Note: Non-Salient Region Object Mining for Weakly Supervised Semantic Segmentation

sgc

May 28, 2022

1 Introduction

This paper proposed a non-salient region object mining method(Potential Object Mining POM, Non-Salient Region Masking NSRM) for WSSS to exploit the distant and disjoint surrounding information. And authors introduce a graph-based global reasoning unit to strengthen the classification network's capability in activating the object features outside the salient region.
Main contributions:
1, This paper introduced a global reasoning unit to capture global relations.
2, This paper proposed a potential object mining module to discover more objects in the non-salient region.
3, This paper proposed a non-salient region masking module with a dilation policy to generate mesked pseudo labels.

2 Method

Framework:

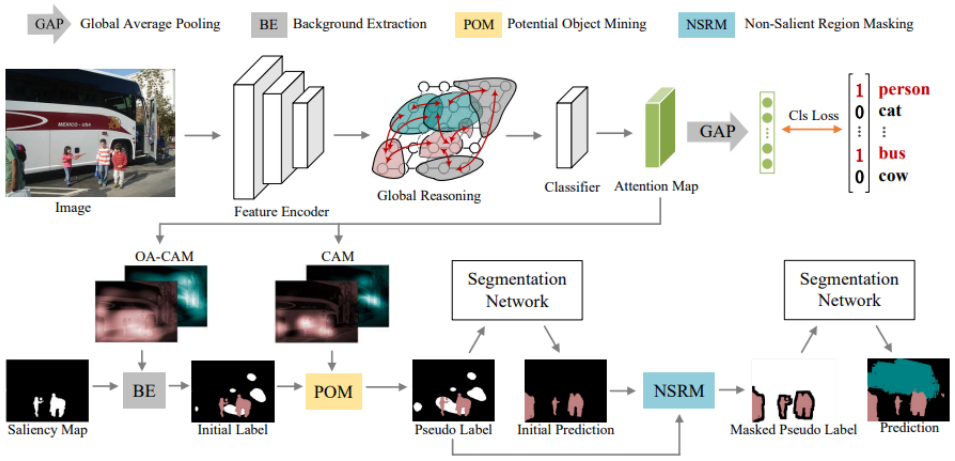

Figure 2. The architecture of our proposed approach. We train a classification network to generate the class activation maps (CAMs) and online accumulated class attention maps (OA-CAMs). A graph-based global reasoning unit is inserted into the classification network to help activate the objects outside the salient region. After obtaining the initial label with background extraction (BE), we utilize the potential object mining module (POM) to discover more objects missed in the initial label. We further leverage the non-salient region masking module (NSRM) to generate masked pseudo labels for the training of the segmentation network. Best viewed in color.

2.1 CAM Generation

A graph based global reasoning unit is introduced before the final classifier. The features $\mathbf{X}$ is generated by the encoder, and then is projected to a latent space through the projection function:

$$v_i - b_i X = \sum_i b_{ij} x_j$$

where b is learnable projection weights. Then a graph convolution is applied to capture the features in the new space:

$$Z = ((I - A_g)V)W_g$$

where A_g is learnable adjacency matrix. W_g denotes the state update function. Reverse projection Y is conducted to project the feature back to the original space:

$$y_i = d_i Z = \sum_j d_{ij} z_j$$

The loss is:

$$L_{cls} = -\frac{1}{C}\sum_{c=1}^{C} y_c \log \sigma\left(p_c\right) + \left(1 - y_c\right) \log\left[1 - \sigma\left(p_c\right)\right]. \tag{4}$$

OAA is also implemented to generate OA-CAMs to have more entire regions.

2.2 Potential Object Mining

Use pseudo labels generated from OA-CAM after the background extraction(BE). To enlarge the activation region, potential object mining(POM) is introduced. The class adaptive threshold:

$$T_c = \begin{cases} MED(v) & \text{if } \exists\,(i,j)\,, \text{s.t. } l_{ij} = c \\ TQ(v) & \text{otherwise} \end{cases}.$$

where v is the attention values of pixels in the CAM, their locations are selected as follows:

$$p = \begin{cases} \{(i,j)|l_{ij} = c\} & \text{if } \exists\,(i,j)\,, \text{s.t.} l_{ij} = c \\ \{(i,j)|a_{ij} > T_{bg}\} & \text{otherwise} \end{cases},$$

where l is the initial label($l = 0$ means it is a background pixel), T_{bg} is the background threshold, MED is the median, TQ denotes the top quartile.
Then adjust the initial label as follows:

$$l_{ij} = \begin{cases} 255 & \text{if } \exists\, c, (i,j)\,, \text{s.t. } l_{ij} = 0, a^c_{i,j} > T_c \\ l_{ij} & \text{otherwise} \end{cases}.$$

The background pixels($l = 255$ white) are ignores while training.

2.3 Non-Salient Region Masking

The final output is generated through 2 segmentation process. as illustrated in Figure 2.
The NSRM: First expand the object region in the initial prediction with the guidance of the pseudo labels. Then extract the object mask from the expanded prediction. Expand the object mask with a dilation map. Finally, a masking operation is applied to the expanded prediction map to get the masked pseudo labels.
All in all, this method get more information the designate some background pixels (out of saliency map) to 225 to exclude them from training. The goal of NSRM is to exclude those useless background information?

3 Result

Results demonstrate the effectiveness of mining the objects in the non-salient region for the task of weakly supervised semantic segmentation.

4 Summary

The main method introduced in this paper:POM(to reduce the false-negative rate), NSRM(to exert the self correction ability from segmentation task itself), and the gragh-based global reasoning unit to help capture the global information.

References