Note: IMPORTANCE SAMPLING CAMS FOR WEAKLY-SUPERVISED SEGMENTATION

sgc

May 22, 2022

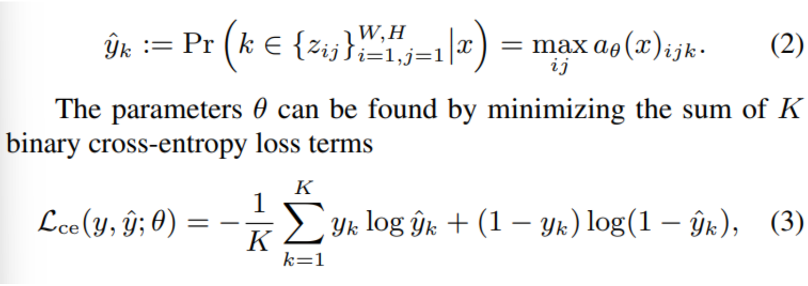
1 Introduction

The conventional CAMs have 2 main drawbacks: 1, mainly focus on discriminative regions;2, to produce CAMs without well-defined predictions contours. This paper solves the first problem by substituting GAP with importance sampling based on the class-wise probability mass function. And to deal with unclear prediction contour, authors formulated a new feature similarity loss term.

2 Method

2.1 Computing CAMs

The Loss:



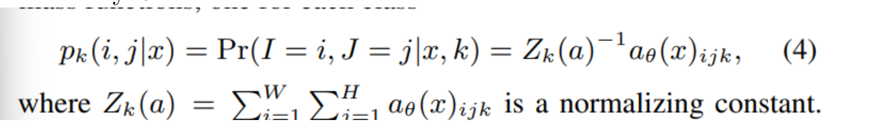
$$\hat{y}_k := \Pr\left(k \in \{z_{ij}\}_{i=1,j=1}^{W,H} \middle| x\right) = \max_{ij} a_\theta(x)_{ijk}. \qquad (2)$$

The parameters θ can be found by minimizing the sum of K binary cross-entropy loss terms

$$\mathcal{L}_{\text{ce}}(y, \hat{y}; \theta) = -\frac{1}{K}\sum_{k=1}^{K} y_k \log \hat{y}_k + (1 - y_k)\log(1 - \hat{y}_k), \qquad (3)$$

2.2 Importance sampling

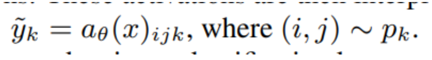
This paper adds a new L_{ce} by sampling a pixel for each class using probability induced by CAM.



$$p_k(i, j|x) = \Pr(I = i, J = j|x, k) = Z_k(a)^{-1} a_\theta(x)_{ijk}, \qquad (4)$$

where $Z_k(a) = \sum_{i=1}^{W}\sum_{i=1}^{H} a_\theta(x)_{ijk}$ is a normalizing constant.

And the new activation is generated by



$\tilde{y}_k = a_\theta(x)_{ijk}$, where $(i, j) \sim p_k$.

And the Final Loss is:

$$\mathcal{L}_{\text{cls}}(y, \hat{y}, \tilde{y}) = (1 - \lambda)\mathcal{L}_{\text{ce}}(y, \hat{y}) + \lambda\mathcal{L}_{\text{ce}}(y, \tilde{y}),$$

2.3 Feature similarity loss

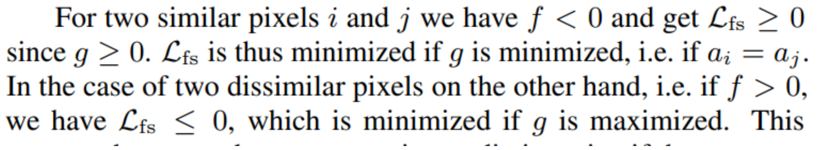
The main idea is to penalize dissimilar predictions of nearby similar pixels and vice versa.
The gating function: $g(a_i, a_j) = \frac{1}{2}\|a_i - a_j\|_2^2$
Function f : $f(\delta(x_i - x_j)) = tanh(\mu + log(\delta/(1 - \delta)))$
Thus the total Loss function is defined as follows:

$$L_{fs}(a, x) = -(HW)^{-2}\sum_{ij}\omega_{ij}g(a_i, a_j)f(x_i, x_j)$$

where ω_{ij} is the weight defined using a Gaussian neighborhood:

$$\omega_{ij} = (2\pi\sigma^2)^{-1}exp(-\|p_i - p_j\|_2^2/(2\sigma^2))$$

The explanation of functions above is detailed in the paper. This makes the network to classify larger regions to the same class.



For two similar pixels i and j we have $f < 0$ and get $\mathcal{L}_{\text{fs}} \geq 0$ since $g \geq 0$. $\mathcal{L}_{\text{fs}}$ is thus minimized if g is minimized, i.e. if $a_i = a_j$. In the case of two dissimilar pixels on the other hand, i.e. if $f > 0$, we have $\mathcal{L}_{\text{fs}} \leq 0$, which is minimized if g is maximized. This

And in this paper, $\mu, \mathbf{and}\sigma$ in f are learned instead of chosen manually.

3 Results

This method reaches similar performance in terms of region similarity compared to current state-of-the-art methods.

4 Summary

This paper proposed 2 new method to improve the performance in WSSS. Importance sampling results CAMs to cover larger regions of object. Newly added feature similarity loss term significantly improves the contour accuracy of the segmentation result.

References