Attention-based Selection Strategy for Weakly Supervised Object Localization

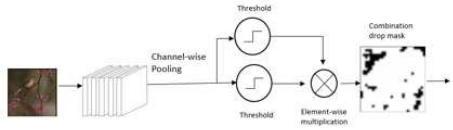
Zhenfei Zhang, Tien D.Bui
Concordia University, Montreal, Canada

Introduction

Weakly Supervised Object Localization aims to locate objects by only using image-level labels. Most techniques uses CAM that can only highlight the most discriminative parts. Nowadays, some hiding methods are proposed to address the issue of CAM [1]. However, they hide the images very blind. In this work, we propose a new method that can dynamically generate drop masks according to different input feature maps.

Proposed Methods

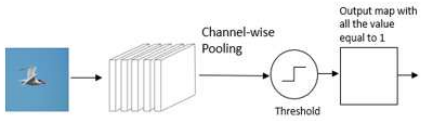
We divided input feature maps into three categories and for each class the drop mask is different.



Algorithm 1 Selection strategy

Input: $x \subseteq R^{H \times W \times C}$: feature map; a: the number of the most discriminative pixels in one featuere map; b: the total pixel number of one feature map; v: the most discriminative pixel dispersion; d: total pixels of target
Output: y : dropped feature maps

```
for i − 0 → batch_size do
    if v >= dispersion threshold and a >= 0.05b then
        selected_map[i] = a drop mask that can remove unhelpful information
    else if d/b < small area threshold or a/d > drop mask proportion threshold then
        selected_map[i] = a feature map that all the values are 1
    else
        selected_map[i] = a drop mask to remove the most discriminative part
    end if
end for
return y = x · selected_map
```



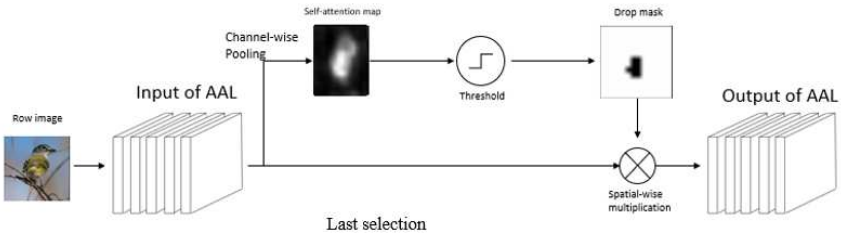


Figure 1: Algorithm and the diagrams

Results

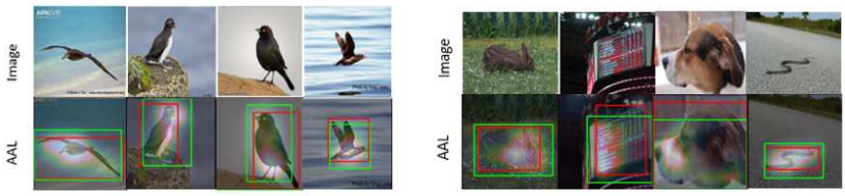


Figure 2: Results

Conclusions

- we propose a simple but effective method for Weakly Supervised Object Localization task.
- Our method dynamically generates the drop masks according to different input feature map.

References

[1] B. Zhou, A. Khosla, A. Lapedriza, A. Olive, and A. Torralba. Learning deep features for discriminative localization. pages 2921–2929. IEEE Conference on computer vision and pattern recognition, 2016.

UNIVERSITÉ Concordia UNIVERSITY