TML ASSIGNMENT 4

TASK 2

TEAM 16

AHRAR BIN ASLAM

MUHAMMAD MUBEEN SIDDIQUI

The aim of this task is to visualize and interpret the internal decision-making process of a pretrained ResNet50 image classifier using Class Activation Mapping (CAM) techniques. We investigate how the model focuses on various regions of an input image to predict its class. In addition to the standard Grad-CAM, we extend our analysis using two other CAM variants: AblationCAM and ScoreCAM.

Model and Target Layer

- **Model**: ResNet50 pretrained on ImageNet
- **Target Layer**: The last convolutional layer was selected for CAM computation, as it retains spatial semantics useful for localization.

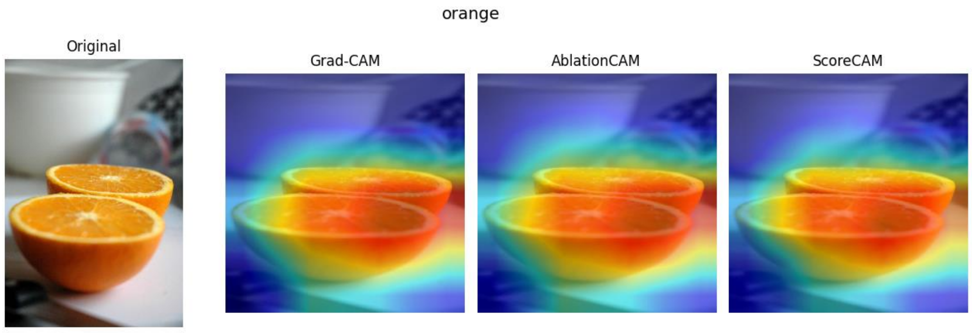
Image Preprocessing

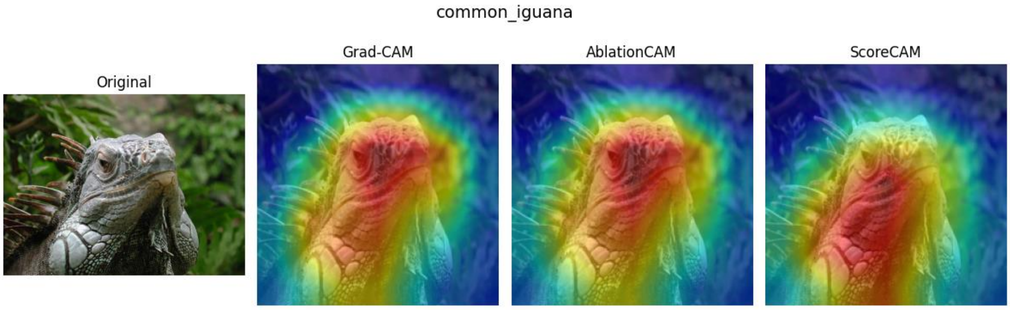
- All images were resized to 224×224 pixels and normalized using ImageNet mean and standard deviation.
- Processed inputs were passed through the model to obtain predicted class scores.

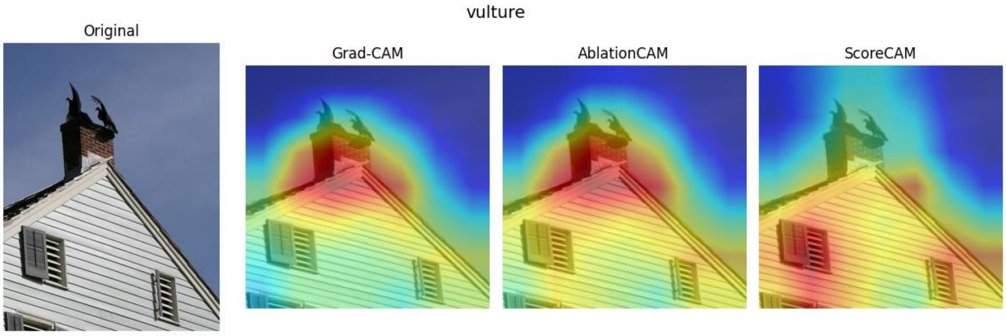
CAM Techniques Applied

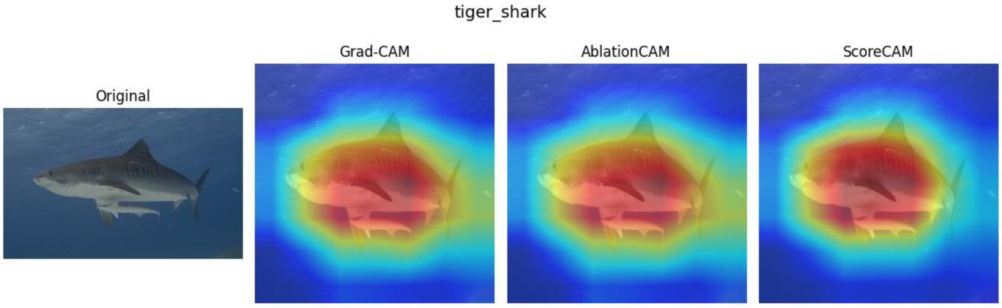
- **Grad-CAM**: Uses gradients of the predicted class with respect to feature maps to generate heatmaps.
- **AblationCAM**: Ablates (removes) one channel at a time to estimate each channel's contribution to the class score.
- **ScoreCAM**: Passes masked inputs and weighs feature maps based on their class scores without using gradients.

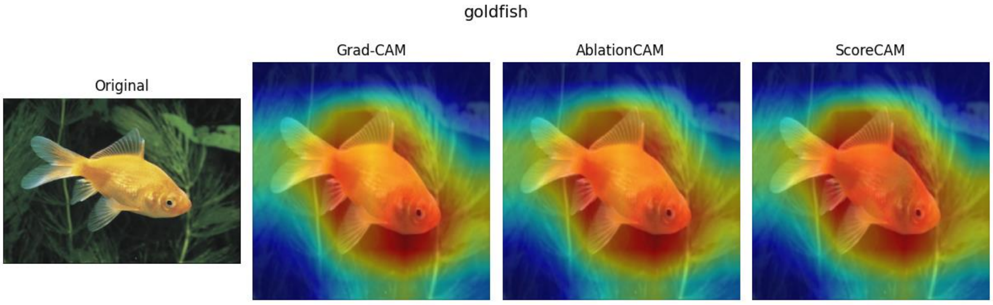
VISUAL OUTPUTS:

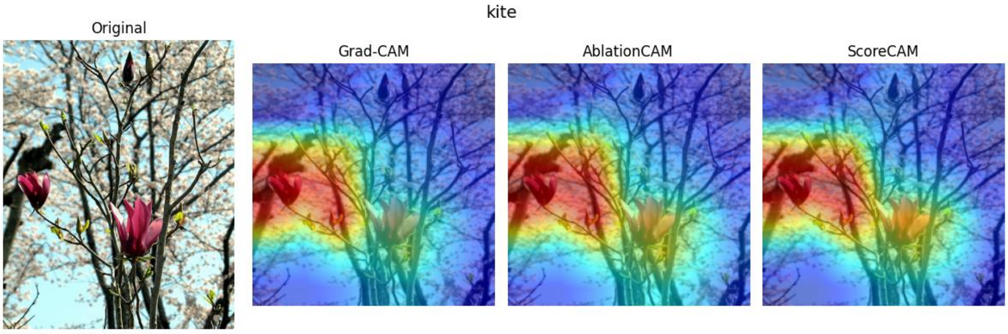


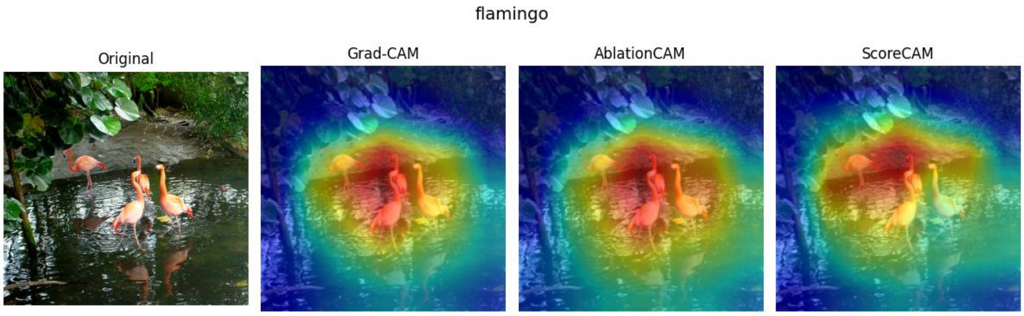

common_iguana
Original
Grad-CAM
AblationCAM
ScoreCAM

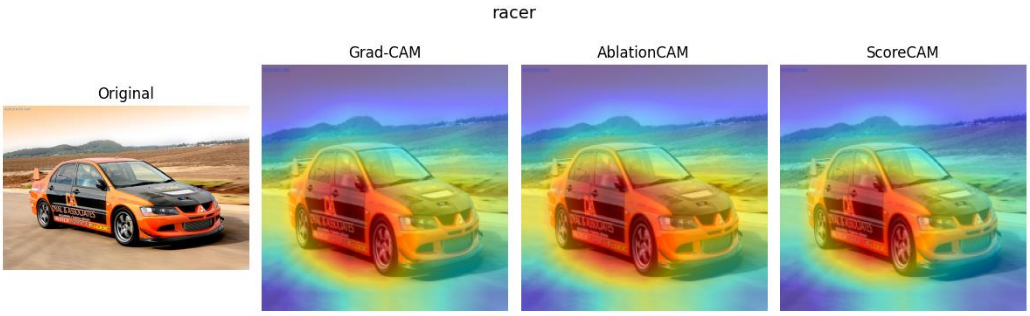

vulture
Original
Grad-CAM
AblationCAM
ScoreCAM

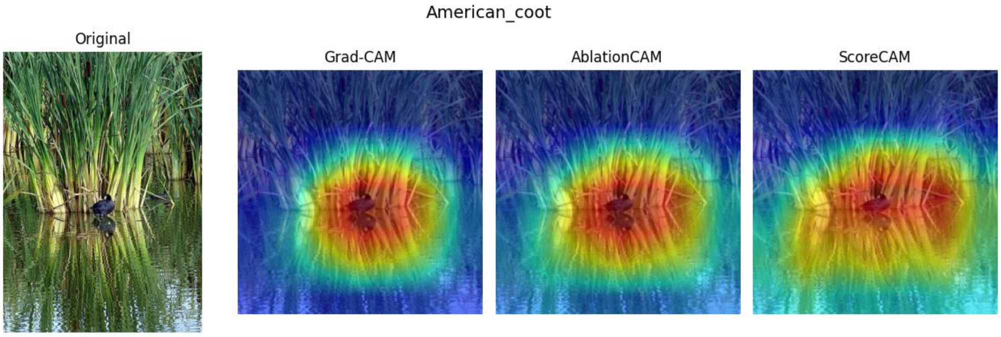

tiger_shark
Original
Grad-CAM
AblationCAM
ScoreCAM

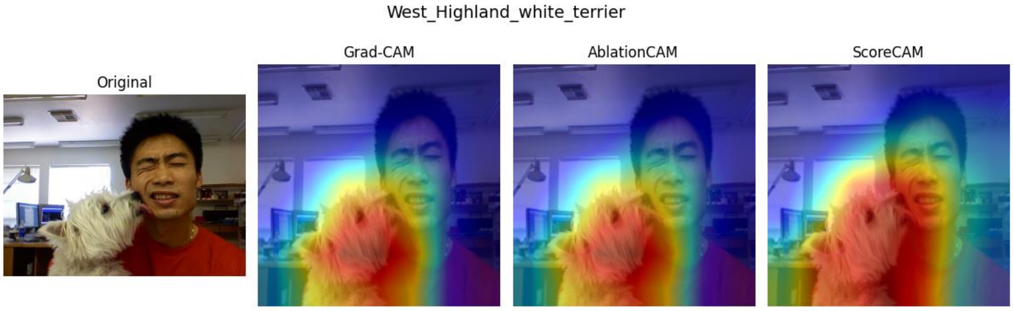

goldfish
Original
Grad-CAM
AblationCAM
ScoreCAM


kite
Original
Grad-CAM
AblationCAM
ScoreCAM









Analysis of Results

The application of Grad-CAM, AblationCAM, and ScoreCAM across the 10 ImageNet images revealed consistent patterns in how each technique highlights class-relevant regions. Each method has strengths and weaknesses that became evident across different image contexts.

Grad-CAM

Grad-CAM provided strong and broad localization of the main object, usually centered around key semantic regions. The strength of Grad-CAM is that it is effective at quickly identifying relevant areas in well-composed images. Whereas, it tends to produce more diffused or less focused attention when multiple salient objects are present.

Example:

- In the “racer” image, Grad-CAM clearly highlights the car body and logos, showing that the model bases its decision on vehicle-related features.
- However, in the “West Highland White Terrier” image, Grad-CAM spreads attention across both the dog and the person, suggesting some confusion or shared feature reliance.

AblationCAM

AblationCAM yielded more focused and spatially precise heatmaps, often isolating the most discriminative object regions. It is excellent at pinpointing the key class-specific features, especially useful in cluttered scenes. However, slightly more computationally expensive due to the channel ablation process.

Example:

- In the “American Coot” image, AblationCAM focused tightly on the bird in the middle of dense reeds, showing strong feature separation from background clutter.
- In the “flamingo” image, it highlighted the central birds more cleanly than Grad-CAM or ScoreCAM, providing superior interpretability.

ScoreCAM

ScoreCAM often produced visually rich and detailed heatmaps, especially good at capturing texture and fine boundaries. It is able to highlight subtle object features and shapes due to forward-based weighting. Whereas, it tends to produce slightly diffuse attention and is slower to compute than the other two methods.

Example:

- In the “goldfish” image, ScoreCAM provided a smooth and high-resolution heatmap covering the full body and fins, emphasizing contours.
- However, in the “kite” image, ScoreCAM, along with the others highlighted red flowers instead of the bird, reflecting model misclassification or attention bias. This shows how CAM methods expose the model’s learned associations even when they are incorrect.