COCO Challenge 2018
Panoptic Segmentation Task

Team name: PKU_360

Team members: Yibo Yang, Xia Li, Hongyang Li, Tiancheng Shen, Zhouchen Lin, Jian Dong, Jiashi Feng, Shuicheng Yan

Peking University
Qihoo Corporation
• Task Analysis

Semantic segmentation

Panoptic segmentation

Instance segmentation
• Task Analysis

- Occlusion between instances

- Occlusion between instance and semantic pixels
• Task Analysis

- Occlusion between instances
  - Non overlapping detector, such as [1]
  - Reasoning to solve occlusion, such as by post processing or learnable NMS.

- Occlusion between instance and semantic pixels

[1] Arnab et al. pixelwise instance segmentation with a dynamically instantiated network, CVPR 2017
• Task Analysis

- Occlusion between instances
  - Non overlapping detector, such as [1]
  - Reasoning to solve occlusion, such as by post processing or learnable NMS.

- Occlusion between instance and semantic pixels
  - Comparison between semantic confidence and objectness score.
  - Thing segments override stuff segments.

[1] Arnab et al. pixelwise instance segmentation with a dynamically instantiated network, CVPR 2017
• Task Analysis

Training methods

• Multi-task in an e2e manner

Instance and semantic segmentation share the same Conv body to extract feature.

Train instance and semantic segmentation separately
• **Instance Segmentation**

- Based on Mask RCNN
- **Backbone**
  - ResNeXt-152 trained on ImageNet 5k provided by Facebook.
- **Best single model performance**
  - 43.5 mask mAP on test-dev (used for our panoptic results)
- **Methods**
  - Non-local module\(^1\)
  - Squeeze and excitation module\(^2\)
  - Bottom-up path aggregation\(^3\) in an alternate updating manner\(^4\)
  - Synchronized BN, multi-scale training/testing, etc.
- **Training details**
  - 300k iterations
  - Single image on each GPU
  - Initial lr: 0.01

---

• **Instance Segmentation**

  ■ Non-local module

  • On backbone (Res4)

  • On FPN (the same level with Res4)

  • On mask head (before each conv of the 4-convs head)

  • Synchronized BN or affine operation with scale parameter initialized as 0
• Instance Segmentation

■ Non-local module
  • On backbone (Res4)
  • On FPN (the same level with Res4)
  • On mask head (before each conv of the 4-convs head)
  • Synchronized BN or affine operation with scale parameter initialized as 0

■ Squeeze-and-excitation module
  • On mask head (after each conv of the 4-convs head)
• Instance Segmentation

- Bottom-up path aggregation

  • Original
• Instance Segmentation

- Bottom-up path aggregation

  - Ours

  ![Diagram showing bottom-up path aggregation with blocks and arrows labeled as Residual and Conv/Deconv]
• Instance Segmentation

- Bottom-up path aggregation

- Ours

![Diagram with layers and arrows indicating Residual and Conv/Deconv pathways]
• Instance Segmentation

- Bottom-up path aggregation

  - Ours

  ![Diagram showing bottom-up path aggregation with residual and convolution/deconvolution layers]
• Instance Segmentation

- Bottom-up path aggregation

  - Ours

![Diagram of bottom-up path aggregation with residual and convolution-deconvolution layers](image_url)
## Instance Segmentation

**Ablation experiments (40000 iterations, no test time augmentation, on val set)**

<table>
<thead>
<tr>
<th></th>
<th>Box map</th>
<th>Mask map</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-50 baseline</td>
<td>33.66</td>
<td>30.76</td>
</tr>
<tr>
<td>+ 4SE mask head</td>
<td>33.83</td>
<td>30.96</td>
</tr>
<tr>
<td>+ nonlocal backbone + 4SE mask head</td>
<td>33.83</td>
<td>31.09</td>
</tr>
<tr>
<td>+ nonlocal backbone + 4SE mask head + 4nonlocal mask head</td>
<td>33.99</td>
<td>31.15</td>
</tr>
<tr>
<td>+ nonlocal backbone + nonlocal FPN</td>
<td>34.02</td>
<td>31.08</td>
</tr>
<tr>
<td>+ nonlocal backbone + nonlocal FPN + path aggregation (original)</td>
<td>34.11</td>
<td>31.28</td>
</tr>
<tr>
<td>+ nonlocal backbone + nonlocal FPN + path aggregation (ours)</td>
<td>34.60</td>
<td>31.75</td>
</tr>
</tbody>
</table>
Semantic Segmentation

Original FPN
By Facebook
Semantic Segmentation

Deconv FPN
- Conv+Bilinear → Cascaded Deconv
- Similar parameter number

Deconv
• **Semantic Segmentation**

Original Labelbank\(^1\)

- Auxiliary branch to determine whether a label occurs in an image
- Multiply with seg map to remove non-existing labels in prediction

---

• Semantic Segmentation

Modified Labelbank (LB)
• Share backbone of two branches
• Simplify the ‘Merge’ operation
• Semantic Segmentation

Comparative Experiment
• Backbone: SE-ResNet50
• Init Learning Rate: 1e-2
• Iteration: 20k
• Optimizer: Adam
• Input size: 512
• Dataset: COCO-stuff 10k

<table>
<thead>
<tr>
<th></th>
<th>mIoU</th>
<th>fIoU</th>
<th>mAcc</th>
<th>pAcc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original FPN</td>
<td>31.19</td>
<td>48.34</td>
<td>42.71</td>
<td>62.54</td>
</tr>
<tr>
<td>Deconv FPN</td>
<td>31.52</td>
<td>49.23</td>
<td>42.74</td>
<td>63.64</td>
</tr>
<tr>
<td>FPN + LB</td>
<td>33.12</td>
<td>50.1</td>
<td>45.01</td>
<td>64.74</td>
</tr>
<tr>
<td>Deeplab[1]</td>
<td>32.37</td>
<td>50.73</td>
<td>43.34</td>
<td>65.2</td>
</tr>
<tr>
<td>PSPNet[2]</td>
<td>32.58</td>
<td>50.41</td>
<td>43.49</td>
<td>64.93</td>
</tr>
<tr>
<td>FPANet[3]</td>
<td>32.14</td>
<td>49.23</td>
<td>43.91</td>
<td>63.69</td>
</tr>
</tbody>
</table>

• Semantic Segmentation

Final Submit
• Backbone: ResNeXt152
• Init Learning Rate:
  • Backbone: 1e-3
  • Seg Head: 1e-2
• Normalization:
  • Backbone: freeze
  • Seg Head: no BN
• Iteration: 60k
• Optimizer: Adam
• Dataset: COCO- Panoptic (Stuff Parts)

<table>
<thead>
<tr>
<th></th>
<th>Original FPN</th>
<th>Deconv FPN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input size</td>
<td>800</td>
<td>732</td>
</tr>
<tr>
<td>mIoU</td>
<td>49.54</td>
<td>49.39</td>
</tr>
<tr>
<td>fIoU</td>
<td>67.53</td>
<td>67.2</td>
</tr>
<tr>
<td>mAcc</td>
<td>62.10</td>
<td>62.38</td>
</tr>
<tr>
<td>pAcc</td>
<td>79.51</td>
<td>79.29</td>
</tr>
</tbody>
</table>

Average the two models for panoptic calculation
• Panoptic Segmentation

- Baseline method (provided by panoptic cocoapi)
  - Filter out instances (objectness score below a threshold)
  - NMS-like procedure (remove pixels which have been assigned to a segment with higher score, accept the non-overlapping portion if sufficient fraction remains)
  - Filter our semantic segments (area below a threshold)
  - Thing override stuff
Panoptic Segmentation

Baseline method (provided by panoptic cocoapi)

- Filter out instances (objectness score below a threshold)
- NMS-like procedure (remove pixels which have been assigned to a segment with higher score, accept the non-overlapping portion if sufficient fraction remains)
- Filter our semantic segments (area below a threshold)
- Thing override stuff

Problem: does not solve occlusion, take object relationships into account

e.g.

Tie -> Person
Spoon -> Bowl -> Dinning table
• Panoptic Segmentation

■ Our method

• Filter out instances (objectness score below a threshold);
• Select the labels that are more likely to be overlapped with other labels according to the frequency;
• For the selected labels, apply the NMS-like procedure within each label (the procedure is valid only when two segments are of the same label);
• For the other labels, apply the NMS-like procedure among them;
• Assign the overlapped pixels according to label prior to solve occlusion;
Panoptic Segmentation

Our method

- Filter out instances (objectness score below a threshold);
- Select the labels that are more likely to be overlapped with other labels according to the frequency;
- For the selected labels, apply the NMS-like procedure within each label (the procedure is valid only when two segments are of the same label);
- For the other labels, apply the NMS-like procedure among them;
- Assign the overlapped pixels according to label prior to solve occlusion;
- Filter out semantic segments (area below a threshold)
- Filter out semantic pixels (confidence below a threshold)
- Assign a semantic pixel to the second highest prediction label when its probability is above a threshold and the highest prediction is void.
- Thing override stuff
- **Panoptic Segmentation**

- Ablation experiments (on val set)

<table>
<thead>
<tr>
<th>Method</th>
<th>PQ</th>
<th>SQ</th>
<th>RQ</th>
<th>PQ-t</th>
<th>SQ-t</th>
<th>RQ-t</th>
<th>PQ-s</th>
<th>SQ-s</th>
<th>RQ-s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>&lt;45.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Method 1</td>
<td>45.6</td>
<td>79.9</td>
<td>55.4</td>
<td>57.2</td>
<td>83.5</td>
<td>67.9</td>
<td>28.2</td>
<td>74.4</td>
<td>36.5</td>
</tr>
<tr>
<td>Method 2</td>
<td>46.02</td>
<td>79.9</td>
<td>55.9</td>
<td>57.8</td>
<td>83.5</td>
<td>68.7</td>
<td>28.2</td>
<td>74.4</td>
<td>36.5</td>
</tr>
<tr>
<td>Method 3</td>
<td>46.06</td>
<td>79.9</td>
<td>55.9</td>
<td>57.9</td>
<td>83.5</td>
<td>68.8</td>
<td>28.2</td>
<td>74.4</td>
<td>36.5</td>
</tr>
</tbody>
</table>

Method 1: Do not apply the procedure on our selected out labels, and apply on the other labels.
Method 2: Apply the procedure within each label for all labels.
Method 3: Apply the procedure within each label for our selected labels, and apply the procedure among the other labels.
• Panoptic Segmentation

- Ablation experiments (on test-dev set)

<table>
<thead>
<tr>
<th>Method</th>
<th>PQ</th>
<th>SQ</th>
<th>RQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>44.2</td>
<td>79.5</td>
<td>53.5</td>
</tr>
<tr>
<td>+ semantic area threshold</td>
<td>45.6</td>
<td>79.8</td>
<td>55.2</td>
</tr>
<tr>
<td>+ semantic area threshold + Method 3</td>
<td>46.3</td>
<td>79.7</td>
<td>56.1</td>
</tr>
</tbody>
</table>

Submitted entry
• Panoptic Segmentation

- Some examples (from val set)

Image  |  Panoptic output  |  Ground truth
---|---|---
![Image](image1.png)  |  ![Panoptic output](panoptic1.png)  |  ![Ground truth](groundtruth1.png)

![Image](image2.png)  |  ![Panoptic output](panoptic2.png)  |  ![Ground truth](groundtruth2.png)

![Image](image3.png)  |  ![Panoptic output](panoptic3.png)  |  ![Ground truth](groundtruth3.png)
• Panoptic Segmentation

- Some examples (from val set)
• Panoptic Segmentation

Image

Panoptic output

Ground truth
Panoptic Segmentation

- Future direction
  
  - Reasoning object relationships in an e2e manner to resolve the overlap between instances.
  - Semantic and instance segmentation output can be unified into a single framework to resolve the overlap between thing and stuff.
Thank you!

For any question, please contact: ibo@pku.edu.cn