SuctionNet: An end to end model for suction

Nikita Demir  
demir@stanford.edu

Jose Miguel Giron  
giron@stanford.edu

Mauricio Wulfsohn  
mwulfs@stanford.edu

Motivation

- In order for robots to robustly perform useful tasks in the real-world—from stocking grocery shelves to assembling complex machinery—they must be able to interact with varied previously-unsensed objects.
- A crucial and necessary first component of such tasks is for robots to automatically detect the best places to grab these items. In this paper, we work with a synthetic dataset provided by Nimble.ai created to train robots that use suction grasping.

Problem Definition

- Input: RGB-D images of objects labeled with their optimal suction location
- Output: predicted optimal suction grasp labels on unseen RGB-D images
- Goal: End-to-end system

Related Works


Data

- Synthetic dataset provided to us by Nimble.ai generated by dropping randomly selected 3D objects on a simulated tray using a physics simulator.

Processing

- Resized to 224x224, split-alpha dimension, divide by 255
- Label: every pixel either 0 (not good suction) or 1 (good suction)
- Depth: Jet-encoding maps near pixels to red, over green, all the way to further pixels which map to green

Results and Analysis

<table>
<thead>
<tr>
<th></th>
<th>Shallow SaNet</th>
<th>Deep SaNet</th>
<th>RGB FCN-6s</th>
<th>Depth FCN-6s</th>
<th>RGB-D FCN-6s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intersection</td>
<td>0.00</td>
<td>0.00</td>
<td>0.15</td>
<td>0.12</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Future Work

- Fine-tuning on real data
- Experimenting with different merge strategies for the FCNN
- Manual learning rate decay