Video Object Segmentation using Deep Learning

Update Presentation, Week 10

Zack While

Advised by: Rui Hou, Dr. Chen Chen, and Dr. Mubarak Shah

July 21, 2017

Youngstown State University
1. Previous Work

2. Current Work

3. Upcoming Work
Previous Work
Figure 1: Training Loss Curve after 189,000 Iterations
Table 1: DAVIS 2016 mIoU Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>FST</th>
<th>CUT</th>
<th>NLC</th>
<th>MSG</th>
<th>KEY</th>
<th>UCF-C3D</th>
<th>CVOS</th>
<th>TRC</th>
<th>SAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>mIoU</td>
<td>0.558</td>
<td>0.552</td>
<td>0.551</td>
<td>0.533</td>
<td>0.498</td>
<td>0.486</td>
<td>0.482</td>
<td>0.473</td>
<td>0.393</td>
</tr>
</tbody>
</table>
Finishing the addition of the PSPNet module.

- **Pyramid Scene Parsing Network** by Zhao et al. (arXiv, April 2017)
- Current version:
  
  ![Diagram of current version]

- Proposed change:

  ![Diagram of proposed change]
Figure 2: Training Loss Curve after 69,000 Iterations
## Quantitative Results

### Table 2: DAVIS 2016 mIoU Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>FST</th>
<th>CUT</th>
<th>NLC</th>
<th>MSG</th>
<th>KEY</th>
<th>CVOS</th>
<th>TRC</th>
<th>SAL</th>
<th>UCF-C3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>mIoU</td>
<td>0.558</td>
<td>0.552</td>
<td>0.551</td>
<td>0.533</td>
<td>0.498</td>
<td>0.482</td>
<td>0.473</td>
<td>0.393</td>
<td>0.306</td>
</tr>
</tbody>
</table>

Comparison data provided by **One-Shot Video Object Segmentation** by Caelles et al. (arXiv, April 2017).
## Quantitative Results

### Table 2: DAVIS 2016 mIoU Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>FST</th>
<th>CUT</th>
<th>NLC</th>
<th>MSG</th>
<th>KEY</th>
<th>CVOS</th>
<th>TRC</th>
<th>SAL</th>
<th>UCF-C3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>mIoU</td>
<td>0.558</td>
<td>0.552</td>
<td>0.551</td>
<td>0.533</td>
<td>0.498</td>
<td>0.482</td>
<td>0.473</td>
<td>0.393</td>
<td>0.306</td>
</tr>
</tbody>
</table>

Comparison data provided by **One-Shot Video Object Segmentation** by Caelles et al. (arXiv, April 2017).
Figure 3: GOAT

IoU: 0.446048927307
**Figure 3: GOAT**

IoU: 0.446048927307

- Actual Clip
- Mask Heatmap
- Ground Truth Mask
- Thresholded Mask
Figure 4: BLACKSWAN

IoU: 0.475382010142
Figure 4: PARAGLIDING-LAUNCH

IoU: 0.475382010142
Figure 5: CAR-SHADOW

IoU: 0.514067745209
Figure 5: CAR-SHADOW

IoU: 0.514067745209

Actual Clip

Mask Heatmap

Ground Truth Mask

Thresholded Mask
Upcoming Work
Plan for Next Week

- Going back to the original architecture for now

- Increasing the size of the dataset
  - JumpCut dataset
  - DAVIS 2017 converted to binary masks

- Depending on results with larger dataset, may look into further fixing the multi-scale maxpooling
Increasing weight values for foreground to balance loss function.

- **One-Shot Video Object Segmentation** by Caelles et al. (arXiv, April 2017)
- More difficult implementation-wise due to wanting weight values based each sequence’s background/foreground ratio.
Questions?