Visual Explanation Tasks - Detailed Report

==========

Task 2: GradCAM, ScoreCAM, and AblationCAM Comparison

This task investigates the interpretability of CNN models using gradient-based and perturbation-based class activation mapping techniques:

- GradCAM: Utilizes gradients of the target class flowing into the final convolutional layer to produce coarse localization maps.
- ScoreCAM: Relies on the forward pass and the model's confidence scores, without requiring gradients.
- AblationCAM: Perturbs the network by removing one channel at a time, measuring the output drop to estimate importance.

Key Findings:

- GradCAM produces sharper boundaries but may be noisy for fine-grained classes.
- ScoreCAM offers smoother visualizations and better localization for high-confidence regions.
- AblationCAM is computationally expensive but often yields interpretable results.

Visualization heatmaps were generated per class and compared. A collective heatmap provided comparative insights.

Task 3: Optimizing LIME Explanations

In this task, LIME is employed to interpret image classification by locally approximating the model with an interpretable surrogate model:

- Optimization of parameters (e.g., number of samples, kernel width) was conducted to enhance output relevance.
- LIME outputs were visualized across multiple class examples.

Key Findings:

- Tuned LIME outputs highlighted meaningful superpixels correlated with model predictions.
- The enhanced explanations improved transparency in model behavior.
- Compared to CAM-based methods, LIME provides instance-specific and model-agnostic explanations.