Deep Learning for large-scale traffic-sign detection and recognition

Summary

1. Argue that current dataset is too easy:
   - small no. of classes
   - do not exhibit property of signs in the wild:
     - large degree of intra-category (within-category) variation
     - low degree of inter-category (between-category) variations
2. Propose modifications to Mask RCNN:
   - Online hard-example mining (decide which ROIs to pass to the classification learning module based on classification locs)
   - Select ROIs to even cover all sizes of the training objects:
     - argue that randomly selecting ROI bias selecting large obj.
     - select the same no. of ROIs for each obj present in the img. (how?)
   - weight backgrounds with smaller weights:
     - Their evaluation shows that Mask R-CNN can not achieve 100% recall due to missing region proposals in certain cases.
3. Propose data augmentation methods (2 classes of distortions):
   - geometric/shape distortions (perspective change, changes in scale)
   - appearance distortions (variations in brightness and contrast)
4. Propose a new dataset: 6957 images with 13239 tightly annotated traffic-sign instances corresponding to 200 categories.
5. They observe miss rate & recall rate as a function of N where N is the no. of top regions selected from the RPN ⇒ a good way to evaluate region proposal.
6. They also observe miss rate & recall rate separately for all valid annotations & for small obj only.
7. The paper also proposes nuanced quantitative analysis of the accuracy of the trained detector.

Questions

1. Did they do ablations for the different modifications?