Note: Class Re-Activation Maps for Weakly-Supervised Semantic Segmentation[1]

sgc

April 23, 2022

1 Introduction

Authors ascribe the generating unsatisfactory pseudo masks to BCE Loss used in CAM. So they proposed Reactivating the converged CAM with BCE by using softmax cross-entropy loss (SCE), dubbed ReCAM.
Contributions:
1, This paper proposed ReCAM to generate pseudo labels

2 Method

2.1 ReCAM Pipeline

Backbone and Multi-Label Features Authors use a standard ResNet-50 as the backbone. Given image x and multi-hot class label y, the output is generated from a encoder.

FC Layer-1 with BCE Loss Just like conventional CAM the prediction logits are denoted as:

$$z = FC_1(GAP(f(x)))$$

then z and teh image level label y are used to compute a BCE Loss:

$$\mathcal{L}_{bce} = -\frac{1}{K}\sum_{k=1}^{K} y\left[k\right]\log\sigma\left(z[k]\right)+\left(1-y[k]\right)\log\left[1-\sigma\left(z[k]\right)\right], \tag{1}$$

Extracting CAM Extract CAM for each class given the feature and weights:

$$\mathrm{CAM}_k(\boldsymbol{x}) = \frac{\mathrm{ReLU}\left(\boldsymbol{A}_k\right)}{\max\left(\mathrm{ReLU}\left(\boldsymbol{A}_k\right)\right)}, \boldsymbol{A}_k = \mathbf{w}_k^{\top} f(\boldsymbol{x}),$$

denote CAM to M.

Single-Label Feature Extract class-specific feature map $f_k(x)$:

$$f_k^c(x) = M_k \bigotimes f^c(x)$$

K is the total number of class.

FC Layer-2 with SCE Loss New prediction logits:

$$z_k' = FC_2(GAP(f_k(x)))$$

And then the problem is converted to SCE Loss based model.
The SCE Loss:

$$\mathcal{L}_{sce} = -\frac{1}{\sum_{i=1}^{K} \boldsymbol{y}[i]} \sum_{k=1}^{K} \boldsymbol{y}[k] \log \frac{\exp(\boldsymbol{z}_k'[k])}{\sum_j \exp(\boldsymbol{z}_k'[j])},$$

where y is the one-hot label.
Therefore the total loss is:

$$L_{ReCAM} = L_{bce} + \lambda L_{sce}$$

Extracting ReCAM ReCAM is generate as:

$$\text{ReCAM}_k(\boldsymbol{x}) = \frac{\text{ReLU}(\boldsymbol{A}_k)}{\max(\text{ReLU}(\boldsymbol{A}_k))}, \boldsymbol{A}_k = \mathbf{w}_k''^{\top} f(\boldsymbol{x}),$$

where w'' has many options: 1) w, 2) w', 3) $w \bigoplus w'$, 4) $w \bigotimes w'$.(because we have 2 FC Layers)

Refining ReCAM 1, AdvCAM: iteratively refines ReCAM by perturbing images x through adversarial climbing:

$$\begin{aligned}\boldsymbol{x}^t &= \boldsymbol{x}^{t-1} + \xi \nabla_{\boldsymbol{x}^{t-1}} \mathcal{L}_{adv}, \\ \mathcal{L}_{adv} &= \boldsymbol{y}^{t-1}[k] - \sum_{j \in K \setminus k} \boldsymbol{y}^{t-1}[j] \\ &\quad - \mu \left\| \mathcal{M} \otimes \left| \text{ReCAM}_k\left(\boldsymbol{x}^{t-1}\right) - \text{ReCAM}_k\left(\boldsymbol{x}^0\right) \right| \right\|_1,\end{aligned} \tag{10}$$

where $M = \mathit{1}(ReCAM_k(x^(t-1)) > 0.5)$ 2, IRN: takes ReCAM as the input and trains an inter-pixel relation network (IRNet) to estimate the class boundary maps B. Then, it applies a random walk to refine ReCAM with B and the transition probability matrix T:

$$\text{vec}(\boldsymbol{M}_k') = \mathbf{T}^t \cdot \text{vec}(\text{ReCAM}_k(\boldsymbol{x}) \otimes (1 - \mathcal{B})),$$

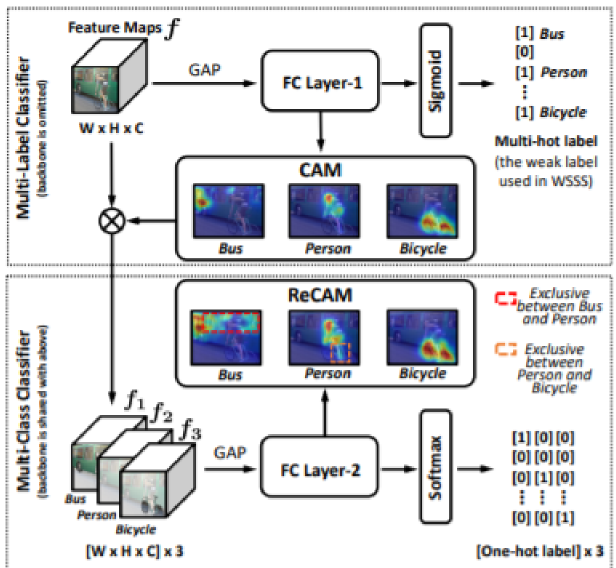

Figure 3. The training framework of ReCAM. In the upper block, it is the conventional training of multi-label classifiers using BCE. The feature extraction via backbone is omitted for conciseness. We extract the CAM for each class and then apply it (as a normalized soft mask) on the feature maps f to obtain the class-specific feature f_k. In the lower block, we use f_k and its single label to learn multi-class classifiers with SCE loss. The gradients of this loss are backpropagated through the whole network including backbone.

2.2 Justification: BCE *vs* CE

The Author gives a brief justification between BCE and CE theoretically and empirically. More details is in the paper.

3 Results

1, SCE only(w/o FC2) works worse and SCE+BCE gain little w/o FC2.
2, 1) w, 2) w', 3) $w \bigoplus w'$, 4) $w \bigotimes w'$ all works better than the baseline. $w \bigotimes w'$ works the best in VOC and w works best in MS COCO(perhaps because FC2 is not trained well).
3, ReCAM is not sensitive to λ, this paper use 1.
4, ReCAM shows consistent advantages over CAM on both VOC and MS COCO.

4 Summary

This paper proposed a simple yet effective method ReCAM by applying SCE into the BCE based model to reactivate the model. And author points that BCE Loss is responsible for low efficiency in CAM and also gives a convincing compare between CE and BCE loss. This paper encourages us to think more when applying BCE Loss.

References

[1] Z. Chen, T. Wang, X. Wu, X.-S. Hua, H. Zhang, and Q. Sun, “Class re-activation maps for weakly-supervised semantic segmentation,” *arXiv preprint arXiv:2203.00962*, 2022. (document)