Single Network Panoptic Segmentation for Street Scene Understanding

DAAN DE GEUS, PANAGIOTIS MELETIS, GIJS DUBBELMAN
DEPARTMENT OF ELECTRICAL ENGINEERING – SPS-VCA – MOBILE PERCEPTION SYSTEMS
E-MAIL: D.C.D.GEUS@TUE.NL

PANOPTIC SEGMENTATION

In our work, we address the computer vision task of panoptic segmentation [1]. For panoptic segmentation, the goal is to predict 1. a class label 2. an instance id
for each pixel in an image. This way, it is possible to identify all stuff (e.g. sky, road, vegetation) and things (e.g. pedestrians, vehicles, traffic signs) classes in an image, and distinguish between different things instances.

A baseline method [1] is proposed which uses separate semantic segmentation and instance segmentation networks and combines the results using heuristics.

METHOD

We present the first single deep neural network for panoptic segmentation. The network makes parallel semantic and instance segmentation predictions, and fuses these using heuristics. A single network has several advantages over the use of separate networks:
- Reduced computational time and resources
- Easier implementation on a device
- Sharing information in the network can improve performance

After a shared feature extractor, the network is split into two branches: one for instance segmentation and the other for semantic segmentation. To train jointly, loss functions from these two branches are weighted and combined. Moreover, we introduce additional information flow to improve the overall performance.

RESULTS

We evaluate our single network on the Mapillary Vistas dataset, and compare to separate networks, using the Panoptic Quality (PQ) [1].

<table>
<thead>
<tr>
<th>Method</th>
<th>PQ</th>
<th>PQ$_{things}$</th>
<th>PQ$_{stuff}$</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separate networks</td>
<td>21.3</td>
<td>13.8</td>
<td>31.4</td>
<td>903 ms</td>
</tr>
<tr>
<td>Single network</td>
<td>21.0</td>
<td>14.9</td>
<td>29.2</td>
<td>532 ms</td>
</tr>
<tr>
<td>+ Our improvements</td>
<td>23.9</td>
<td>15.5</td>
<td>35.0</td>
<td>484 ms</td>
</tr>
</tbody>
</table>

It is found that our single network greatly reduces the required computational time for prediction. Moreover, our network achieves improved performance with respect to the implementation using separate networks.

REFERENCES