Translation-aware Fully Convolutional Instance Segmentation

Jifeng Dai*, Haozhi Qi*, Yi Li*
Microsoft Research Asia
Visual Computing Group

(*Equal contribution. This work was done when Haozhi Qi and Yi Li were interns at MSRA)
Highlights

• The **first pure FCN-based** method for instance segmentation

• **1st place** in COCO segmentation challenge 2016, 11% better than 2nd
  • 33% better than the 2015 championship entry (MNC)
  • We won the challenge back-to-back!

• **Fastest** CNN-based method for instance segmentation
  • 0.24 sec/img using ResNet-101 on K40 GPU
  • ~6x faster than MNC

Conventional FCNs

- Successful end-to-end solution for semantic segmentation
  - Each pixel on score maps is a per-pixel category classifier
Conventional FCNs

• Unclear application in instance segmentation
  • Instance segmentation: *translation variant* output
  • Conventional FCNs: by nature *translation invariant*
  • Some endeavors exist, but no pure FCN-based end-to-end solution

Shu Liu, Xiaojuan Qi, Jianping Shi, Hong Zhang, Jiaya Jia. Multi-scale Patch Aggregation (MPA) for Simultaneous Detection and Segmentation. In CVPR 2016.
Prevalent Methods for Instance Segmentation

• Composed of two sub-networks divided by RoI pooling & masking
  • Issues
    • Detailed spatial info is damaged and even lost in feature warping & fc sub-network
    • Region-specific fc sub-network computation


Can we make FCNs aware of translation for end-to-end instance segmentation?
Translation-aware Fully Convolutional Networks (TA-FCNs)

- Translation-aware fg/bg score maps + position-sensitive RoI pooling
- Shared score maps between foreground mask estimation and categorization

TA-FCNs

- Network design

\[
\begin{align*}
\text{conv} & \rightarrow \text{conv} \\
\text{RPN} & \rightarrow \text{position-sensitive RoI pooling} \\
\text{translation-aware } fg/bg \text{ score maps} & \rightarrow \text{pixel-wise softmax} \\
\text{c-th category} & \rightarrow \text{aggregate & vote} \\
\end{align*}
\]

\[k=7 \text{ in experiments}\]

TA-FCNs

• Backbone architecture
  • Based on ResNet-101
  • “Algorithme atrous” on conv5 (output stride: 32 pixels->16 pixels)

• Inference
  • Negligible per-RoI computation
  • Well preserved spatial info: no feature warping/fc layers

• Learning
  • Cost-free online hard example mining (OHEM)


TA-FCNs

• Implementation details (very few hacks)
  • Finetuning on COCO 2014 trainval from ImageNet pretrained net
  • Image shorter side of 600 pixels in training
  • 12 anchors, 300 proposals per image in RPN
  • Mask voting on the initial/refined RoIs

same as MNC
Experiments

• Comparison with MNC on COCO test-dev (using ResNet-101)

<table>
<thead>
<tr>
<th>method</th>
<th>OHEM on initial/refined RoIs?</th>
<th>train time /img</th>
<th>test time /img</th>
<th>mAP</th>
<th>mAP@0.5</th>
<th>mAP (small)</th>
<th>mAP (mid)</th>
<th>mAP (large)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNC</td>
<td>No</td>
<td>2.05s</td>
<td>1.37s</td>
<td>24.6</td>
<td>44.3</td>
<td>4.7</td>
<td>25.9</td>
<td>43.6</td>
</tr>
<tr>
<td>TA-FCN</td>
<td>No</td>
<td>0.53s</td>
<td>0.24s</td>
<td>28.8</td>
<td>48.7</td>
<td>6.8</td>
<td>30.8</td>
<td>49.5</td>
</tr>
<tr>
<td>MNC</td>
<td>Yes</td>
<td>3.22s</td>
<td>1.37s</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>TA-FCN</td>
<td>Yes</td>
<td>0.54s</td>
<td>0.24s</td>
<td>29.2</td>
<td>49.5</td>
<td>7.1</td>
<td>31.3</td>
<td>50.0</td>
</tr>
</tbody>
</table>
## Experiments

- COCO Segmentation Challenge 2016 Entry

<table>
<thead>
<tr>
<th>Method</th>
<th>mAP</th>
<th>mAP@0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAIRCNN (2015 2nd)</td>
<td>25.0</td>
<td>45.6</td>
</tr>
<tr>
<td>MNC+++ (2015 1st)</td>
<td>27.9</td>
<td>51.2</td>
</tr>
<tr>
<td>G-RMI (2016 2nd)</td>
<td>33.8</td>
<td>56.9</td>
</tr>
<tr>
<td><strong>TA-FCN</strong></td>
<td>29.2</td>
<td>49.5</td>
</tr>
<tr>
<td>+ multiscale testing</td>
<td>32.0</td>
<td>51.9</td>
</tr>
<tr>
<td>+ horizontal flip</td>
<td>32.7</td>
<td>52.7</td>
</tr>
<tr>
<td>+ multiscale training</td>
<td>33.6</td>
<td>54.5</td>
</tr>
<tr>
<td>+ ensemble (2016 1st)</td>
<td><strong>37.6</strong></td>
<td><strong>59.9</strong></td>
</tr>
</tbody>
</table>
Results on the first 5k images from the COCO test set is available at https://github.com/daijifeng001/TA-FCN
Results on the first 5k images from the COCO test set is available at https://github.com/daijifeng001/TA-FCN
• **TA-FCN**
  • The first pure FCN-based end-to-end solution for instance segmentation
  • Fast and accurate

• **Our team**
  
  Jifeng Dai*
  Haozhi Qi*
  Yi Li*

(*Equal contribution. This work was done when Haozhi Qi and Yi Li were interns at MSRA)

• **Resources**
  • More details would be available in our tech report
  • Code of TA-FCN would be released
  • Results on the first 5k images from the COCO test set is available now!

  [https://github.com/daijifeng001/TA-FCN](https://github.com/daijifeng001/TA-FCN)