1 Summary

The paper tackles the problem of object detection across domains by incorporating two domain adaptation components (losses) on feature level and instance level into SOTA Faster R-CNN model to minimize domain discrepancy. For each component, domain classifiers are trained using adversarial learning strategy to learn domain invariant features. Furthermore, the two classifiers are linked with a consistency loss to encourage the RPN to be domain-invariant. Experiments section reveal that each of these losses are helpful in enhancing model performance individually and greatly outperform baseline together.

2 Strengths

- (i) Paper provides a logical analysis of the domain shift problem from a probabilistic perspective. (ii) The idea of applying a consistency regularization to alleviate the bias in estimating $P(B|I)$ is clever. (iii) Since batch size is usually small for training an object detection network, authors employ a patch-based domain classifier. This is helpful to increase the number of training samples for training the domain classifier. (iv) The model cam be trained end-to-end using the standard SGD optimization technique.

3 Weaknesses

- The paper does not provide any metric for explicitly checking if domain discrepancy is actually getting reduced using the method. A comparison of some discrepancy metric between baseline and proposed method would be ideal. tSNE visualizations of features before and after domain alignment would also be helpful.
- Figure 4 shows that proposed method performs best when target images are twice the size of source images. However, intuitively performance should be best when image sizes across domains are similar with a very slow decrease in performance as sizes differ. Further explanation behind the plot would have been appreciated.

4 Critique of Experiments

- The authors evaluate their model for object detection in three different domain shift scenarios.
- (i) SIM 10k dataset to Cityscapes dataset for detecting car category. Ablation studies show that both Image level and Instance level adaptation individually increase model's performance. Combining these together with consistency loss gives a performance boost of 9% over source only Faster R-CNN baseline.
- (ii) Clear Cityscapes dataset to Foggy Cityscapes dataset. Proposed model beats the baseline by significant margins across various object classes. This shows the proposed method is not biased towards a particular object category. (iii) KITTI to Cityscapes dataset and vice-versa. The proposed model significantly outperforms baseline both ways, which is a good sign.
- Study of effect of discrepancy in size of images from source to target domain reveals that proposed model is highly robust as compared to baseline, whose performance degrades quickly as images sizes differ. Further, model with Image-level alignment is more robust compared to Instance-level alignment which can be attributed to the fact that global domain shift is mainly tackled by image-level alignment.

5 Follow Ups/Extensions

- The idea of adding image level and instance level alignment can be extended to other tasks/architectures. For example, these can be used on Mask R-CNN for the task of human pose estimation across domains.
- To find an explicit domain discrepancy measure, classifier features can be used to find MMD loss with and without proposed domain adaptation method. tSNE plots can also be visualized on these features to find domain alignment changes these features.