Annotating Object Instances with a Polygon-RNN

Lluís Castrejón,1,2 Kaustav Kundu,1,2 Raquel Urtasun,3 Sanja Fidler1,2

University of Toronto,1 Vector Institute,2 Uber ATG Toronto3

Annotating object instances is expensive and time-consuming

• We propose a model that segments instances automatically using polygons
• Our model easily integrates user corrections and can be used as an annotation tool

Object Instance Annotation

Polygon-RNN

Interactive Object Annotation Tool

Add box
To correct the prediction, drag and drop a point

Predicition Mode

We investigate what the agreement (in % IOU) with the GT annotations in Cityscapes
• Our inputs are the GT instance boxes
• No user corrections are needed

Annotating Object Instances with a Polygon-RNN

<table>
<thead>
<tr>
<th>Model</th>
<th>Bicycle</th>
<th>Bus</th>
<th>Person</th>
<th>Train</th>
<th>Truck</th>
<th>Motorcycle</th>
<th>Car</th>
<th>Rider</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Square Box</td>
<td>35.41</td>
<td>53.44</td>
<td>26.36</td>
<td>39.34</td>
<td>54.75</td>
<td>39.47</td>
<td>46.04</td>
<td>26.09</td>
<td>40.11</td>
</tr>
<tr>
<td>Dilatation10</td>
<td>46.80</td>
<td>48.35</td>
<td>49.37</td>
<td>44.18</td>
<td>35.71</td>
<td>26.97</td>
<td>61.49</td>
<td>38.21</td>
<td>43.89</td>
</tr>
<tr>
<td>DeepMask</td>
<td>47.19</td>
<td>69.82</td>
<td>47.93</td>
<td>62.20</td>
<td>63.15</td>
<td>47.47</td>
<td>61.64</td>
<td>52.20</td>
<td>56.45</td>
</tr>
<tr>
<td>SharpMask</td>
<td>52.08</td>
<td>73.02</td>
<td>53.63</td>
<td>64.06</td>
<td>65.49</td>
<td>51.92</td>
<td>65.17</td>
<td>56.32</td>
<td>60.21</td>
</tr>
<tr>
<td>Ours</td>
<td>52.13</td>
<td>69.53</td>
<td>63.94</td>
<td>53.74</td>
<td>68.43</td>
<td>52.97</td>
<td>71.17</td>
<td>60.58</td>
<td>61.40</td>
</tr>
</tbody>
</table>

Annotating Object Instances with a Polygon-RNN

GT Cityscapes (813 clicks)
Ours (28 boxes, no corrections)

Conclusions

• Our method is competitive with SOTA instance segmentation models given GT boxes
• x4.7 speed-up when annotating Cityscapes with our model
• With Polygon-RNN an annotator can get more accurate annotations with less clicks compared to GrabCut
• Our model can be applied out-of-the-box in other domains

Annotation Mode: Comparison with GrabCut

GrabCut is a method to segment images commonly used as an annotation tool

Key differences from Polygon-RNN:
• Produces a dense pixel-wise annotation
• Users segment instances by painting strokes

Cross-Domain Results on KITTI

We test on KITTI models trained on Cityscapes without fine-tuning
• Without any user corrections, our model performs comparably to SOTA
• With <6 clicks on average our model is at the human agreement level