Units in layers controlling adjacent FOA-units has to become activate only if they are adjacent with respect to image locations. This constraint implements the neighbourhood preserving mapping in the content network. The "Appendix" documents the mathematical implementation of these constraints within the framework of the energy minimization approach. The resulting connections are illustrated in Fig. 4. Each layer has an overall inhibitory connections between units implemented the first constraint (competitive process). Units between layers are connected via excitatory connections along the diagonals implementing the second constraint (co-operative process). In addition, the gradient procedure applied to the overall energy function introduces the input from the matching network. This input results from terms in the energy function which ensure that VS-SAIM's behaviour including the behaviour of the
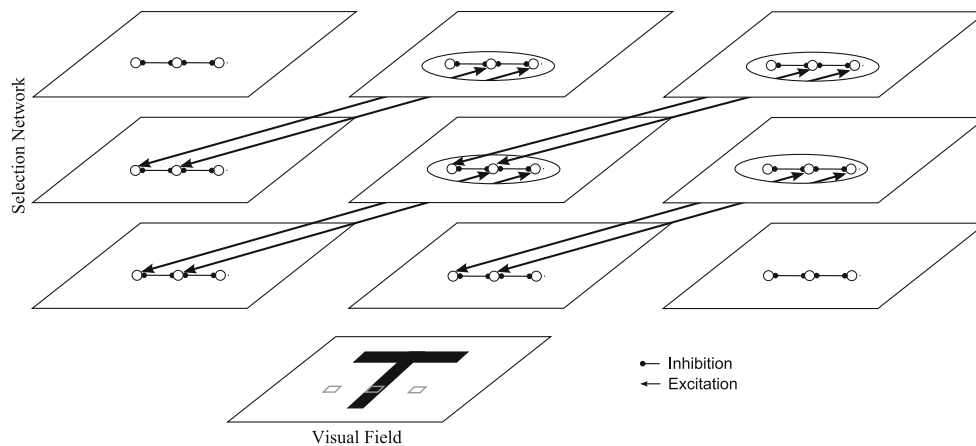
Fig. 4 The structure of the selection network. The selection network consists of several layers. There are inhibitory connections within each layer and excitatory connections between the layers. As explained in the text, connections ensure that the selection network forms a veridical representation of objects in the FOA. Note that for illustration purposes adjacent pixels are depicted further apart

selection network is constraint by the input image and the template-knowledge implemented in the knowledge network.

Knowledge Network

The knowledge network implements the object identification in VS-SAIM. A unit in the knowledge network represents an object by being associated with a template of this object. The template is a copy of the object, as it would appear in the FOA. In order to determine which object is represented in the FOA, the template units compare their template with the FOA activation in a matching process. The similarity measure in this template matching is based on the Euclidian distance commonly used in connectionist networks. In order to determine which of the template units represents the best matching template, the units interact in a competitive process similar to the one implemented in the selection network (see "Appendix" for mathematical details). The output activation of the template units represent the output of VS-SAIM. A high output activation indicates that VS-SAIM has successfully identified the content of the FOA.

In VS-SAIM the knowledge network introduces not only a identification stage as an output stage, but also adds a general knowledge-based constraint on VS-SAIM's behaviour. In order to fully integrate this additional constraint the knowledge network also influences the behaviour of the selection network via the matching network. This top-down pathway is a direct outcome of the energy minimization procedure employed in VS-SAIM (see "Appendix" for details). In general, this knowledge biases the VS-SAIM's behaviour towards selecting locations in the input image that matches best the templates. Moreover, if the initial activation in the knowledge network is biased

towards one template unit, VS-SAIM's overall behaviour is biased towards selecting the item associated with this template. In this paper, we use this property to implement the fact that the visual search experiment requires the search for a set target. Hence, we will bias VS-SAIM towards the selection of the target item. If the target is not present, VS-SAIM is expected to overcome the initial bias and select a distractor item.

Matching Network

The function of the matching network is to mediate the feedback from the knowledge network to the selection network. It implements this function in two stages (see Fig. 5). First a weighted representation of the templates is formed, termed *matching template*. The weighting is determined by the output activation of the template units. As a consequence of the energy minimization approach (see "Appendix" for details), the weighted representation slowly builds up starting from an unbiased overlay of all templates. As a simulation progresses the template reflects more and more a biased overlay between the templates determined by the output activation of the template units. For instance, if the L template units has an output activation of 0.4 and the T template unit an output of 0.6, the resulting matching template is made up of 40% of an L and 60% of an T.

In the second stage, the matching template is compared with the feature maps from the EVPS and the result of this comparison feeds into the selection network. Again, like in the knowledge network the matching is based on the Euclidian distance. The usage of this distance is a direct outcome of the energy function minimization approach. It reflects the necessity that the matching in the bottom-up pathway needs to be consistent with the matching in the
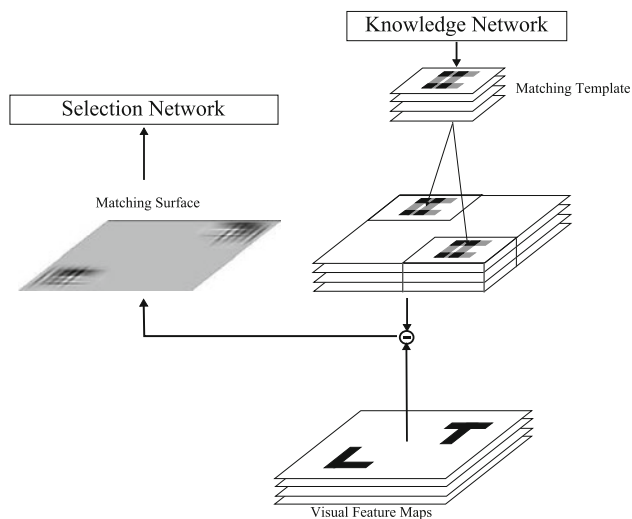
Fig. 5 Structure of the matching network. The matching network implements the top-down modulation of the selection network from the knowledge network. The top-down modulation is implemented in two stages: First the matching template is formed (*top right*). Second the matching template is compared with the feature maps ("minus circles"). This comparison is translation-invariant. This is illustrated by depicting copies of matching templates in the block of layers on the right. Also note that the output of the matching process, termed matching surface, plays an important role in VS-SAIM's capability to simulate the experimental findings

top-down pathway to ensure an overall consistent behaviour. Note that the matching network also mirrors the translation-invariant mapping of the bottom-up pathway by implementing the comparison between matching template and feature maps in a location-by-location fashion. Figure 5 illustrates this implementation for 'L' and 'T' as templates and an 'L' and 'T' in the input image. Figure 5 also shows the result of the matching process. Since the outcome plays an important role in this paper we introduced a special term, the *matching surface*. Bright pixels stand for highly matching locations and dark pixels represent no matching. The matching surface forms the input to the selection network, where the competitive processes activate units at locations with high matching values.

Discussion

This section presented details on how VS-SAIM achieves translation-invariant object identification in a multiple object display. Crucial for achieving this objective are three mechanisms: competitive interactions for selection and identification of items; similarity-based matching in the bottom-up and top-down pathway to direct the selection process and identify the selected item; and an interaction between top-down and bottom-up pathways to ensure consistency between both levels. To implement the search for a target in visual search, the initial activation in the

knowledge network is biased towards one template unit, biasing VS-SAIM's overall behaviour towards selecting the target.

It should be noted that VS-SAIM is part of an ongoing project. Some of the mechanisms presented here have already been validated against experimental evidence other than data from visual search. For instance, the layered structure in the selection network, turned out to be crucial for simulating attentional disorder, such as extinction and object-based neglect [16]. The excitatory connections in the selection network were useful in simulating proximity-based grouping [16]. A first step towards the integration of similarity-based grouping was presented in [18]. Also, SAIM proved robust enough to process natural images [20]. Compared to the version published in 2003, the main extensions here are a different similarity measure (Euclidian distance instead of scalar product) and the introduction of an early visual processing stage.

VS-SAIM falls into a class of models that conceptualize visual attention as mapping details of an input image into a new representation. The most prominent representative of this class is the Selective Tuning (ST)-model by Tsotsos et al. [38]. Similar to VS-SAIM, the ST-model uses competitive processes controlled by bottom-up and top-down pathways to guide the mapping process. Interestingly, in a recent extension of the ST-model Tsotsos et al. [39] stressed the importance of considering interactions between recognition and attention when modelling visual attention. This type of integrative approach is also taken by VS-SAIM and its earlier version, SAIM.

However, for the remainder of this discussion and in keeping with the theme of this paper we will focus on the most prominent theories and models of visual search in experimental psychology. Similar to VS-SAIM, all these models and theories postulate that an interaction between top-down and bottom-up influences plays a role in human performance in visual search. Moreover, all models suggest that at some stage a "featureless" encoding of the search display. For instance, in the Guided-search model [40] this representation is called "saliency map" or "master map". In MORSEL [30] the input to the attentional module represents the contents at locations in search display "featureless". In Deco and Zihl's biased-competition model of visual attention [10] a location map receives inputs from all feature maps in a retinotopic fashion. In VS-SAIM the selection network and its input, the matching surface, are "featureless" maps. However, the Guided-search model and MORSEL suggest that this "featureless" map is static and is no longer modified during the search process. In contrast, Deco and Zihl's model [10] and VS-SAIM postulate that the "featureless" map is dynamic and changes during the selection process. Especially, in VS-SAIM the dynamic "featureless" map, the matching surface, is an
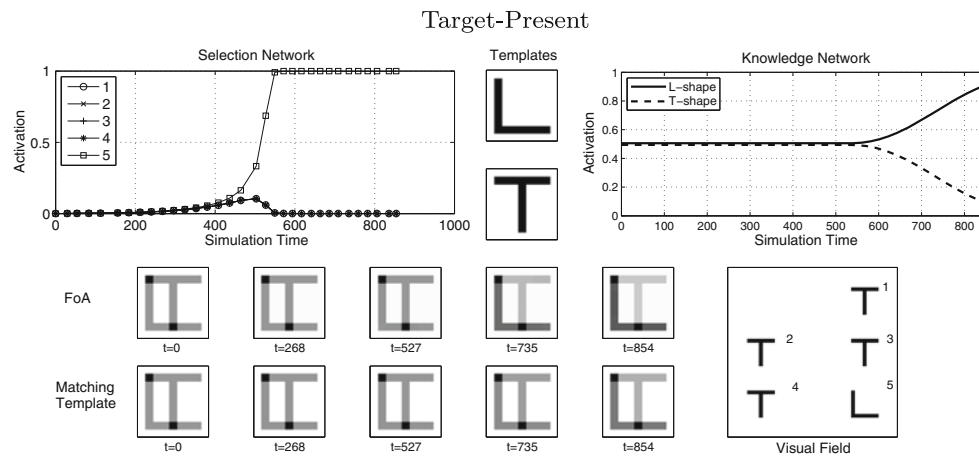
Target-Present

Fig. 6 Time course of activations for a target-present display. The image at the *bottom-right* shows the search display with "L" as target and "T" as distractor. Note that the numbers in the visual field display are not presented. They correspond to the number in the legend of the time course plot of the selection network. This plot shows the time course of the activation of units in the central layer at the centre of the items. A detailed discussion of this figure can be found in the text

integral part of interactions between the selection process and the identification process.[1] Intuitively, this seems to be a more biologically plausible approach to modelling processes during visual search tasks. Finally, VS-SAIM also shares with the seminal Attentional Engagement Theory (AET [12]) the assumption that similarity-based matching plays a crucial role.

Another point to note is that VS-SAIM implements visual search in completely parallel manner. This contrasts with earlier versions of SAIM [16, 19] and also with most other models of visual search. For instance and most prominently the Guide-Search model postulates an entirely serial search process. Even the models with a competitive approach assume that there is some sort of serial rechecking mechanism (see the Search via Recursive Rejection (SERR)-model, [22]; for an example). However, our implementation of VS-SAIM does not imply that visual search is performed entirely in parallel. Instead, the work of VS-SAIM focuses on contributions of competitive processes to visual search which we, nevertheless, consider to play a crucial part in visual search. On the same token, the visual search mechanism proposed in this paper are assumed still to play an important role even when a serial mechanism is added to VS-SAIM in future versions.

Study 1: Basic Behaviour

This study does not primarily aim at simulating experimental results but to illustrate the interworking of VS-SAIM, such as time courses of activations in the

selection network, knowledge network and matching network (see Figs. 6, 7). These time courses are representative of the processes in all simulations in this paper. The simulations used 'L' as target and 'T' as distractor (see [12]; for corresponding experimental results). The search display contained 5 items. To encode the target, the 'L' template unit was initialized with higher activation (0.506) than the distractor unit (0.494). Also, it is important to note that the reaction time (RT) of the model is the simulation time it takes for one template unit to pass a set threshold. Passing the threshold is interpreted as the model having recognized an item. Moreover, VS-SAIM does not make any mistakes. Compared to human performance in visual search this assumption is not realistic. However, often error rates are not statically significant in visual search tasks and human performance is typically influenced in terms of reaction times. Therefore the simulations focus on reaction times as dependent variable.

Figure 6 shows the simulation result with target being present. The simulation was terminated after the knowledge network produced a clear-cut winner (see time plot of the knowledge network). At this point of time, activations in VS-SAIM were dominated by the target item. FOA and matching template show a stronger representation of 'L' than of 'T'. The time plot of the selection network (top left) shows only the time course of the activations in the centre layer at the central locations of the items in the search display. The time plot illustrates that the target item (Item 5) in the visual field won the competition. This successful selection of the target item began in knowledge network where the initial activation of the two template units is biased towards the target item. This bias drove the matching network from a unbiased matching template (both templates are equally weighted) towards a matching
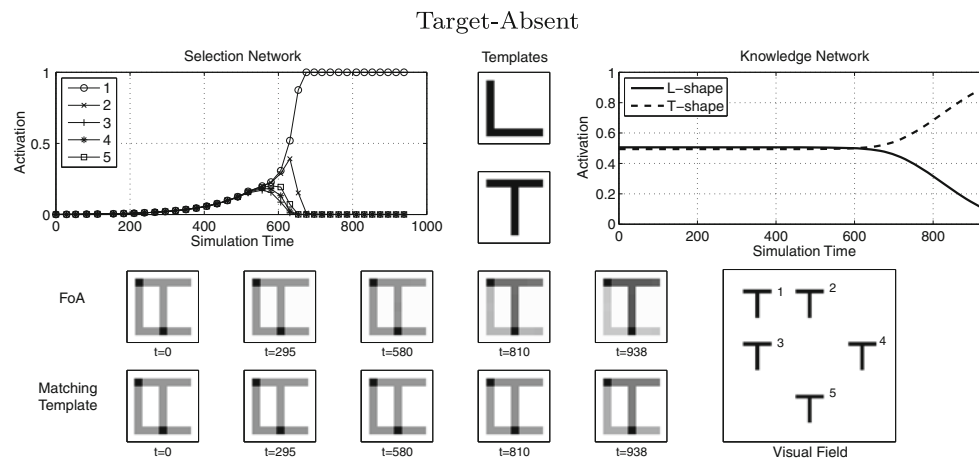
[1] Indeed simulations not included in this paper suggest that the concept of a "dynamic saliency map" can improve our understanding of visual search tasks.

Target-Absent

Fig. 7 Time course of activations for a target-absent display. The image at the bottom-right shows the search display with "T" as distractor and "L" as target if it were present. Note that the numbers in the visual field display are not presented. They correspond to the number in the legend of the time course plot of the selection network. This plot shows the time course of the activation of units in the central layer at the centre of the items. A detailed discussion of this simulation result can be found in the text

template that is biased towards the target template. This, in turn, led to better matching values at the target location in the matching surface. Therefore, the selection network began selecting the target item which resulted in a stronger representation of the target item in the FOA. This improvement reinforced the initial bias in the knowledge network, eventually making the target template unit the winner unit.

When the target is absent (Fig. 7), the initial bias towards the target unit is overcome and the distractor template eventually wins the competition. Analogous to the present trial, the identity of this winner item was eventually reflected in all parts of the model. However, VS-SAIM reaches this state later than in the present trial. Hence, the initial bias in the knowledge network contributes to the delay of reaction times compared to the present trials. Moreover, in the absent condition the matching surface does not produce a clear winner early on, as in the present condition. Instead, the noise added in the EVPS generates a small difference between distractor items which, eventually, allows the selection network to randomly chose an item.

It is interesting to note that the delay in VS-SAIM's reaction time in the target presence condition compared to the absent condition mimics typical experimental findings in visual search tasks [42]. However, these simulation results go beyond the focus of the present paper. The strategy with which participants treat absent trials represents a entirely different issue (see [5]; for a rare example of modelling absent trials). Further simulations will need to explore whether this treatment of absent trials constitutes a valid approach.

Study 2: Search Asymmetry

This paper focuses on two asymmetries found in oriented line searches. First, if a tilted line is searched among vertical lines, search is more efficient than a vertical line among tilted lines [37]. Second, if a diagonal line is searched among vertical lines, search is more efficient than a vertical line among diagonal lines [3].

Method

Stimuli

The input display were grey-value pixel pattern of value range of [0;1]. All items were 9×9 pixels of size and were placed in a fixed 3×3 grid within the input display evenly spaced. The total pixel size of the input display was 43×43. The search items were a vertical line, 30°-line and 45°-line (see Fig. 8 for examples). To generate the tilted lines, a vertical line of seven pixels was rotated using the Matlab function *imrotate* with a bi-linear interpolation method.

Procedure

Displays were generated with set-sizes of 2, 3, 4, 5, 6, 7 and 8 items. Each condition was run 5 times amounting to 70 trails in total. Only templates for the items present in a particular experiment were included in the knowledge network. At the beginning of each simulation run the template unit of the target was biased to a higher activation 0.506 than the distractor unit 0.494.
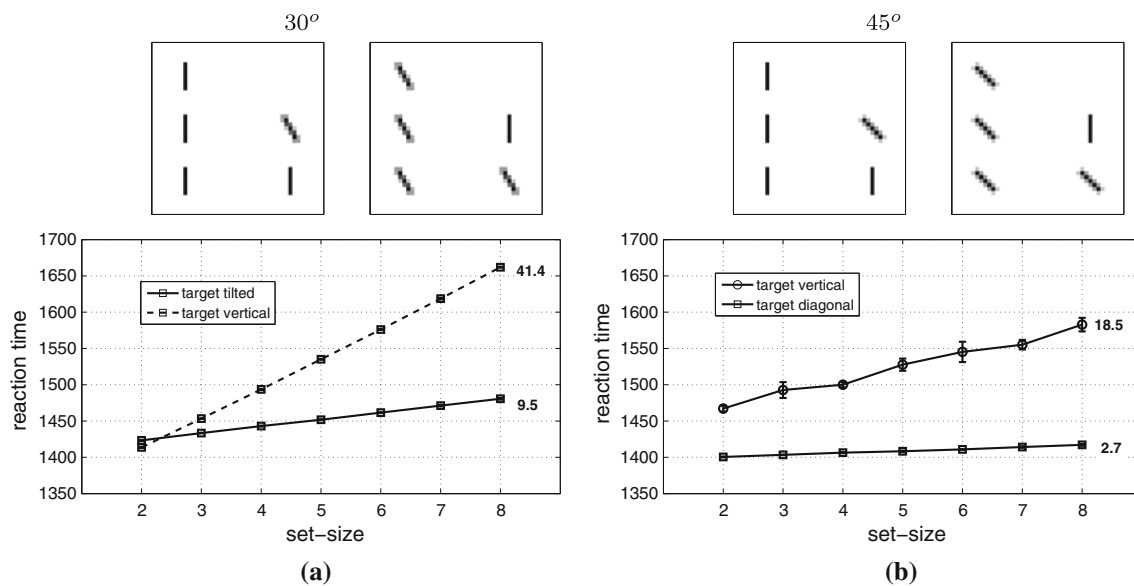
Fig. 8 Search asymmetries of line orientations. The search functions document the simulation results for **a** tilted line and vertical line, and for **b** diagonal line and vertical line. The results show that search is more efficient when the oriented line (diagonal or title) is the search target compared to a search with the vertical line as target. These effects constitute a search asymmetry and mimics experimental findings with the same search times (see text for details)

Reaction Time

The reaction time (RT) of the model is the simulation time it takes for one template unit to pass a set threshold 0.7. Passing the threshold was interpreted as the model having recognized an item.

Data Analysis

VS-SAIM's reaction times were analysed with an ANOVA as well as a linear regression to obtain search slope and intercept. In the search function plots, the search slope is depicted next to the average reaction time for highest set-size.

Results

Figure 8a, b show the RT functions for both orientation differences, 30° and 45°. For each orientation difference a separate two-way ANOVA with set-size and target-type as independent variables was carried out. The ANOVA for 30°-difference revealed significant main effects of set-size ($F(6, 69) = 2631.6$, $p < 0.001$) and target-type ($F(1, 69) = 49939.0$, $p < 0.001$) The interaction between the two factors was also significant ($F(6, 69) = 812.45$, $p < 0.001$). Figure 8a shows that overall reaction times increased with increasing set-size and that search for a vertical target was slower compared to search for the tilted target. The significant interaction resulted from a higher search efficiency when the tilted line was the target compared to when the vertical line was the target. This finding is also confirmed by the different slopes shown in Fig. 8a.

The results for the 45°-orientation difference were similar. The main effects of set-size ($F(6, 279) = 90355.0$, $p < 0.001$) and target-type ($F(1, 69) = 16700.0$, $p < 0.001$) were both significant. Also, the interaction between the two factors was significant ($F(6, 69) = 4755.4$, $p < 0.001$). Figure 8b shows that overall reaction times increased with increasing set-size and that search for a vertical target was slower compared to search for the diagonal target. The significant interaction resulted from a higher search efficiency when the diagonal line was the target compared to when the vertical line was the target. This finding is also confirmed by the different slopes shown in Fig. 8b.

Discussion

The simulation results show that VS-SAIM is able to qualitatively reproduce the central result of asymmetric visual search tasks, that of an altered search efficiency when target and distractor roles are swapped. A vertical line target among tilted lines is searched less efficient compared to a tilted line among vertical lines. There are three interesting aspects of these results. First the results demonstrate that the competition processes can produce set-size effects. Second the set-size effect is modulated by target-distractor similarity. Third target-distractor similarity is not the only factor influencing search efficient as otherwise search asymmetry would not have been possible.

The first two results were expected and are briefly discussed here. The third finding needs more explanation and will be discussed in the best part of this discussion.

As discussed earlier, the fact that competition process can produce set-size effects has been by our earlier work [19] by other such as a biased-competition model of visual search [10] and MORSEL [30]. A good way of conceptualizing the reason for this behaviour is that the speed of convergence of the competitive process in the selection network by and large determines the VS-SAIM's reaction times.[2] Moreover, the speed of convergence is proportional to the contrast between activations in the matching surface. The contrast is the difference between the highest input activation (target position) and all other input activations (distractor locations and background). For instance, the contrast would be highest, if there was only one item in the display. The contrast diminishes the more items are present in the search display leading to the set-size effect. Furthermore, the search slope depends on the target-distractor similarity, because the more similar target and distractor are the more the contrast diminishes with each additional item.

However, it is not obvious why the search asymmetry is simulated by VS-SAIM as well. In order to explain this result, it is necessary to examine the matching surface closer. Figure 9 shows two illustrations of a matching surface. The input stimuli were a vertical line and a titled line. The matching template for both illustrations was constructed from a equally weighted vertical and titled line. The resulting matching surface has three important characteristics: First, the highest match is obtained at the item's central location. This is to be expected, as the matching template and the item are aligned at this location. Second, the display background has a lower match than the central locations. Thirdly and interestingly, the area in the immediate vicinity of the items exhibits an even lower match than the background. This ''mismatch'' surrounding the item plays an important role in VS-SAIM's behaviour and is a direct consequence of the matching process in the matching network. Figure 10 shows a schematic illustration of the matching process. The graphs at the top show a one-dimensional illustration of the matching surface as found in real simulations (e.g. Fig. 9). The illustration at the bottom of the figure depicts three positions of the matching template (framed 'L') and relates them to the resulting matching values in the matching surface. Crucially, the mismatch occurs at the second location where there is only a partial overlap between matching template and item. The left part of the matching template is compared to the display background whereas the right part (background in the matching template) is compared to parts
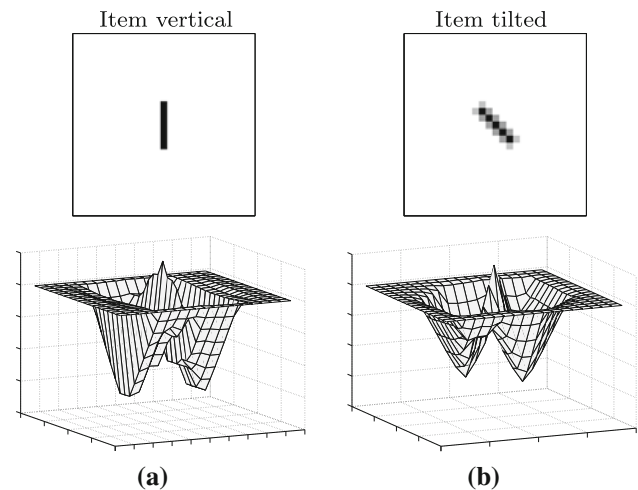

Fig. 9 Examples of two matching surfaces. The matching template consisted of two equally weighted template items (*top row*). This matching template was matched with the single vertical line and tilted line. The centre of the matching surfaces has the highest matching value consistent with the location of the respective items. Importantly the area surrounding the location of the best match shows a strong mismatch. This mismatch area plays an important role VS-SAIM's behaviour (see text for details)

of the L. So the left part produces the same matching value compared to when the matching template is entirely located on the background, but the right part generates a lower matching value than on the background leading to an overall matching value lower than the match against the background. Interestingly the shape of the matching surface is reminiscent of recent findings of behavioral performances surrounding the focus of attention (e.g. [2, 4, 8, 29, 32]) and on-centre-off-surround receptive fields in the early visual system (e.g. [1, 6, 23, 35]). However, in contrast to the mismatch occurring in the matching network the on-centre-off-surround effect in these early visual areas is assumed to be unaffected by knowledge-based influences. We will return to this interesting aspect in the general discussion of this paper.

It is also important to note that the amplitude of the mismatch is influenced by the absolute activation in the feature maps, as opposed to the relative activation resulting from the matching between matching template and item. This is illustrated in Fig. 10. For simplicity this effect is depicted for the intensity feature map. However, it should be noted that each feature map leads to the same effect. In Fig. 10 b) the input item is brighter than in Fig. 10 a). Hence, when matching template and input item partially overlap, the mismatch is larger when the input item is brighter than when the input item is dimmer since the match is mainly performed against the background in the matching template. In VS-SAIM, this matching is implemented with the Euclidian distance. Returning to the simulation results, it is important to note that this Euclidian distance for the

---

[2] In fact, the dynamics of the activations in the matching surface also play a role, but are not crucial for the simulation results in this paper.
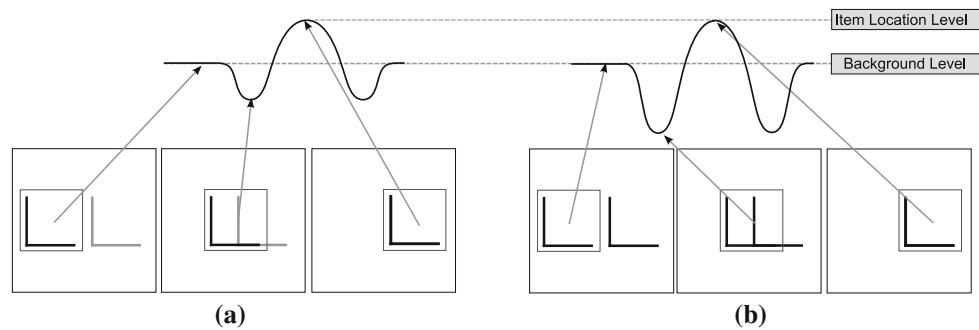
Fig. 10 Mismatch area in the matching surface. These two graphs explain the origin and the properties of the mismatch surrounding search items. The bottom row illustrate the matching process for three spatial distances from for input items a dim L on the *left side* and a

dark L (without frame) on the *right side*. The matching template is depicted as a framed L. The *top row* shows one-dimensional illustrations of the outcome of the matching process (see main text for further explanations)

vertical line is larger than for the diagonal and tilted line. This results from the fact that the diagonal feature map is less weighted than the feature map for vertical orientations. In turn this difference leads to a smaller mismatch for the diagonal line compared to the vertical line. Moreover this leads to a smaller decrease in contrast in the matching surface for the vertical line as distractor compared to the diagonal line as distractor. Therefore the property of the mismatch surround a search item explains the search asymmetry found in the simulations.

Study 3: Background Orientation

The explanation of the previous simulations highlighted that VS-SAIM's search efficiency depends not only on target-distractor similarity but also on the orientations of the distractors. Interestingly there is experimental evidence supporting this assumption. Foster and Westland [14] reported that search performance was also modulated by the absolute distractor orientation ('background orientation'). Search performance peaked at horizontal and vertical distractor orientations and fell towards oblique orientations, while the relative orientation between target and distractor was kept constant. In contrast an increase in the relative orientation only improved the overall performance (see Fig. 11 for an illustration). This study tests whether VS-SAIM is able to simulate this specific modulation of search performance through the background orientation.

Method

In this simulation, the distractor item was one out of 0 to 180° rotated counter clockwise from the vertical rotated lines, with a step-size of 30°. The target object was either a 30° or 45° counter clockwise from the vertical rotated line with respect to the background orientation. All distractors in a display had the same orientation. The rotated lines
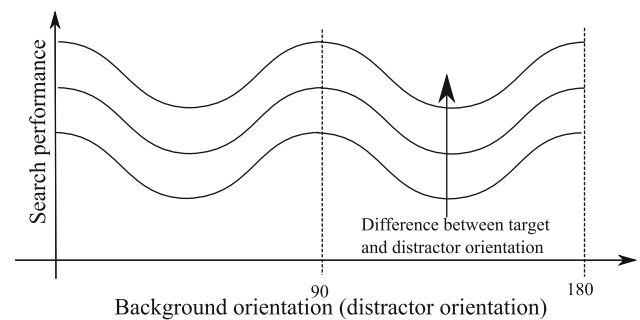
Fig. 11 Illustration of visual search results from Foster and Westland[14].Foster & Westland [14] varied line orientations of targets and distractors. They manipulated orientation difference between target and distractors together with the orientation of the distractors (background orientation). They found that both factors affect search performance

were created with Matlab routine *imrotate* and a bi-linear interpolation. Display size was five.

Results and Discussion

Figure 12 shows the mean reaction times across all background orientations for both relative orientation conditions successfully mimicking the findings by Foster and Westland [14]. Hence, VS-SAIM can generalize to additional orientations compared to Study 2. They also suggest that the way the mismatch area surrounding the distractors changes (Euclidian distance of the distractor features from the background) represents a good approximation of factors influencing visual search performance.

Study 4: Symmetric Search

So far the simulations concentrated on mimicking asymmetric search patterns. Indeed, the simulations seem to imply that the asymmetric search pattern is the standard finding and there should be no symmetric search pattern.

Fig. 12 General orientation search for 5 items. There were five distractor orientation ('background orientation') and two orientation differences between target and distractor (30° and 45°). The *top row* shows examples of search displays for the 30°-differences. These results mimic experimental findings ([14]; see also Fig. 11)

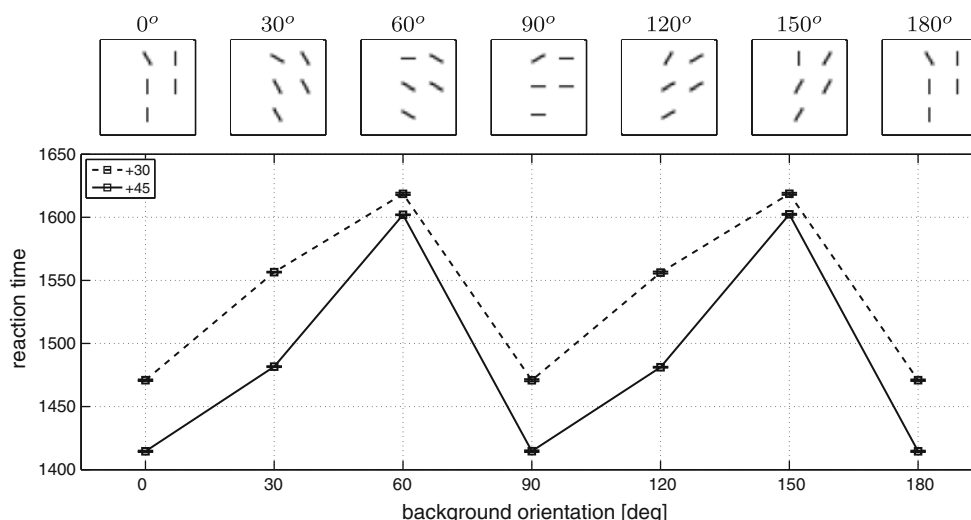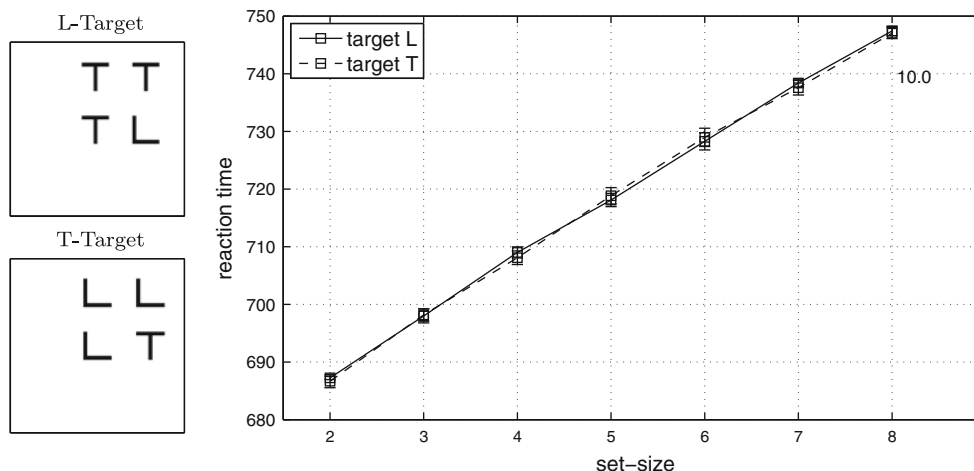**Fig. 13** Search symmetry. The search function show the simulation results for "T" and "L" as search items. In line with experimental findings VS-SAIM shows that search efficiency does not depend on whether the "L" or the "T" is the target item. These results are important as they highlight that not all simulations exhibit asymmetries. The reasons for this result and the theoretical implications are discussed in the text

However, there is some empirical evidence for symmetric search as well. For instance, Egeth and Dagenach [13] showed that in a search with 'L' and 'T' items, the swap of target and distractor has no significant effect on the participants' search performance. These simulations tests whether VS-SAIM can also simulate these symmetric experimental results.

Method

The same method as in Study 2 was used. The only difference were the search items. In this simulation 'L' and 'T' were used (see Fig. 13).

Results and Discussion

Figure 13 shows the search function produced by the model. A three-way ANOVA revealed a significant main effect of set-size ($F(6, 69) = 455.9$, $p < 0.001$) and no significant main effect in target-type ($F(1, 69) = 1.07$, $p = 0.31$) reflecting the symmetric search behaviour. The interaction between set-size and target-type was not significant ($F(6, 69) = 0.84$, $p = 0.55$). The results show that there is no modulation of search efficiency by swapping the target and distractor roles of the items. The reason for the successful simulation of the symmetric search pattern is that both, L and T, are mainly made up of vertical and

horizontal strokes. Only at cross points and end points the diagonal feature map shows some responses (see Fig. 2 for an illustration). Therefore, the mismatch area does not change much when L and T are swapped because the Euclidian distance from the background for both items does not differ. In other words, VS-SAIM suggests that when the item are predominately made of similarly weighted features, e.g. vertical and horizontal strokes, the search results should be symmetrical.

General Discussion

The Selective Attention for Identification Model (VS-SAIM) is a model of translation-invariant object recognition in a multiple object scene. In a first step, a early visual processing stage generates feature maps of vertical, horizontal and diagonal orientations. Then translation-invariance is achieved by mapping the content of the feature maps through to an attention window (FOA). Object recognition is implemented by a similarity-based (Euclidian distance) matching between stored templates for objects and activation in the FOA. With the issue of multiple objects, VS-SAIM deals with a mix of competitive and co-operative processes which are controlled by bottom-up and top-down influences. In the present paper, we simulated important findings from visual search experiments. Study 2 utilized search displays consisting of oriented lines (vertical, diagonal and titled lines). Each of these lines were either target or distractor in the simulations. The simulations demonstrated that VS-SAIM was able to mimic the typical increase in reaction times with increasing numbers of items (search slope). This result originates from the competitive processes in the selection network. As discussed in the introduction, this explanation has been put forward by several biologically plausible models, e.g. MORSEL [30], a biased-competition model of visual search [10] and our own work (e.g. [19]). Compared to these earlier works, the main progress is that, despite complex interactions between several competitive layers, VS-SAIM still produces a linear increase in reaction times. Hence, VS-SAIM suggests that, despite the fact that several competitive processes must interact in the brain, it is still possible that linear search function can emerge from these interactions. Furthermore, the slope of the search function is proportional to the similarity between target and distractor, in terms of orientation. For instance, search for the diagonal line among vertical lines is more efficient than search for a titled line among vertical lines. This is not unexpected as similarity-based matching plays a large role in VS-SAIM's behaviour. This outcome also fits to one of the central hypotheses put forward by the Attentional Engagement Theory [12].

Finally and most unexpectedly for a similarity-based approach, the simulation results mimicked the experimental findings of search asymmetries for oriented lines. For instance, search for a diagonal line among vertical lines is more efficient than search for a vertical line among diagonal lines. As explained in detail in the result section, crucial for these results are the contrast in VS-SAIM's matching surface which is modulated by the mismatch surrounding each item. Because the profile of the activation is reminiscent of receptive fields found in early visual processing in the brain and the fact that the profile is generated in the VS-SAIM's top-down path we termed the response profile, *knowledge-based on-centre-off-surround receptive field*. The centre of this new type of receptive field is dominated by the influence of the knowledge and the surround by the featural responses of the input stimuli (see Fig. 14 for an illustration). We will return this concept at the end of this discussion. Finally, the analysis revealed that biologically plausible unbalanced weighting of feature maps (e.g. [7]) is crucial for simulating search asymmetries of line orientation, with vertical and horizontal orientation weighted higher than diagonal orientation. By combining behavioural findings with this physiological evidence VS-SAIM's approach is validated further.

The following two studies tested two implications of Study 2. Study 3 showed that VS-SAIM cannot only simulate the influence of distractor orientation on visual search performance in general, but also the specific modulation found by Foster and Westland [14]. This success is mainly due to how the Euclidian distance between distractor items and background changes with item orientation. Second, Study 2 seem to imply that search asymmetry is the standard finding. However, there is also evidence for symmetric search pattern [13]. With its successful simulation VS-SAIM suggests that symmetry occurs when search items are formed by similarly weighted features, such as
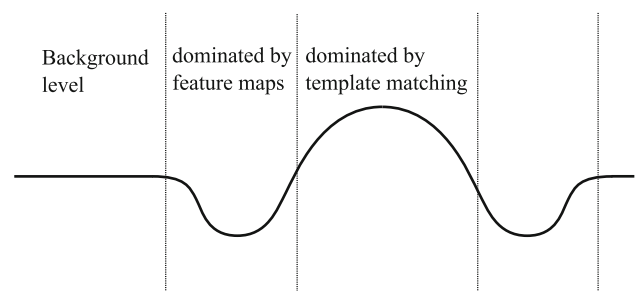
Fig. 14 The knowledge-based on-centre-off-surround receptive field. The graph illustrates the activation profile in the matching surface near a search item. The labels above indicate the origin of the activation levels. The profile is term knowledge-based receptive field because on one hand the shape of the profile is reminiscent of classical receptive fields in the early visual system and, on the other hand, the matching surface is the result of the matching between feature maps and top-down modulation

'L' and 'T'. Future research needs to follow up this prediction. For now it is important to notice the simulations presented in this paper underline the validity of VS-SAIM. Moreover, the results go beyond a simple "proof of existence" and make a specific prediction of what is crucial for these simulation results, the knowledge-based centre-on-surround-off receptive field. Because of its novelty and, to some extend, its counterintuitivity this concept is discussed for the remainder of this paper.

In general VS-SAIM suggests that search asymmetries can be the outcome of knowledge-based influence. Interestingly, this is consistent with behavioral evidence that can be interpreted as knowledge-based influences, e.g. mirrored letters versus normal letters [24] or "inverted elephants" versus normal elephants [43]. On the other hand, search asymmetries are often seen as diagnostic for the existence of feature maps (e.g. [37]). However, this apparent contradiction is resolved in VS-SAIM by the fact that the top-down influence is modulated by the featural properties. Moreover, the simulations presented here suggest that this Euclidian-based modulation in VS-SAIM presents a good approximation for searches among lines.

But in how far are the spatial properties of this top-down modulation, the knowledge-based on-centre-off-surround receptive field, plausible? To begin with, it is intuitive to suggest that top-down influence effects search not only exactly at locations of items but also in the vicinity of items. If this top-down influence consists of some kind of matching processes as assumed in VS-SAIM, this matching should not drop off rapidly, as the system has to be robust against noise, distortion, etc. Now, the matching could either tail off to the level of the background level or go below the background level and then increase again as it is the case in our simulation results. The latter option has the advantages that it improves the contrast against the background and makes it more detectable for following processing stages. Moreover and importantly, apart from these theoretical considerations, there is also empirical support for the on-centre-off-surround shape of the matching surface: the well-known response characteristic of receptive fields in the early visual system (e.g. [1, 6, 23, 35]) and recent findings of behavioral performances surrounding the focus of attention (e.g. [2, 4, 8, 29, 32]). The classical findings on on-centre-off-surround receptive fields are usually interpreted as a purely feature-based process located in the retina or the LGN. VS-SAIM generalizes this type of spatial response to a knowledge-based on-centre-off-surround receptive field. The location of such receptive fields in the brain is unclear. It could be that the receptive fields in the early visual field indeed are influenced by knowledge. This has not been tested, but there are indications that responses in early visual processing are influenced by top-down modulation (e.g. see [25] for evidence

on the effect of spatial attention in V1 in an animal study). An obvious alternative could be regions in which fMRI studies have shown indication of object processing, e.g. lateral occipital cortex (e.g. [15]). It is also worth noting that such a generalization from a model of low-level processes to higher-level processes is not uncommon. For instance, models based on the principle component analysis (PCA) have been applied to model the formation of low-level receptive fields (e.g. [33]) and human face recognition (see [27]; for a recent example). A similar transfer of a mechanism from low-level processes to high-level process is suggested by VS-SAIM for the on-centre-off-surround receptive field.

The second supporting evidence for VS-SAIM's on-centre-off-surround receptive field comes from behavioural experiments on visual attention. In these experiments, the location of the focus of attention is manipulated by target locations in visual search [2, 4, 29], spatial cue [8] or identification of letters at a pre-defined location [32]. The spatial profile of the focus of attention is determined by measuring the success of detecting a simple probe stimulus [2, 4], comparing the identity of the probe letter (same colour) with the target letter [8] and identifying the probe stimulus [29, 32]. The experiments show that the probe performance exhibits a similar on-centre-off-surround profile as VS-SAIM. Interestingly, even some details of the response characteristics are consistent with VS-SAIM's profile. The profile is influenced by the saliency of the target whereby the inhibitory zone is deeper when the target is more salient [29]. Second, Boehler et al. [2] showed that the exact shape of the profile depends on the task performed, i.a. simple target detection vs. detecting a feature on the target. This finding can be interpreted that the profile is affected by top-down processes as in VS-SAIM. However, future research needs to test whether this attentional profile is affected only by the task setting or whether properties of distractors influence the profile, e.g. by applying a probe task to asymmetric and symmetric search tasks. Furthermore, these experimental findings are normally conceptualized as profiling the focus of attention. Hence in VS-SAIM this can be construed as activation profile in the selection network. On the other hand these experiments can also be interpreted as tapping into the control mechanism of attention (see [29] for a similar point). This interpretation is consistent with VS-SAIM's prediction that the centre-on-surround-off profile is produced by the matching network. Future experiments need to tease these two hypotheses apart.

Finally, the simulations with VS-SAIM suggest that search is strongly influenced by bottom-up properties of the distractors, especially highlighted by Study 3. In other words VS-SAIM's simulations suggest that, apart from the target-distractor similarity, the properties of distractors

play an important role in the efficiency of visual search. This point is interesting, because it is in contrast to most classical theories on visual search, where the focus is on the properties of the target rather than the distractors. In some sense VS-SAIM's suggestion seems intuitively plausible as there are simply more distractors present in the search display, consequently, exerting stronger influence on human behaviour. Future experiments need to explore this novel suggestion.

Appendix: Mathematical Details of VS-SAIM

General Modelling Principle

As explained in the main text, VS-SAIM's design is based on the principle of minimization of an energy function in order to implement a soft-constraint satisfaction. The core idea is that activation patterns fulfilling a set of constraints constitute minima in an energy function. In order to determine the minimum, VS-SAIM performs a gradient descent in the energy function following a suggestion by Hopfield and Tank [21]:

$$\tau \dot{x}_k = -g \frac{\partial E(\mathbf{y})}{\partial y_k}; \quad y_k = f(x_k) \tag{1}$$

where τ is a time constant, g a gain factor, $\mathbf{x}$ the inner activation of the model and $\mathbf{y}$ the model output activation. $f(x)$ is the output function of the individual units. In VS-SAIM each module, knowledge network, contents network, etc. pursuits a different set of constraints depending on its task, e.g. identifying objects in the knowledge network. Hence, for each network an energy function is defined with different minima reflecting a correct completion of its task. In order ensure that VS-SAIM, as a whole, satisfies all constraints at the same time, the individual energy functions are added together to a overall energy function. Apart from the energy function as such, the choice of the output function $f(x)$ depends on the implemented constraints. Here, we used either the sigmoid function,

$$f(x) = \frac{1}{1 + \exp(-m(x - s))} \tag{2}$$

or a linear function

$$f(x) = m \cdot x - s. \tag{3}$$

whereby m is the slope and s the intercept. The sigmoid function is more suitable if the attractor states have to be either zero or one. In contrast, the linear function is appropriate, if the final state should reflect continuous values, e.g. the activation in the feature maps. Therefore, the sigmoid function was used in the selection network and the knowledge network. The linear function was used in the content network and the matching network. Finally, to achieve some degree of biologically plausible in VS-SAIM, the differential equations can also include a leaky integrator so that the resulting differential equation turns into:

$$\tau \dot{x}_k = -x_k - g \frac{\partial E(\mathbf{y})}{\partial y_k} \tag{4}$$

Before the equations of the individual networks are introduced, we will present the equations for a simple WTA-network in a separate section. This illustrates the energy function approach in a simple example and also introduces the central building block for VS-SAIM.

Winner Take All (WTA)

The WTA energy function was suggested by [28]:

$$E^{WTA}(\mathbf{y}) = \alpha \left(\sum_{k=1}^{K} y_k - 1 \right)^2 - \beta \sum_{k=1}^{K} I_k y_k \tag{5}$$

where y_k is the output activation of the *kth* neuron and I_k its input. Every neuron has an internal activation x_k and its output activation y_k is calculated via the sigmoid function. The WTA energy function is minimal when all y_is are zero except one (first term), and the corresponding I_i has the maximal value of all I_is (second term). The parameter α and β weights the two terms or constraints against each other. The gradient descent together with a leaky integrator for each neuron results in the following equations:

$$\tau \dot{x}_k = -x_k - \frac{\partial E^{WTA}(\mathbf{y})}{\partial y_k} = -x_k - 2\alpha \left(\sum_{k=1}^{K} y_k - 1 \right) + \beta I_k \tag{6}$$

Finally, note that the term $\sum_{k=1}^{K} y_k$ implements a global inhibition in the WTA- network.

Early visual processing stage (EVPS)

The early visual processing stage (EVPS) consists of Gabor-filters with four orientations (0°, 90°, 45° and 135°), modelling V1 operations. The following equation describes the normalization:

$$I_{ij}^{norm} = \frac{I_{ij}}{\sqrt{\sum_{u=-U}^{U} \sum_{v=-V}^{V} I_{i+u,j+v}^2}} \qquad (7)$$

with a window size of $(2U + 1) \times (2V + 1)$. The term I_{ij} represents the input image while $I^{norm}{}_{ij}$ is the normalized intensity image. The Gabor-filter implementation follows the model of the receptive field of simple cells in the visual cortex proposed by [9]:

$$x(s,r) = s \cdot \cos\theta + r \cdot \sin\theta \qquad (8)$$

$$y(s,r) = r \cdot \cos\theta - s \cdot \sin\theta \qquad (9)$$

$$G(s,r,\theta,k) = \frac{1}{A}\exp\left(-0.5\left(\left(\frac{x(s,r)}{\sigma}\right)^2 + \left(\frac{y(s,r)}{\sigma}\right)^2\right)\right)$$
$$\times \cos\left(2\pi kx(s,r)\right) \qquad (10)$$

$$A = \left|\sum_{sr} G(s,r,\theta,k)\right| \qquad (11)$$

with θ being the Gabor kernel orientation and k its frequency.

The Gabor-filters are convolved with the normalized intensity image. Consequently, the output of the EVPS consists of a five feature maps: The first feature map, $n = 1$, is the normalized intensity image:

$$f_{ij}^{(1)} = p^{(1)}I_{ij}^{norm} \qquad (12)$$

while the other four feature maps $n = 2$, $n = 3$, $n = 4$, $n = 5$ are the convolution results of the normalized intensity input with a Gabor-filter mask:

$$f_{ij}^{(n)} = p^{(n)}\left|\sum_{s=-S}^{S} \sum_{r=-R}^{R} G(s,r,\theta_n,k_n)I_{i+s,j+r}^{norm}\right| \qquad (13)$$

In the cases $i + s$ or $j + r$ exceeds the visual field boundaries, the intensity values are set to zero (boundary handling). The parameter $p^{(n)}$ weights the feature maps and were altered in the "exploration of the parameter space".

Noise

In order to model noise in the visual system, noise was added to the input display and was based on the following equation implemented in an earlier version of VS-SAIM [19]:

$$\ddot{x} + \gamma\dot{x} + \sin x = A\cos(\omega t + p) \qquad (14)$$

This equation was inspired by the motion equation of a periodically driven pendulum where γ is the damping constant and the right side describes a driving torque with amplitude A, angular frequency ω and phase p.

This equation was chosen on merely technical grounds and exhibits a chaotic behaviour or quasi-stochastic

temporal behaviour. Since this "noise" is described with a differential equation, it fits seamlessly into the differential equations derived from the energy function approach.

To ensure that each retinal unit of the input receives a different signal, each retinal unit has its own pendulum equation:

$$\ddot{x}_{ij} + \gamma\dot{x}_{ij} + \sin x_{ij} = A\cos(\omega t + p_{ij}) \qquad (15)$$

For each retinal unit (i, j) an initial state $x_{ij}(0)$ and phase p_{ij} is randomly chosen drawn from an equal distribution.

To limit the amplitude of the noise x_{ij} was fed into the following equation:

$$y_{ij}^{noise}(t) = 0.5\big((max - min) \cdot \sin\left(x_{ij}(t)\right) + (max + min)\big) \qquad (16)$$

where max and min are the limits of the noise amplitude. The term $y^{noise}{}_{ij}$ is then added to each feature map f_{ij}^n.

Contents Network

The contents network aims to enable a translation-invariant mapping from the EPVS to the smaller focus of attention. The energy function for the contents network is defined as,

$$E^{CN}\big(\mathbf{y}^{CN}, \mathbf{y}^{SN}\big) = \alpha^{CN}\sum_{lm}\sum_{ij}\sum_{n}\left(y_{lmn}^{CN} - f_{ij}^{(n)}\right)^2\left(y_{lmij}^{SN}\right)^q \qquad (17)$$

where α^{CN} is the weight of the energy term in the overall energy function of VS-SAIM. The indices i and j refer to image locations, and the indices l and m refer to FOA-locations. n indexes the feature map. The variable y_{lmn}^{CN} is the output activation of the contents network while y_{lmij}^{SN} is the output activation of the selection network.

The term $\left(y_{lmn}^{CN} - f_{ij}^{(n)}\right)^2$ ensures this energy function is minimal when the FOA matches the activations of the features maps. However, because of the multiplication with the output activations from the selection network $\left(y_{lmij}^{SN}\right)^q$ this match is only required at selected locations ($y_{lmij}^{SN} = 1$), whereas deselected locations ($y_{lmij}^{SN} \approx 0$) do not contribute to the minimization of the energy function. The parameter q, chosen to a value larger than one, enhances the selection effect of the selection network, since it decreases small activations more than larger activations. This implementation of the contents network was initially design for the Grouping-SAIM (G-SAIM; [18]). In G-SAIM we simulated simple effects of with- and between-object grouping. In this version, the contents network was crucial for its success. However, for the aim of this paper it is only relevant the contents network operates similar to a sigma-pi

unit, as explained in the main text (see also [16]). Here, we will demonstrate this equivalence. The differential equation for the contents network units is described by,

$$\tau \dot{x}_{lmn}^{CN} = -x_{lmn}^{CN} - \frac{\partial E^{CN}(\mathbf{y}^{SN}, \mathbf{y}^{CN})}{\partial y_{lmn}^{CN}} \tag{18}$$

with the derivation term defined as,

$$\frac{\partial E^{CN}(\mathbf{y}^{SN}, \mathbf{y}^{CN})}{\partial y_{lmn}^{CN}} = 2\alpha^{CN} \sum_{ij} \left(y_{lmn}^{CN} - f_{ij}^{n}\right)\left(y_{lmij}^{SN}\right)^{q} \tag{19}$$

The contents network uses a linear output function, so that y_{lmn}^{CN} is equal with x_{lmn}^{CN}. So once the differential equations are converged, $\dot{x}_{lmn}^{CN}$ is zero and Eq. 18 turns into:

$$0 = -y_{lmn}^{CN} - 2\alpha^{CN} \sum_{ij} \left(y_{lmn}^{CN} - f_{ij}^{n}\right)\left(y_{lmij}^{SN}\right)^{q} \tag{20}$$

The solution of this equation for y_{lmn}^{CN} is:

$$y_{lmn}^{CN} = 2\alpha^{CN} \frac{\sum_{ij} f_{ij}^{n} \cdot \left(y_{lmij}^{SN}\right)^{q}}{1 + \sum_{ij} \left(y_{lmij}^{SN}\right)^{q}} \tag{21}$$

Since $\sum_{ij} y_{lmij}^{SN}$ converges to one (see selection network), the converged contents network operates like a sigma-pi unit:

$$y_{lmn}^{CN} = \alpha^{CN} \sum_{ij} f_{ij}^{n} \cdot \left(y_{lmij}^{SN}\right)^{q} \tag{22}$$

Selection Network

The selection network aims to select an item in the input image by generating an appropriate activation pattern which ensures a veridical mapping of this item into the FOA. The selection network is structured into layers whereby each layer controls the routing for a different FOA-pixel. To ensure a veridical representation, the selection network has to fulfil two constraints (see main text): (a) one unit in the FOA should not receive an input from more than one retinal unit and (b) neighbourhood relations should be preserved during the mapping process. Constraint (a) is implemented as a WTA:

$$E^{SN1}(\mathbf{y}^{SN}) = \alpha^{SN} \sum_{lm} \left(\sum_{ij} y_{lmij}^{SN} - 1\right)^{2} \tag{23}$$

where α^{SN} weights the constraint. Constraint (b) is realized with excitatory connections between layers:

$$E^{SN2}(\mathbf{y}^{SN}) = -\beta^{SN} \sum_{lm} \sum_{ij} \sum_{\substack{s=-S \\ s\neq 0}}^{S} \sum_{\substack{r=-R \\ r\neq 0}}^{R} g(s,r) y_{l+s,m+r}^{SN}{}_{i+s,j+r} y_{lmij}^{SN} \tag{24}$$

and is weighted by β^{SN}. The strength of the connection

$g(s, r)$ decreases with the distance between units and weakens the co-operation between units further apart from each other (see also [16]). The neighbourhood function $g(s, r)$ is defined by,

$$g(s, r) = \frac{1}{A} e^{-\frac{s^2 + r^2}{\sigma^2}} \tag{25}$$

where A is a normalization factor with

$$A = \sum_{s=-S}^{S} \sum_{r=-R}^{R} e^{-\frac{s^2 + r^2}{\sigma^2}} \tag{26}$$

The differential equation for a selection network unit is given with

$$\tau \dot{x}_{lmij}^{SN} = -x_{lmij}^{SN} - \frac{\partial E^{SN1}(\mathbf{y}^{SN})}{\partial y_{lmij}^{SN}} - \frac{\partial E^{SN2}(\mathbf{y}^{SN})}{\partial y_{lmij}^{SN}} \tag{27}$$

where the individual terms are:

$$\frac{\partial E^{SN1}(\mathbf{y}^{SN})}{\partial y_{lmij}^{SN}} = 2\alpha^{SN} \left(\sum_{ij} y_{lmij}^{SN} - 1\right) \tag{28}$$

$$\frac{\partial E^{SN2}(\mathbf{y}^{SN})}{\partial y_{lmij}^{SN}} = -\beta^{SN} \sum_{\substack{s=-S \\ s\neq 0}}^{S} \sum_{\substack{r=-R \\ r\neq 0}}^{R} g(s,r) y_{l+s,m+r,i+s,j+r}^{SN} \tag{29}$$

Knowledge Network

The knowledge network implements the object identification in VS-SAIM. The energy function for the knowledge network is defined as the following,

$$\begin{aligned} E^{KN}(\mathbf{y}^{KN}, \mathbf{y}^{CN}) = \alpha^{KN} \left(\sum_{k=1}^{K} y_k^{KN} - 1\right)^2 \\ + \beta^{KN} \sum_{k=1}^{K} \left(M(\mathbf{y}^{CN}, \mathbf{w}^k) \right. \\ \left. - \frac{1}{K} \sum_{k'=1}^{K} M(\mathbf{y}^{CN}, \mathbf{w}^{k'})\right) y_k^{KN} \end{aligned} \tag{30}$$

with $M(\mathbf{y}^{CN}, \mathbf{w}^k)$ being the Euclidean matching function between the FOA and an individual template,

$$M(\mathbf{y}^{CN}, \mathbf{w}^k) = \sum_{lm} \sum_{n} \left(y_{lmn}^{CN} - w_{lmn}^k\right)^2 \tag{31}$$

where the index k refers to template units whose template features are stored in the weights w_{lmn}^k. The parameter K is the total number of templates in the model. The term $\left(\sum_k y_k^{KN} - 1\right)^2$ implements the WTA-constraint. The matching term $\sum_{lmn} \left(y_{lmn}^{CN} - w_{lmn}^k\right)^2$ ensures that only the template unit is activated which gives the best match with the FOA contents. The term $\frac{1}{K} \sum_{k'=1}^{K} M(\mathbf{y}^{CN}, \mathbf{w}^{k'})$ calculates the mean distance of the FOA from all templates and will be explained in the context of the differential

equations. The parameter α^{KN} and β^{KN} weight the constraints against each other.

The energy function is differentiated with respect to y_k^{KN} which creates the following differential equation (without the leaky integrator) for a knowledge network unit,

$$\tau \dot{x}_k^{KN} = -\frac{\partial E(\mathbf{y}^{KN}, \mathbf{y}^{CN})}{\partial y_k^{KN}} \tag{32}$$

with the partial derivation term defined as,

$$\frac{\partial E(\mathbf{y}^{KN}, \mathbf{y}^{CN})}{\partial y_k^{KN}} = 2\alpha^{KN}\left(\sum_k y_k^{KN} - 1\right)$$
$$+ \beta^{KN}\left(\sum_{lmn}\left(y_{lmn}^{CN} - w_{lmn}^k\right)^2 \right. \tag{33}$$
$$\left. -\frac{1}{K}\sum_{k'}\sum_{lmn}\left(y_{lmn}^{CN} - w_{lmn}^{k'}\right)^2\right)$$

With this implementation, the knowledge network "waits" during the "first" phase of the selection process until the FOA begins to represent information about the input image. In other words the knowledge network is only influenced by the "real" selection process and pilot studies showed that this approach made VS-SAM's search performance more robust. This waiting status ($\dot{x}_k^{KN} = 0$) results from the following conditions at the beginning of simulations: First, the sum of the initial values of the output activation is one (see parameters), thus, the WTA-term is zero. Second, the FOA output is initialized with average between the two templates (unbiased template) and, consequently, the matching term is zero too, because the mean matching value is subtracted from the matching values.

Top-Down Feedback and Matching Network

So far, we only discussed the following partial derivatives of the energy functions: $\frac{\partial E(\mathbf{y}^{KN}, \mathbf{y}^{CN})}{\partial y_k^{KN}}, \frac{\partial E^{CN}(\mathbf{y}^{SN}, \mathbf{y}^{CN})}{\partial y_{lmn}^{CN}}$. In fact, these terms define the bottom-up pathway. However, for the gradient descent approach to be successful, it requires to consider partial derivatives to all dynamical variables: $\frac{\partial E(\mathbf{y}^{KN}, \mathbf{y}^{CN})}{\partial y_k^{CN}}, \frac{\partial E^{CN}(\mathbf{y}^{SN}, \mathbf{y}^{CN})}{\partial y_{lmn}^{SN}}$. These partial derivative construct the top-down pathway.

In extensive tests, we found that "strict" application of the gradient descent procedure showed that does not lead to a reliable visual search in VS-SAIM. Subsequently, we modified the network architecture and introduced the *"matching network"*. Nevertheless, the resulting topology was inspired by a strict application of the gradient descent method. Therefore, we will first present its correct application and, then, introduce and discuss the changes which led to the network architecture, as it is presented in the main text.

For the knowledge network, the top-down path is derived by a partial derivative with respect to x_{lmn}^{CN}:

$$\tau \dot{x}_{lmn}^{CN} = -x_{lmn}^{CN} - \frac{\partial E^{KN}(\mathbf{y}^{CN}, \mathbf{y}^{KN})}{\partial y_{lmn}^{CN}} \tag{34}$$

with the derivation termed defined as,

$$\frac{\partial E^{KN}(\mathbf{y}^{CN}, \mathbf{y}^{KN})}{\partial y_{lmn}^{CN}} =$$
$$= 2\beta^{KN}\sum_k\left(\left(y_{lmn}^{CN} - w_{lmn}^k\right) - \frac{1}{K}\sum_{k'}\left(y_{lmn}^{CN} - w_{lmn}^{k'}\right)\right)y_k^{KN} \tag{35}$$

This term would have been added to Eq. 19 and would have introduced a direct feedback from the knowledge network into the contents network. For the contents network the energy function is derived with respect to y_{lmij}^{SN}:

$$\frac{\partial E^{CN}(\mathbf{y}^{SN}, \mathbf{y}^{CN})}{\partial y_{lmn}^{CN}} =$$
$$\alpha^{CN} \cdot q \sum_{lm}\sum_{ij}\sum_n\left(y_{lmn}^{CN} - f_{ij}^n\right)^2\left(y_{lmij}^{SN}\right)^{(q-1)} \tag{36}$$

This term would have been added to Eq. 27 and would have introduced a input from the contents network into the selection network.

The pilot simulations revealed that the direct feedback from the knowledge network into the contents network represented a major problem for a successful visual search, because the directness of the feedback loop made it difficult to balance bottom-up and top-down influence. Therefore, we replaced the contents network in the top-down path with the matching network x_{lmn}^{MN}, so that the top-down influence became less immediate and the feedback loop is closed via the selection network. Expressed in mathematical terms, Eq. 34 turned into

$$\tau \dot{x}_{lmn}^{MN} = -x_{lmn}^{MN} - \frac{\partial E^{KN}(\mathbf{y}^{MN}, \mathbf{y}^{KN})}{\partial y_{lmn}^{NM}} \tag{37}$$

with the derivation termed defined as,

$$\frac{\partial E^{KN}(\mathbf{y}^{MN}, \mathbf{y}^{KN})}{\partial y_{lmn}^{MN}} =$$
$$2\beta^{KN}\sum_k\left(\left(y_{lmn}^{MN} - w_{lmn}^k\right) - \frac{1}{K}\sum_{k'}\left(y_{lmn}^{MN} - w_{lmn}^{k'}\right)\right)y_k^{KN} \tag{38}$$

Like the contents network, the initial values of the matching network were the averaged templates. The way the matching network projects into the selection network following Eq. 37. However, our pilot studies showed that the factor $\left(y_{lmij}^{SN}\right)^{(q-1)}$ often prevented a successful selection, as this

term is often close to zero and therefore blocks any top-down modulation from the matching network. There we set $q = 1$ in the top-down pathway and eliminated this factor.

The pilot studies revealed a further problem with the matching network. The matching term $(y_{lmn}^{CN} - w_{lmn}^{k'})$ does not take into account any the spatial neighbourhood relations between pixels in items. Even though the neighbourhood constraint in the selection network should be able to solve this problem, the pilot simulations indicated that this is not sufficient to ensure a successful operation of VS-SAIM for all items used here. Therefore, we introduced a spatial matching window for each feature map location with the same size as the matching template. This heuristics was successfully employed in natural images [20]. The resulting equation is the following:

$$E^{MN}(\mathbf{y}^{SN}, \mathbf{y}^{MN}) =$$
$$\alpha^{MN} \sum_{lm} \sum_{ij} \left(\sum_{s=-S}^{S} \sum_{r=-R}^{R} \sum_{n} \left(y_{l+s,m+r,n}^{MN} - f_{i+s,j+r}^{n} \right)^2 \right) y_{lmij}^{SN} \tag{39}$$

The term $\sum_{s=-S}^{S} \sum_{r=-R}^{R} \left(y_{l+s,m+r,n}^{MN} - f_{i+s,j+r}^{n} \right)^2$ constitutes the additional matching window. The partial derivative with respect to y_{lmij}^{SN} is

$$\frac{\partial E^{MN}(\mathbf{y}^{SN}, \mathbf{y}^{MN})}{\partial y_{lmij}^{SN}} = \alpha^{MN} \sum_{s,r} \sum_{n} \left(y_{l+s,m+r,n}^{MN} - f_{i+s,j+r}^{n} \right)^2 \tag{40}$$

The pilot studies highlighted a final problem. The activation amplitudes of the input into the selection $\left(\frac{\partial E^{MN}(\mathbf{y}^{SN}, \mathbf{y}^{MN})}{\partial y_{lmij}^{SN}} \right)$ can greatly vary depending on the search items. Occasionally, these activations were too high for the WTA-constraint in the selection network to work successfully, especially for large displays. In other words, the global inhibition was not able to restrict the number of winners to one. This situation particularly occurred at the beginning of the selection process, termed Phase 1 in the main text, when the unbiased matching templates matches equally well all items. A possible consequence of this failure is that several distractors are being selected before the knowledge network induces the target template in the matching template. Therefore, we added a normalization to input of the selection network:

$$\tau \dot{x}_{lmij}^{SN} = -x_{lmij}^{SN} - \frac{\partial E^{SN1}(\mathbf{y}^{SN})}{\partial y_{lmij}^{SN}} - \frac{\partial E^{SN2}(\mathbf{y}^{SN})}{\partial y_{lmij}^{SN}}$$
$$- norm\left(\frac{\partial E^{MN}(\mathbf{y}^{SN}, \mathbf{y}^{MN})}{\partial y_{lmij}^{SN}} \right) \tag{41}$$

It is important to note the result of the normalization is termed *matching surface* in the main text. The function $norm(\bullet)$ is defined as,

$$norm(z) = v * \frac{z - \|\mathbf{y}^{MN}(t=0)\|^2}{\|\mathbf{y}^{MN}(t=0) - \mathbf{w}^1\|^2} + \xi \tag{42}$$

where the term $\mathbf{y}^{MN}(t = 0)$ is the activation of the matching template at initialization (unbiased matching template), while $\mathbf{w}^1$ is the first template in the knowledge network template. The normalization ensures that the matching level is ξ in the background and $v + \xi$ at the centre of an item. ξ and v were set to 0.5 and 1.0 respectively. The normalization subtracts the matching value of the unbiased matching templates with the background $\|\mathbf{y}^{MN}(t=0)\|^2$ from the matching levels (z). Hence, the normalization transforms the matching level in the background to ξ. The best matching level at the beginning of the selection process is $\|\mathbf{y}^{MN}(t=0) - \mathbf{w}^1\|^2$ which, in fact, is the same value as $\|\mathbf{y}^{MN}(t=0) - \mathbf{w}^2\|^2$, as the unbiased matching template has the same Euclidian distance from the two templates. By dividing $z - \|\mathbf{y}^{MN}(t=0)\|^2$ through this highest value the normalization restricts the input activation to selection network to $v + \xi$.

Simulation Parameter

See Tables 1, 2, 3, 4, 5, and 6.

Table 1 Contents network

Parameter	Value	Description
$f(x)$	Linear	Type of output function
m	1.0	
s	0.0	
τ	1.0	Time constant
α^{CN}	1.0	Mapping factor
q	2.0	Mapping power factor

Table 2 Feature extraction

Parameter	Value	Description
I × J	43 × 43	Size of input
L × M	9 × 9	Size of FoA
N	5	Feature dimensions
k	0.3	Gabor frequency
σ	3.0	Gabor sigma
θ	[0, 90, 45, 135]	Tuned orientations in [°]
p^n	[0.3,1.0,1.0,0.1,0.1]	Feature weighting
S & R	2	Half window size Gabor filter

Table 3 Selection Network

Parameter	Value	Description
$f(x)$	Sigmoid	Type of output function
m	15.0	
s	1.0	
τ	0.5	Time constant
α^{SN}	120.0	Weighting WTA
β^{SN}	20.0	Weighting neighbourhood function
σ	7.0	Sigma in neighbourhood function
S & R	4	Half size of neighbourhood function

Table 4 Knowledge network

Parameter	Value	Description
$f(x)$	Sigmoid	Type of output function
m	15.0	
s	0.5	
τ	0.1	Time constant
α^{KN}	0.1	Weighting WTA
β^{KN}	1.0	Weighting matching
$Decision - thresh$	0.7	Search termination threshold
$Target - init$	0.506	Biased initialization of target unit
$Distractor - init$	0.494	Biased initialization of distractor unit

Table 5 Noise

Parameter	Value	Description
Tau	1.0	Time constant
Min	0.0	Minimum noise amplitude
Max	0.0001	Maximum noise amplitude
γ	0.00001	Damping constant
ω	10.0	Frequency
A	7.0	Amplitude

Table 6 Matching network

Parameter	Value	Description
$f(x)$	Linear	Type of output function
m	1.0	
s	0.0	
τ	1.0	Time constant
α^{MN}	1.0	Feedback factor
S & R	4	Half window size
ξ	1.0	Background matching surface
v	0.5	Item location to blank background contrast

References

1. Barlow HB. Summation and inhibition in the frog's retina. J Physiol. 1953;119:69–88.
2. Boehler CM, Tsotsos JK, Schoenfeld MA, Heinze HJ, Hopf JM. The centre-surround profile of the focus of attention arises from recurrent processing in visual cortex. Cereb Cortex. 2009;19: 982–91.
3. Carrasco M, McLean T, Katz S, Frieder K. Feature asymmetries in visual search: effects of display duration, target eccentricity, orientation and spatial frequency. Vision Res. 1998;38(3): 347–74.
4. Cave KR, Bichot NP. Visuo-spatial attention: beyond a spotlight model. Psychon Bull Rev. 1999;6:204–23.
5. Chun M, Wolfe J. Just say no: how are visual searches terminated when there is no target present? Cogn Psychol. 1996;30:39–78.
6. Cook P, McReynolds J. Lateral inhibition in the inner retina is important for spatial tuning of ganglion cells. Nat Neurosci. 1998;1:714–9.
7. Coppola D, White L, Fitzpatrick D, Purves D. Unequal representation of cardinal and oblique contours in ferret visual cortex. Neurobiology. 1998;95(5):2621–3.
8. Cutzu F, Tsotsos JK. The selective tuning model of attention: psychophysical evidence for a suppressive annulus around attended item. Vision Res. 2003;43:205–19.
9. Daugman J. Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Opt Soc Am A. 1985;2:1160–9.
10. Deco G, Zihl J. Top-down selective visual attention: a neurodynamic approach. Visual Cogn. 2001;8(1):119–40.
11. Dick M, Ullman S, Sagi D. Parallel and serial processes in motion detection. Science. 1987;237:400–2.
12. Duncan J, Humphreys GW. Visual search and stimulus similarity. Psychol Rev. 1989;96(3):433–58.
13. Egeth H, Dagenbach D. Parallel versus serial processing in visual search: further evidence from subadditive effects of visual quality. J Exp Psychol Hum Percept Perform. 1991;17(2):551–60.
14. Foster D, Westland S. Orientation contrast vs. orientation in line-target detection. Vision Res. 1995;35(6):733–8.
15. Grill-Spector K. The neural basis of object perception. Curr Opin Neurobiol. 2003;13:159–66.
16. Heinke D, Humphreys GW. Attention, spatial representation and visual neglect: Simulating emergent attention and spatial memory in the Selective Attention for Identification Model (SAIM). Psychol Rev. 2003;110(1):29–87.
17. Heinke D, Mavritsaki E (ed). Computational modelling in behavioural neuroscience: closing the gap between neurophysiology and behaviour. London: Psychology Press; 2009.
18. Heinke D, Sun YR, Humphreys GW. Modeling grouping through interactions between top-down and bottom-up processes: the grouping and selective attention for identification model (G-SAIM). In: Paletta L, Tsotsos JK, Rome E, Humphreys GW (ed) Attention and performance in computational vision, lecture notes in computer science, vol 3368. Springer, New York, pp. 148–58; 2005.
19. Heinke D, Humphreys G, Tweed C. Top-down guidance of visual search: a computational account. Vis Cogn. 2006; 14(4/5/6/7/8):985–1005.
20. Heinke D, Backhaus A, Sun YR, Humphreys GW. The Selective Attention for Identification model SAIM): simulating visual search in natural colour images. In: Paletta L, Rome E (eds) Attention in cognitive systems, lecture notes in computer science 4840, pp 141–54; 2008.
21. Hopfield J, Tank D. 'Neural' computation of decisions in optimization problems. Biol Cybern. 1985;52:141–52.

22. Humphreys G, Müller HJ. SEarch via Recursive Rejection (SERR): a connectionist model of visual search. Cogn Psychol. 1993;25:43–110.

23. Jones, Andolina I, Oakely N, Murphy P, Sillito A. Spatial summation in lateral geniculate nucleus and visual cortex. Exp Brain Res. 2000;135:279–84.

24. Malinowski P, Häner R. The effect of familiarity on visual-search performance: evidence for learned basic features. Percept Psychophys. 2001;63:458–63.

25. McAdams CJ, Reid RC. Attention modulates the responses of simple cells in monkey primary visual cortex. J Neurosci. 2005;25(47):11,023–33.

26. McClelland J, Rumelhart D, Hinton G. The appeal of parallel distributed processing. In: Parrallel distributed processing: explorations in the microstructure of cognition, volume I: foundations. MIT Press/Bradford Books, Cambridge; 1986.

27. Meytlis M, Sirovich L. On the dimensionality of face space. IEEE Trans Pattern Anal Mach Intell. 2007;29(7):1262–7.

28. Mjolsness E, Garrett C. Algebraic transformations of objective functions. Neural Netw. 1990;3:651–69.

29. Mounts JRW. Evidence for suppressive mechanisms in attentional selection: Feature singletons produce inhibitory surrounds. Percept Psychophys. 2000;62(5):969–83.

30. Mozer M, Sitton M. Computational modeling of spatial attention. In: Pashler H (ed) Attention. London UCL Press, London, pp 341–93; 1998.

31. Muller H, Krummenacher J (eds). Visual search and attention, visual cognition, vol 14. Psychology Press, Essex; 2006.

32. Müller NG, Mollenhauer M, Rösler A, Kleinschmidt A. The attentional field has a Mexican hat distribution. Vision Res. 2005;45:1129–37.

33. Olshausen BA, Field DJ. Natural image statistics and efficient coding. Netw Comput Neural Syst. 1996;7:333–9.

34. Royden C, Wolfe J, Klempen N. Visual search asymmetries in motion and optioc flow fields. Percept Psychophys. 2001;63:436–44.

35. Schiller P. The on and off channels of the visual system. Trends Neurosci. 1992;15:86–92.

36. Treisman A. Features and objects: the 14th bartlett memorial lecture. Q J Exp Psychol. 1988;40A:201–37.

37. Treisman A, Gormican S. Feature analysis in early vision: evidence from search asymmetries. Psychol Rev. 1988;95(1):15–48.

38. Tsotsos JK, Culhane SM, Wai WYK, Lai Y, Davis N, Nuflo F. Modeling visual attention via selective tuning. Artif Intell. 1995;78:507–45.

39. Tsotsos JK, Rodriguez-Sanchez AJ, Rothenstein AL, Simine E. The different staes of visual recognition need different attentional binding strategies. Brain Res. 2008;1225:119–32.

40. Wolfe JM. Guided Search 2.0 A revised model of visual search. Psychomon Bull Rev. 1994;1(2):202–38.

41. Wolfe JM. Visual search: a review. In: Pashler H(ed) Attention. Psychology Press, pp 13–74; 1998a.

42. Wolfe JM. What can 1 million trials tell us about visual search. Psychol Sci. 1998b;9(1).

43. Wolfe JM. Asymmetries in visual search: an introduction. Percept Psychophys. 2001;63(3):381–9.