Instance Segmentation

Shu Kong
CS, ICS, UCI
1. problem definition
2. object proposal based instance refinement
3. FCN architecture with smarter label
4. others
5. Conclusion
Outline

1. problem definition
2. object proposal based instance refinement
3. FCN architecture with smarter label
4. others
5. Conclusion
semantic segmentation -- find regions belonging to category-level labels by grouping pixels
instance segmentation -- find out all the instances by grouping pixels
or similar things for instance segmentation?
or similar things for instance segmentation?

Yes
or similar things for instance segmentation?

Yes

Where?
or similar things for instance segmentation?

We had that?!

face detection

Yes
Where?
Here!
or similar things for instance segmentation?

Yes
Where?
Here!

But only those faces are left, we need more instances!
for instance segmentation, here is a starter---
for instance segmentation, here is a starter---
find out instances in a class-agnostic way, or object proposals

Semantic segmentation

Instance segmentation
for instance segmentation, here is a starter---find out instances in a class-agnostic way, or object proposals how to find the individual instances in the picture?
for instance segmentation, here is a starter---
find out instances in a class-agnostic way, or object proposals
how to find the individual instances in the picture?

**philosophy** -- crop image (sliding window?), highlight the instance
centered in the crop, and zero out the pixels/regions outside the instance
Methods --
1. implement the idea described above
Methods --
1. implement the idea described above
2. fancier output for instance **inference**

upper boundary

left boundary

right boundary

so on and so forth............
1. problem definition
2. object proposal based instance refinement
3. FCN architecture with smarter label
4. others
5. Conclusion
crop image (sliding window?), highlight the instance centered in the crop, and zero out the pixels/regions outside the instance.
top branch -- predicting the mask for the instance centered at the patch

bottom branch -- predicting a score to indicate whether there is a “valid” instance in the patch
samping data for training -- triplet input (input image, mask, score)
samping data for training -- triplet input (input image, mask, score)

input image -- reshaped into 224x224x3
Proposal based Instance Segmentation

**samping data for training** -- triplet input (input image, mask, score)

input image -- reshaped into 224x224x3

mask -- binary map of size 224x224
Proposal based Instance Segmentation

**Sampling data for training** -- triplet input (input image, mask, score)

- input image -- reshaped into 224x224x3
- mask -- binary map of size 224x224
- score -- binary label, 1 for valid patch (**green**), -1 for invalid patch (**red**)

Pinheiro et al., "Learning to Segment Object Candidates", NIPS, 2015
Proposal based Instance Segmentation

**Sampling data for training** -- triplet input (input image, mask, score)

- input image -- reshaped into 224x224x3
- mask -- binary map of size 224x224
- score -- binary label, 1 for valid patch (green), -1 for invalid patch (red)

**Constraints** --

1. the patch contains an object roughly centered in the patch
2. the object is fully contained in the patch and in a given scale range
Proposal based Instance Segmentation

**objective function** -- a sum of binary logistic regression losses

\[
\mathcal{L}(\theta) = \sum_k \left( \frac{1+y_k}{2w^o h^o} \sum_{ij} \log(1 + e^{-m_k^{ij} f^{ij}_{segm}(x_k)}) + \lambda \log(1 + e^{-y_k f_{score}(x_k)}) \right)
\]

- \(x_k\) the k-th patch
- \(m_k\) its mask
- \(y_k\) its objectness score
- \(i,j\) the pixel location

\[\lambda = \frac{1}{32}\]

the output of the classification layer to be \(h^o \times w^o\)

Pinheiro et al., "Learning to Segment Object Candidates", NIPS, 2015
Proposal based Instance Segmentation

objective function -- a sum of binary logistic regression losses

\[
\mathcal{L}(\theta) = \sum_k \left( \frac{1+y_k}{2w^o h^o} \sum_{ij} \log(1 + e^{-m_{k}^i j f_{segm}^{i j}(x_k)}) + \lambda \log(1 + e^{-y_k f_{score}(x_k)}) \right)
\]

remarks --
1. negative samples do not contribute segmentation loss (critical)
2. alternating backpropogating the two branches
3. for scoring branch, sampling data with equal number of positive\&negative
4. can be deployed in a fully convolutional manner
5. sampling data includes translation shift, scale deformation, horizontal flip
6. non-trainable upsampling layer (bilinear upsampling)
Proposal based Instance Segmentation

qualitative results -- pretty visualization on model generalization

Pinheiro et al., "Learning to Segment Object Candidates", NIPS, 2015
Proposal based Instance Segmentation

quantitative results -- seems awesome

metrics -- Intersection over Union (IoU), Average Recall (AR) btwn IoU 0.5~1.0

Pinheiro et al., "Learning to Segment Object Candidates", NIPS, 2015
Proposal based Instance Segmentation

**quantitative results** -- seems awesome

metrics -- Intersection over Union (IoU), Average Recall (AR) btwn IoU 0.5~1.0

<table>
<thead>
<tr>
<th></th>
<th>AR@10</th>
<th>AR@100</th>
<th>AR@1000</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>EdgeBoxes</td>
<td>.074</td>
<td>.178</td>
<td>.338</td>
<td>.139</td>
</tr>
<tr>
<td>Geodesic</td>
<td>.040</td>
<td>.180</td>
<td>.359</td>
<td>.126</td>
</tr>
<tr>
<td>Rigor</td>
<td>-</td>
<td>.133</td>
<td>.337</td>
<td>.101</td>
</tr>
<tr>
<td>SelectiveSearch</td>
<td>.052</td>
<td>.163</td>
<td>.357</td>
<td>.126</td>
</tr>
<tr>
<td>MCG</td>
<td>.101</td>
<td>.246</td>
<td>.398</td>
<td>.180</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>AR@10</th>
<th>AR@100</th>
<th>AR@1000</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepMask20</td>
<td>.139</td>
<td>.286</td>
<td>.431</td>
<td>.217</td>
</tr>
<tr>
<td>DeepMask20*</td>
<td>.152</td>
<td>.306</td>
<td>.432</td>
<td>.228</td>
</tr>
<tr>
<td>DeepMaskZoom</td>
<td>.150</td>
<td>.326</td>
<td>.482</td>
<td>.242</td>
</tr>
<tr>
<td>DeepMaskFull</td>
<td>.149</td>
<td>.310</td>
<td>.442</td>
<td>.231</td>
</tr>
<tr>
<td>DeepMask</td>
<td>.153</td>
<td>.313</td>
<td>.446</td>
<td>.233</td>
</tr>
</tbody>
</table>

~1.5s per image

Pinheiro et al., "Learning to Segment Object Candidates", NIPS, 2015
1. problem definition
2. object proposal based instance refinement
3. FCN architecture with smarter label
4. others
5. Conclusion
fancier output for instance **inference**

upper boundary  
left boundary  
right boundary

so on and so forth............

Dai et al., "Instance-sensitive Fully Convolutional Networks", ECCV, 2016
from FCN to InstanceFCN

Dai et al., "Instance-sensitive Fully Convolutional Networks", ECCV, 2016
from FCN to InstanceFCN

FCN for semantic segmentation

InstanceFCN for instance segment proposal

Dai et al., "Instance-sensitive Fully Convolutional Networks", ECCV, 2016
**InstanceFCN** -- differentiate left from right regions

Dai et al., "Instance-sensitive Fully Convolutional Networks", ECCV, 2016
Instance Assembling Module - producing instance based on maps

say, 9 output maps, mosaic them w.r.t relative positions, similar to **mosaic upsampling**
FCN with Fancier Label

training

vgg16 as base model
modify it with reduced stride at pool4, “hole algorithm” at conv5_1 and conv5_3
two fc branches for segmentation and scoring objectness

Dai et al., "Instance-sensitive Fully Convolutional Networks", ECCV, 2016
FCN with Fancier Label

\[ \sum_i (\mathcal{L}(p_i, p_i^*)) + \sum_j \mathcal{L}(S_{i,j}, S_{i,j}^*) \]

training

sampling for training

\[ 600 \times 1.5 \{ -4, -3, -2, -1, 0, 1 \} \]

8-GPU, each for one image with 256 sampled windows -- batch-8

\~1.5s for testing one image

NMS (0.8) for final set of proposals

Dai et al., "Instance-sensitive Fully Convolutional Networks", ECCV, 2016
Quantitative Results

Table 2. Ablation comparisons between ~DeepMask and our method on the PASCAL VOC 2012 validation set. "~DeepMask" is our implementation based on controlled settings (see more descriptions in the main text).

<table>
<thead>
<tr>
<th>method</th>
<th>train</th>
<th>test</th>
<th>AR@10 (%)</th>
<th>AR@100 (%)</th>
<th>AR@1000 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>~DeepMask</td>
<td>crop 224x224</td>
<td>sliding fc</td>
<td>31.2</td>
<td>42.9</td>
<td>47.0</td>
</tr>
<tr>
<td>ours</td>
<td>crop 224x224</td>
<td>fully conv.</td>
<td>37.4</td>
<td>48.4</td>
<td>51.4</td>
</tr>
<tr>
<td></td>
<td>fully conv.</td>
<td>fully conv.</td>
<td><strong>38.9</strong></td>
<td><strong>49.7</strong></td>
<td><strong>52.6</strong></td>
</tr>
</tbody>
</table>

Table 3. Comparisons with state-of-the-art segment proposal methods on the PASCAL VOC 2012 validation set. The results of SS [6] and MCG [12] are from the publicly available code, and the results of MNC [20] is provided by the authors of [20].

<table>
<thead>
<tr>
<th>method</th>
<th>AR@10 (%)</th>
<th>AR@100 (%)</th>
<th>AR@1000 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS [6]</td>
<td>7.0</td>
<td>23.5</td>
<td>43.3</td>
</tr>
<tr>
<td>MCG [12]</td>
<td>18.9</td>
<td>36.8</td>
<td>49.5</td>
</tr>
<tr>
<td>~DeepMask</td>
<td>31.2</td>
<td>42.9</td>
<td>47.0</td>
</tr>
<tr>
<td>MNC [20]</td>
<td>33.4</td>
<td>48.5</td>
<td><strong>53.8</strong></td>
</tr>
<tr>
<td>ours</td>
<td><strong>38.9</strong></td>
<td><strong>49.7</strong></td>
<td><strong>52.6</strong></td>
</tr>
</tbody>
</table>
Quantitative Results

**Table 5.** Comparisons of instance segment proposals on the first 5k images [8] from the MS COCO validation set. DeepMask’s results are from [8].

<table>
<thead>
<tr>
<th>segment proposals</th>
<th>AR@10 (%)</th>
<th>AR@100 (%)</th>
<th>AR@1000 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOP [29]</td>
<td>2.3</td>
<td>12.3</td>
<td>25.3</td>
</tr>
<tr>
<td>SS [6]</td>
<td>2.5</td>
<td>9.5</td>
<td>23.0</td>
</tr>
<tr>
<td>MCG [7]</td>
<td>7.7</td>
<td>18.6</td>
<td>29.9</td>
</tr>
<tr>
<td>DeepMask [8]</td>
<td>12.6</td>
<td>24.5</td>
<td>33.1</td>
</tr>
<tr>
<td>DeepMaskZoom [8]</td>
<td>12.7</td>
<td>26.1</td>
<td>36.6</td>
</tr>
<tr>
<td><strong>ours</strong></td>
<td><strong>16.6</strong></td>
<td><strong>31.7</strong></td>
<td><strong>39.2</strong></td>
</tr>
</tbody>
</table>
Qualitative Results
NO conclusion.
Thank you

Question & Answer
leaving blank

Content after this page is not suitable for people to watch!