Human, Soar-in-the-loop

Visual Guidance through Reasoning

Mohamed El Banani

University of Michigan - Ann Arbor
37th Soar Workshop

June 7, 2017
Overview

1. Motivation

2. Hybrid Intelligence

3. Application: Viewpoint Estimation

4. Conclusion
Motivation
Can an agent use its reasoning and interaction capabilities to improve its perception?
Hybrid Intelligence
What is Hybrid Intelligence?

Hybrid Intelligence has been studied under different names:

- **Hybrid Intelligent Systems**: "Computational architectures integrating neural and symbolic processes." [4]

- **Human-in-the-loop Systems**: "The system asks humans to make judgments whenever the computer is less confident – resulting in the most accurate, trustworthy system." [1]

- **Symbiotic Autonomy**: " [...] a robot reasons about, plans for, and overcomes its limitations by proactively asking humans in the environment for help". [6]
Reasons for Hybrid Intelligence

Different intelligent systems have different strengths:

- **Symbolic Rule-based systems** work very well when the environment can be abstracted in such a way that allows rules to be applied.
- **Connectionist systems** work very well when a general pattern exists in the data.
- **Humans** can flexibly deal with anomalies and novel cases, and they think out of the "box".

Systems that are able to leverage the strengths of all those systems are likely to perform better than any single one of those systems.
Viewpoint Estimation
Task Description

- Estimate the agent’s viewpoint from a 2D image.
- A richer description than location or object class.

Source: PASCAL 3D+ Dataset [7]
Typical Approaches

- Match image to a 3-D model.[2]
- Train a neural network.[3]

Human-in-the-loop Viewpoint Estimation

Source: Click-Here CNN [5]
Why use Soar?

1. Interface to minimize expected human input.
2. Provide an autonomous agent with more control over its perception.
Part detectors combined with a parts model could allow for reasoning about part relations.

SVS would support the parts model and spatial reasoning.
Conclusion
Conclusion

Nuggets:

1. Hybrid Intelligence allows one to leverage different intelligent systems (including humans).
2. Auxiliary input can be used to improve the performance of a deep learning vision system.
3. Soar could use its reasoning and interaction capabilities to improve its perception.
4. Integrating deep learning with Soar.

Coal:

1. To be implemented!
References


Ron Sun and Lawrence A Bookman.  

Ryan Szeto and Jason J Corso.  
Click here: Human-localized keypoints as guidance for viewpoint estimation.  

Manuela M Veloso, Joydeep Biswas, Brian Coltin, and Stephanie Rosenthal.  
Cobots: Robust symbiotic autonomous mobile service robots.  
In *IJCAI*, page 4423, 2015.
Yu Xiang, Roozbeh Mottaghi, and Silvio Savarese.
Beyond pascal: A benchmark for 3d object detection in the wild.
### Human-in-the-loop Performance

<table>
<thead>
<tr>
<th></th>
<th>bus</th>
<th>car</th>
<th>motor</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>R4CNN [22]</td>
<td>92.4</td>
<td>78.5</td>
<td>81.4</td>
<td>84.1</td>
</tr>
<tr>
<td>R4CNN [22], fine-tuned</td>
<td>90.6</td>
<td>82.4</td>
<td>84.1</td>
<td>85.7</td>
</tr>
<tr>
<td>Keypoint features (Gaussian fixed attn.)</td>
<td>88.9</td>
<td>81.3</td>
<td>82.8</td>
<td>84.4</td>
</tr>
<tr>
<td>Keypoint features (uniform fixed attn.)</td>
<td>90.6</td>
<td>82.0</td>
<td>83.7</td>
<td>85.4</td>
</tr>
<tr>
<td>CH-CNN (keypoint map only)</td>
<td>90.6</td>
<td>82.0</td>
<td>84.2</td>
<td>85.6</td>
</tr>
<tr>
<td>CH-CNN (keypoint class only)</td>
<td>90.9</td>
<td>86.3</td>
<td>83.1</td>
<td>86.8</td>
</tr>
<tr>
<td>CH-CNN (keypoint map + class)</td>
<td><strong>96.8</strong></td>
<td><strong>90.2</strong></td>
<td><strong>85.2</strong></td>
<td><strong>90.7</strong></td>
</tr>
</tbody>
</table>

Source: Click-Here CNN [5]