1 Title:
Domain Adaptive Faster R-CNN for Object Detection in the Wild

2 Summary:
This paper proposes to enhance Faster R-CNN to support unsupervised domain adaptation at both instance and image level for object detection task. Adversarial domain adaptation is applied on convolution outputs to achieve image level adaptation and similar adaptation on ROI feature vectors yields instance level adaptation. Further, a consistency term ties the probability of belonging to a particular domain as predicted by domain classifier at instance level and image level. Experimental results are presented on SIM10k->Cityscapes, Cityscapes<->KITTI and Cityscapes -> Foggy Cityscapes adaptations.

3 Strengths:
i) The consistency between image and instance level adaptation is not easy to achieve when MMD/CORAL are used as the target domain bounding boxes are unavailable. On the other hand, the dual problem of consistency of domain probabilities with and without bounding box neatly ties up with the ROI and Conv Layer domain classifiers. ii) Instance level adaptation and image level adaptation tackle sufficiently different tasks namely, instance level focuses on adapting for object sizes, pose etc. while image level focuses on illumination, style. Hence, adapting both of these in conjunction makes intuitive sense. Further, this is supported by the conditional probabilities under covariate shift assumption. iii) Adapts upon one of the successful object detection modules, and hence suitable for practical use.

4 Weaknesses:
i) Covariate shift assumption is made while developing the theory. However, it's not necessarily valid in many practical scenarios. For example, foggy weather can have lesser and different type of traffic. ii) The algorithm is tied up to the the specific architecture and may not be scalable. For example, if Faster-RCNN is replaced by a deep network that doesn't have explicit region proposals, this method is not applicable by definition. iii) H-divergence is used instead of the $H\Delta H$. The error between minimizing H- divergence and target error is not direct as compared to $H\Delta H$

5 Analysis of Experiments:
i) Tables 1.,2.,3. AP results show that adding all the proposed losses yields best results in most cases. However, the marginal improvement upon adding either instance level or image level adaptation is higher as compared to additional improvement upon adding all terms. ii) Instance level adaptation is more useful than image level adaptation for Sim10->Cityscapes, Cityscapes->Foggy Cityscapes. But on Kitty<->Cityscapes, image level is better. This is because of the scale drift between cameras in the latter case. iii) Table 4. suggests that consistency term helps in overlap of region proposals with ground-truth bounding boxes which agrees with the premise of the dual problem $P(D|I) = P(D|B, I) <=> P(B|I) = P(B|D, I)$

6 Possible Extensions:
i) Relaxing covariate shift assumption by adding class weight balancing terms. ii) Aligning $H\Delta H$ by leveraging classifier discrepancy and imposing consistency indirectly by connecting the two bounding box regressors and the two object classifiers for $H\Delta H$.