A Connectionist Model for Visual Search via Excitation

Kevin Leung

Stanford University

Abstract

Several important discoveries have been made in visual search tasks, including the difference in reaction time between single feature search and conjunctive search and between target trials and blank trials. The most popular models for visual search depend on inhibition of distractors and explicit evaluation of items to find the target in a visual field. This paper presents a model driven only by excitation and probabilistic attention. Although the model does replicate the linear reaction time for conjunctive search, it cannot account for blank trials. It does, however, present a possible explanation for errors in visual search.

A Connectionist Model for Visual Search via Excitation

Background

As we look around the world, we focus on many different objects. When we're looking for something specific, like a red sweater in the closet or an atlas on the bookshelf, we're also attending to the various stimuli, trying to pick out the one subject that matches all of our requirements. Sometimes it feels like the target pops out at us, and other times, we must procedurally look at each item in turn.

Visual search tasks attempt to replicate these normal activities experimentally with simple searches over a few features. A typical setup has several items in a visual field laid out in a grid, where each item is a colored letter, and the subject must find some target item, which is present in target trials and absent in blank trials. In either case, the subject is asked only to report whether the target is present, not its exact location. In single feature search, the items only vary by one feature. For example, in the left example below, the subject must find the red X among the red O's, which are distractors. In conjunctive feature search, the distractors will vary by both features. For example, in the right example below, the subject must find the red X among green X's and red O's.

O		O	
	O		
	X		O
		O	

X			X
	X	O	
O		O	
		X	

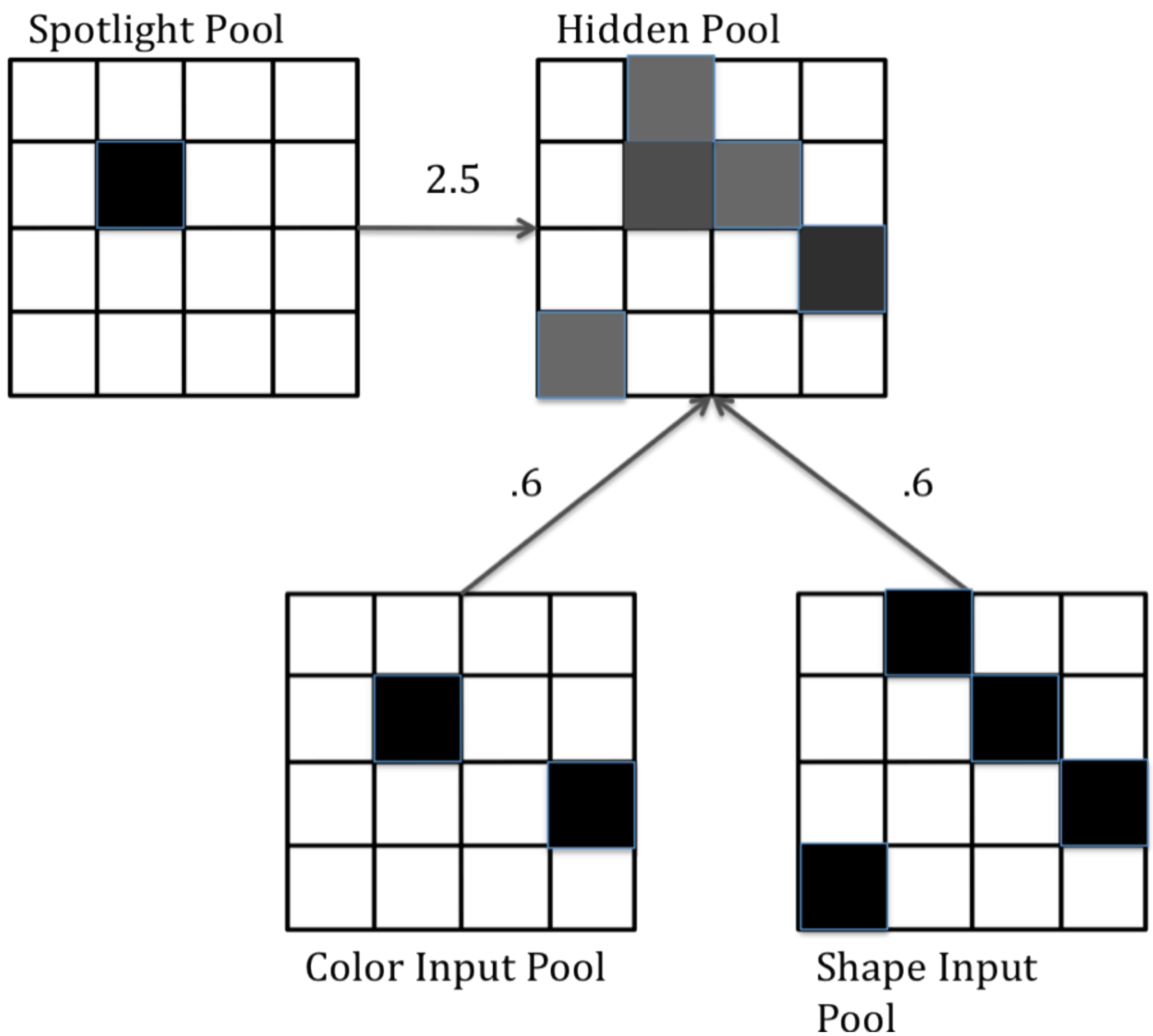
The most notable result of these studies is that single feature search takes constant time regardless of the number of distractors (subjects report that the target appears to "pop-out"), while conjunctive search takes linear time proportional to the number of distractors. Another claim that is the slope of reaction times to the number of items is roughly twice as steep for blank trials as for target trials. The intuition is that it takes twice as long to analyze all objects to determine whether the target is present.

Many models provide possible explanations for how we do these visual search tasks. According to Feature Integration Theory, or FIT, we first do a single-feature parallel search over all items, followed by serial search on multiple features if necessary (Treisman & Gelade 1980). According to Guided Search, we use the information from the parallel search to guide serial search towards likely candidates, improving upon FIT by accounting for faster reaction times in triple-conjunction searches (Wolfe, Cave, & Franzel 1989). According to SERR, a connectionist model for visual search, we group similar items together, then rejecting groups one by one (Humphreys & Müller 1993).

A common element in all of these models is that they find the target item by inhibiting distractors until the target is considered. Since attention can be seen as both the excitation and inhibition of various stimuli, the model tries to do visual search using only excitation. Instead of exciting and picking the target and inhibiting and ignoring distractors, all items are excited. This much simpler model intends to capture the basic, unconscious attentional aspects of visual search without relying on higher-level mechanisms for assessment.

Architecture

The model is built off the interactive activation model, like the Jets and Sharks model (McClelland & Rumelhart 1985). It has 4 pools: each pool represents a 4x4 grid of the 16 possible locations for items in a visual field, shown below in the figure using grayscale values for levels of activation.



Like FIT and Guided Search, the model first processes single features in parallel. There are 2 input pools: 1 for color, and 1 for the shape. The units in these pools receive an input of either 0 or 1, depending on whether the target feature is present at that location. In the figure above, units 6 and 12 are activated in the color pool, and units 2, 7, 12, and 13 are activated in the shape pool, for a total of 4 distractors (units 2, 6, 7, and 13). The target is at unit 12, where both inputs are on.

Both of these pools feed into a hidden pool, with only a single active connection from any unit to its corresponding unit in the hidden pool. For example, color unit 4 and shape unit 4 only excite hidden unit 4, and hidden unit 4 is only excited by those 2 units. All of these weights are equal so that location, color, and shape are all weighted equally. The hidden pool forms a combined activation map for these 2 features. If no item is present at a location, that unit is off. If there's a distractor at a location, that unit is partially activated. The location for the target item, if present, is the most active.

The last pool represents attention on any particular location. Like the spotlight in Guided Search, attention goes from item to item serially. Each of the units is connected to only its corresponding unit in the hidden layer. Because the interactive activation model only has bidirectional connections, the spotlight units need a strong negative input, or else the spotlight units will remain active from resonance with the hidden layer. The negative input into the spotlight layer then represents the decay of attention. All of these weights are equal.

Note that these weights were all set by hand, and the model does not do any learning.

Functionality of the model

Given some set of inputs into the shape and color pools representing a visual field, the model first runs for 20 cycles to allow the hidden layer to mostly settle from its initial activation. Next, the model picks a unit for attention every 2 cycles by exciting a unit randomly with probability proportional to its hidden unit activation. If a unit is picked, its corresponding spotlight unit receives a positive input instead of its normal negative input. This in turn activates the spotlight unit, which increases the activation of the corresponding hidden unit on the subsequent cycle. After those 2 cycles, that unit's spotlight input is reset to its negative value, and a new unit is randomly chosen for attention. Once any hidden unit's activation goes above a threshold value, the trial ends, and the model believes it has found the target at some location.

This model depends on two properties to correctly find the target. First, because the target item's unit is more highly activated, it is initally the most likely candidate for selection by the spotlight. As the model runs, the target's equilibrium activation is higher than any other unit, so it consistently remains a likely target for the spotlight.

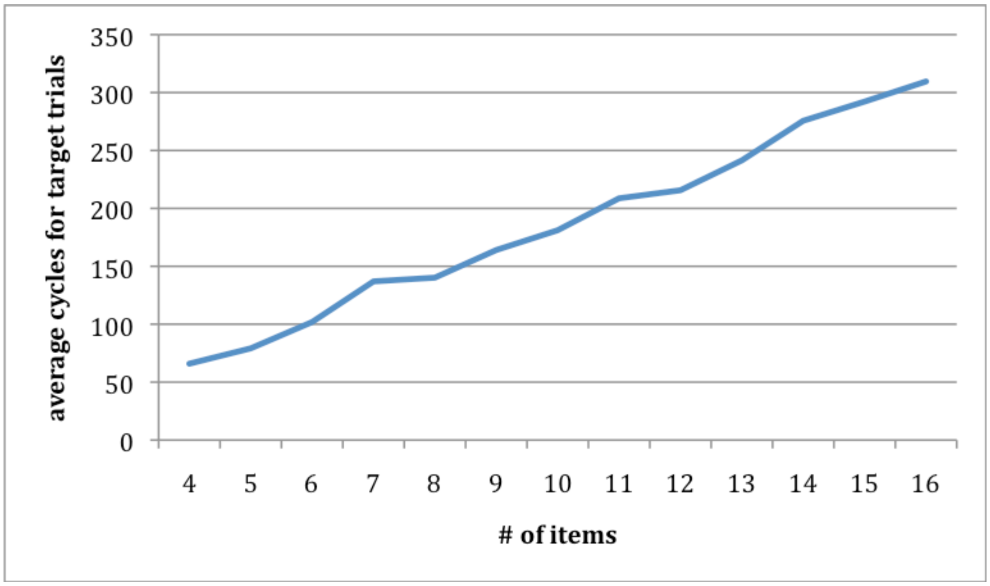
Second, the spotlight tends to reactivate the same units it recently activated. Because the spotlight increases the activation of a hidden unit and the hidden unit activation is the criterion for spotlight selection, that unit is more likely to be picked again. This loop allows a unit's activation to slowly creep up towards the threshold. The natural decay of the model prevents the total activation of the model from growing too rapidly.

Experiment

To determine the effect of the number of items present in the visual field, I presented the model with between 4 and 16 items on each trial. For each number of items, the model ran 1000 randomized target trials and 1000 randomized blank trials. Each trial terminated when a unit's activation goes above the threshold, when the model believes it has found the target at that location, or until it reaches a maximum cycle. Because the model doesn't have a mechanism for determining whether a target item is present (discussed below), a trial terminates after (40 x number of items) cycles. This formula was intended to be roughly twice the average reaction time for target trials, like the 2-to-1 slope for target and blank trials. On each trial, the completion cycle and correctness of its output were recorded.

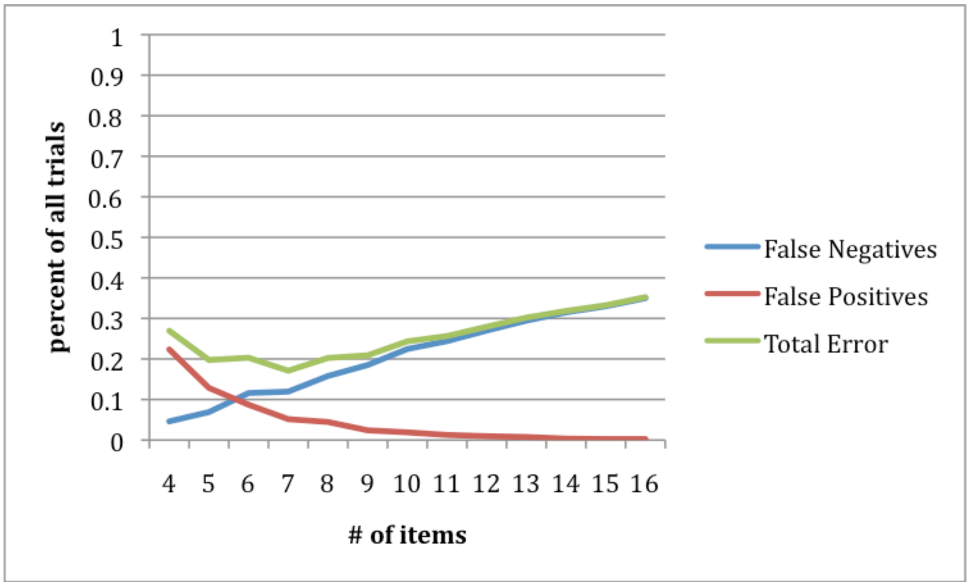
Results

For each number of items in the visual field, I averaged the number of cycles it took for it to find a target item in the target trials when it terminated before the maximum cycle. Results are in the chart below:



The data appears to be linear (r =.997). Because the model cannot determine when the target is absent, the model cannot show the 2 to 1 slope for reaction times between target and blank trials. The model does sometimes pick the wrong item as the target, but this occurred no more than 5% of the time for any number of items. Actual studies of reaction times only require subjects to indicate the presence of the target, not the target's actual location, so the trials where it picked the wrong location are marked as true positives.

The model also demonstrated a relationship between the number of errors and the number of items. Depending on the exact stimuli, error rates for actual subjects vary, with error rates going as high as 20% in conjunctive trials (Treisman & Gelade 1980). That study, however, also dropped subjects who had error rates above 30%. In this model, false negatives increase linearly with the number of items (r = .995), and false positives converge to 0.



Discussion

Overall, the model does not capture most findings of visual search. It does replicate the constant reaction time for single feature search and linear reaction time for conjunctive search. Although not shown in the results, single feature search occurs in constant time by design. For example, if we search only over shape, we would ignore the input from the color pool by setting it to 0, and only a single unit would be active in the shape pool. With only one option, the spotlight will always pick the same item until for attention it reaches activation. In this case, distractors are no different than nonexistent items.

The model doesn't have a good mechanism for giving up. In the blank trials, I assumed that reaching the upper bound on the number of cycles is the same as absence, but the upper bound was determined by hand. Models that depend on inhibition have a better explanation for determining when the target is absent: all of the units are below a threshold. In that case, everything has been rejected, and no items are left to look at.

This spotlight mechanism is not intended to be biologically plausible. It does provide a method for guiding repeated attention in repeated activation of certain units. A consequence of it is the pattern in false negatives and in false positives. It seems unlikely that false positives should converge to 0 and false negatives should grow linearly as the model indicates. Changing free variables or adding more complex weights to spotlight selection may allow the model to more reasonably generate errors. These problems may also be inherent to a probabilistic, excitatory spotlight.

The model, however, does provide a more plausible explanation for how these error occur. As a whole, this simple excitation works better than inhibition. When other models, such as Guided Search, attend to an item, they check it for correctness. Since that's a simple check for equality, the model should be able to recognize a target every time. To account for errors, these models add noise and will return the wrong answer with some small probability. That chance of error has little justification and exists entirely outside of the model.

This model allows false positive to occur naturally when the wrong unit is excited too many times. Since this model never checks a unit for correctness and simply excites it whenever the spotlight is put on it, it can make mistakes in judgment. False negatives also occur when the target unit isn't activated frequently enough within the given time.

Given the limitations, the model doesn't appear to have modeled most of the basic behavior in visual search for it to be a promising model for visual search as a whole. There are, however, several possible places for extensions. Other models address other known issues, including triple-conjunctions of features, the salience of the target versus distractors.

The most significant addition would be cognitive control of the task. The model currently considers attention on a basic, unconscious level, perhaps even below saccades and covert attention. Other models have active control in attending to items and evaluating them, which would correspond to subjects consciously looking at and considering items.

Another mechanism for the model could act as cognitive control and potentially fix current problems. First, this mechanism could end trials when it determines that the target is absent. This model would separate the levels where accepting and rejecting happen: while the model would still accept a target based on activation, it would reject the presence of a target based on its meta-knowledge of the visual field as a hole. This may seem odd but might make sense for the subjective experience of visual search. When the target is found, it just appears from unconscious and becomes hard to unsee, whereas rejection requires a conscious action.

Second, cognitive control can also give more direction to the spotlight by weighting probabilities for items in a manner similar to a methodical search over locations. That could improve the error rates by affecting long-term attention in whether the spotlight returns to a particular item.

References

Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology*, 25(1), 43-110. doi: 10.1006/cogp.1993.1002

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114, 159–188.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136. doi:10.1016/0010-0285(80)90005-5

Wolfe, J.M., Cave, K. R., Franzel, S.L. (1989). Guided Search: An Alternative to the Feature Integration Model for Visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 419-433.