This is our supplementary material which includes more experiments on semantic consistency and performance analysis to show the effectiveness of our work.

A Analysis on Semantic Consistency Module

We visualize the dynamic consistency at different epoches to see how semantic consistency affects on the learning targets. As shown in Fig. 8, the sampled points in each epoch with both high classification scores among categories and high IoU scores are highlighted, named high consistency samples. The low consistency samples are appear in dark colors. Part of sample points at initial stage is not locate at the instance, as the model is not robust at the beginning. With the semantic consistency module, the learned positive samples are progressively distributed at the semantic area of the instance. As the training going on, high consistency samples become robust and appear in lighter colors. We also evaluate to see how inconsistency problem be solved by our method. Some qualitative results are presented in fig. 9, the typical inconsistency in which center-like annotations cannot handle presented in Fig. 1 are improved to a large extent. By utilizing the segmentation annotations, we found that the proportion of samples locate on background reduced around 15% (from 51.7% at initial to 36.1% when training finished) with the semantic consistency module.

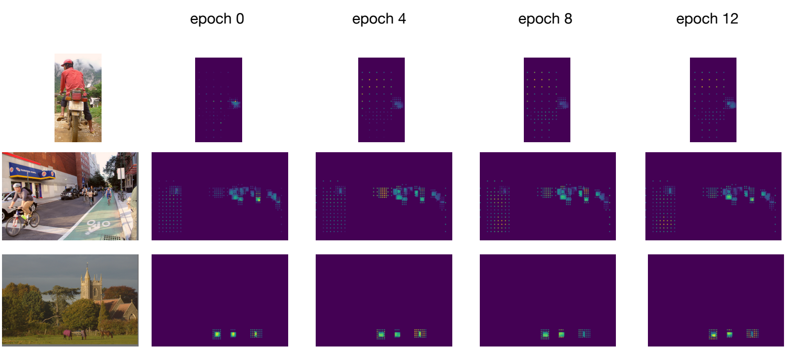
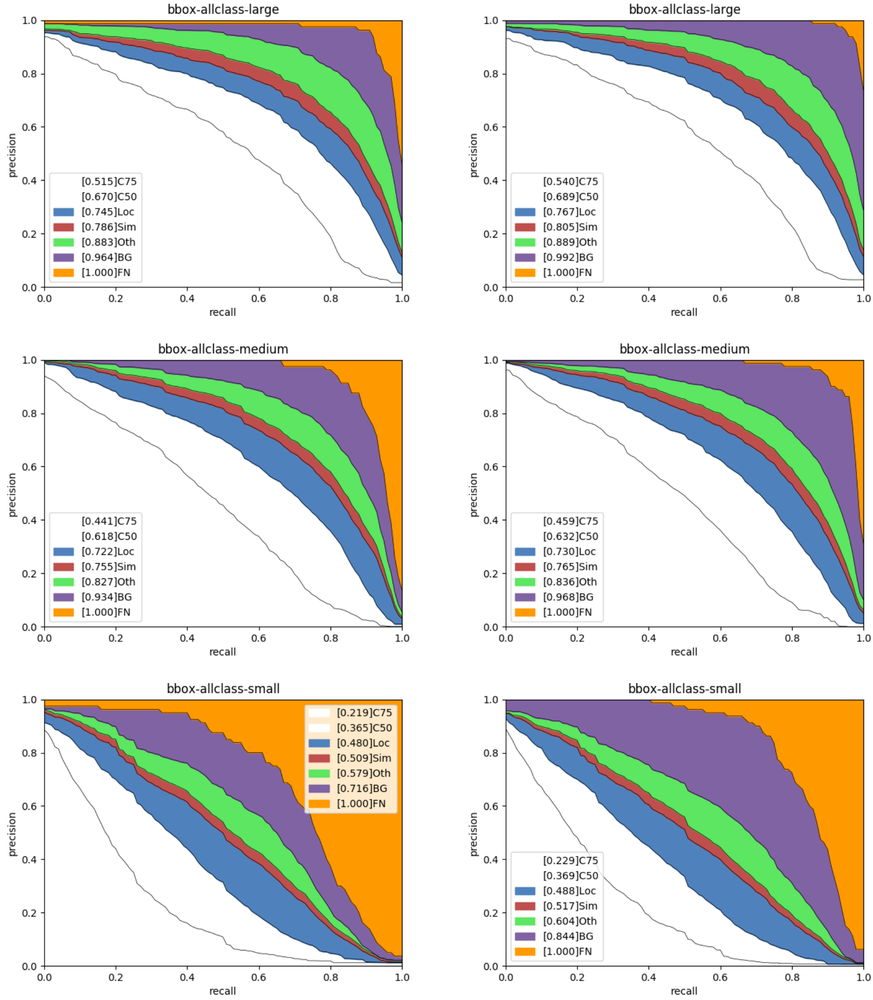


Fig. 8: Visualized examples of semantic consistency module. The left image of each row is the training data select from COCO *trainval*35*k*. Rest images on the right are the heatmaps of sampled points with semantic consistency module at different training epoch. Note that, entries of heatmaps represent the product results of IoU scores and classification scores. Sampled points with high IoU scores and high classification scores are highlighted in the heatmaps. Sampled points with low IoU scores or low classification scores are in dark colors. *Better viewed in colors and zoom in.*

Fig. 9: Visualization on center inconsistency examples. Sample points with high IoU scores and high classification scores are highlighted. The corresponding areas preferred by the semantic consistency module are marked as red boxes on the images. Images are select from COCO *trainval*35*k* and evaluated with the trained model with ResNet-50 as the backbone.

B Precision-Recall curves

The precision-recall(PR) curves of FCOS [27] and DDBNet under different evaluation settings provided by [18] on the *minival* split are shown in Fig. 10. PR curves were plotted for small-, medium- and large-scale objects in two models. The area in orange indicates the false negative(FN) portion of the evaluated dataset, which can be considered as the PR with all errors removed. The purple area presents the falsely detected objects. We can see that the area of orange in DDBNet is much lower than the one in FCOS [27], which means DDBNet is much robust after all background and class confusions removed.



(e) (f)

Fig. 10: **Precision Recall Curves.** Precision-recall(PR) curves of FCOS [27] and DDBNet under different evaluation settings provided by [18] on the *mini-val* split with ResNet-50 as backbone. (a)(c)(e): Evaluation results in FCOS. (b)(d)(f): Evaluation results in DDBNet. DDBNet gets better performance under the strict evaluation settings. Especially, we find out that DDBNet works much robust after all background and class confusions removed.