How is visual search guided by shape?
Using features from deep learning to understand preattentive “shape space”

Krista A. Ehinger[1,2] & Jeremy M. Wolfe[1,2]

[1]Harvard Medical School, [2]Brigham & Women’s Hospital

How does shape guide search?

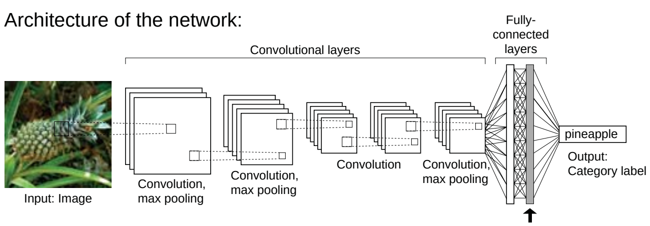
The shape of a target guides attention in seach, but not all shape features guide equally: for example, search for curvature (eg, ● among ■) is faster than search for intersection type (eg, **T** among **L**). [1]

Understanding visual search for shape requires understanding how the visual system represents object shape: what makes a distractor shape similar to a target?

Neural network representation of shape

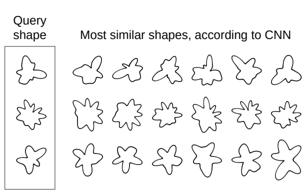
We use features from a “deep” convolutional neural network (CNN) [2] to represent object shape. The network was trained to classify photos into 1000 ImageNet categories, but the image representation learned for this task generalizes well to many other visual tasks [3].

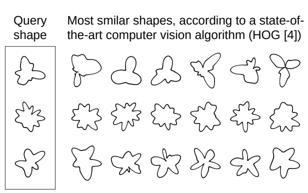
Architecture of the network:



Shape descriptor taken from this layer

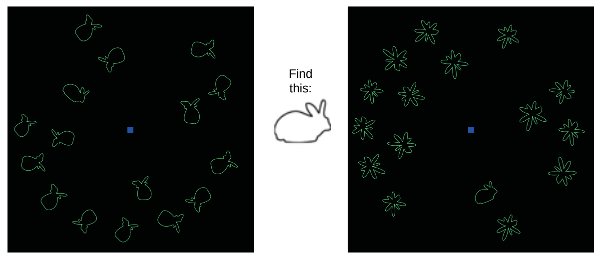
The representation of shape in the CNN seems to be a good match to human perception: shapes which look similar to the CNN also look similar to humans.





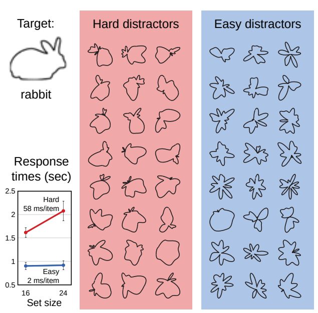
Making search for shape hard or easy

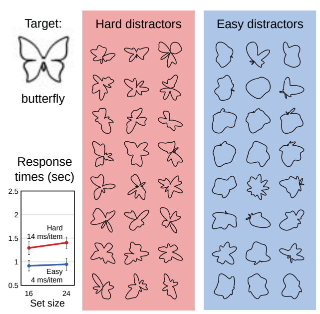
Task: Find and click on a target shape in an array of distractors
Target: Either a rabbit or butterfly silhouette
Distractors: Random radial frequency patterns (one distractor per trial, different on every trial)



Goal: Find distractors that make search harder or easier
After each block of 20 distractors, we identify the ones which produced slowest (or fastest) response times, average their CNN shape features, and generate more distractors from around this average. After eight of these “tuning” blocks, the hardest (or easiest) distractors are repeated in a final “validation” block.
Subjects: 8 participants, within-subjects design

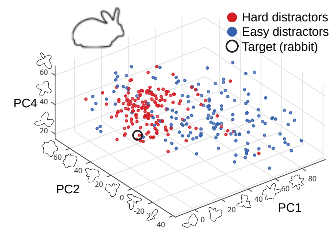
Hardest and easiest distractors for each subject and target:

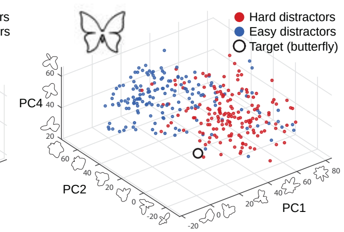




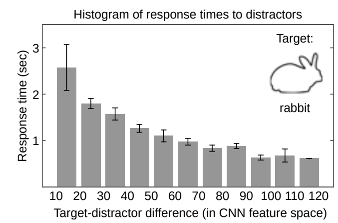
Shape similarity predicts search times

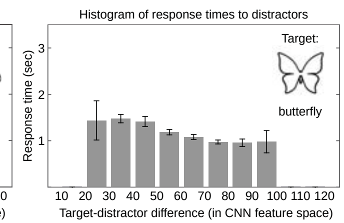
The hard distractors discovered by our algorithm were closer to the target (in CNN features) than the easy distractors.





Response times for individual distractors were higher for distractors that were more similiar (= closer in CNN space) to the target.





Conclusion

Features from a “deep” CNN trained on natural image classification are a good proxy for human perception of shape. We can use these features to identify similar shapes and find distractors which will make search for a specific target harder or easier.

[1] Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In Gray, W.D., Ed.: *Integrated Models of Cognitive Systems*.
[2] Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *NIPS*.
[3] Razavian, A. S., Azizpour, H., Sullivan, J., Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. *CVPR*.
[4] Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. *CVPR*.