**Lime Explanation – Analysis of the results**

The goal of this task was to get a holistic view of the regions that causes the model to output a certain prediction on an image input using the LIME method, hence enhancing the explanability and interpretation of the model. This report presents an analysis of the visual explanations generated for 10 provided ImageNet images, highlighting the regions of input that are considered as the most important.

**A Brief Overview of the Methodology:**

For this task, the ResNet50 model, pre-trained on ImageNet, was utilized. The lime package was used to implement the LIME methodology. The mask, highlighting the important regions for the explanation is plotted, alongside with the original image, and the confidence of classification.

**Image-by-Image LIME Analysis**  
While LIME explanations were generated for all 10 images, and the model was able to predict 9 out of the 10 images correctly (however, the racer class was predicted as sports car). To understand the explanations, we will focus on three illustrative cases, which are important. One of high confidence, other wrongly classified and one with a 50% confidence.

**American Coot**  
**Predicted Class:** American coot (Correct)  
**Confidence:** 0.603

For this image, LIME segmented the input into ~12 superpixels and then estimated a local linear model around the prediction. The **top positive contributors** (i.e., superpixels whose presence increased the log‑odds of “American coot”) were: covering the bird’s head and upper body, the main torso region and the patch on the wing.

Although, like GradCAM and AblationCAM the reflection was in the bounded region of importance, the main superpixel were associated with the main body of the Bird.

No strongly negative superpixels were reported among the top six, indicating that most of the influential patches lay directly on the bird itself. The resulting LIME heat map shows warm (reddish) patches tightly over the coot, with little activation elsewhere, suggesting that LIME honed in on the bird’s silhouette and texture rather than on background reflections.

**2. Kite**   
**Predicted Class:** Spoonbill (Incorrect)  
**Confidence:** 0.117

LIME’s explanation reveals that the **highest positive weights** were essentially **zero** for the first six superpixels—many of which corresponded to sky and foliage —indicating that those background regions neither helped nor hurt the spoonbill score. Instead, the first truly positive contributor was only the 7th or 8th superpixel, which captured the two pink flowers. Conversely, the sparse, “spotty” heat map underscores LIME’s ability to point out precisely which localized regions (flowers!) drove the erroneous prediction.

**3. Tiger Shark**  
**Predicted Class:** Tiger shark  
**Confidence:** 0.572

LIME divided this underwater scene into ~12 superpixels and found that:

* Shark’s head and snout was by far the strongest positive contributor.
* The dorsal fin also boosted the class score.
* The most negative contributor, lay in the water background, indicating that removing that patch actually increases the tiger shark confidence score.

The resulting LIME map shows warm highlights that connect head and fin regions—mirroring the continuous attention seen in other methods—along with cool (blueish) shading over empty water. This bipolar pattern of positive vs. negative superpixels provides a richly detailed, locally faithful explanation for the correct “tiger shark” classification.

**General Observations**

* **Positive vs. Negative Contributions:** LIME’s ability to highlight negative‐weight regions (as in the kite and shark examples) helps diagnose not only what pushed the model *toward* a class, but also what pushed it *away* from competing classes.
* **Interpretability Trade‑Offs:** While LIME often yields more **interpretable** and **locally faithful** explanations than CAM, its reliance on superpixel segmentation can sometimes obscure fine details—e.g., when a misclassification arises from tiny texture cues rather than large contiguous regions. Hence, it is very good to explain predictions from the images that has the object away from many cluttered details.
* **Computation Time & Overlap:**
  + **Average IoU** between predicted and ground‐truth masks (using the optimized LIME segmentation) was **0.3046**, indicating moderate alignment.
  + **Average explanation time** was **1.8659 s** per image, reflecting LIME’s higher computational cost compared to gradient‐based CAMs.

**Conclusion**

This task demonstrated that LIME provides highly granular, locally faithful explanations by assigning explicit weights to superpixels—revealing both what the model leans on and what it rejects. For correctly classified images like the American coot and tiger shark, LIME cleanly isolates the object’s key parts (head, body, and fin). For misclassifications like the kite → spoonbill, it pinpoints the misleading patches (pink flowers) and shows which background elements detracted from other classes. However, the sparsity and dependence on superpixels, sometimes yield very fragmented masks, this can be specifically highlighted in the **common iguana** image.