Revisiting Multi-Task Learning with ROCK: A Deep Residual Auxiliary Block for Visual Detection

Taylor Mordan, Nicolas Thome, Gilles Henaff, Matthieu Cord

Context: Multiple Tasks
- Solution to data starving: Transfer Learning
- parallel transfer: Multi-Task Learning [2, 3]

Primary MTL
- Flat MTL
- Primary task (focus) ≠ Auxiliary tasks (help)

Multi-Modal Object Detection
- Input image
- Object detection

Analysis and Results
- Experiments on NYUv2 dataset [7]

Ablation Study

<table>
<thead>
<tr>
<th>Model</th>
<th>mAP@0.5</th>
<th>mAP@0.75</th>
<th>mAP@0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSD [6]</td>
<td>31.2</td>
<td>15.8</td>
<td>16.2</td>
</tr>
<tr>
<td>+ flat MTL</td>
<td>34.3</td>
<td>16.0</td>
<td>17.4</td>
</tr>
<tr>
<td>+ intensive pool.</td>
<td>35.7</td>
<td>16.2</td>
<td>17.4</td>
</tr>
<tr>
<td>+ fusion</td>
<td>37.6</td>
<td>17.1</td>
<td>18.5</td>
</tr>
</tbody>
</table>

Multi-Modal Object Detection
- Auxiliary information: depth only

Model
- Pose CNN [10] D* N S + SYN 38.8

Auxiliary information: all available
- ROCK D N S 39.8
- ROCK D N S + SYN 46.8

Effectiveness of Additional Supervision
- Auxiliary supervision ~ 30% images here

References
- Gupta et al., Factors of Transferability, TPAMI (2016)
- Caruana, Multitask Learning, Machine Learning (1997)
- Kokkinos, UberNet, CVPR (2017)
- Vapnik and Vashist, LUPI, Neural Networks (2009)
- Hoffman et al., Modality Hallucination, CVPR (2016)
- Liu et al., SSD, ECCV (2016)
- Silverman et al., NYUv2, ECCV (2012)
- Gupta et al., RGB-D Features, ECCV (2014)