OVERVIEW

Motivation: This implementation uses our new popular object detection & classification models on an existing, large, dataset and evaluate their performance on a dataset that reflects real-world images. Object detection and classification is an essential preprocessing step for feeding a Siamese network [11]. This project will focus on the detection and classification and compare the performance of a YOLOv3 and SSD model.

Summary: This project focuses on object detection and object classification, based on the AI City Challenge 2019 dataset. Object images were manually relabeled and used for training. The 2018 results are chosen as benchmark. YOLOv3 and SSD model performance was compared - with a 0.67 IDF1 score for the YOLOv3 model and a 0.56 IDF1 score for the SSD model.

MODELS; INDIVIDUAL DISCUSSION

YOLOv3: A YOLOv3 was implemented using three different architectures (S20, 416, 608):

- Loss function: The YOLOv3 loss function is comprised of squared errors with scale parameters to balance the prediction of objects / no objects.
- SSD: A PyTorch based SSD implementation was built for the purpose of this project. The implementation supports multiple input formats.

RESULTS (1): OUTPUT BOXES

Both the SSD and YOLOv3 performed reasonably well on large vehicles, but not so good on small vehicles, where often no bounding box or low quality bounding box was presented (figures have been cropped for clarity - original figures in report). YOLOv3 overall captured objects better and more accurately.

RESULTS (2): IDF1/MAP SCORE

The YOLOv3 model, after hyperparameter tuning, performed better than the SSD score, in particular for the map score:

<table>
<thead>
<tr>
<th>Model</th>
<th>IDF1 score (test)</th>
<th>MAP score (test)</th>
<th>IDF1 score (train)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSD-baseline</td>
<td>0.44</td>
<td>0.87</td>
<td>0.62</td>
</tr>
<tr>
<td>SSD-optimized</td>
<td>0.50</td>
<td>0.48</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Both models converged well, shown by the loss functions (log scale) versus nr of epochs.

DISCUSSION AND FUTURE WORK

Conclusion:
- The goal was to compare SSD with YOLOv3. The YOLOv3 performed better.
- Overall I did not manage to beat the AI City challenge baseline set by the three used models (YOLOv3, SSD, and Faster R-CNN). Most likely this is because my training set was much too small – a training set of 400 images is small for Computer Vision problems.
- Winning 2018 benchmark of AI City challenge scored a F1 score of 0.86, but did a lot of relabeling and data augmentation.
- For the limited scope and dataset, results are reasonable good.

Future work:
- Low SSD label performance should be investigated
- The YOLOv3 model showed most promising results to continue with.
- A larger training set is essential to meet benchmark (AI City Challenge 2019) standards.

REFERENCES

[1] Xinchao Liu, Wu Liu, Tao Mei, and Huadong Ma, A Deep Learning-Based Approach to Progressive Vehicle Re-identification for Urban Surveillance

Further references are in main report.