Introduction
Is attention useful for object recognition?
Attention is believed to be useful for object recognition by selecting individual parts of a scene and thereby serializing visual information. How does this work in practice? How much improvement does selective attention provide for object recognition? Using David Lowe’s recognition system [1], we measured the improvement in a series of computer experiments.


Objects in Cluttered scenes
Methods
- 21 objects placed into 21 natural images each, one for training, 20 for testing.
- Process repeated for various relative object sizes (0.05%, 0.2%, 0.6%, 1.08%, 2.78%, 5%).
- Objects are learned from only one image with clutter, using the attended regions.
- Recognition is attempted at the attended regions of the test images.
- Human validation: attended regions classified by human operator, replacing the object recognition system.
- Images of the other objects are used as negative examples.

Results
- Images of the other objects are used as negative examples.

Landmark learning
Methods
- 1749 frame video recorded by autonomously moving robot.
- Aim: Objects which can be learned unsupervised and afterwards recognized many times (>10 times) are useful landmarks.
- 3 Attentional fixations per frame to learn and recognize objects (unsupervised).
- Control Random selection of Objects.

Results
- Without attention, 14 useful landmarks were found (1.9% of all objects learned).
- With attention, 87 useful landmarks were found (10% of all objects learned).
- Thus: attentionally selected patches are more likely to contain objects that can be learned unsupervised and later be recognized.

Multiple objects
Methods
- 102 digital photographs of various objects in mixed environments (labs, offices), at 1280x960 pixels resolution.
- One image selected for training, 101 images for testing.
- For learning and recognition using the first 30 fixations.
- Learning unsupervised from all fixations.

Results
- Two objects learned from the training image and recognized in the test images - a box and a book.
- In the 101 test images, the box is present in 23, and the book in 24, of those 4 images contain both objects.
- It is difficult to compare this to some baseline, because the task is impossible for the recognition system alone.
- Results:
<table>
<thead>
<tr>
<th>object</th>
<th>box</th>
<th>book</th>
</tr>
</thead>
<tbody>
<tr>
<td>hits</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>misses</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>false</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Conclusions
Yes, attention is useful for object recognition:
- Selective attention improves performance of object recognition in highly cluttered scenes.
- Selective attention enables learning of multiple objects from only one image.
- We only demonstrated bottom-up attention - deployment of top-down attention can be modeled in a similar fashion.

Acknowledgements:
This work was funded by NSF-ITR, NSF-ITR, NH, the Keck Foundation, and a Sloan Research Fellowship in U.S. The region selection code was developed by the authors as part of the N.Y.U. community effort (http://lab.ics.nyu.edu/). Thanks to Revolution Robotics Inc. for making their robot vision software available to us, and to NSF Human Factors Research Institute, the Netherlands, for the background images.