Agile bioinformatics - Literature review by Trang Le

Weakly-Supervised Semantic Segmentation by Iteratively Mining Common Object Features

Xiang Wang[1] Shaodi You[2,3] Xi Li[1] Huimin Ma[1]*

[1] Department of Electronic Engineering, Tsinghua University

[2] DATA61-CSIRO 3 Australian National University

Background:

Bridging the gap between high-level semantic (segmentation label) to low-level appearance (image detail) is a challenging task, especially when training with weakly labeled images. Since no pixel-to-pixel mask is available for training, previous models have to rely only on image labels to localize objects. Inaccurate and coarse discriminative object region detection can harm these model performances. To solve this problem, the authors proposed an iterative bottom-up and top-down framework which alternatively expands object regions and optimizes segmentation network.

Method:

The author proposed a Mining Common Object Features (MCOF) framework which contains RegionNet (top-down), PixelNet (bottom-up) and a saliency guide to iteratively produce more refined segmentation mask (details in Figure2).

RegionNet: Classification networks can produce initial localization of regions (hypothetically) containing significant common features about objects. After each training round (epoch), the refined object regions (output of PixelNet) are used as training data to predict object masks (new initial localization, higher accuracy).

PixelNet: From these initial localizations outputted from RegionNet, common object features were mined. The object regions were expanded with mined features.

Saliency-guide: To segment non-discriminative regions, saliency maps are then considered under Bayesian framework to refine the object regions.

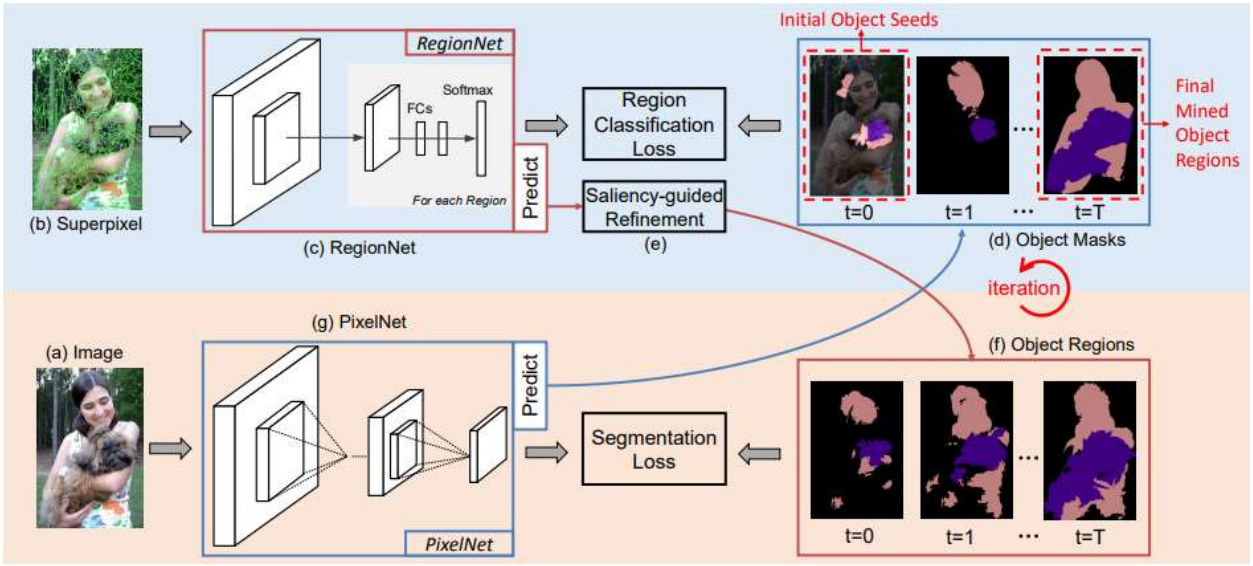

Figure 2. Pipeline of the proposed MCOF framework. At first (t=0), we mine common object features from initial object seeds. We segment (a) image into (b) superpixel regions and train the (c) region classification network RegionNet with the (d) initial object seeds. We then re-predict the training images regions with the trained RegionNet to get object regions. While the object regions may still only focus on discriminative regions of object, we address this by (e) saliency-guided refinement to get (f) refined object regions. The refined object regions are then used to train the (g) PixelNet. With the trained PixelNet, we re-predict the (d) segmentation masks of training images, are then used them as supervision to train the RegionNet, and the processes above are conducted iteratively. With the iterations, we can mine finer object regions and the PixelNet trained in the last iteration is used for inference. [Wang et al., 2018]

Experimental setup:

Data: Pascal VOC 2012 dataset: 20 class objects + 1 background class. For the segmentation task, it contains 1464 training, 1449 validation and 1456 test images. They used augmented data of 10,582 images as training set.

Hardware: not mentioned!

Results:

The MCOF framework outperformed previous state-of-the-art weakly-supervised semantic segmentation models by large margin. The performance metrics for comparison is mIOU (mean Intersection over Union of all 21 classes), and benchmark models included CCNN, EM-Adapt, MIL-sppxl, STC, DCSM, BFBP, AF-SS, SEC, CBTS and AE-PSL.

Conclusion: the MCOF framework has two main advantages over previous approached. First, the iterative bottom-up and top-down framework tolerates inaccurate initial object localization. By iteratively mining common object features, the model can progressively produce segmentation masks with improving accuracy. Second, saliency-guided refinement method can allow for non-discriminative regions in initial localization, hence increase segmentation accuracy. This framework established new state-of-the-art performance.