Analyzing Grad-GAM and related methods

This report gives an overview of the insights obtained on implementing Task 3 from Assignment 4. The main idea of the report is to analyze the visualizations (for the 10 given ImageNet data points) responsible for the main prediction using Grad-CAM and related methods with a ResNet-50 model. Grad CAM is a gradient-based, CNN-specific technique that generates a heatmap of the important regions in an image for a specific class prediction. Similarly, Ablation CAM is also model-specific, however, instead of using gradients, it removes or nullifies parts of the feature map and observes the effect on the class scores. Score CAM, on the other hand, does not rely on the gradients and uses the model's actual scores to determine the important regions in the activation map. The visualizations obtained using Grad-CAM, Ablation CAM, and score-CAM methods can be obtained here. The implementation can be found here.

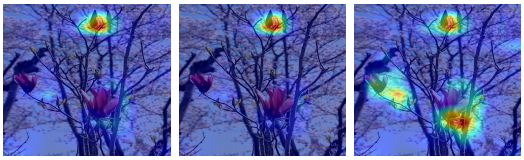

Figure 1: Ablation,Grad, Score CAM

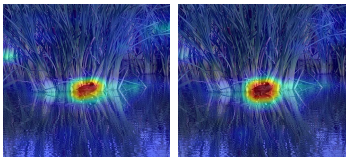

Figure 2: Score and Ablation CAM

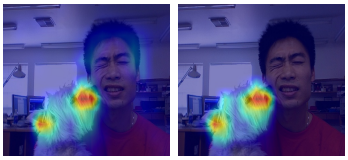

Figure 3: Without and with smoothing

Figure 4: Ablation, Grad, Score CAM

Based on the images in Figure 1 of the kite, we see that Score CAM specifically highlights all the flowers in the heat map with warm colors like red and orange, suggesting higher attention and importance to those features. The yellow and green areas are less important, yet significantly influence the model's prediction. In the case of the first image in Figure 1, which is generated using the Ablation CAM technique, it is noted that similar to Score CAM, it does highlight the blooming flowers, but with more focus on the one at the top. The intensity of the red and orange colors is also relatively higher than that in Score CAM, suggesting more precision. The second image in Figure 1, which is generated using Grad CAM, follows a similar pattern as Ablation CAM. In totality, we see that Grad CAM, Ablation CAM, and Score CAM identify and highlight the key features of the image correctly, with a slight variation and attention, based on their methodology. For Figure 2, with images of an American coot, in the case of ablation CAM, the coot is highlighted with more precision than in the score CAM. Using this information and heat maps for other images, it can be generalized that Ablation CAM tries to highlight the most important areas with precision, whereas Score CAM does so, for a broader context sometimes, with slightly less precision.

During the experiments, eigen smooth and aug smooth were also set to true, thus removing the background noise in the image and better centering the CAM around the objects. The first image in Figure 3 shows that the human's face is also highlighted (though with lesser intensity) along with the white terrier. No smoothing technique is used for the generation of this heat map, thus it highlights even the objects that are not of interest. However, in the second image, smoothing techniques are applied and this is reflected in the generated heat map, such that only the key features of the white terrier are highlighted, as it is our main object of interest.On observing the images in Figure 4, we see that for Ablation CAM, the front part of the racer is hardly highlighted, emphasizing that very low attention is given by the model to these parts of the car while doing the prediction. For the heat map in the middle that's generated using Grad CAM, some of that part is mildly highlighted suggesting that though it is not significantly affecting the model's prediction, it still does have some influence. Lastly, the heat map generated using Score CAM highlights the front part of the car with more concentration and precision. Besides that, all three heat maps consistently highlight the wheels of the car and the structure of the car at the back, which means the model considers the wheels as one of the key features along with its structure. To summarize,for all the generated heat maps, the key features in the image are highlighted consistently by all three methods, with some additional features highlighted by few methods more prominently than the others.