VIOLA uses object proposal priors to learn a closed-loop visuomotor policy to make coffee with only 50 raw demonstrations

**Motivation**
- Imitate closed-loop visuomotor policies efficiently for manipulation
- Reduce spurious correlation in visuomotor learning
- Improve policy generalization from a handful of demonstrations

**Insights**
- Object-centric priors facilitate more efficient and robust policy inference
- Pre-trained RPN from large-scale image datasets captures general object priors
- Transformer self-attention mechanism selects task-relevant object features

**Experiments**

<table>
<thead>
<tr>
<th>Task</th>
<th>Success rates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dining-PlateFork</td>
<td>36.7/76.7</td>
</tr>
<tr>
<td>Dining-Bowl</td>
<td>20.0/60.0</td>
</tr>
<tr>
<td>Make-Coffee</td>
<td>0.0/60.0</td>
</tr>
</tbody>
</table>

Success rates (%): BC-RNN, VIOLA

**Laundry List**
- Impedance controllers for behavior cloning
- Suboptimal data actually helps to learn close-loop behaviors
- Generalization is limited to same task setup from training

**VIOLA Model**

VIOLA: Imitation Learning for Vision-Based Manipulation with Object Proposal Priors
Yifeng Zhu*, Abhishek Joshi, Peter Stone, Yuke Zhu

PRESENTER: Yifeng Zhu