Dual-attention Guided Dropblock Module for Weakly Supervised Object
Junhui Yin, Siqiq Zhang, Dongliang Chang, Zhanyu Ma*, and Jin Guo
Beijing University of Posts and Telecommunications

Abstract: Attention mechanisms is frequently used to learn the discriminative features for better feature representations. Attention mechanisms is frequently used to learn the discriminative features for better feature representations. In this paper, we extend the attention mechanism to the task of weakly supervised object localization (WSOL), and propose the dual-attention guided dropblock module (DGDMM), which aims at learning the informative and complementary visual patterns for WSOL. This module contains two key components, the channel attention guided dropout (CAGD) and the spatial attention guided dropout (SAGD). To model channel dependencies, the CAGD ranks the channel attentions and treats the top-k attentions with the largest magnitudes as the important ones. It also keeps some low-valued elements to increase their value if they become important during training. The SAGD can efficiently remove the most discriminative information by erasing the contiguous regions of feature maps rather than individual pixels. This guides the model to capture the less discriminative parts for classification. Furthermore, it can also distinguish the foreground objects from the background regions to alleviate the attention misdirection. Experimental results demonstrate that the proposed method achieves new state-of-the-art localization performance.

1. Introduction
Weakly supervised object localization (WSOL) requires less detailed annotations to identify the object location in a given image compared to the fully-supervised learning. WSOL is a challenging task since neural networks have access to only image-level labels (‘cat’ or ‘no cat’) that confirms the existence of the target object, but not the guidance of the expensive bounding box annotations in an image.

2. Method
Deep networks implemented with DGDMM incorporate the image classification and WSOL. In an end-to-end learning manner, the proposed method captures the complementary and discriminative visual features for precise object localization and achieves good result of image classification.

3. Results

Table 2. Quantitative evaluation results on CUB-200-2011 listed set with the state-of-the-art results.

4. Conclusions
In this paper, we proposed a simple yet effective dual-attention guided dropblock module (DGDMM) for weakly supervised object localization (WSOL). We designed two key components of DGDMM, namely the channel attention guided dropout (CAGD) and the spatial attention guided dropblock (SAGD), and integrated them with the deep learning framework. The proposed method hides the most discriminative part and then encourages the CNNs model to discover the less discriminative part. We defined a pruning strategy so that CAGD can be adapted to model the interdependencies across the channels. In addition, SAGD can not only efficiently remove the information by erasing the contiguous regions of feature maps rather than the independent individual pixels, but also sense the target objects and background regions to alleviate the attention misdirection. Compared to some existing WSOL techniques, the proposed method is lightweight, and can be easily employed to different CNNs classifiers. We also achieve new SOTA localization accuracy on CUB-200-2011, Stanford Cars, and ILSVRC.

References

Acknowledgements: This work was supported in part by the National Key R&D Program of China under Grant 2019YFF0303300 and under Subject II No. 2019YFF0303302.

* Please contact mushay@bupt.edu.cn for further information.