Note: Pseudo-mask Matters in Weakly-supervised Semantic Segmentation

sgc

May 26, 2022

1 Introduction

This paper proposed several designs to push the performance in WSSS ti SOTA:
(i) Coefficient of Variation Smoothing to smooth the CAMs adaptively;
(ii) Proportional Pseudo-mask Generation to project the expanded CAMs to pseudo-mask based on a new metric indicating the importance of each class on each location, instead of the scores trained from binary classifiers.
(iii) Pretended Under-Fitting strategy to suppress the influence of noise in pseudo-mask;
(iv) Cyclic Pseudo-mask to boost the pseudo-masks during training of fully supervised semantic segmentation (FSSS). The main contributions are presented below:
1, Authors generated high-quality pseudo-masks by the proposed Proportional Pseudo-mask Generation with Coefficient of Variation Smoothing. This expands the activation area of CAMs.(Basically all newly proposed methods in WSSS tend to enlarge the are of activation or clarify the boundary of object)
2, Utilization of pseudo-masks via reducing the influence of noise by our Pretended Under-fitting Strategy, and narrow the gap between ground-truths and pseudo-masks via Cyclic Pseudo-masks.
3, The proposed method actually achieves new SOTA.

2 Method

Preliminary: $\mathbf{X}$ is the CAM, the goal is project CAM to pseudo-mask $\mathbf{Y}$. The image is $\mathbf{I}$.

2.1 Pseudo-mask Generation with CRF

Min-Max Normalization is applied on $\mathbf{X}$:

$$norm(\boldsymbol{X}_{c,h,w}) = \frac{\boldsymbol{X}_{c,h,w} - min(\boldsymbol{X}_{c,:,:})}{max(\boldsymbol{X}_{c,:,:}) - min(\boldsymbol{X}_{c,:,:})}, \forall c, h, w \tag{1}$$

where h and w are the coordinates on the CAMs, and c represents the channel index.

Then the normalized pixel on background matrix is constructed from the normalized foreground pixels:

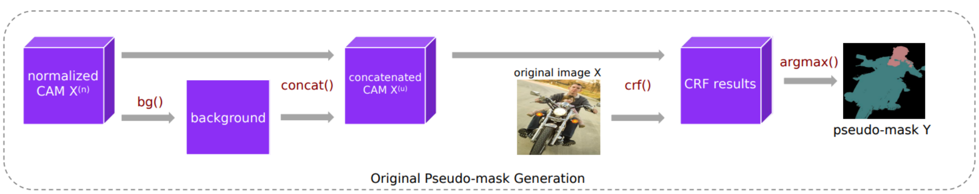
$$bg(\boldsymbol{X}^{(n)}_{:,h,w}) = (1 - max(\boldsymbol{X}^{(n)}_{:,h,w}))^{\alpha}, \forall h, w$$

Then concatenate the foreground and background.

$$\mathbf{X}^{(u)} = concat(bg(\mathbf{X}^{(n)}), \mathbf{X}^{(n)})$$

Finally the pseudo-mask is identified via argmax operation:

$$\mathbf{Y} = argmax(crf(\mathbf{I}, \mathbf{X}^(u)))$$



2.2 Coefficient of Variation Smoothing

Firstly, the authors introduce the the Coefficient of Variation (c_v) as the metric for the foreground pixels $X^{(f)}_{c,:,:}$ whose scores are higher than threshold t in normalized metric $X^{(n)}_{c,:,:}$ at channel c. the cv function is defined:

$$cv(\boldsymbol{X}^{(n)}_{c,:,:}) = \frac{\sqrt{\mathbb{D}(\boldsymbol{X}^{(f)}_{c,:,:})}}{\mathbb{E}(\boldsymbol{X}^{(f)}_{c,:,:})},$$

where $\mathbb{D}$ counts the deviation δ^2 and $\mathbb{E}$ counts the mean μ.
The cvs function:

$$cvs(\boldsymbol{X}^{(n)}_{c,h,w}, (c_v)_c) = (X^{(n)}_{c,h,w})^{(1-s\times(c_v)_c)}, \forall c, h, w \tag{6}$$

Experimentally, CVS works better with strong augmentations.

2.3 Proportional Pseudo-mask Generation

The proportion function p is defined as:

$$p(\boldsymbol{X}^{(n)}_{c,h,w}) = \frac{\boldsymbol{X}^{(n)}_{c,h,w}}{sum(\boldsymbol{X}^{(n)}_{(c,:,:)} \cdot \boldsymbol{X}^{(m)}_{c,:,:})}, \forall c, h, w \qquad (8)$$

Algorithm 1 Proportional Pseudo-mask Generation

Input: image $\boldsymbol{I}$ and CAM $\boldsymbol{X} \in \mathbb{R}^{C \times H \times W}$

Output: pseudo-mask $\boldsymbol{Y} \in \mathbb{R}^{H \times W}$

1: normalize the CAM: $\boldsymbol{X}^{(n)} = norm(\boldsymbol{X})$

2: count c_v for each class: $c_v = cv(\boldsymbol{X}^{(n)})$

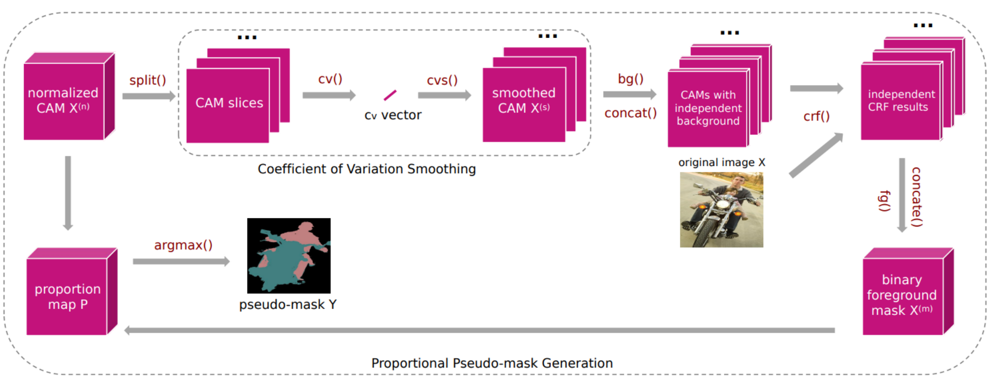
3: smooth the CAM: $\boldsymbol{X}^{(s)} = cvs(\boldsymbol{X}^{(n)}, c_v)$

4: compute binary mask with crf:
$\boldsymbol{X}^{(m)}_{c,:,:} = fg(crf(\boldsymbol{I}, bg(\boldsymbol{X}^{(s)}_{c,:,:}))), \forall c$

5: count the proportion map: $\boldsymbol{P} = p(\boldsymbol{X}^{(n)})$

6: generare pseudo-mask: $\boldsymbol{Y} = argmax(\boldsymbol{X}^{(m)} \cdot \boldsymbol{P})$

Note that $\mathbb{X}^m$ is a binary mask matrix.



2.4 Pretended Under-fitting Strategy

The proposed method attempts to suppress the noise during the model training:

$$\ell(\boldsymbol{L}) = \begin{cases} mean(\boldsymbol{L}) & mean(\boldsymbol{L}) >= \beta \\ mean(pus(\boldsymbol{L})) & mean(\boldsymbol{L}) < \beta \end{cases}$$

where the $\mathbf{L}$ is the loss map for pixels from the BCE loss.

$$pus_{clamp}(\boldsymbol{L}) = \begin{cases} \boldsymbol{L}_{h,w} & \boldsymbol{L}_{h,w} < \kappa \\ \boldsymbol{L}_{h,w} \cdot \frac{\kappa}{\boldsymbol{L}_{h,w}} & \boldsymbol{L}_{h,w} >= \kappa \end{cases} \forall h, w \quad (11)$$

$$pus_{pow}(\boldsymbol{L}) = \boldsymbol{L}^{\kappa} \quad (12)$$

$$pus_{ignore}(\boldsymbol{L}) = \begin{cases} \boldsymbol{L}_{h,w} & \boldsymbol{L}_{h,w} < \kappa \\ 0 & \boldsymbol{L}_{h,w} >= \kappa \end{cases} \forall h, w \quad (13)$$

2.5 Cyclic Pseudo-mask and Overall Pipeline

Replace the pseudo-mask on training dataset as the predictions from the trained model on it.

3 Result

These methods introduce more detailed information and achieved new SOTA.

4 Summary

To solve the problems occurring th WSSS, these methods are introduced. CVS is designed to smooth the CAM by its distribution statistics, Pretended Under-fitting Strategy is designed to deal with noisy CAMs. Cyclic Pseudo-mask is effective. Finally the proposed method achieved SOTA.

References