Note: Threshold Matters in WSSS: Manipulating the Activation for the Robust and Accurate Segmentation Model Against Thresholds

sgc

May 22, 2022

1 Introduction

Conventional CAMs only concentrate the most discriminative parts of the image, incurring the bottle neck of the existing methods. Authors think it is caused by the global threshold.
This paper proposed AMN(activation manipulation network) to generate th pseudo masks.

2 Method

The proposed AMN have 2 main learning objectives. 1, a per-pixel classification loss to reduce the activation balance inside the foreground and provides the large gap between the foreground and the background. 2, a lable conditioning module to eliminate the activation from the non-target classes.

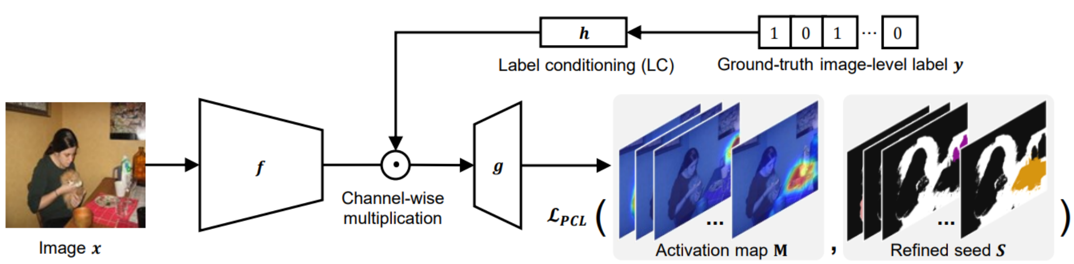

Figure 3. The overall framework of activation manipulation network (AMN). The refined seed from a classification network is used as noisy supervision to train AMN. The per-pixel classification loss (PCL) and label conditioning (LC) improve the pseudo-mask quality.

2.1 Overall training procedure

The training of the proposed WSSS framework consists of three stages: 1) seed generation, 2) pseudo-mask generation with the proposed AMN, and 3) final segmentation. Notably that this model changed the GAP and the last classification layer.
For seed generation, this module obtains noisy pixel-level annotations from image-level labels by applying CAM. Then, applying conditional random field (CRF) to refine the seeds. And the seeds are the target output of AMN.

2.2 Per-pixel classification

This is designed to solve the first problem aforementioned. PCL enforced a 2-level activation(e.g., 0 or 1). This has 2 main advantages:
1, this reduces the activation imbalance in side the foreground because the foreground should be assigned to the same activation value(0).And the two-level activation naturally retains the large activation gap between the foreground and background.
2, Getting rid of GAP.

2.3 Label conditioning

Compared with conventional classifier, LC has less classes (K+1, K is the number of classes appearing in the GT image level label.) while conventional classifier contains the total number of classes.
This is effective because:
1, it helps distinguish objects with similar appearances unless they really appear together in the image. It prevents false predictions due to confusing textures by allowing activation only if its class is corresponding to any of the input ground-truth image-level label.
2, LC can work as additional supervision information.
Notably LC should work on a higher feature map or it will cause unnecessary bias.
The final activation map is:

$$M = \sigma(g(f(x).h(y_{gt})))$$

3 Ablation Study

From the ablation study, authors show that LC not only subsidize the non-target activation but also improved the activation of target activation.
And they also find implementing LC to high level of the network helps.

4 Results

This method achieved the new SOTA gaining 1.6% over RIB. And in segmentation works it produces better results than models like IRN.

5 Summary

1, Limitations: This method cannot overcome co-occurring objects and inaccurate boundary problem. Additionally this method consists of 3 stages which harden the power consumption.
2, This paper proposed a new method AMN to generate CAMs. Mainly to solve 2 problem caused by GAP or the global threshold. This method set the SOTA.

References