1 Summary

The paper deals with the problem of domain adaptation for semantic image segmentation of urban scenes. The approach consists of two-fold method of first finding easy and useful tasks which are helpful in inferring label distribution of target images and landmark superpixels. This information is then used by pixel-wise discriminative segmentation network to ensure its predictions are consistent. Image-level label distribution inform the segmentation network how to update the predictions whereas landmark superpixel distribution tell the network where to update. In this way, the learned knowledge on easy tasks helps regularize the network predictions.

2 Strengths

- Unlike usual DA methods, this method does not depend on the assumption of existence of a common prediction function for both domain in transformed feature space, ie, they do no assume constant marginals across domain and hence more robust.
- (i) The idea of solving an easy task whose knowledge is then used to guide another prediction network is clever. The interesting part is we can come up with many such independent 'easy' tasks to further improve prediction network performance. (ii) The method is very simple with no feature level adaptation and is still very effective for the task, as can be seen from results table.

3 Weaknesses

- There seems to be a lack of analysis of feature level discrepancy measure. For example, tSNE plots with and without their DA procedure could give insights as to how well the network is adapting domain shift.
- The paper mentions using MMD and adversarial based losses on feature spaces does not lead to significant performance boost. Deeper insights as to why these standard DA methods failed for their task would be appreciated.

4 Critique of Experiments

- For predicting target label distributions, a multinomial logistic regression is learned because its chi^2 distance between groundtruth and predicted label distributions is smallest as compared to different methods.
- (i) The model trained with full-image label distribution loss gives a significant improvement over other methods. This shows that it is able to correct some obvious errors of the baseline network. (ii) Label distribution loss over full-image and landmark superpixels gives largest improvement. (iii) Loss over all superpixels has minimal improvement because of very high regularization at every superpixel.
- Results show that superpixel based method miss small objects and very accurate for objects that occupy large portions of image. Full image distribution loss based method performs better on small objects. The two methods complement each other and it is evident from results table where combination of two performs best.

5 Follow Ups/Extensions

- The paper uses multinomial logistic regression for target label prediction which gives a chi^2 of 0.27. This method seems very basic which can be built upon to get better predictions.
- We can divide the scene image into different spatial regions, and match the label distributions from the same spatial region respectively. In this way, we can hope to align features from two domains with similar spatial properties.