Check for updates

A Neurodynamic Model of Feature-Based Spatial Selection

Mateja Marić and Dražen Domijan*

Department of Psychology, Faculty of Humanities and Social Sciences, University of Rijeka, Rijeka, Croatia

Huang and Pashler (2007) suggested that feature-based attention creates a special form of spatial representation, which is termed a Boolean map. It partitions the visual scene into two distinct and complementary regions: selected and not selected. Here, we developed a model of a recurrent competitive network that is capable of state-dependent computation. It selects multiple winning locations based on a joint top-down cue. We augmented a model of the WTA circuit that is based on linear-threshold units with two computational elements: dendritic non-linearity that acts on the excitatory units and activity-dependent modulation of synaptic transmission between excitatory and inhibitory units. Computer simulations showed that the proposed model could create a Boolean map in response to a featured cue and elaborate it using the logical operations of intersection and union. In addition, it was shown that in the absence of top-down guidance, the model is sensitive to bottom-up cues such as saliency and abrupt visual onset.

Keywords: boolean map, feature-based attention, lateral inhibition, neural network, winner-take-all

INTRODUCTION

In the literature on visual attention, significant progress has been made in characterizing the principles of selection. Visual attention can be allocated flexibly to a circumscribed region of space, the whole object or feature dimensions such as color and orientation (Nobre and Kastner, 2014). Indeed, early work suggested that a restricted circular region of space is a representational format of attentional selection. Posner (1980) proposed that attention operates like a spotlight that highlights a single circular region of space with a fixed radius. All locations that fall inside the spotlight are selected, and everything outside is left out. An extension of this proposal, which is called the zoom-lens model, suggests that the spotlight of attention can change its radius depending on the spatial resolution that one wants to achieve (Eriksen and St. James, 1986). If high resolution is required, the spotlight can be narrowed to capture details in the selected region, whereas the radius of the spotlight can be widened when a lower resolution is sufficient.

Other studies point to an object as a unit of selection. Duncan (1984) showed that it is easier to report two attributes if they appear on the same object, relative to the scenario in which each attribute appears on a different object. This finding implies that the object is selected as a whole and has been replicated many times using different stimuli and behavioral paradigms (Scholl, 2001). This effect cannot be explained by spatial attention because objects were spatially superimposed, that is, they shared the same locations. More recently, it was shown that attention can also be allocated to a visual feature such as color or direction of motion independent of spatial location (Saenz et al., 2002, 2003). Single-unit recordings have shown that feature-based attention is accompanied by the global location-independent modulation of neural response in a range of areas in the visual cortex. Attentional modulation was described as a

multiplicative gain change that increases responses of neurons that are selective to attended feature values and decreases responses of neurons that are tuned to unattended feature values (Treue and Martinez-Trujillo, 1999; Martinez-Trujillo and Treue, 2004).

Object-based attention, however, is not necessarily detached from spatial representation. There is behavioral and neurophysiological evidence that object-based attention involves selection of all spatial locations that are occupied by the same object. Specifically, it was suggested that attention selects a grouped array of locations (O'Grady and Müller, 2000). In other words, attention spreads from one spatial location along the shape of the object and highlights all locations that belong to the object (Richard et al., 2008; Vatterott and Vecera, 2015). Neurophysiological studies showed that object-based selection is indeed achieved by the spreading of the enhanced firing rate along the shape of the object (Roelfsema, 2006; Roelfsema and de Lange, 2016).
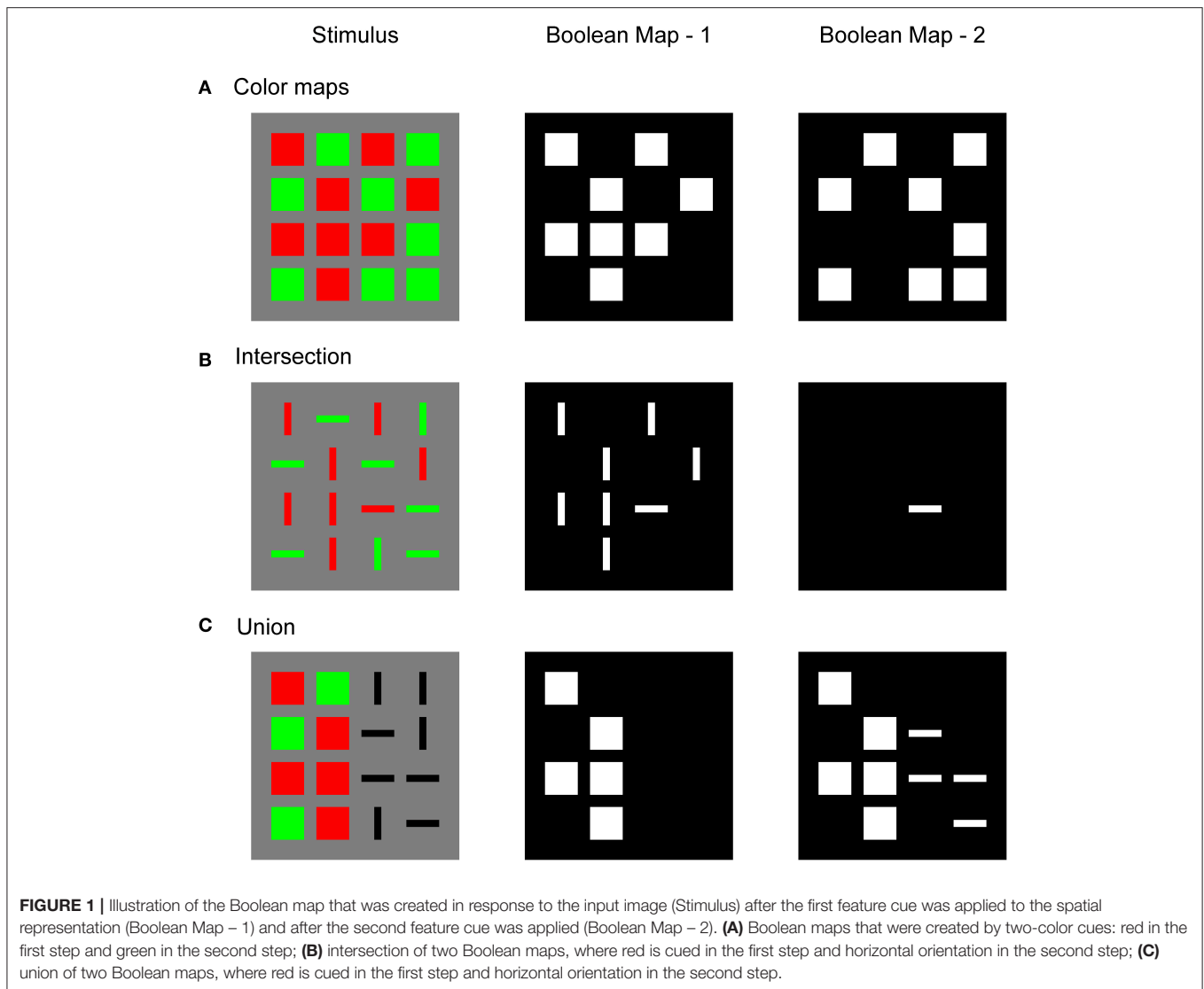
In a similar way, feature-based attention might involve the selection of all locations that are occupied by the same feature value, as shown by Huang and Pashler (2007). They proposed that attention is limited because it may access only one feature value (e.g., red) per dimension (e.g., color) at any given moment. However, the accessed feature value is bound to space in parallel, without capacity limits. Feature-based attention is allocated in space via the formation of a binary or Boolean map. When a conscious decision is made to attend to a specific feature value, the Boolean map indicates all spatial locations that are occupied by the chosen feature value because they are labeled by a positive value (e.g., 1), while all other locations are labeled with zero. In each selection process, selected locations need not be contiguous in space, but they must share the same feature value. After a Boolean map is formed, it is possible to operate on its output by applying the set operations of intersection and union. Recent work suggests that a spatial representation, such as a Boolean map, might mediate perceptual grouping by similarity (Huang, 2015; Yu and Franconeri, 2015). Moreover, the idea has been recently applied successfully in the computer vision literature on developing algorithms for saliency detection (Zhang and Sclaroff, 2016; Qi et al., 2017).

Figure 1 illustrates a Boolean map that is formed in response to three different stimulus configurations and sequential application of two top-down feature cues. Figure 1A shows a simple stimulus that consists of red and green squares. An observer might attempt to isolate only red or only green items. To do so, a top-down cue should be supplied to the feature map that encodes the desired feature value. For example, when attention is directed to the red color, the top-down cue highlights all locations that are occupied by red squares. The Boolean map picks up on this feature cue and forms a spatial representation in which cued locations are labeled with 1 (white) and non-cued locations are labeled with 0 (black). In terms of a neural network, these labels correspond to the active (excited) and inactive (inhibited) states of the corresponding nodes in the network (Boolean Map – 1). Later, the observer might wish to switch to green color (Boolean Map – 2). Again, in a response to a new feature cue, the Boolean map now shows all locations that are occupied by green squares.

Figure 1B shows a typical stimulus that is used in visual search experiments. It consists of red and green horizontal and vertical bars. The task is to find a red horizontal bar. This is an example of a conjunction search task in which two feature dimensions should be combined to find the target object. According to Huang and Pashler (2007), the conjunction task is solved in two steps. In the first step, a Boolean map is formed by top-down cueing of red items, irrespective of their orientations. In the second step, only horizontal items are cued. However, since red items have already been selected, the second Boolean map will correspond to the intersection of red and horizontal items. There is only one item that satisfies these selection criteria: the target. In this way, visual search is substantially faster compared to the strategy of sequentially visiting each item by moving the attentional spotlight across the visual field. It is also possible to reverse the order of the applied feature cues. In the first step, horizontal items might be cued, and the intersection is formed by highlighting red items in the second step. Importantly, there is behavioral evidence that observers indeed implement such a *subset selection* strategy in conjunction search tasks (Egeth et al., 1984; Kaptein et al., 1995). Moreover, Huang and Pashler (2012) showed that the same strategy is used in the perception of spatial structure in a stimulus that is composed of multiple items that differ in several dimensions.

Figure 1C illustrates an example of the union of two Boolean maps. As in the previous example, the observer starts by cueing red items and creating a Boolean map that consists of a representation of their locations. In the second step, the observer wishes to combine red with horizontal items. Therefore, in the second step, one should cue horizontal items but simultaneously maintain locations of the remaining items in memory. The resulting new Boolean map now represents the locations of all red and all horizontal items that were found in the image. Computing with Boolean maps might not be restricted to only two steps, as Figure 1 suggests. It is possible to incorporate more feature dimensions, such as motion, texture, or size, that can also be engaged in creating Boolean maps that are more complicated.

Feature-based spatial selection, as illustrated by the Boolean map, provides a strong constraint on the computational models of visual attention because it requires simultaneous selection of arbitrarily many locations based on an arbitrary criterion that is set by the observer. Computational models of attention often rely on a winner-take-all (WTA) network to select a single, most salient location from the input image (Itti and Koch, 2000, 2001). The WTA network consists of an array of excitatory nodes that are connected reciprocally with inhibitory interneurons. This anatomical arrangement creates lateral inhibition among excitatory nodes that lead to the selection of a single node that receives maximal input and the suppression of all other nodes, which receive non-maximal input. However, when faced with the input where multiple (potentially many) nodes share the same maximal input level, the typical WTA network tends to suppress all winning nodes due to a strong mutual inhibition among them instead of selecting them together. For example, Usher and Cohen (1999) showed that, under the conditions of strong recurrent excitation and weak lateral inhibition, the WTA network reaches a steady state with multiple active winners.

FIGURE 1 | Illustration of the Boolean map that was created in response to the input image (Stimulus) after the first feature cue was applied to the spatial representation (Boolean Map – 1) and after the second feature cue was applied (Boolean Map – 2). **(A)** Boolean maps that were created by two-color cues: red in the first step and green in the second step; **(B)** intersection of two Boolean maps, where red is cued in the first step and horizontal orientation in the second step; **(C)** union of two Boolean maps, where red is cued in the first step and horizontal orientation in the second step.

Importantly, activation of the winning nodes decreases linearly toward zero as their quantity increases. In other words, this network design suffers from the capacity limitation. This is a useful property in modeling short-term memory and frontal lobe function (Haarmann and Usher, 2001) but it is inadequate for understanding how the Boolean map might arise in a large retinotopic map, as exemplified by **Figure 1**.

Another problem is that the dynamics of the WTA network are not sensitive to transient changes in the input amplitude. Due to strong self-excitation and the resulting persistent activity, the WTA network settles into one of its memory states (fixed points). Importantly, each memory state is independent of later inputs. If self-excitation is weakened, the network will become sensitive to input. However, at the same time, it will lose its ability to form a memory state and will behave like a feedforward network (Rutishauser and Douglas, 2009). One way to solve this problem is to apply an external reset signal to the network before a new input is processed (Grossberg,

1980; Kaski and Kohonen, 1994; Itti and Koch, 2000, 2001). However, this is not sufficient in the context of feature-based attention. An intersection or union operation between two Boolean maps requires that the currently active memory state (formed after the first feature cue) be updated by taking into account new input (the second feature cue). Therefore, the dynamics of the WTA network should allow uninterrupted transition between memory states that are governed by external inputs. In other words, the WTA network should be capable of state-dependent computation (Rutishauser and Douglas, 2009).

To summarize, a WTA network that is capable of computing with Boolean maps should simultaneously satisfy two computational constraints:

1. It should be able to select together all locations that share a common feature value. This should be achieved without degrading the representation of the winners.
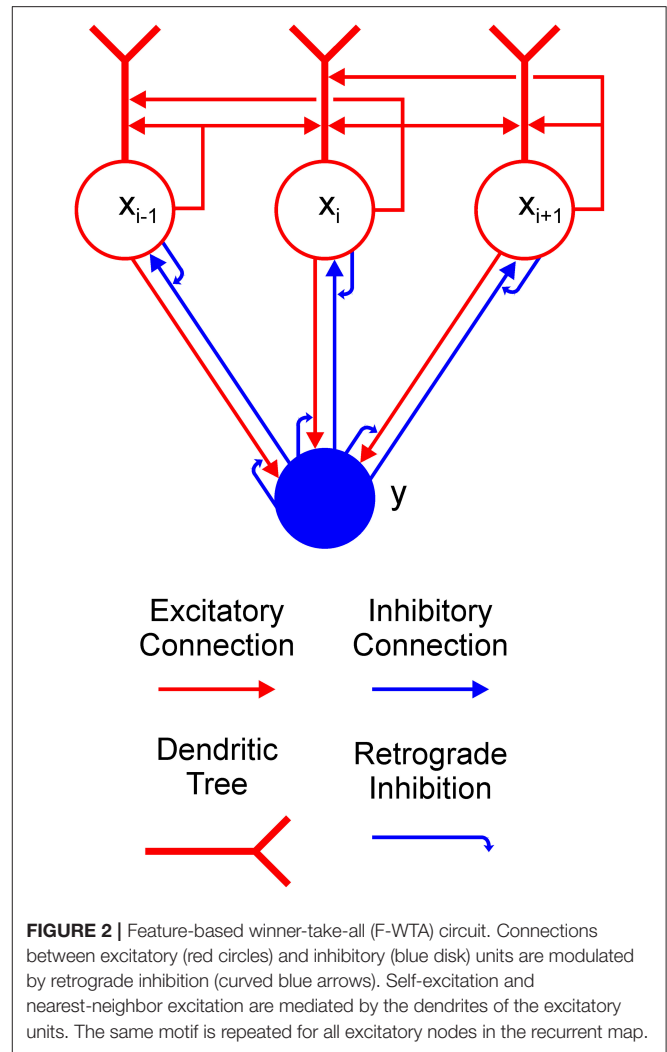
2. It should exhibit state-dependent computation, in which new inputs are combined with the current memory state to produce a new resultant state (e.g., intersection or union).

Here, we have developed a new WTA network that satisfies these constraints and provides the neural implementation of the Boolean map theory of attention (Huang and Pashler, 2007).
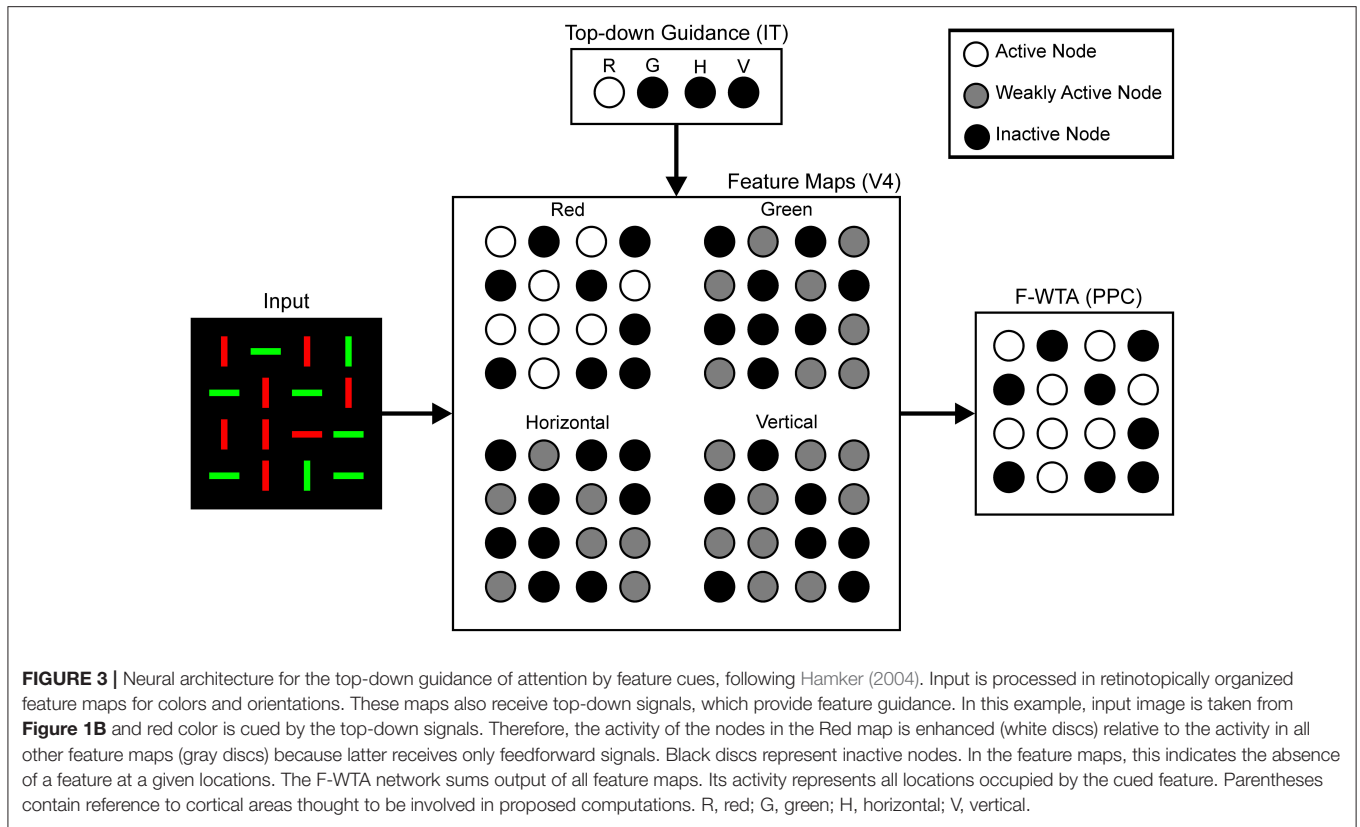
MODEL DESCRIPTION

The aim of the current work is to provide an explanation of how a Boolean map may be formed in a recurrent competitive network that can implement feature-based winner-take-all (F-WTA) selection. To this end, we have extended the previously proposed network model based on the linear-threshold units (Hahnloser, 1998; Hahnloser et al., 2003; Rutishauser and Douglas, 2009). Concretely, the model circuit is presented in **Figure 2**. It consists of a single inhibitory unit, which is reciprocally connected to a group of excitatory units. In addition to these basic elements, we introduce two processing components into the WTA circuit to expand its computational power. The first is a dendritic non-linearity, which prevents excessive excitation that arises from self-recurrent and nearest-neighbor collaterals. We modeled the dendritic tree as a separate electrical compartment with its own non-linear output that is supplied to the node's body (Häusser and Mel, 2003; London and Häusser, 2005; Branco and Häusser, 2010; Mel, 2016). The second is modulation of synaptic transmission by retrograde inhibitory signaling (Tao and Poo, 2001; Alger, 2002; Zilberter et al., 2005; Regehr et al., 2009). This is a form of presynaptic inhibition, where postsynaptic cells release a neurotransmitter that binds to the receptors that are located on the presynaptic terminals. Retrograde signaling creates a feedback loop that dynamically regulates the amount of transmitter that is released from the presynaptic terminals. Here, we have hypothesized that such interactions occur in recurrent pathways from the excitatory nodes to the inhibitory interneuron and back from the interneuron to the excitatory nodes. In the excitatory-to-inhibitory pathway, retrograde signaling enables the inhibitory interneuron to compute the maximum instead of the sum of its inputs. Computation of the maximum arises from the limitation that the activity of the inhibitory interneuron cannot grow beyond the maximal input that it receives from the excitatory nodes. Furthermore, retrograde signaling in the inhibitory-to-excitatory pathway enables the excitatory nodes that receive maximal input to protect themselves from the common inhibition. In this way, the network can select all excitatory nodes with maximal input, irrespective of their quantity or arrangement in visual space.

At first sight, it might appear strange to propose that an excitatory unit can inhibit its input by releasing a neurotransmitter that binds to the presynaptic terminal. However, several signaling molecules have been identified to support such interactions, including endogenous cannabinoids (Alger, 2002). Moreover, Zilberter (2000) found that glutamate is released from dendrites of pyramidal neurons in the rat neocortex and suppresses the inhibition that impinges on them. In addition, similar action has been found for GABA (Zilberter

FIGURE 2 | Feature-based winner-take-all (F-WTA) circuit. Connections between excitatory (red circles) and inhibitory (blue disk) units are modulated by retrograde inhibition (curved blue arrows). Self-excitation and nearest-neighbor excitation are mediated by the dendrites of the excitatory units. The same motif is repeated for all excitatory nodes in the recurrent map.

et al., 1999), which suggests that conventional neurotransmitters can engage in retrograde signaling.

To situate the proposed F-WTA circuit in a larger neural architecture that describes the cortical computations that underlie top-down attentional control, we have adopted the model that was proposed by Hamker (2004). He showed how attentional selection of a target arises from the recurrent interactions within a distributed network that consists of model cortical area V4, the inferotemporal cortex (IT), the posterior parietal cortex (PPC), and the frontal eye fields (FEF). **Figure 3** illustrates part of these interactions that are involved in feature-based attentional guidance. Top-down signals that provide feature cues originate in the IT, which contains a spatially invariant representation of relevant visual features. The IT sends feature-specific feedback projections to the V4, where topographically organized feature maps for each feature value are located. For simplicity, we consider only maps for two colors (red and green), and two orientations (vertical and horizontal). We do not explicitly model IT and V4 dynamics. Rather, they serve here as a tentative explanation of how input to the F-WTA network

FIGURE 3 | Neural architecture for the top-down guidance of attention by feature cues, following Hamker (2004). Input is processed in retinotopically organized feature maps for colors and orientations. These maps also receive top-down signals, which provide feature guidance. In this example, input image is taken from **Figure 1B** and red color is cued by the top-down signals. Therefore, the activity of the nodes in the Red map is enhanced (white discs) relative to the activity in all other feature maps (gray discs) because latter receives only feedforward signals. Black discs represent inactive nodes. In the feature maps, this indicates the absence of a feature at a given locations. The F-WTA network sums output of all feature maps. Its activity represents all locations occupied by the cued feature. Parentheses contain reference to cortical areas thought to be involved in proposed computations. R, red; G, green; H, horizontal; V, vertical.

arises within the ventral visual pathway. Also, we omitted the contribution of the FEF and its spatial reentry signals to the V4 activity.

We hypothesize that the feature-based WTA network resides in the PPC, where it receives summed input over all feature maps from the V4. Top-down guidance is implemented by a temporary increase in activity in one of the V4 feature maps. For example, when the decision is made to attend to the red color, the IT representation of red color sends feedback signals to the Red Map in the V4. Top-down signals to the feature map are modeled as a multiplicative gain of neural activity, which is consistent with neurophysiological findings (Treue and Martinez-Trujillo, 1999; Martinez-Trujillo and Treue, 2004; Maunsell and Treue, 2006).

The following neural network equations represent the quantitative description of the model. Each unit is defined by its instantaneous firing rate (Dayan and Abbott, 2000). The time evolution of the activity of excitatory node x at position i in the recurrent map is given by the following differential equation:

$$\tau_x \frac{dx_i}{dt} + x_i = \left[I_i(t) + \alpha f(x_i + x_{i+1} + x_{i-1}) - \beta_1 g(y - x_i - T_y) \right]^+.$$

$$(1)$$

The time evolution of the activity of inhibitory interneuron y is given by

$$\tau_y \frac{dy}{dt} + y = \left[\beta_2 \sum_i g(x_i - y - T_x) \right]^+.$$

$$(2)$$

Parameters τ_x and τ_y are integration time constants for excitatory and inhibitory nodes, respectively. We assume that inequality $\tau_x > \tau_y$ holds, which accords with the observation in electrophysiological measurements that inhibitory cells exhibit faster dynamics than excitatory cells (McCormick et al., 1985). The second term on the left-hand side of Equations (1) and (2) describes the passive decay that drives the unit's activity to the resting state in the absence of external input. Firing rate activation function $[u]^+$ is a non-saturating rectification nonlinearity, which is defined by

$$[u]^+ = \max(u, 0).$$

$$(3)$$

Following Hamker (2004), we assume that feedforward input I_i at time t to the excitatory node x_i in the F-WTA network is given by the sum over activity in all V4 feature maps $I_i^{(m)}$,

$$I_i(t) = \sum_m I_i^{(m)} G^{(m)}(t).$$

$$(4)$$

In Equation (4), m denotes available feature maps with $m \in \{red, green\}$ in the simulation that is reported in section Simulation of the Formation of a Single Boolean Map and $m \in \{red, green, horizontal, vertical\}$ in the simulation that is reported in section Simulation of the Intersection and Union of Two Boolean Maps. Parameter G^m refers to the feature-specific, global multiplicative gain that all units $I_i^{(m)}$ within the same feature map m receive via top-down projections. As shown in **Figure 2**,

these projections arrive from the feature representation in the IT. Multiplicative gating is generally consistent with previous models that describe the effect of feature-based attention on the responses of neurons in the early visual cortex (Boynton, 2005, 2009). Equation (4) ensures that the F-WTA network is not particularly sensitive to any feature value. Rather, it signals the behavioral relevance of locations in a spatial map. Here, the relevance can be set according to differences in the bottom-up input $I_i^{(m)}$ that arise from competitive interactions in the early visual cortex. Alternatively, relevance can be signaled by the top-down feature cues G^m that change the gain of all locations that are occupied by the same feature value.

Dendritic output $f(u)$ is described by the sigmoid response function

$$f(u) = \frac{S_d}{1 + e^{-\lambda(u - T_d)}} \quad (5)$$

where λ and T_d control the shape of the sigmoid function and S_d is its upper asymptotic value. We set λ to a high value to achieve a steep rise of the dendritic activity immediately after its input crosses the dendritic threshold, which is denoted as T_d. Such strong non-linearity is justified by experimental data, which show all-or-none behavior in real dendrites (Wei et al., 2001; Polsky et al., 2004). In Equation (1), parameter α controls the strength of the impact that the dendritic compartment exerts on the soma.

Self-recurrent x_i and nearest-neighbor collaterals x_{i-1} and x_{i+1} arrive on the dendrite of the excitatory node, which is consistent with the anatomical observation that most recurrent excitatory connections are made on the dendrites of the excitatory cells (Spruston, 2008). Nodes at the edge of the network receive excitation only from a single available neighbor. That is, node x_1 receives excitation only from x_2, and x_N receives excitation only from x_{N-1}. Nearest-neighbor excitatory interactions enable feature cues to spread activity enhancement automatically to all connected locations that contain a given feature value. This is not essential for the simulation of Boolean maps but we included it in our model because recurrent connections among nearby neurons are prominent feature of the synaptic organization of the cortex (Douglas and Martin, 2004). Also, we wanted to show that the proposed model is capable of simulating object-based attention (Roelfsema, 2006; Roelfsema and de Lange, 2016). Moreover, Wannig et al. (2011) found direct evidence for activity spreading among neurons that encode the same feature value in the primary visual cortex.

The output of the presynaptic interactions $g(u)$ is defined by the rectification non-linearity of the form

$$g(u) = [u]^+ = \max(u, 0). \quad (6)$$

In Equation (1), the term $-g(y - x_i - T_y)$ describes the output of the presynaptic terminal that delivers inhibition from interneuron y to excitatory node x_i (**Figure 4A**). However, we did not explicitly model the dynamics of retrograde signaling. We assumed that the release of the retrograde transmitter occurs simultaneously with the activation of the postsynaptic node and that it is proportional to its firing rate. Therefore, it is represented by the term $-x_i$.

FIGURE 4 | Retrograde inhibitory signaling (blue curved arrows) from excitatory node x_i to the presynaptic terminal of inhibitory interneuron y **(A)** and from the inhibitory interneuron to the presynaptic terminal of the excitatory node **(B)**. Both terminals compute half-wave rectification $g(u)$ of their input. Terminals release respective inhibitory **(A)** or excitatory **(B)** neurotransmitter (straight horizontal arrows) only when they receive net positive input.

Function $g(u)$ ensures that the presynaptic terminal will release the inhibitory transmitter only when the electrical signal from node y exceeds the inhibitory retrograde signal $-x_i$ and the threshold for presynaptic activation, which is denoted as T_y. In other words, node x_i will be inhibited only if $y > x_i + T_x$. If this is not the case, node x_i will effectively isolate itself from the inhibitory influence of node y. This is always the case for the winning node because $x(t) > y(t)$ for $t > 0$. Moreover, this result extends to all other nodes whose input magnitude is sufficiently close to the maximal input. The strength of the inhibition is determined by parameter β_1. In a similar vein, in Equation (2), the term $-g(x_i - y - T_x)$ describes the action of the retrograde signal that is released from inhibitory interneuron y on the presynaptic terminal that delivers excitation from node x_i (**Figure 4B**). Here, parameter T_x describes the threshold for the

activation of the presynaptic terminal of the excitatory node and β_2 determines the strength of the excitation.

We have proposed a model of a one-dimensional network, although it attempts to simulate phenomena that occur in 2-D, as illustrated by **Figure 1**. We have chosen to work with the 1-D version of the network simply because we want to focus on the analysis of its temporal dynamics and its ability to combine information over time. Without loss of generality, the computer simulations that are reported in section Computer Simulations should be considered as a cross-section of a 2-D network.

For simplicity, the thresholds that control the activation of the excitatory and inhibitory nodes are all set to zero and are omitted from the model description. Parameters were set as follows: $\tau_x = 5$; $\tau_y = 2$; $\alpha = 1$; $\beta_1 = 1$; $\beta_2 = 10$; $S_d = 1$; $\lambda = 100$; $T_d = 0.1$; $T_x = 0.1$; and $T_y = 0.1$. Parameters were chosen in a way to simultaneously achieve intersection and union. Systematic variations on the parameters α, β_1 and β_2 showed that intersection is observed when $1 \leq (\alpha, \beta_1) \leq 5$. In contrast, union is observed when $0.8 \leq (\alpha, \beta_1) \leq 1$. Parameter β_2 can be set to any value above the default without changing the results.

MODEL EXTENSIONS

The network that is defined by Equations (1) and (2) is chosen in a way that achieves the desired behavior with the minimal number of computational elements. This simplicity heuristic is important for understanding model properties without adding extra neuroscientific complexity (Ashby and Hélie, 2011). However, at the same time, this approach sacrifices anatomical and biophysical plausibility of the proposed model. In this section, we present several extensions and generalizations of the basic model that bring it closer to satisfying the neurobiological constraints.

Inhibitory Pool

The model has just one inhibitory interneuron for computational convenience, which is not realistic. It is known that excitatory neurons outnumber inhibitory neurons by a factor of four in the cortex (Braitenberg and Schüz, 1991). However, it is possible to design an F-WTA network with a pool of inhibitory interneurons and the appropriate ratio between excitatory and inhibitory nodes that achieves the same behavior as the original model. An extended F-WTA network is presented in **Figure 5A**. Here, each inhibitory interneuron receives input from a subset of the excitatory nodes. We depicted each excitatory subset as a vertical arrangement of four nodes that do not overlap in their projections to the inhibitory pool. Therefore, each excitatory node projects to just one inhibitory node. Naturally, this does not need to be the case. It is possible that each excitatory node projects to more than one node without compromising the network output. Importantly, all inhibitory interneurons are mutually connected. In addition, each inhibitory interneuron projects its output to all excitatory nodes (denoted by thick blue arrow). As in the original model, we assume that all

inhibitory and excitatory nodes are endowed with the capability of retrograde signaling on their synaptic contacts.

Within the pool of inhibitory nodes, retrograde signaling enables computation of the MAX function, as in the original model. To see this, consider the inhibitory node that receives maximal input. Due to the retrograde signaling, it will reach a steady state that corresponds to the computation of the MAX function over input from its excitatory subset. Moreover, it will not receive inhibition from the other members of the pool. All other inhibitory nodes, which receive less excitatory support, will be silenced because their retrograde signaling is not sufficiently strong to prevent lateral inhibition from the winning node. However, if there are multiple inhibitory nodes with the same level of activity, they will remain active together. Finally, the winning nodes send inhibition to all excitatory subsets. Since excitatory nodes also engage in retrograde signaling, the nodes that receive maximal input will block inhibition and remain active. Therefore, the network output will look much like the original model because the MAX computation on the inhibitory nodes makes irrelevant the number of them that are active simultaneously.

Localized Inhibition

An important shortcoming of the previous model is that it assumes that inhibitory projections extend across the whole network of excitatory units. This is clearly not the case in real neural networks, where the spatial spread of inhibition is limited. To account for this property, we have constructed a more elaborate version of the basic model, which is shown in **Figure 5B**. It contains a new pool z_j of excitatory nodes with long-range projections. The z_j nodes receive input from the subset of the x_i nodes. Additionally, each z_j node sends its projection to at least one y_j node from the pool of inhibitory nodes. The number of z nodes must equal the number of inhibitory nodes y_j so that they can be indexed by the same subscript j. Again, we assume that the z_j nodes are equipped with the ability of retrograde signaling on their synapses. Therefore, they also compute the MAX function over all their inputs, including feedforward input from the corresponding subset of x_i nodes and recurrent input from other z_j nodes. In this design, the maximum level of activity that is sensed by the x_i nodes in one part of the network is easily propagated via z_j nodes to all other parts of the network. Furthermore, z_j nodes transfer this activity to inhibitory nodes. Therefore, each inhibitory node will eventually receive the maximal level of activity and apply it to the subset of x_i nodes to which it is connected. In this design, it is not necessary for inhibitory nodes to interact with one another. The excitatory nodes x_i that receive maximal input will block inhibition by their retrograde signaling and remain active in the same manner as described in the previous section. In this way, the proposed circuit achieves the same result as the original model.

Output Functions

The model employs threshold-linear output functions for the soma and the logistic sigmoid function for dendrites. This is inconsistent with the observation that somatic output also saturates and is also often modeled by the sigmoid function.
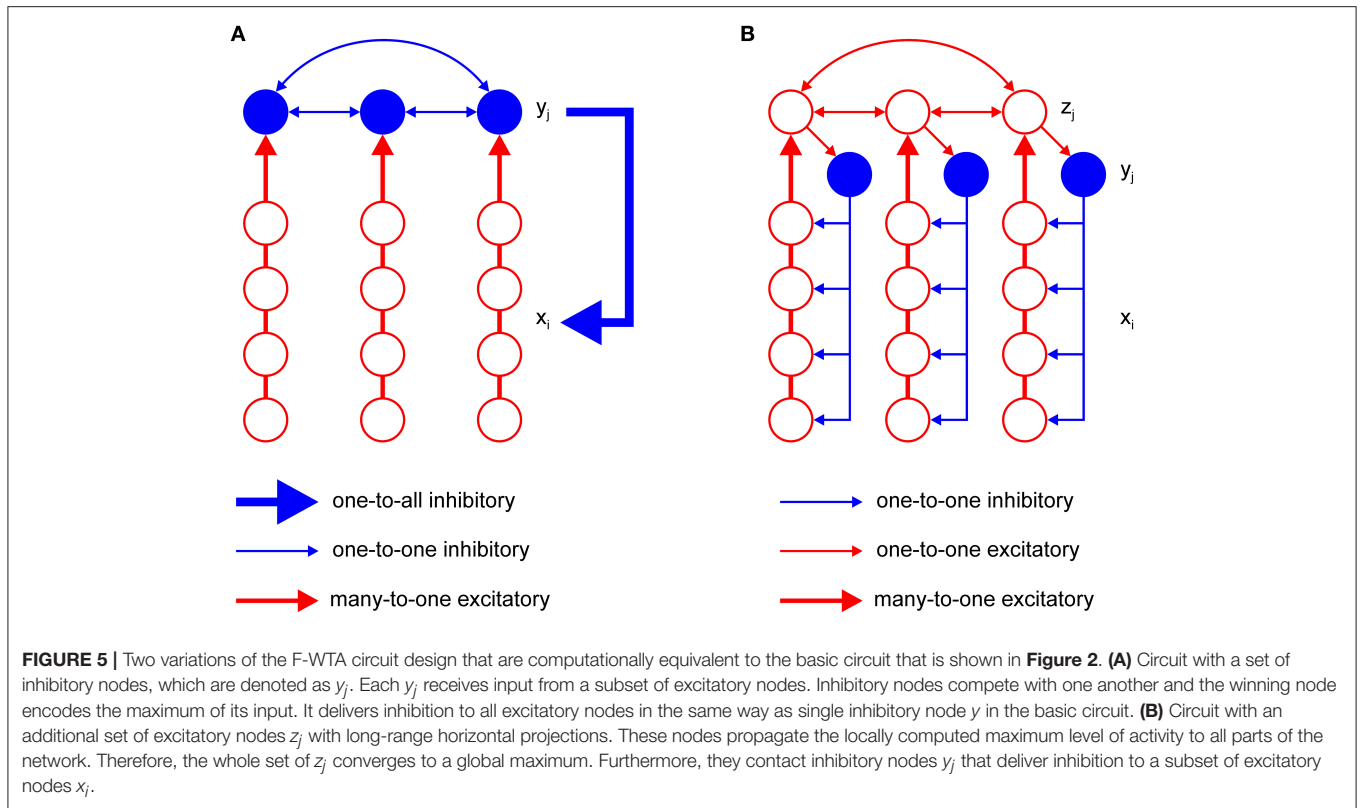
FIGURE 5 | Two variations of the F-WTA circuit design that are computationally equivalent to the basic circuit that is shown in **Figure 2**. **(A)** Circuit with a set of inhibitory nodes, which are denoted as y_j. Each y_j receives input from a subset of excitatory nodes. Inhibitory nodes compete with one another and the winning node encodes the maximum of its input. It delivers inhibition to all excitatory nodes in the same way as single inhibitory node y in the basic circuit. **(B)** Circuit with an additional set of excitatory nodes z_j with long-range horizontal projections. These nodes propagate the locally computed maximum level of activity to all parts of the network. Therefore, the whole set of z_j converges to a global maximum. Furthermore, they contact inhibitory nodes y_j that deliver inhibition to a subset of excitatory nodes x_i.

However, in normal circumstances, neurons operate in a linear mode that is far from their saturation level (Rutishauser and Douglas, 2009). To provide a more systematic approach to the output functions that are used in the model, we introduce a piecewise-linear approximation to the sigmoid function $s_q(u)$ of the form

$$s_q(u) = \begin{cases} 0 & if & u \leq 0 \\ u & if & 0 < u < S_q \\ S_q & if & u \geq S_q \end{cases} \quad (7)$$

where S_q denotes the upper saturation point, which can be set differently for different computational units $q \in \{c, d, p\}$, which correspond to the somatic, dendritic, and presynaptic terminal outputs, respectively. With the output function $s_q(u)$ applied to all computational elements of a single node, the model equations, namely, Equations (1) and (2), can be restated as

$$\tau_x \frac{dx_i}{dt} + x_i = s_c \left[I_i(t) + \alpha s_d(x_i + x_{i+1} + x_{i-1} - T_d) \right. $$
$$\left. - \beta_1 s_p(y - x_i - T_y) \right] \quad (8)$$

and

$$\tau_y \frac{dy}{dt} + y = s_c \left[\beta_2 \sum_i s_p(x_i - y - T_x) \right]. \quad (9)$$

An important constraint of the model that is defined by Equations (8) and (9) is that saturation point for the dendritic

output S_d should be chosen to be smaller than S_c, which is the saturation point of the somatic output. In this way, feedforward input I_i can be combined with the dendritic output without causing saturation at the output of the node. In contrast, if dendrites are allowed to saturate at the same activity level as the node, the dendritic output will overshadow the feedforward input. Consequently, the network will lose its sensitivity to the input changes. This is undesirable with respect to the requirements that are imposed by the sequential formation of the multiple Boolean maps. Therefore, the choice between the linear or the sigmoid output function for the node is not important if the dendritic output is restricted to a smaller interval relative to the output of the node itself.

LINEAR STABILITY ANALYSIS

Fixed Points

Fixed point is found iteratively starting from the set of nodes receiving maximal input, x_M. We assume that the winning nodes and inhibitory interneuron are activated above their thresholds, so we set $[u]^+ = u$. Next, we observe that the winning nodes do not receive inhibition from the interneuron y since $x_M(t) > y(t)$ for $t > 0$. This holds because the activity of the inhibitory node is bounded above by $x_M + T_x > y$ where T_x is a positive constant. Then, retrograde signaling ensures that $g(y - x_M - T_y) = 0$ for all times t. Consequently, nodes receiving maximal input are driven solely by excitatory terms. Since the recurrent excitation is bounded above by its asymptotic value S_d, dendritic output

function $f(u)$ in Equation (1) is replaced with S_d. This yields the following approximation to the steady state of the winning nodes:

$$x_M \approx I_M + \alpha S_d. \tag{10}$$

After the x_M, inhibitory interneuron y also reaches its steady state because its activity is driven primarily by the input from x_M. As the activity of y grows, terms $g(x_i - y - T_x)$ in Equation (2) vanish for all nodes that do not receive maximal input x_i where $i \notin M$. In contrast, the presynaptic terminals of x_M are above the threshold for their activation just before y reaches equilibrium, that is, $g(x_M - y - T_x) > 0$. Therefore, the output function of the presynaptic terminal $g(u)$ can be replaced by u. Then, Equation (2) is solved as

$$y = \frac{\beta_2 k (x_M - T_x)}{\beta_2 k + 1}, \tag{11}$$

where k is the number of x_M. When β_2 is chosen to be sufficiently large, and/or there are many nodes with maximal input x_M, then

$$y \to x_M - T_x. \tag{12}$$

Continuity of the function defined by Equation (2) implies that y cannot grow above $x_M - T_x$, that is, $y(t) > x_M(t) - T_x$ cannot hold at any time t unless $y(t_0) = x_M(t_0) - T_x$ at some earlier time $t_0 < t$. However, equality $y(t_0) = x_M(t_0) - T_x$ implies that $dy/dt = 0$ at time t_0 because $g(x_M(t_0) - y(t_0) - T_x) = 0$. In other words, node y loses all its excitatory drive when it reaches $x_M - T_x$. This is true irrespective of the number k of x_M. Thus, node y computes the maximum over its input.

The x_M nodes, together with the inhibitory node, create a quenching threshold (QT) for the network, which is defined by

$$QT = y - T_y = x_M - T_x - T_y. \tag{13}$$

Grossberg (1973) introduced the concept of the quenching threshold to describe the property of contrast enhancement in recurrent competitive networks. Nodes whose activity is above QT are enhanced and stored in the memory state, while all nodes whose activity is below QT are suppressed and removed from the memory representation. In the same manner, the remaining excitatory nodes converge to one of two states, depending on whether they exceed QT or not:

$$x_{i \notin M} \approx \begin{cases} I_i + \alpha S_d & \text{if } x_i \geq QT \\ 0 & \text{if } x_i < QT. \end{cases} \tag{14}$$

QT and its relationship with the activity of the winning and non-winning nodes and inhibitory interneuron is illustrated in **Figure 6**. According to Equations (10), (11), and (14), the fixed-point linearly combines input and recurrent excitation. As maximal input increases or decreases, the fixed point will move up or down and track these changes. Moreover, the input may cease, and the winning nodes will settle into the activity level that is provided by the recurrent excitation alone, which is expressed as αS_d. In other words, the network remembers who the last winner was. The same is true in the case where the winner is determined by transient cues that are applied sequentially on a sustained input. This is a protocol that is used in the computer simulations that are reported in section Computer Simulations.
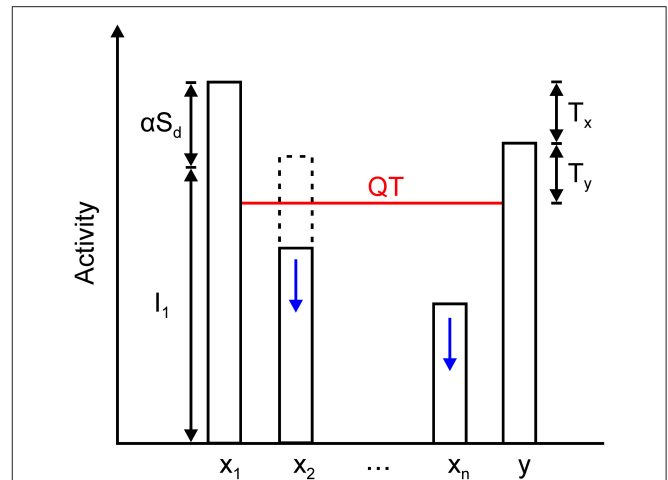

FIGURE 6 | Relationship among the steady state of the winning node x_1, inhibitory node y, and all other excitatory nodes in the network, $x_2 \ldots x_n$. The activity of the winning node is given by the sum of its feedforward input I_1 and the output of its dendrite mediating self- and nearest neighbor excitation, which is expressed as αS_d. Inhibitory node y approximately converges to $x_1 - T_x$. It sets the quenching threshold (QT) that separates excitatory nodes into two sets. Nodes $x_2 \ldots x_n$ are spared from inhibition if their activity is above the QT (dashed line); otherwise, they are silenced to zero (solid line). QT equals $y - T_y$ (or $x_1 - T_x - T_y$) because the activity of the inhibitory node must exceed the threshold on its presynaptic terminals that contact the excitatory nodes.

Linearization Near Fixed Points

To simplify the stability analysis, we consider an F-WTA network with two excitatory nodes and one inhibitory node: $[x_1, x_2, y]$. This system has three fixed points: x_1 is the only winner, x_2 is the only winner, and both excitatory nodes are winners. To which fixed point the network will converge depends on the relationship between inputs I_1 and I_2.

Local stability of the fixed point is estimated from the eigenvalues of the Jacobian matrix, which is the matrix of partial derivatives of the system of equations. If the real parts of all eigenvalues of the Jacobian are negative, the fixed point will be asymptotically stable (Rutishauser and Douglas, 2009). However, before we can compute the Jacobian matrix, we note that a linear-threshold function is continuous, but not differentiable. To sidestep this problem, we follow the approach that was described by Rutishauser et al. (2011) of inserting dummy terms that correspond to the derivate. That is, we need three separate dummy terms: c_i and p_{xi}, which correspond to the somatic and presynaptic output functions of excitatory node i, and a set of p_{yi} dummy terms that describe the presynaptic output function of inhibitory node y. The dummy terms are defined as.

$$c_i = p_{xi} = p_{yi} = \frac{d}{du}[u_i(t)]^+ = \begin{cases} 0 & \text{if } u_i(t) \leq 0 \\ 1 & \text{if } u_i(t) > 0. \end{cases} \tag{15}$$

Based on the above definition of the dummy terms, we have constructed the Jacobian matrix of

the system that consists of Equations (1) and (2):

$$J = \begin{bmatrix} \tau_x^{-1}\left(c_1\left(\alpha D_1 f + \beta_1 p_{y1}\right) - 1\right) & \tau_x^{-1} c_1 \alpha D_2 f & \tau_x^{-1} c_1 \beta_1 p_{y1} \\ \tau_x^{-1} c_2 \alpha D_1 f & \tau_x^{-1}\left(c_2\left(\alpha D_2 f + \beta_1 p_{y2}\right) - 1\right) & \tau_x^{-1} c_2 \beta_1 p_{y2} \\ \tau_y^{-1} \beta_2 p_{x1} & \tau_y^{-1} \beta_2 p_{x2} & -\tau_y^{-1}\left(\beta_2\left(p_{x1} + p_{x2}\right) - 1\right) \end{bmatrix} \quad (16)$$

where $D_1 f$ and $D_2 f$ denote the partial derivatives of the sigmoid function with respect to x_1 and x_2. Now, we examine the Jacobian matrix at the three fixed points that are mentioned above. If x_1 is the only winner, then $c_1 = 1$. However, $D_{x1} f \approx 0$ because the recurrent excitation of the winning node approaches its asymptotic value, which is S_d. In addition, $p_{y1} = 0$ because the winning node blocks inhibition from node y, as discussed above. Node x_2 is inhibited below its somatic threshold, that is, $c_2 = 0$. Presynaptic signaling by inhibitory node y blocks excitation from x_1 and x_2 is inactive, so $p_{x1} = p_{x2} = 0$. Consequently, the Jacobian matrix at the fixed point reduces to a diagonal matrix of the form

$$J_{W1} = J_{W2} = J_{W12} = \begin{bmatrix} -\tau_x^{-1} & 0 & 0 \\ 0 & -\tau_x^{-1} & 0 \\ 0 & 0 & -\tau_y^{-1} \end{bmatrix}. \quad (17)$$

All eigenvalues of the J_{W1} are negative, and the fixed point is asymptotically stable. In the case when x_2 is the sole winner, the same arguments are applied to set the dummy terms, thereby leading to the same diagonal matrix J_{W2} as shown in Equation (17). Moreover, if both excitatory nodes are winners, then $c_1 = c_2 = 1$, $D_{x1} f = D_{x2} f \approx 0$ and $p_{x1} = p_{x2} = 0$. Again, the Jacobian matrix J_{W12} is diagonal. Thus, all three fixed points are asymptotically stable.

The same analysis can be generalized to a network of arbitrary size and arbitrarily many fixed points. Retrograde signaling and dendritic saturation will ensure that the Jacobian matrix of any size will be diagonal and that the network dynamics will be independent of the network parameters, namely, α, β_1, and β_2. Local stability analysis suggests that the system behaves much like a feedforward network that is driven by the input. However, an important difference is that the F-WTA network has memory states like the recurrent network (Usher and Cohen, 1999; Rutishauser and Douglas, 2009).

COMPUTER SIMULATIONS

We performed a set of computer simulations to illustrate the model behavior. We employed a vector of 200 excitatory units and one inhibitory unit. Differential Equations (1) and (2) were solved numerically using MATLAB's *ode15s* solver. The simulations were run for 250 time steps. In subsequent figures, we followed the convention that activity of the node at position i as a function of time is depicted by a shade of gray, with white representing the maximal value and black representing zero.

Simulation of the Formation of a Single Boolean Map

First, we demonstrate how a Boolean map arises in the F-WTA network in response to the presentation of the color cue, as illustrated by **Figure 1A**. In **Figure 7A**, we recreate a similar stimulus condition in the 1-D map. The input consists of red and green items of equal sizes, which are intermixed in space on a black background. Input magnitude I was set to 1 in both maps and to 0.2 in the empty space around items to represent spontaneous activity in the absence of visual stimulation. Initially, the top-down or attentional gain is set to $G^m = 1$ in both feature maps $m \in \{red, green\}$. At $t = 50$, the red color is attended, which is reflected in the input to the network by increasing the gain for all nodes in the Red map ($G^{red} = 2$) and simultaneously reducing the gain in the Green map by the same factor ($G^{green} = 1/G^{red} = 1/2$). Top-down gain is also applied to the empty space between items, which is consistent with the finding that feature-based attention spreads across the whole visual field (Saenz et al., 2002, 2003; Serences and Boynton, 2007). The duration of the top-down cue is 50 simulated time steps. For simplicity, top-down signals are suddenly switched on and off without exponential decay. At $t = 150$, the green color is cued in the same way.

At the beginning of the simulation, before the top-down signals are applied, the F-WTA network simply selects all presented items together, irrespective of their color. Next, when the red color is cued by applying top-down signals to the corresponding feature map, the network responds to the new input by selectively increasing and sustaining the activity of nodes that encode locations of red items in the input and suppressing locations that encode green items. That is, the network creates a Boolean map by highlighting the spatial pattern that is associated with the red color. Furthermore, due to a self-excitation, the network maintains locations of the cued feature value in working memory after the top-down signals cease to influence the feature map. When the observer decides to switch attention to another feature value, the network can select the locations of the new feature value and suppress the locations that are associated with the previously cued value without requiring an external reset. Namely, the network is sensitive to input changes even though it also exhibits activity persistence.

Importantly, the activity level at selected locations is invariant with respect to the number of active nodes. At the beginning of the simulation, the number of active nodes was four times larger than after the cue was delivered. However, the active nodes remained at the same activity level as they were at the beginning of the simulation. This is a consequence of retrograde inhibitory signaling in recurrent pathways. It prevents unbounded growth of inhibition due to the dynamic regulation of its strength. To illustrate this point further, we run another simulation with items that are almost double in size (**Figure 7B**). Even though the total size of the cued items is increased, the activity of the cued nodes converges to the same level as before. In this simulation, we also checked that the network successfully operates even if we remove gain reduction from the non-attended feature map.
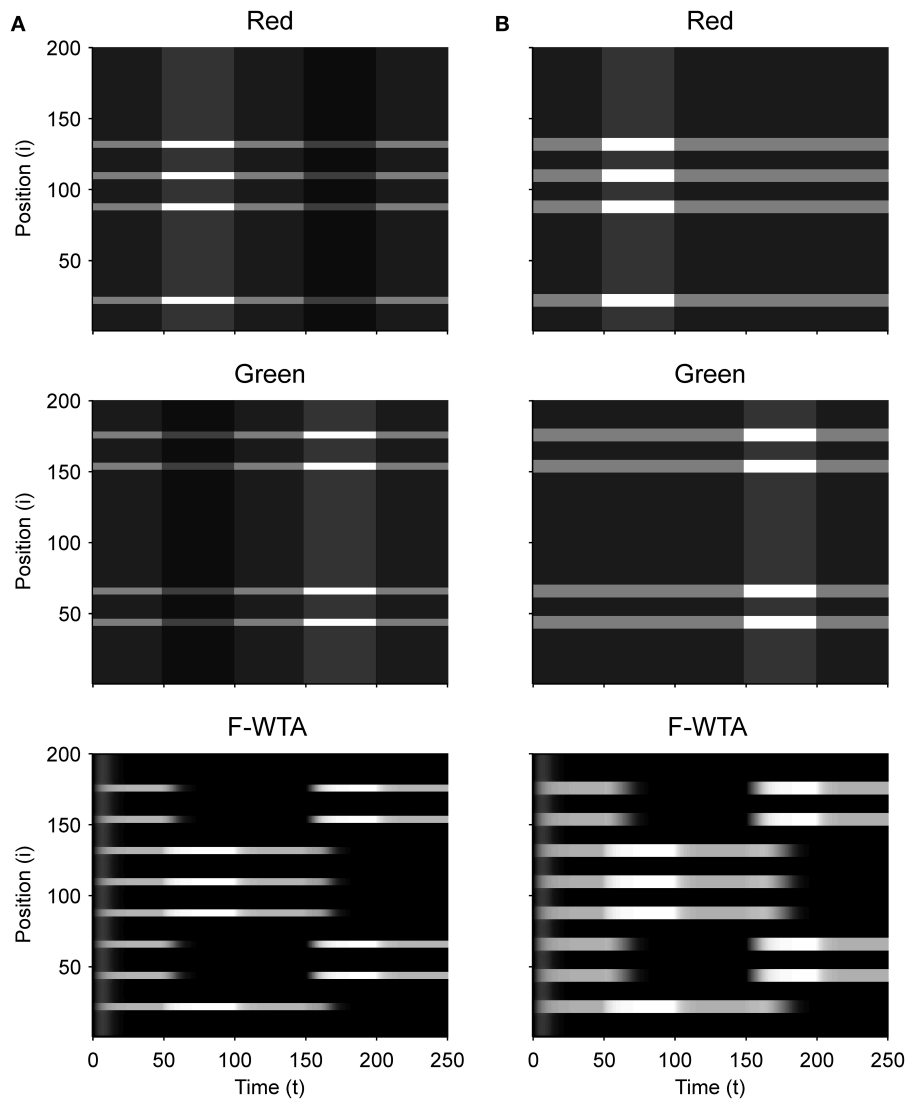
FIGURE 7 | (A) Simulation of the Boolean map formation in the F-WTA network in response to the sequential presentation of two color cues (red appeared between 50th and 100th and green appeared between 150th and 200th simulated time unit). **(B)** The same simulation with larger items and without gain modulation applied on the unattended feature map.

Next, we determined the minimal feature gain that must be applied on the input to produce the desired behavior. When the gain modulation is applied simultaneously on attended feature map G^A and on unattended feature map G^{NA} (where $G^{NA} = 1/G^A$), we found that $G^A \geq 1.7$ is sufficient for creating a Boolean map and switching to another one. In contrast, when the gain modulation is not applied on the unattended feature map, as shown in **Figure 7B**, the feature gain in the attended map should be set to $G^A \geq 2$ to achieve the same behavior.

Figure 8 illustrates that the F-WTA network can support space- and object-based attention alongside feature-based attention. When the spatial cue is applied to a single location in one of the feature maps, the network responds by selecting only this location. Neighboring nodes are not selected even

though they are reciprocally connected to the cued node. The reason is that they receive weaker input relative to the cued node. Furthermore, recurrent excitation that arrives from the cued node is bound by the dendritic non-linearity. Thus, it is not sufficiently strong to keep them active. Interestingly, when the spatial cue is removed, the network activity starts to propagate from the cued node toward the boundary of the whole item. In this case, the network selects not just the cued location, but all locations that are connected to it. Therefore, the F-WTA network exhibits object-based selection, which is consistent with neurophysiological studies that show spreading of enhanced activity along the shape of the object (Roelfsema, 2006). This property arises because the removal of the cue equalizes the input magnitude along the object, which allows activity enhancement to propagate via local lateral connections.
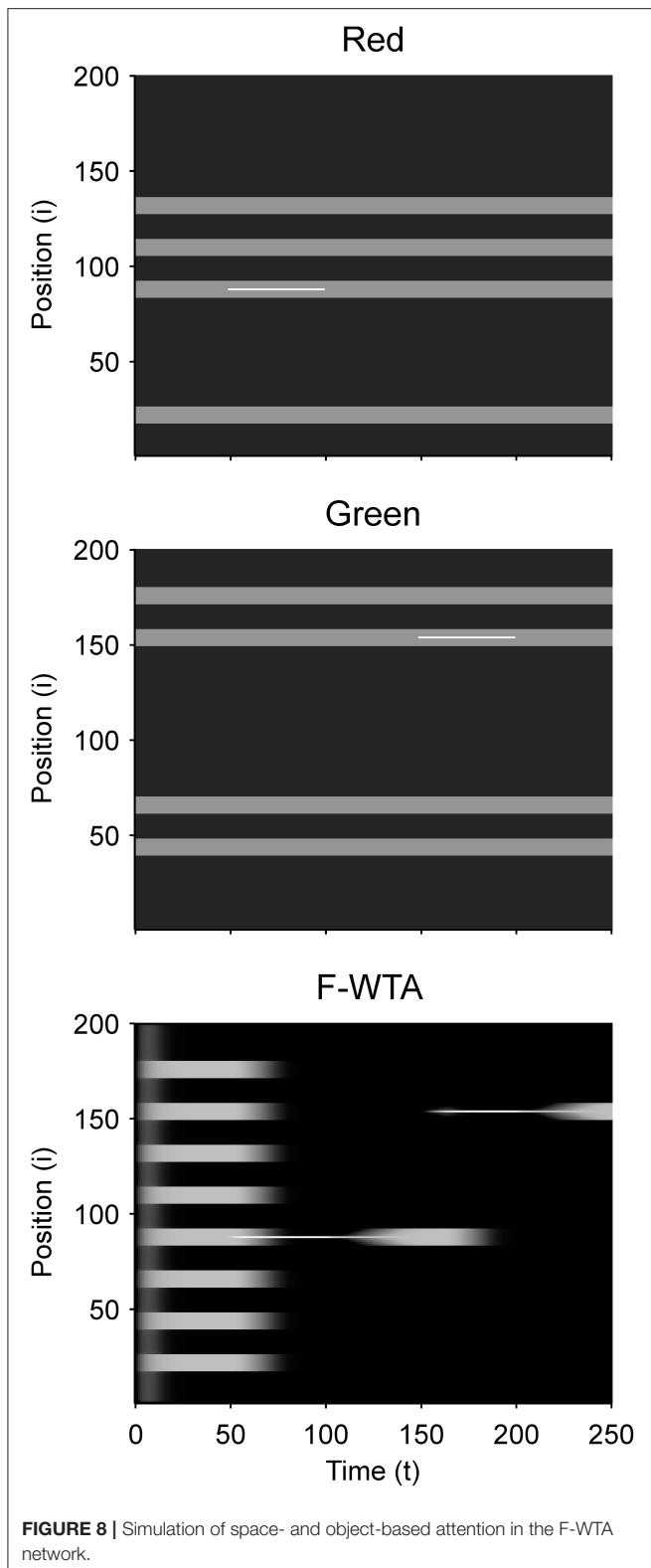
FIGURE 8 | Simulation of space- and object-based attention in the F-WTA network.

In addition, this simulation shows that spatial attention can be easily oriented toward a new location in a single jump without the need for attentional pointers that move attention across the map (Hahnloser et al., 1999).

Simulation of the Intersection and Union of Two Boolean Maps

Figure 9 illustrates that the model can sequentially combine two Boolean maps when the network is cued by top-down signals from two separate feature dimensions. In this simulation, we have employed a visual input that consists of red and green horizontal and red and green vertical bars, like those that are illustrated in **Figure 1B**. First, the F-WTA network is cued to select red bars, irrespective of their orientation. In the second step, it is cued to select horizontal bars, irrespective of their color. However, green vertical bars are already suppressed and the top-down signal that is supplied to them is not sufficient to override the inhibition that arises from red vertical bars. The net result is the selection of a subset of red horizontal bars. In other words, the network activity converges to an intersection between a set of red bars and a set of horizontal bars, thereby resulting in the selection of red horizontal bars.

Next, we examined how the network achieves the union of two Boolean maps (**Figure 10**). Here, we assumed that the input consists of two non-overlapping components: colored squares that activate color maps but do not activate orientation maps, and achromatic horizontal and vertical bars that activate orientation maps but do not activate color maps, as shown in **Figure 1C**. Red-colored items occupy locations between 1 and 100 and oriented bars occupy locations between 101 and 200. This closely resembles the stimulus that is used by Huang and Pashler (2007) to demonstrate the union of color and texture. Taken together, the data show that the union of two Boolean maps is possible only when two top-down cues overlap in time or when the second cue closely follows the withdrawal of the first cue. In **Figure 10**, the cue for the red map is applied in the interval [50, 100] and the cue for the horizontal map is applied in the interval [110, 160]. In this case, the F-WTA network converges to the union of red and horizontal items. However, when top-down cues do not overlap, as shown in **Figure 11**, the second cue overrides the network activity that remains from the first cue. We suggest that this property partly explains why the union is difficult to achieve, as observed by Huang and Pashler (2007).

In addition, we examine the boundary conditions on the choice of the feature gain parameter. We parametrically vary the feature gain in steps of 0.1 starting from $G = 2$ and moving below and above to determine when the ability to form the intersection or union breaks down. When the gain modulation is applied simultaneously on attended (G^A) and unattended (G^{NA}) feature maps, we find that G^A should be chosen from the interval [1.5, 2.1] to achieve the intersection between two maps. When $G^A < 1.5$, the network fails to segregate cued from non-cued locations in the first step. In contrast, when $G^A > 2.1$, the network successfully segregates cued from non-cued locations in the first step. However, the gain is too high, so all horizontal items are selected together in the second step. That is, the representation of red horizontal items is merged with the representation of green horizontal items. When $G^{NA} = 1$ throughout the simulation, G^A should be chosen from the interval [1.8, 2.0] to achieve intersection.
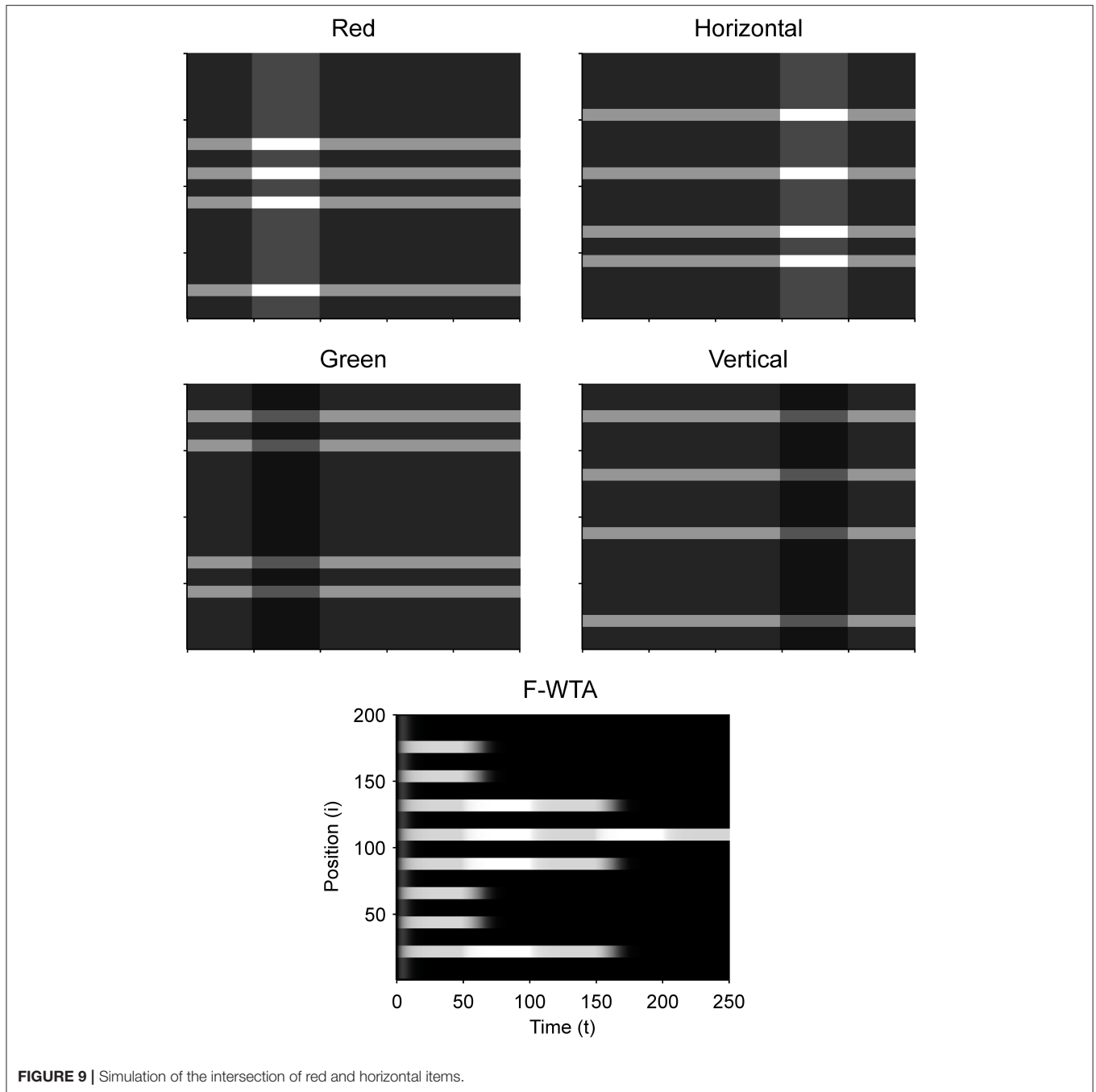
FIGURE 9 | Simulation of the intersection of red and horizontal items.

With respect to the union of two maps, the feature gain G^A should be chosen from the interval [1.4, 2.0] when G^{NA} = 1/G^A and from the interval [1.6, 2.0] when G^{NA} = 1. When G^A is chosen below the suggested intervals, feature gain is too weak, and the second cue will not be able to raise the activity level of the nodes that represent horizontal items above the quenching threshold. Therefore, the network ends up with the Boolean map of red items that is formed in the first step. When G^A is chosen above the suggested interval, the network switches between the representation

of the red items in the first step to the representation of the horizontal items in the second step. In this case, the feature cue is too high, and the activity of the nodes that represent horizontal items simply overrides the activity of the nodes that represent the red items. These constraints are derived from the situation in which the two top-down cues overlap in time. As shown above, temporal lag of the second cue relative to the first cue also destroys the ability of the network to form the union of two Boolean maps.
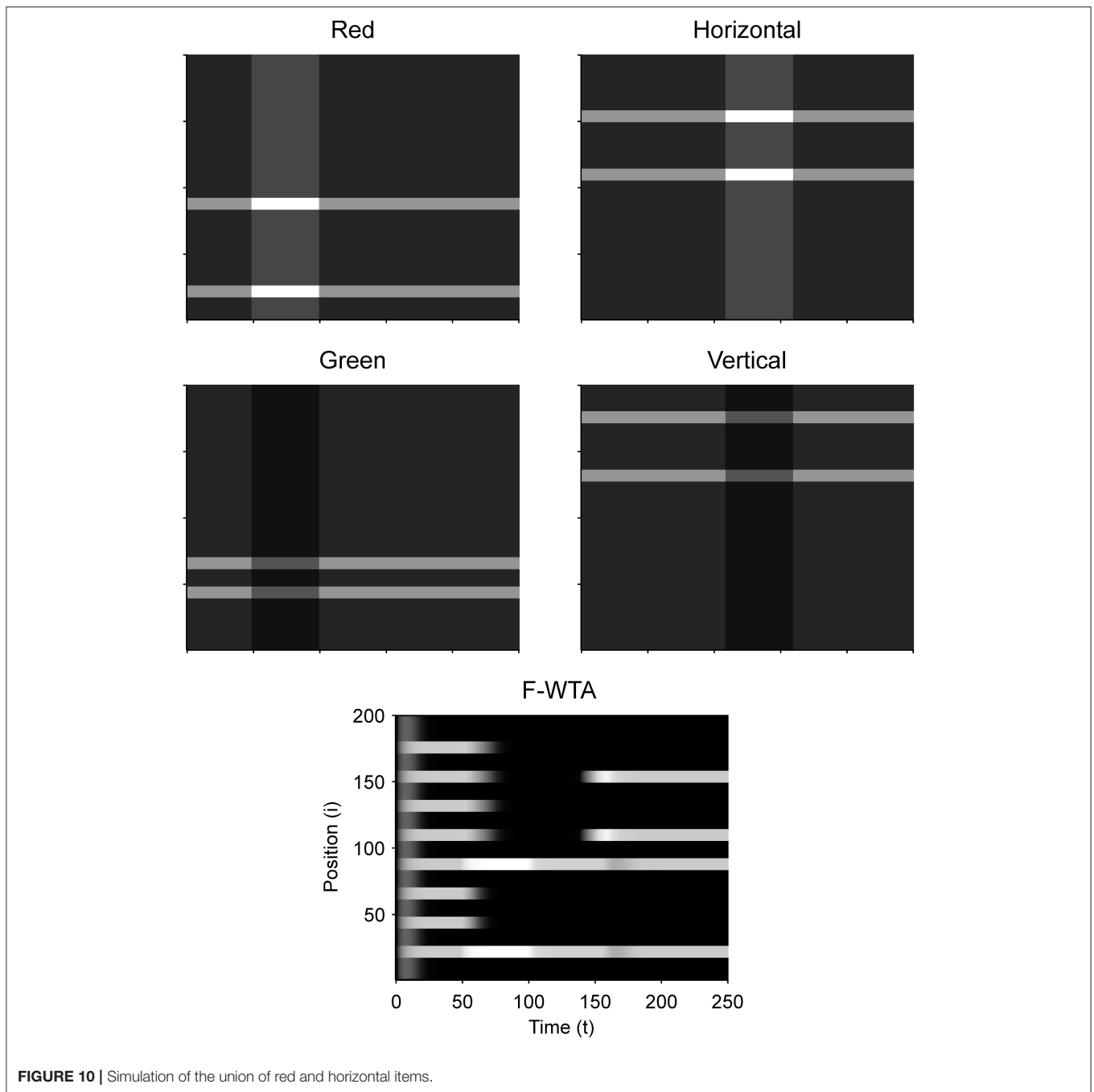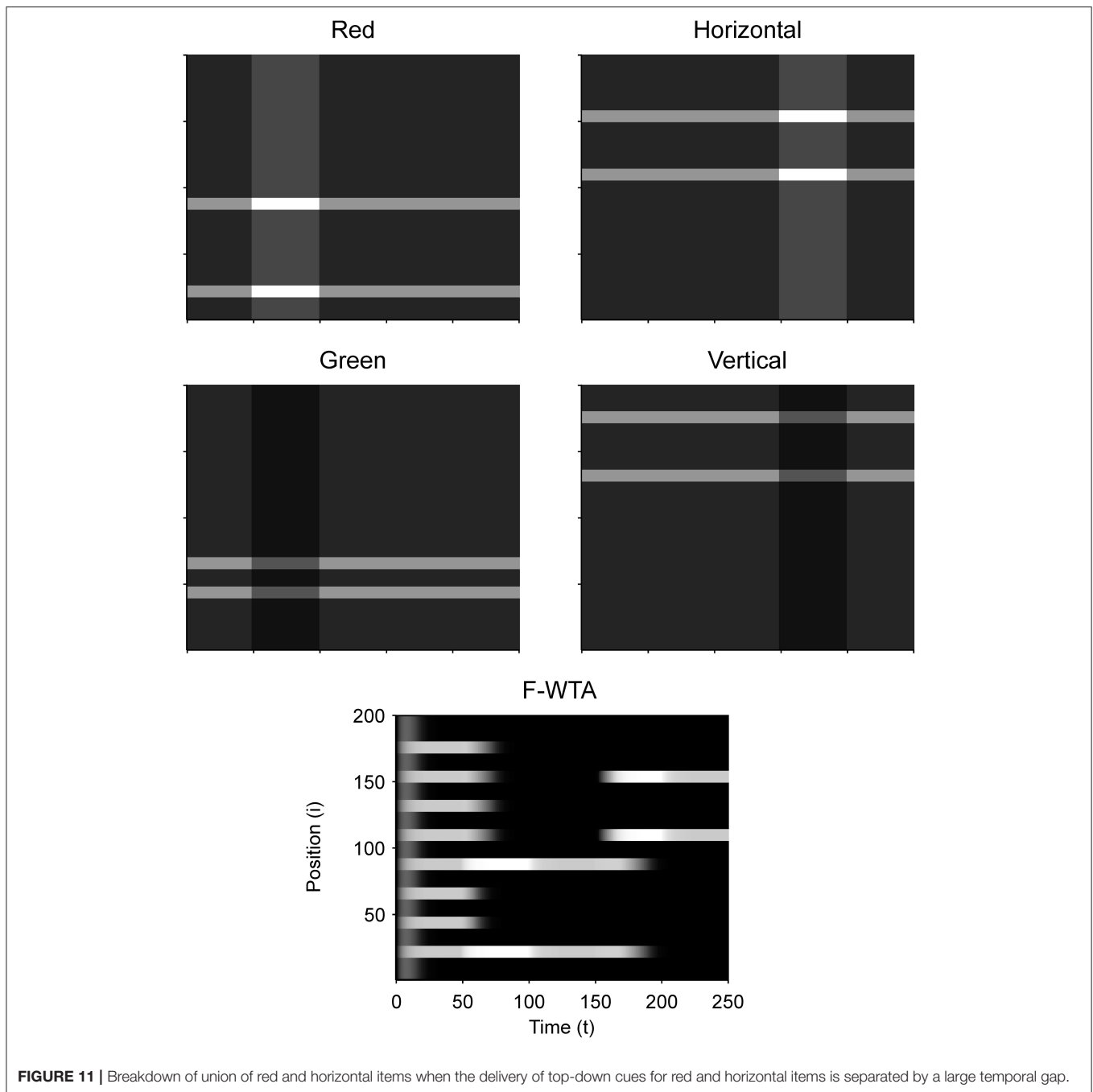
FIGURE 10 | Simulation of the union of red and horizontal items.

Simulation of Bottom-Up Spatial Selection

Finally, we have shown that when there is no top-down guidance, the network selects the most-salient locations based on the bottom-up salience that is computed within feature maps (**Figure 12**). We did not explicitly model competition among maps, but it is reasonable to assume that in a scene with many multi-featured objects, their input magnitudes (i.e., saliencies) will be different. Therefore, we arbitrarily assigned different input magnitudes to different items. As shown in **Figure 12A**, the F-WTA network selects the most salient object if the difference in

input magnitude between the two most active nodes is sufficiently large. However, when this difference is small, as shown in **Figure 12B**, the F-WTA model chooses two most salient items together. Furthermore, in both examples, the network activity retains the input amplitude of the winning item (or items), thereby illustrating the ability to compute the function maximum (Yu et al., 2002).

The precision of saliency detection depends on the threshold for the activation of synaptic receptors on the inhibitory interneuron. In all reported simulations, it was set to $T_y = 0.1$.

FIGURE 11 | Breakdown of union of red and horizontal items when the delivery of top-down cues for red and horizontal items is separated by a large temporal gap.

If smaller values were chosen, the network would improve in terms of precision and be able to separate the two objects that are presented in **Figure 12B**. However, this comes at the price of losing the ability to form a union of two Boolean maps. Therefore, there is a trade-off between the precision of saliency detection and the ability to form Boolean maps.

An important aspect of stimulus-driven attentional control is attentional capture by peripheral cues. Behavioral studies have shown that the abrupt onset of a new object in a visual scene can automatically capture attention even if it is irrelevant

for the current goal (Theeuwes, 2010). **Figure 13** illustrates the sensitivity of the F-WTA network to abrupt visual onset. To simulate this effect, we have made the additional assumption that the network receives input not only from a sustained channel that is comprised of feature maps in V4 but also from a transient channel that responds vigorously only to changes in input (Kulikowski and Tolhurst, 1973; Legge, 1978). Thus, when the abrupt onset is accompanied by a strong transient signal that exceeds the activity level of the currently attended item, the F-WTA network temporarily switch activity toward the location of
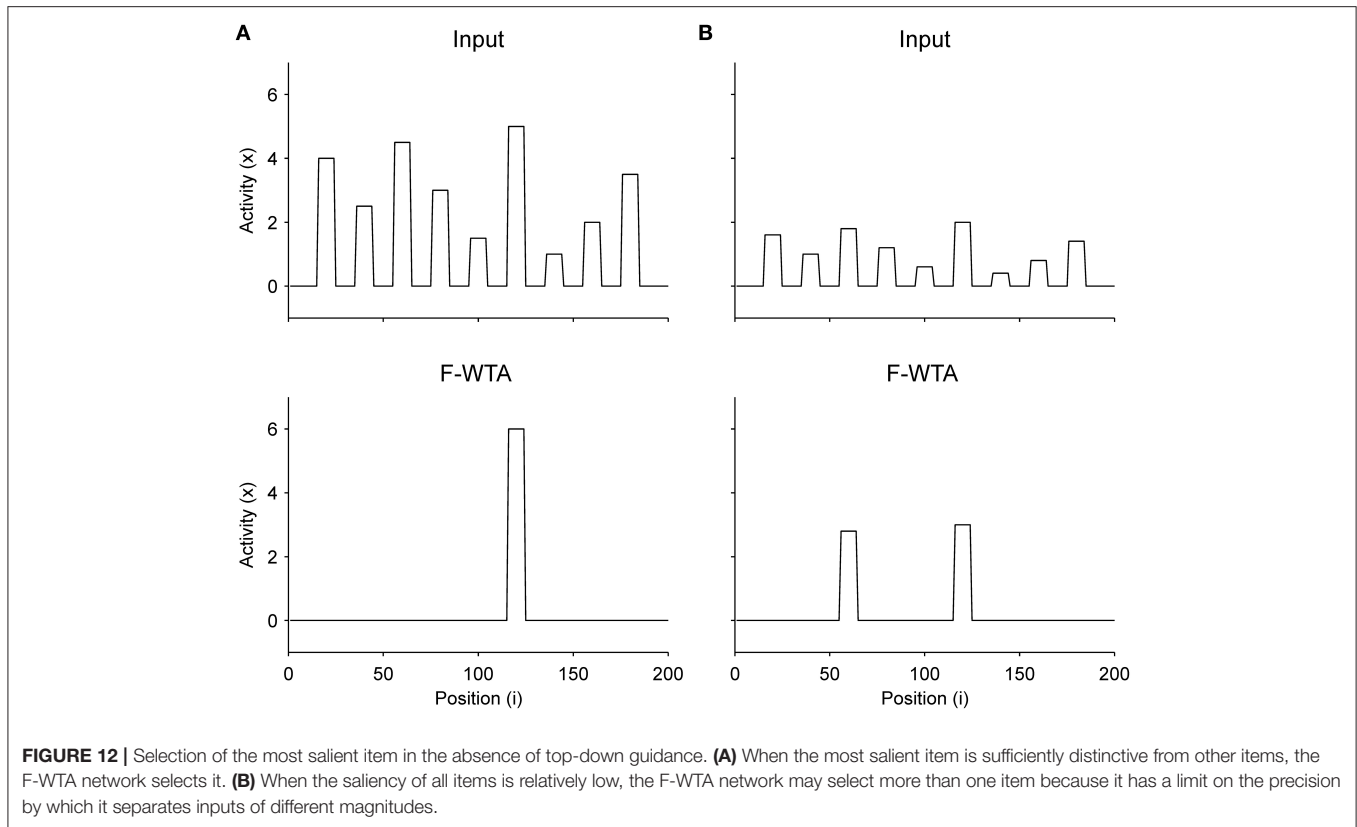
FIGURE 12 | Selection of the most salient item in the absence of top-down guidance. **(A)** When the most salient item is sufficiently distinctive from other items, the F-WTA network selects it. **(B)** When the saliency of all items is relatively low, the F-WTA network may select more than one item because it has a limit on the precision by which it separates inputs of different magnitudes.

the onset (**Figure 13A**). Here, the input at the locations that are occupied by the winning item in the center of the map was set to $I_W = 2$. Input to all other items was set to $I_i = 1$. Finally, the transient input that appears on the sides of the map was set to $I_T = 4$. It is sufficient to set $I_T \geq I_W + 0.8$ to achieve sensitivity to abrupt onsets. Moreover, the same relation holds even if we choose a larger value for I_W.
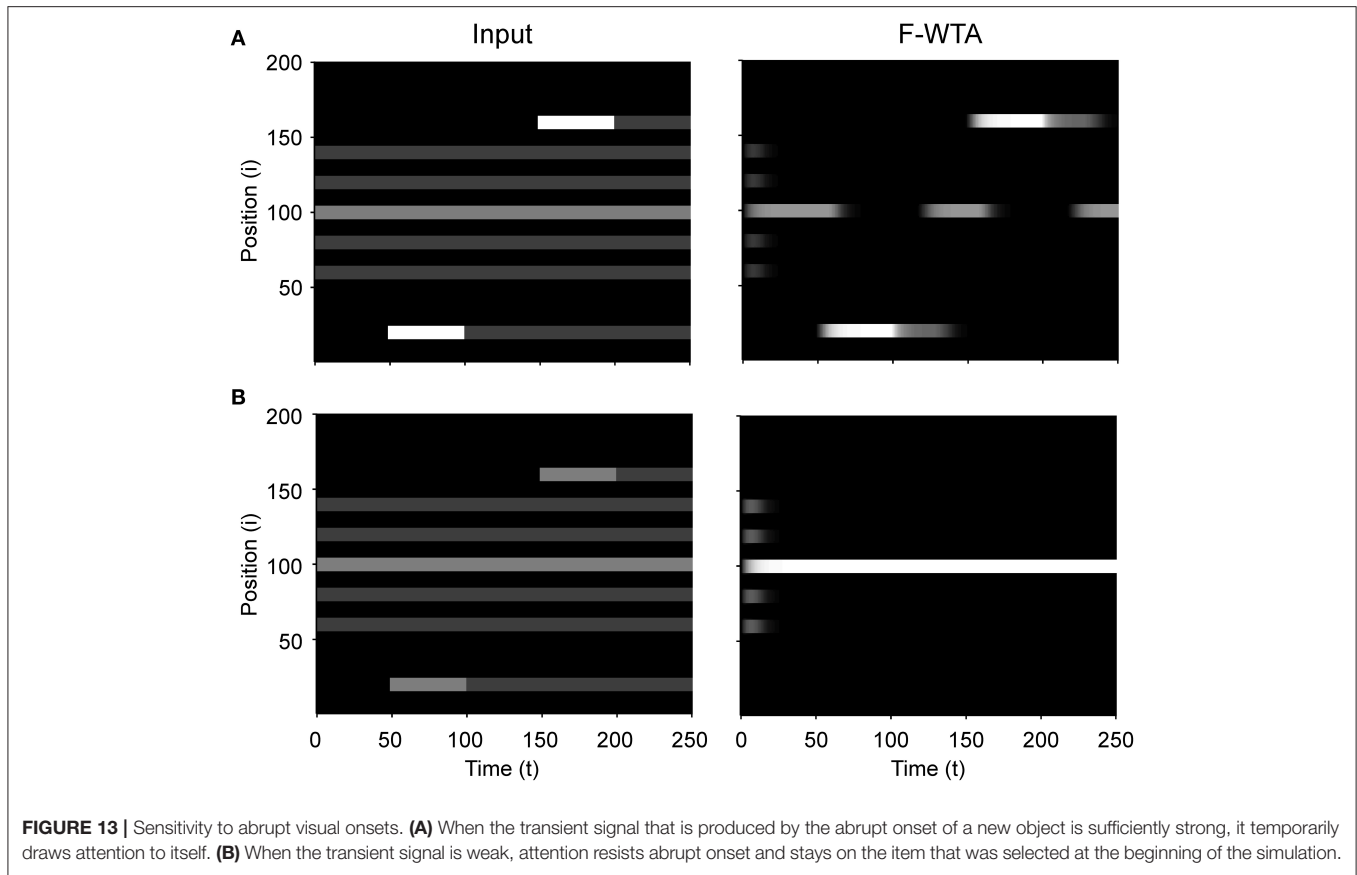
Next, when abrupt onset produces only weak transient signals ($I_T = 2$) that do not satisfy the inequality that is stated above ($I_W = 2$), the activity in the F-WTA network resists abrupt onset and stays on the previously attended item (**Figure 13B**). This observation is consistent with behavioral findings that abrupt onset can be ignored (Theeuwes, 2010), perhaps by attenuating the response of the transient channel. Another possibility is that the top-down gain for the attended location can be increased so that it exceeds the activity of the transient channel. In this case, intense focus on the current object prevents attentional capture, which is consistent with the psychological concept of the attentional window (Belopolsky and Theeuwes, 2010).

DISCUSSION

We have proposed a new model of the WTA network that can simultaneously select multiple spatial locations based on a shared feature value. We named the model the feature-based WTA (F-WTA) network because the unit of selection is not a point in space or object, but rather an abstract feature value that is set by the top-down signals. We have demonstrated how

the F-WTA network implements the central proposal of the Boolean theory of visual attention that there exists a spatial map that divides the visual space into two mutually exclusive sets. One set represents all locations that are occupied by the chosen feature value. The other set contains all other locations, which are not of interest. The Boolean map controls spatial selection and access to the consciousness (Huang and Pashler, 2007). Moreover, we have shown that the network successfully integrates information across space and time to form the intersection or union of two maps that are defined by different feature cues. Previous models of the WTA network are not capable of such integration because they require that the current winner be externally inhibited to allow attentional focus to move from one location to another (Kaski and Kohonen, 1994; Itti and Koch, 2000, 2001). Another possibility to move activity across locations in the network is to introduce dynamic thresholds that simulate habituation or fatigue in individual neurons. In this case, current winner loses its competitive advantage due to the raise of its threshold. This allows non-winners to gain access to working memory (Horn and Usher, 1990). However, both approaches are not suitable for forming the intersection or union of a set of previous winners and a set of later winners.

Another important property of the F-WTA network that sets it apart from previous models of WTA behavior is the ability to select and store arbitrarily many locations in the memory. This is achieved by inhibitory retrograde signaling, which effectively isolates winning nodes from mutual inhibition. First, the amount

FIGURE 13 | Sensitivity to abrupt visual onsets. **(A)** When the transient signal that is produced by the abrupt onset of a new object is sufficiently strong, it temporarily draws attention to itself. **(B)** When the transient signal is weak, attention resists abrupt onset and stays on the item that was selected at the beginning of the simulation.

of inhibition in the network is significantly reduced because the inhibitory interneuron computes the maximum instead of the sum of the recurrent input that it receives from the excitatory nodes. Second, the winning excitatory nodes release their retrograde signals and block inhibition from the interneuron. Consequently, arbitrarily many winners can participate in representing the selected locations without degrading their activation. In other words, there is no capacity limit on the number of objects that can be simultaneously selected. This is consistent with recent behavioral findings that suggest that our ability to select multiple objects is not fixed. Rather, spatial attention should be considered a fundamentally continuous resource without a strict capacity limit (Davis et al., 2000, 2001; Alvarez and Franconeri, 2007; Liverence and Franconeri, 2015; Scimeca and Franconeri, 2015).

In addition, the network is sensitive to the sudden appearance of a new object in the scene, which suggests that it can also be guided by bottom-up feature cues (Theeuwes, 2013). We hypothesize that the network receives strong input from the transient channel. Such input overrides the network's current memory state, thereby making it sensitive to abrupt onsets. Moreover, the transient channel can be activated by any type of change in the spatiotemporal energy of the input, and not just by the sudden appearance (or disappearance) of objects. For example, it will be activated by a sudden change in the direction of motion (Farid, 2002). When the network simultaneously receives

transient input from different locations, they all will be selected together. In this way, the network achieves temporal grouping of synchronous transient input. That is, the network can discover spatial structures that are defined purely by temporal cues (Lee and Blake, 1999; Rideaux et al., 2016).

Biophysical Considerations

As noted above, the model of the F-WTA network rests upon three key computational elements: the dendrite as an independent computational unit, retrograde signaling on synaptic contacts, and computing the maximum over inputs. Here, we review supporting neuroscientific evidence that suggests that all three biophysical mechanisms are plausible candidates for computation in real neural networks.

There is a growing body of evidence that the excitatory pyramidal cell should not be viewed as a single electrical compartment. Rather, it consists of multiple independent synaptic integration zones arranged in a two-layer hierarchy (Häusser and Mel, 2003; London and Häusser, 2005; Branco and Häusser, 2010; Mel, 2016). Using a detailed biophysical model of the pyramidal neuron, Poirazi et al. (2003) showed that its output is well approximated by a two-layer neural network. In the first layer of the network, dendrites independently integrate their synaptic input and produce sigmoidal output. In the second layer, the dendritic output is summed at the soma to produce the neuron's firing rate. Importantly, the somatic and dendritic

output functions need not be the same (Jadi et al., 2014). For example, Behabadi and Mel (2014) showed that the soma of the model neuron generates nearly linear output, while the dendritic output is sigmoid. In our model, the dendrite conveys recurrent excitation to the node. Due to the dendritic non-linearity, there is no risk of unbounded activity growth in the node. Furthermore, the dendritic output is summed with the external input at the soma of the node. By using a linear output function at the soma, we have ensured that the F-WTA network remains sensitive to input fluctuations.

Synaptic transmission can be dynamically regulated in an activity-dependent manner, as shown by the existence of depolarization-induced suppression of inhibition (DSI) (Pitler and Alger, 1992) and depolarization-induced suppression of excitation (DSE) (Kreitzer and Regehr, 2001). DSI (DSE) refers to the reduction in inhibitory (excitatory) post-synaptic potentials following depolarization of the postsynaptic cell. These processes have been observed in various brain regions, including the cerebellum, hippocampus, and neocortex. A retrograde messenger that is released from postsynaptic cell due to its depolarization mediates DSI and DSE. After release, the retrograde messenger binds to the receptors at the presynaptic axon terminals and suppresses the release of the transmitter. Based on these properties, Regehr et al. (2009) suggested that a possible physiological function of DSI and DSE is to provide negative feedback that reduces the impact of the synaptic input on the ongoing neural activity.

The model behavior rests upon the assumption that the inhibitory interneuron computes the maximum instead of the sum of its inputs. There is some direct physiological evidence that real cortical neurons indeed compute the MAX function. For example, Sato (1989) examined responses of neurons in the primate inferior temporal cortex to the presentation of one or two bars in their receptive field. He concluded that the responses to two bars that were presented simultaneously were well described by the maximum of the responses to each separately. In a similar vein, Gawne and Martin (2002) recorded the activity of neurons in primate V4 and found that their firing rate in response to the combination of stimuli is best described by the maximum function over the firing rates that are evoked by each stimulus alone. Furthermore, Lampl et al. (2004) directly measured membrane potentials in the complex cells of the cat primary visual cortex and found evidence for the MAX-like behavior in response to the pair of optimal bars.

Indirectly, the importance of the MAX-like operation in cortical information processing can be appreciated by considering the many computational models of visual functions that have employed it in simulating rich and complex datasets. For example, Riesenhuber and Poggio (1999) employed hierarchical computation of the MAX function in a model of invariant object recognition. Spratling (2010, 2011) used it in simulating a large range of classical and non-classical receptive field properties of V1 neurons. Moreover, Tsui et al. (2010) used MAX-like input integration to explain diverse properties of MT neurons and Hamker (2004) used it in his model of top-down guidance of spatial attention. Furthermore, Kouh and Poggio (2008) developed a canonical cortical circuit that is capable of

many non-linear operations, including computation of the MAX function. Here, we have shown that a single inhibitory node that is endowed with retrograde signaling can compute the maximum.

Based on the proposed model, we have derived two testable predictions. The cortical network that is involved in spatial selection will contain inhibitory interneurons that can compute the MAX function. Moreover, both the excitatory and inhibitory neurons in this network will be endowed with the anatomical structures that support retrograde signaling (presynaptic receptors and postsynaptic transmitter release sites).

Comparison With Other WTA Network Models

Several models of biophysical mechanisms have been proposed for implementing WTA behavior in a neural network, including linear-threshold units (Hahnloser, 1998; Rutishauser and Douglas, 2009), non-linear shunting units (Grossberg, 1973; Fukai and Tanaka, 1997), and oscillatory units (Wang, 1999; Borisyuk and Kazanovich, 2004).

A simple model of a competitive network that is based on linear-threshold units has been extensively studied. Stability analysis revealed that this network requires fine-tuning of the connectivity to achieve stable dynamics that can perform cognitively relevant computations, such as choice behavior (Hahnloser, 1998; Hahnloser et al., 2003; Rutishauser et al., 2015). Recently, Binas et al. (2014) showed that a biophysically plausible learning mechanism could tune the network connections in a way that keeps the network dynamics in the stable regime. Here, we have shown how dendritic and synaptic non-linearities ensure that the network dynamics near fixed points depends only on the time constants of the nodes and not on the parameters that control recurrent excitation and lateral inhibition. Therefore, a precise balance between excitation and inhibition is not necessary for achieving a stable memory state. Moreover, the network is sensitive to the input and can iteratively combine the current memory state with new input to form the intersection or union of them.

An important problem for WTA networks that are based on the linear-threshold or sigmoid output functions is that they lack a mechanism for controlling inhibition between the winning nodes. Therefore, they have limited capacity to represent multiple winners. Usher and Cohen (1999) showed that their activation decreases up to the point of complete inactivation as the number of winning nodes increases. This is due to the increased amount of mutual inhibition. The problem cannot be solved simply by reducing the strength of the lateral inhibition because it is not known in advance how many locations will be cued. On the other hand, feature-based spatial selection requires that the network be able to adjust automatically the amount of inhibition to accommodate the selection of a very small or very large number of winners.

Grossberg (1973) proposed a recurrent competitive map model that was based on shunting non-linear interaction between the synaptic input and the membrane potential. The output of the model depends on the exact form of the signal function that is used to convert membrane potential into the firing rate.

When the signal function is chosen to grow faster than linear, the network exhibits WTA behavior. By contrast, when the signal function is sigmoid, the network can select multiple winners if they have similar activity levels. The most important property of this model is the existence of the quenching threshold. All nodes whose activity is above QT are enhanced and all nodes whose activity is below QT are suppressed. This behavior is similar to the operation of the F-WTA network that was proposed here. However, an important difference is that in the shunting model, QT is fixed and dependent on the parameters of the network. In contrast, the feature-based WTA network exhibits dynamic QT that depends on the input to the network and not on its parameters. In this way, the F-WTA network rescales its sensitivity to the input fluctuations.

More recently, a version of the recurrent competitive map was applied in modeling object-based attention (Fazl et al., 2009). It was shown that sustained network activity in the model PPC encompasses the whole object as an attentional shroud around it. Such spatial representation of a single object supports view-invariant object recognition within a larger neural architecture, namely, ARTSCAN. In an extension of the model, Foley et al. (2012) proposed two separate competitive networks that account for distinct properties of object- and space-based attention. A network with strong inhibition is limited to the selection of a single object. The other network utilizes weaker inhibition to support multifocal spatial selection. To increase the capacity of this network to represent multiple objects, Foley et al. (2012) suggested that the amount of lateral inhibition could be controlled externally. As the number of objects that should be selected together increases, the lateral inhibition should become weaker to counteract the effect of the larger number of nodes that participate in the competition. In contrast, the F-WTA network does not require such external adjustments of the strength of the lateral inhibition to accommodate the selection of arbitrarily many objects of arbitrary size. Moreover, in the F-WTA network, object-based and multifocal spatial attention coexist within the same circuit. Whether the network exhibits object-based spatial selection depends on the type of cue that is presented to the network and not on its parameters.

Wang (1999) proposed a model of object-based attention that relies on the phase synchronization and desynchronization among oscillatory units. At each location of the recurrent map, there is a pair of excitatory and inhibitory units with distinct temporal dynamics that creates a relaxation oscillator. Excitatory units are also mutually connected with their nearest neighbors and with a global inhibitor. The network is initialized with random phase differences between oscillators at different network locations. The activity of the global inhibitor further enforces phase separation among excitatory units. However, local excitatory interactions among nearest neighbors oppose global inhibition and result in phase synchronization that spreads among nodes that encode the same object. The net result of these interactions is temporal segmentation and selection of one active object representation at a time in a multi-object input image. Importantly, the network can switch its activity from one object representation to another. However, this transition is generated internally by the oscillator dynamics. It is not

possible to drive the object selection by external cues such as top-down gain control or bottom-up cues such as abrupt onsets. Moreover, it is not possible to enforce simultaneous selection of more than one object by a joint feature value because the global inhibitor will desynchronize all nodes that encode non-connected items. Therefore, it is not clear how synchronous oscillations could support feature-based attentional selection. Taken together, it is still an open issue whether they are relevant for perception and cognition (Ray and Maunsell, 2015).

Limitations

The proposed model of spatial selection successfully simulates the formation of the Boolean map and its elaboration by the set operations of intersection and union but does not fully implement all aspects of the theory that was proposed by Huang and Pashler (2007). Precisely, it does not explain why attention is limited to only one feature value per dimension or how the observer sequentially chooses one feature value after another or combines feature dimensions into intersections or unions of Boolean maps. It is likely that this severe limitation arises from some form of the WTA network. However, this constraint requires a more elaborate model of the interactions among the spatially invariant representation of the feature values in the IT cortex and the interactions between the IT and the prefrontal cortex, where decisions and plans are made.

In all simulations that are reported here, we kept items segregated in space. This was not the case in the stimuli that were used by Huang and Pashler (2007). They employed a matrix of colored squares that were connected to one another. This is because activity spreading can occur among adjacent nodes even if they encode different feature values. Activity spreading is observed after top-down signals stop favoring one feature value over the other. In this case, all feature maps contribute equally to the input of the F-WTA network and the network is no longer able to discriminate between selected and unselected feature values. One way to solve this issue is to assume that the top-down signals are constantly present during the whole trial. In this way, the activity magnitude on the cued locations is kept above that on the non-cued locations. Therefore, non-cued locations are treated as background noise and suppressed, despite their proximity to the cued locations. Another possibility is to impose boundary signals that act upon recurrent collaterals of the nodes in the F-WTA network in a way that is similar to how activity spreading is stopped in the network models of brightness perception (Grossberg and Todorović, 1988), visual segmentation (Domijan, 2004), and figure-ground organization (Domijan and Šetić, 2008).

Finally, input to the network does not follow the distance-dependent activity profile that is usually observed in the visual cortex. However, this is not a critical issue for the model's performance because the precision of selection depends on the thresholds for presynaptic terminal activation, namely, T_x, and T_y. If they are set to very small values, the network will tend to select the centers of the objects when the input pattern is convolved with a Gaussian filter. In contrast, if they are set to larger values, the network will be able to select extended parts of

the objects and possibly even the whole objects. In the same way, the model achieves resistance to the input noise. As thresholds are set to larger values, the network can tolerate a larger amount of noise. However, this comes at a cost of less-precise selection, as demonstrated by the simulation that is shown in **Figure 12**.

CONCLUSIONS

We have demonstrated how the feature-based WTA network achieves spatial selection of all locations that are occupied by the same feature value without suffering from capacity limitations. The network responds to the top-down cue by storing in memory spatial pattern that corresponds to the cued feature value, while non-cued feature values are suppressed. In this way, we have shown how the Boolean map is formed. In addition, we have shown that it is possible to create more complex spatial representations that involve the intersection or the union of two or more Boolean maps. In this way, the F-WTA network goes beyond the capabilities of previous models of the competitive neural network, which cannot integrate information across space and time. Our work suggests that dendritic non-linearity and retrograde signaling are biophysically plausible mechanisms that are essential for model success.

AUTHOR CONTRIBUTIONS

DD designed the study and write the manuscript. MM performed computer simulations and write the manuscript.

FUNDING

REFERENCES

Alger, B. E. (2002). Retrograde signaling in the regulation of synaptic transmission: focus on endocannabinoids. *Prog. Neurobiol.* 68, 247–286. doi: 10.1016/S0301-0082(02)00080-1

Alvarez, G. A., and Franconeri, S. L. (2007). How many objects can you track? evidence for a resource-limited attentive tracking mechanism. *J. Vision* 7, 14.1–14.10. doi: 10.1167/7.13.14

Ashby, F. G., and Hélie, S. (2011). A tutorial on computational cognitive neuroscience: modeling the neurodynamics of cognition. *J. Math. Psychol.* 55, 273–289. doi: 10.1016/j.jmp.2011.04.003

Behabadi, B. F., and Mel, B. W. (2014). Mechanisms underlying subunit independence in pyramidal neuron dendrites. *Proc. Natl. Acad. Sci. U.S.A.* 111, 498–503. doi: 10.1073/pnas.1217645111

Belopolsky, A. V., and Theeuwes, J. (2010). No capture outside the attentional window. *Vision Res.* 50, 2543–2550. doi: 10.1016/j.visres.2010.08.023

Binas, J., Rutishauser, U., Indiveri, G., and Pfeiffer, M. (2014). Learning and stabilization of winner-take-all dynamics through interacting excitatory and inhibitory plasticity. *Front. Comput. Neurosci.* 8:68. doi: 10.3389/fncom.2014.00068

Borisyuk, R. M., and Kazanovich, Y. B. (2004). Oscillatory model of attention-guided object selection and novelty detection. *Neural Netw.* 17, 899–915. doi: 10.1016/j.neunet.2004.03.005

Boynton, G. M. (2005). Attention and visual perception. *Curr. Opin. Neurobiol.* 15, 465–469. doi: 10.1016/j.conb.2005.06.009

Boynton, G. M. (2009). A framework for describing the effects of attention on visual responses. *Vision Res.* 49, 1129–1143. doi: 10.1016/j.visres.2008.11.001

Braitenberg, V., and Schüz, A. (1991). *Anatomy of the Cortex. Statistics and Geometry,* Vol. 18. Berlin: Springer-Verlag.

Branco, T., and Häusser, M. (2010). The single dendritic branch as a fundamental functional unit in the nervous system. *Curr. Opin. Neurobiol.* 20, 494–502. doi: 10.1016/j.conb.2010.07.009

Davis, G., Driver, J., Pavani, F., and Shepherd, A. (2000). Reappraising the apparent costs of attending to two separate visual objects. *Vision Res.* 40, 1323–1332. doi: 10.1016/S0042-6989(99)00189-3

Davis, G., Welch, V. L., Holmes, A., and Shepherd, A. (2001). Can attention select only a fixed number of objects at a time? *Perception* 30, 1227–1248. doi: 10.1068/p3133

Dayan, P., and Abbott, L. F. (2000). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* Cambridge, MA: MIT Press.

Domijan, D. (2004). Recurrent network with large representational capacity. *Neural Comput.* 16, 1917–1942. doi: 10.1162/0899766041336422

Domijan, D., and Šetić, M. (2008). A feedback model of figure-ground assignment. *J. Vis.* 8, 1–27. doi: 10.1167/8.7.10

Douglas, R., and Martin, K. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451. doi: 10.1146/annurev.neuro.27.070203.144152

Duncan, J. (1984). Selective attention and the organization of visual information. *J. Exp. Psychol.* 113, 501–517. doi: 10.1037/0096-3445.113.4.501

Egeth, H. E., Virzi, R. A., and Garbart, H. (1984). Searching for conjunctively defined targets. *J. Exp. Psychol.* 10, 32–39. doi: 10.1037/0096-1523.10.1.32

Eriksen, C. W., and St. James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Percept. Psychophys.* 40, 225–240. doi: 10.3758/BF03211502

Farid, H. (2002). Temporal synchrony in perceptual grouping: a critique. *Trends Cogn. Sci.* 6, 284–288. doi: 10.1016/S1364-6613(02)01927-7

Fazl, A., Grossberg, S., and Mingolla, E. (2009). View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. *Cogn. Psychol.* 58, 1–48. doi: 10.1016/j.cogpsych.2008.05.001

Foley, N. C., Grossberg, S., and Mingolla, E. (2012). Neural dynamics of object-based multifocal visual spatial attention and priming: object cueing, useful-field-of-view, and crowding. *Cogn. Psychol.* 65, 77–117. doi: 10.1016/j.cogpsych.2012.02.001

Fukai, T., and Tanaka, S. (1997). A simple neural network exhibiting selective activation of neuronal ensembles: from winner-take-all to winners-share-all. *Neural Comput.* 9, 77–97. doi: 10.1162/neco.1997.9.1.77

Gawne, T. J., and Martin, J. M. (2002). Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J. Neurophysiol.* 88, 1128–1135. doi: 10.1152/jn.2002.88.3.1128

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Stud. Appl. Math.* 52, 217–257. doi: 10.1002/sapm1973523213

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychol. Rev.* 87, 1–51. doi: 10.1037/0033-295X.87.1.1

Grossberg, S., and Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena. *Percept. Psychophys.* 43, 241–277. doi: 10.3758/BF03207869

Haarmann, H., and Usher, M. (2001). Maintenance of semantic information in capacity limited item short-term memory. *Psychon. Bull. Rev.* 8, 568–578. doi: 10.3758/BF03196193

Hahnloser, R. H., Seung, H. S., and Slotine, J.-J. (2003). Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Comput.* 15, 621–638. doi: 10.1162/089976603321192103

Hahnloser, R. L. (1998). On the piecewise analysis of networks of linear threshold neurons. *Neural Netw.* 11, 691–697. doi: 10.1016/S0893-6080(98)00012-4

Hahnloser, R., Douglas, R. J., Mahowald, M., and Hepp, K. (1999). Feedback interactions between neuronal pointers and maps for attentional processing. *Nat. Neurosci.* 2, 746–752. doi: 10.1038/11219

Hamker, F. H. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Res.* 44, 501–521. doi: 10.1016/j.visres.2003.09.033

Häusser, M., and Mel, B. W. (2003). Dendrites: bug or feature? *Curr. Opin. Neurobiol.* 13, 372–383. doi: 10.1016/S0959-4388(03)00075-8

Horn, D., and Usher, M. (1990). Excitatory–inhibitory networks with dynamical thresholds. *Int. J. Neural Syst.* 1, 249–257. doi: 10.1142/S0129065790000151

Huang, L. (2015). Grouping by similarity is mediated by feature selection: evidence from the failure of cue combination. *Psychon. Bull. Rev.* 22, 1364–1369. doi: 10.3758/s13423-015-0801-z

Huang, L., and Pashler, H. (2007). A boolean map theory of visual attention. *Psychol. Rev.* 114, 599–631. doi: 10.1037/0033-295X.114.3.599

Huang, L., and Pashler, H. (2012). Distinguishing different strategies of across-dimension attentional selection. *J. Exp. Psychol.* 38, 453–464. doi: 10.1037/a0026365

Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506. doi: 10.1016/S0042-6989(99)00163-7

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203. doi: 10.1038/35058500

Jadi, M., Behabadi, B. F., Poleg-Polsky, A., Schiller, J., and Mel, B. W. (2014). An augmented two-layer model captures nonlinear analog spatial integration effects in pyramidal neuron dendrites. *Proc. IEEE Inst. Electr. Electron. Eng.* 102, 782–798. doi: 10.1109/JPROC.2014.2312671

Kaptein, N. A., Theeuwes, J., and van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *J. Exp. Psychol.* 21, 1053–1069. doi: 10.1037/0096-1523.21.5.1053

Kaski, S., and Kohonen, T. (1994). Winner-take-all networks for physiological models of competitive learning. *Neural Netw.* 7, 973–984. doi: 10.1016/S0893-6080(05)80154-6

Kouh, M., and Poggio, T. (2008). A canonical neural circuit for cortical nonlinear operations. *Neural Comput.* 20, 1427–1451. doi: 10.1162/neco.2008.02-07-466

Kreitzer, A. C., and Regehr, W. G. (2001). Retrograde inhibition of presynaptic calcium influx by endogenous cannabinoids at excitatory synapses onto Purkinje cells. *Neuron* 29, 717–727. doi: 10.1016/S0896-6273(01)00246-X

Kulikowski, J. J., and Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *J. Physiol.* 232, 149–162. doi: 10.1113/jphysiol.1973.sp010261

Lampl, I., Ferster, D., Poggio, T., and Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J. Neurophysiol.* 92, 2704–2713. doi: 10.1152/jn.00060.2004

Lee, S. H., and Blake, R. (1999). Visual form created solely from temporal structure. *Science* 284, 1165–1168. doi: 10.1126/science.284.5417.1165

Legge, G. E. (1978). Sustained and transient mechanisms in human vision: temporal and spatial properties. *Vision Res.* 18, 69–81. doi: 10.1016/0042-6989(78)90079-2

Liverence, B. M., and Franconeri, S. L. (2015). "Resource limitations in visual cognition," in *Emerging Trends in the Social and Behavioral Sciences*, eds R. Scott and S. Kosslyn (Hoboken, NJ: John Wiley and Sons), 1–13.

London, M., and Häusser, M. (2005). Dendritic computation. *Annu. Rev. Neurosci.* 28, 503–532. doi: 10.1146/annurev.neuro.28.061604.135703

Martinez-Trujillo, J. C., and Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.* 14, 744–751. doi: 10.1016/j.cub.2004.04.028

Maunsell, J. H. R., and Treue, S. (2006). Feature-based attention in visual cortex. *Trends Neurosci.* 29, 317–322. doi: 10.1016/j.tins.2006.04.001

McCormick, D. A., Connors, B. W., Lighthall, J. W., and Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J. Neurophysiol.* 54, 782–806. doi: 10.1152/jn.1985.54.4.782

Mel, B. W. (2016). "Towards a simplified model of an active dendritic tree," in *Dendrites, 3rd Edn*, eds G. Stuart, N. Spruston, and M. Häusser (Oxford: Oxford University Press), 465–486.

Nobre, A. C., and Kastner, S. (2014). *The Oxford Handbook of Attention*. Oxford: Oxford University Press.

O'Grady, R. B., and Müller, H. J. (2000). Object-based selection operates on a grouped array of locations. *Percept. Psychophys.* 62, 1655–1667. doi: 10.3758/BF03212163

Pitler, T. A., and Alger, B. E. (1992). Postsynaptic spike firing reduces synaptic GABAA responses in hippocampal pyramidal cells. *J. Neurosci.* 12, 4122–4132.

Poirazi, P., Brannon, T. M., and Mel, B. W. (2003). Pyramidal neuron as two-layer neural network. *Neuron* 37, 989–999. doi: 10.1016/S0896-6273(03)00149-1

Polsky, A., Mel, B. W., and Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nat. Neurosci.* 7, 621–627. doi: 10.1038/nn1253

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25. doi: 10.1080/00335558008248231

Qi, W., Han, J., Zhang, Y., and Bai, L. (2017). Saliency detection via Boolean and foreground in a dynamic Bayesian framework. *Vis. Comput.* 33, 209–220. doi: 10.1007/s00371-015-1176-x

Ray, S., and Maunsell, J. H. R. (2015). Do gamma oscillations play a role in cerebral cortex? *Trends Cogn. Sci.* 19, 78–85. doi: 10.1016/j.tics.2014.12.002

Regehr, W. G., Carey, M. R., and Best, A. R. (2009). Activity-dependent regulation of synapses by retrograde messengers. *Neuron* 63, 154–170. doi: 10.1016/j.neuron.2009.06.021

Richard, A. M., Lee, H., and Vecera, S. P. (2008). Attentional spreading in object-based attention. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 842–853. doi: 10.1037/0096-1523.34.4.842

Rideaux, R., Badcock, D. R., Johnston, A., and Edwards, M. (2016). Temporal synchrony is an effective cue for grouping and segmentation in the absence of form cues. *J. Vis.* 16:23. doi: 10.1167/16.11.23

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* 29, 203–227. doi: 10.1146/annurev.neuro.29.051605.112939

Roelfsema, P. R., and de Lange, F. P. (2016). Early visual cortex as a multiscale cognitive blackboard. *Ann. Rev. Vision Sci.* 2, 131–151. doi: 10.1146/annurev-vision-111815-114443

Rutishauser, U., and Douglas, R. J. (2009). State-dependent computation using coupled recurrent networks. *Neural Comput.* 21, 478–509. doi: 10.1162/neco.2008.03-08-734

Rutishauser, U., Douglas, R. J., and Slotine, J.-J. (2011). Collective stability of networks of winner-take-all circuits. *Neural Comput.* 23, 735–773. doi: 10.1162/NECO_a_00091

Rutishauser, U., Slotine, J.-J., and Douglas, R. J. (2015). Computation in dynamically bounded asymmetric systems. *PLoS Comput. Biol.* 11:e1004039. doi: 10.1371/journal.pcbi.1004039

Saenz, M., Buracas, G. T., and Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.* 5, 631–632. doi: 10.1038/nn876

Saenz, M., Buracas, G. T., and Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Res.* 43, 629–637. doi: 10.1016/S0042-6989(02)00595-3

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Exp. Brain Res.* 77, 23–30. doi: 10.1007/BF00250563

Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition* 80, 1–46. doi: 10.1016/S0010-0277(00)00152-9

Scimeca, J. M., and Franconeri, S. L. (2015). Selecting and tracking multiple objects. *Wiley Interdisc. Rev. Cogn. Sci.* 6, 109–118. doi: 10.1002/wcs.1328

Serences, J. T., and Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55, 301–312. doi: 10.1016/j.neuron.2007.06.015

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *J. Neurosci.* 30, 3531–3543. doi: 10.1523/JNEUROSCI.4911-09.2010

Spratling, M. W. (2011). A single functional model accounts for the distinct properties of suppression in cortical area V1. *Vision. Res.* 51, 563–576. doi: 10.1016/j.visres.2011.01.017

Spruston, N. (2008). Pyramidal neurons: Dendritic structure and synaptic integration. *Nat. Rev. Neurosci.* 9, 206–221. doi: 10.1038/nrn2286

Tao, H. W., and Poo, M. (2001). Retrograde signaling at central synapses. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11009–11015. doi: 10.1073/pnas.191351698

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychol.* 135, 77–99. doi: 10.1016/j.actpsy.2010.02.006

Theeuwes, J. (2013). Feature-based attention: it is all bottom-up priming. *Philos. Trans. R. Soc. B* 368:20130055. doi: 10.1098/rstb.2013.0055

Treue, S., and Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399, 575–579. doi: 10.1038/21176

Tsui, J. M. G., Hunter, J. N., Born, R. T., and Pack, C. C. (2010). The role of V1 surround suppression in MT motion integration. *J. Neurophysiol.* 103, 3123–3138. doi: 10.1152/jn.00654.2009

Usher, M., and Cohen, J. D. (1999). "Short term memory and selection processes in a frontal-lobe model," in *Connectionist Models in Cognitive Neuroscience*, eds D. Heinke, G. W. Humphreys, and A. Olson (London: Springer-Verlag), 78–91.

Vatterott, D. B., and Vecera, S. P. (2015). The attentional window configures to object and surface boundaries. *Vis. Cogn.* 23, 561–576. doi: 10.1080/13506285.2015.1054454

Wang, D. L. (1999). Object selection based on oscillatory correlation. *Neural Netw.* 12, 579–592. doi: 10.1016/S0893-6080(99)00028-3

Wannig, A., Stanisor, L., and Roelfsema, P. R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nat. Neurosci.* 18, 1243–1244. doi: 10.1038/nn.2910

Wei, D. S., Mei, Y. A., Bagal, A., Kao, J. P., Thompson, S. M., and Tang, C. M. (2001). Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons. *Science* 293, 2272–2275. doi: 10.1126/science.1061198

Yu, A. J., Giese, M. A., and Poggio, T. A. (2002). Biophysically plausible implementations of the maximum operation. *Neural Comput.* 14, 2857–2881. doi: 10.1162/089976602760805313

Yu, D., and Franconeri, S. L. (2015). Similarity grouping as feature-based selection. *Vis. Cogn.* 23, 843–847. doi: 10.1080/13506285.2015.1093234

Zhang, J., and Sclaroff, S. (2016). Exploiting surroundedness for saliency detection: a Boolean map approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 889–902. doi: 10.1109/TPAMI.2015.2473844

Zilberter, Y. (2000). Dendritic release of glutamate suppresses synaptic inhibition of pyramidal neurons in rat neocortex. *J. Physiol.* 528(Pt 3), 489–496. doi: 10.1111/j.1469-7793.2000.00489.x

Zilberter, Y., Harkany, T., and Holmgren, C. D. (2005). Dendritic release of retrograde messengers controls synaptic transmission in local neocortical networks. *Neuroscientist* 11, 334–344. doi: 10.1177/1073858405275827

Zilberter, Y., Kaiser, K. M., and Sakmann, B. (1999). Dendritic GABA release depresses excitatory transmission between layer 2/3 pyramidal and bitufted neurons in rat neocortex. *Neuron* 24, 979–988. doi: 10.1016/S0896-6273(00)81044-2