Task 2 – Grad-CAM, AblationCAM & ScoreCAM

1. **Grad-CAM** – uses the gradients flowing into the last convolutional layer to produce a coarse localization heatmap.
2. **AblationCAM** – systematically "ablates" (masks out) each feature map in the target layer to measure its effect on the output score, producing a more diffuse attribution map.
3. **ScoreCAM** – generates attribution by scoring the effect of each upsampled activation map on the output, yielding sharper, higher-contrast explanations.

Setup & Method

- **Model**: Pretrained `resnet50` on ImageNet, evaluated in `eval()` mode on CUDA when available.
- **Target Layer**: the final block of `layer4` (`model.layer4[-1]`), chosen for its high-level semantic features.
- **Image preprocessing**: resize to 224×224, normalize with ImageNet means/stds.
- **CAM generation**:

```
def apply_cam(cam, input_tensor, pred_class, original_image):
"""Apply CAM and return visualization"""
try:
    targets = [ClassifierOutputTarget(pred_class)]


    if not input_tensor.requires_grad:
        input_tensor.requires_grad_(True)

    grayscale_cam = cam(input_tensor=input_tensor, targets=targets)

    # Check if CAM was computed successfully
    if grayscale_cam is None or len(grayscale_cam) == 0:
        raise ValueError("CAM computation failed - no gradients computed")

    grayscale_cam = grayscale_cam[0, :]
    normalized_image = normalize_image_for_cam(original_image.copy())

    cam_image = show_cam_on_image(normalized_image, grayscale_cam, use_rgb=True)

    return cam_image
```

- **Output**: The output is saved in the directory `task2/output_gradcam_v2` separated by the CAM method.

Visualizations & Analysis

General Observations

Across the ten images, Grad-CAM and Ablation-CAM consistently produce more focused and informative heatmaps than Score-CAM. In many instances, Score-CAM either fails to highlight the object of interest or produces a very diffuse or minimal activation map, suggesting it may be less effective in these specific cases. Grad-CAM and Ablation-CAM tend to generate similar heatmaps, often highlighting the core features of the objects, though with slight variations in focus and extent.

- **Image 0: West Highland White Terrier**: Both Grad-CAM and Ablation-CAM successfully highlight the dog's face and body, which are the key features for identification. In contrast, Score-CAM completely fails, showing no activation on the dog itself, indicating a failure to localize the object of interest.

- **Image 1: American Coot**: Grad-CAM and Ablation-CAM generate strong, focused heatmaps centered on the coot. Score-CAM, while correctly locating the bird, produces a much weaker and less distinct activation map, showing only a faint highlight.

- **Image 2: Sports Car**: Grad-CAM and Ablation-CAM both effectively highlight large, representative parts of the car. Score-CAM, however, only activates on a small, specific area (the rear wheel and a portion of the door), failing to capture the entire object as the basis for the prediction.

- **Image 3: Flamingo**: Grad-CAM and Ablation-CAM produce very similar and accurate heatmaps that cover the cluster of flamingos. Score-CAM again underperforms, showing only a very faint and scattered activation over the birds.

- **Image 4: Kite**: This is a case of misclassification, as the image is predicted as "spoonbill." Interestingly, Grad-CAM and Ablation-CAM focus their attention on the central flower bud. Score-CAM's heatmap is displaced, focusing on a different flower to the right, which suggests it may be interpreting the model's reasoning differently and less accurately.

- **Image 5: Goldfish**: For the goldfish, Grad-CAM and Ablation-CAM produce nearly identical, well-focused heatmaps on the fish's body. Score-CAM once again fails to produce any meaningful activation, leaving the object entirely unmarked.

- **Image 6: Tiger Shark**: The prediction is accurate. Grad-CAM and Ablation-CAM generate

strong and clear heatmaps over the main body of the shark. Similar to other examples, Score-CAM fails to highlight the shark, demonstrating its inconsistency.

- **Image 7: Vulture**: In this image, Grad-CAM and Ablation-CAM correctly highlights. Score-CAM, true to its pattern in this set, shows no activation.
- **Image 8: Common Iguana**: The heatmaps from Grad-CAM and Ablation-CAM are tightly focused on the iguana's head and crest, which are distinguishing features. Score-CAM produces a puzzling result, with a small, intense activation spot on the background above the iguana, completely missing the subject.
- **Image 9: Orange**: For the correctly identified "orange," Grad-CAM and Ablation-CAM provide excellent localization, highlighting the flesh of the cut orange. As with many of the other images, Score-CAM fails to produce a visible heatmap.

Conclusion

From the analysis of these ten images, Grad-CAM and Ablation-CAM consistently outperform Score-CAM in generating meaningful and localized visual explanations of the model's predictions.

Both Grad-CAM and Ablation-CAM reliably focus on semantically important regions of the object, such as the face of animals, structural contours of vehicles, or textures of objects. This consistency not only validates the correctness of the predictions but also provides valuable insights even in cases of misclassification—highlighting the features that may have led the model astray.

In contrast, Score-CAM demonstrates limited effectiveness across this dataset. Its heatmaps are frequently weak, spatially off-target, or missing entirely. While Score-CAM is known for being gradient-free and theoretically less sensitive to noise, in practice here it fails to produce interpretable or relevant visualizations in most cases. Only in isolated examples (e.g., goldfish, racer) does it gives relatively better results.