Counterfactual Attention Learning for Fine-Grained Visual Categorization and Re-identification
Supplementary Material

A. More Visual Results

To have an intuitive understanding of our counterfactual attention learning method, we compare the attention maps of our models and the baselines models on CUB [3], Stanford Cars [1] and Aircraft datasets [2]. The more visual results are presented in Figure 1. We see our method helps the attention models make correct predictions by 1) reducing the misleading and scatter attentions and 2) encouraging the model to focus on the main clues for classification and explore more discriminative regions.

B. More Implementation Details

Different types of counterfactual attentions. We compared four different counterfactual attentions in our experiments. The details about how to generate them are described as follows.

- **Random Attention.** We use randomly generated attention maps as the counterfactual attentions. The attention value for each location is sampled from a uniform distribution $\mathcal{U}(0, 2)$.
- **Uniform Attention.** We simply set the attention value for each location to the average value of the real attention maps.
- **Reversed Attention.** We reverse the attention maps by subtracting the original attention from the maximal attention value of each sample.
- **Shuffle Attention.** We randomly shuffle the attention maps along the batch dimension.

Attention Regularization Strategy. We investigated several regularization strategies on the baseline attention model to verify the effectiveness of our method. The details about these regularization strategies are described as follows.

- **Attention Dropout.** We apply the Dropout method to the attention maps.
- **Entropy Regularization.** We add an extra term to the loss function to maximize the entropy of the attention maps.
- **Attention Normalization.** We add ℓ_2 normalization to the attention maps.

References

[1] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *ICCVW*, pages 554–561, 2013. 1

[2] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013. 1

[3] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 1

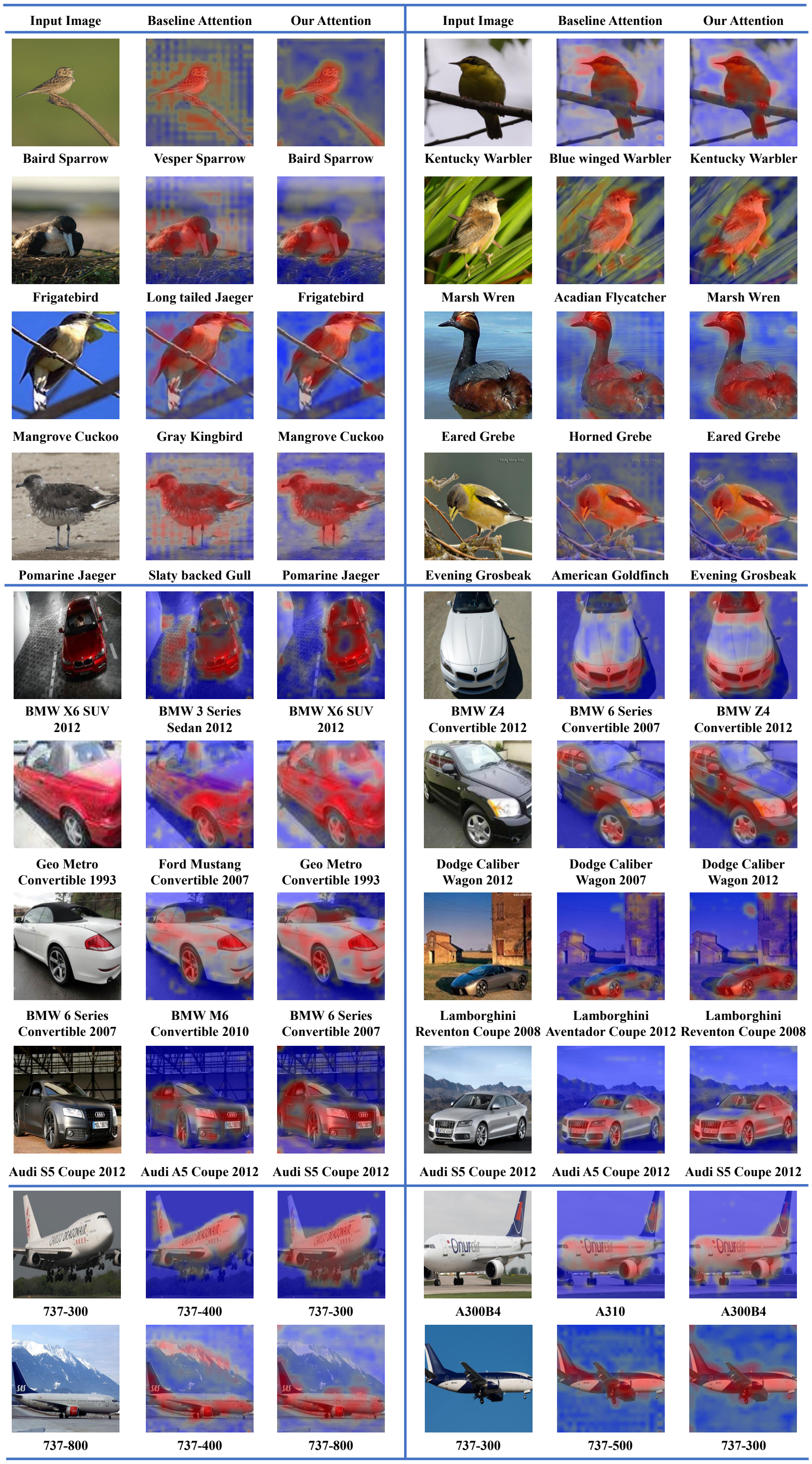

Figure 1: Visualization of the attention maps of our models and the baseline models. We see our method helps the attention models make correct predictions by 1) reducing the misleading and scatter attentions and 2) encouraging the model to focus on the main clues for classification and explore more discriminative regions. Best viewed in color.