Adapting Grad-CAM for Embedding Networks
Supplementary Material

Lei Chen[1,2] Jianhui Chen[1] Hossein Hajimirsadeghi[1] Greg Mori [1,2]
[1]Borealis AI [2]Simon Fraser University
{lei.chen,jimmy.chen,hossein.hajimirsadeghi,greg.mori}@borealisai.com

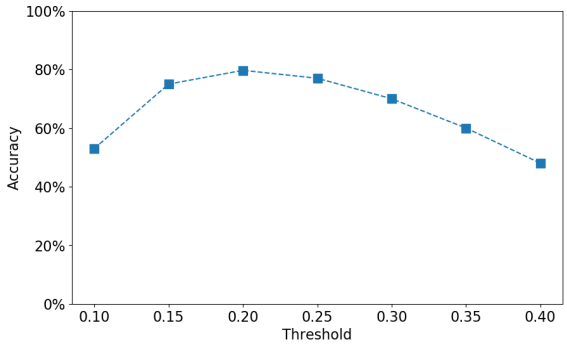

Figure 1. Localization accuracy ($\text{IoU}_{0.5}$) with different thresholds on CUB testing set.

1. More results

Quantitative results: On the CUB200 dataset, the original images have different resolutions. For visual attention score (*i.e.* bounding box score and mask score) evaluation, we scale the images such that their shorter side is 256 pixels and then center crop the image to 224×224.

For the weakly supervised localization accuracy [1, 2], the accuracy is measured by IoU (intersection over union) values between the ground truth bounding box and a bounding box generated from the heatmap on the full-size image (no image scale/crop). To generate a bounding box from the heatmap, we first use a threshold to binarize the heatmap. Then, we take the bounding box that covers the largest connected component in the binary image. Figure 1 shows the $\text{IoU}_{0.5}$ accuracy as a function of the threshold. Our method achieves the accuracy of 79.7% when the threshold is 0.2, which is much higher (79.7% vs. 50.6%) than a recent work [2]. Moreover, the accuracy is quite stable (above 75%) when the threshold is in the range of $[0.15, 0.25]$.

Qualitative results: Figure 2 shows more qualitative results of our method. Our method successfully visualizes important regions in the images. For example, in the last row, our method highlights the *peak* and the *neck* of the birds as these parts distinguish them from other species.

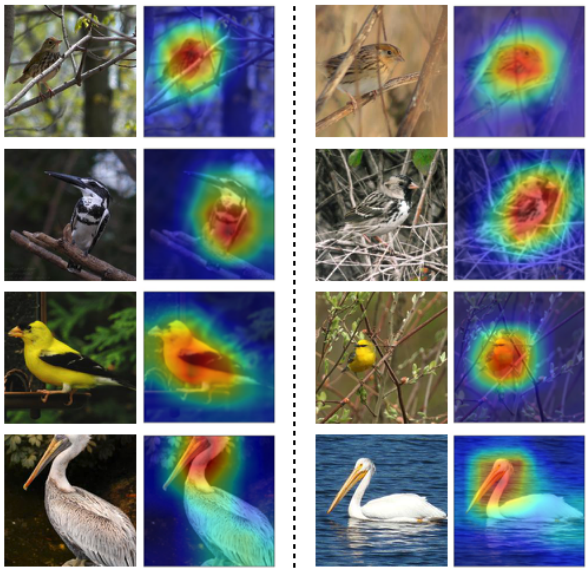

Figure 2. Qualitative results on CUB200. Left: from training set; right: from testing set. In each row, the grad-weights of the testing image are transferred from the training image.

References

[1] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[2] S. Zhu, T. Ynag, and C. Chen. Visual explanation for deep metric learning. *arXiv preprint arXiv:1909.12977*, 2019.