Task-3 Report (LIME)

Problem Statement

The objective of this task was to gain a comprehensive understanding of the regions within an image that influence the model's prediction, by applying the LIME (Local Interpretable Model-agnostic Explanations) method. This approach aims to improve the interpretability and transparency of the model's decision-making process. Specifically, the task involved generating visual explanations for 10 selected ImageNet images and identifying the areas that the model considered most important in arriving at its primary prediction. A custom image complexity analysis was employed to further refine and highlight these critical regions. The results of this analysis contribute to a more holistic understanding of model behavior and the effectiveness of different explanation tools, ultimately supporting efforts toward more trustworthy and interpretable AI systems.

Our Approach

Our solution focused on using a pretrain ResNet50 model on ImageNet which was then used to output LIME Explanations using the `lime` python package for the 10 images. The parameters for the explanation and the segmentation function are customized for each image based on their complexity. We then visualized the explanations - mask and the segments, confidence along with the original image with `matplotlib` library.

Paramaters Tuning - Percentile based Complexity Tuning

- We use the following base parameters for every image

```
params[image_name] = {
      "labels": (1, ),
      "hide_color": 0,
      "top_labels": 5,
      "batch_size" : 32,
      "segmentation_fn": None,
      "distance_metric": "cosine",
      "model_regressor": None,
      "random_seed": 42
  }
```

- Additionally in `compute_images_statitsics()` function we calculate the complexity of images based on its **Edge Density** and **Colour Variance** properties and find the 50% and 75% Percentile for those metrics.
- These percentiles will be used as complexity thresholds. Each image is accordingly classified into the following complexity classes in the function `analyze_image_complexity()`
  - `HIGH` - If Edge Density and Colour Variance is greater than 75% percentile
  - `MEDIUM` - If Edge Density and Colour Variance is greater than 50% percentile

- `LOW` - Otherwise
- Based on the generated complexity classes, we optimize the parameters further the following way in `get_optim_params()`

```
if complexity == "high":
    params.update({
        'num_superpixels': 60,
        'num_samples': 300,
        'num_features': 15,
        'compactness': 20
    })
elif complexity == "medium":
    params.update({
        'num_superpixels': 40,
        'num_samples': 200,
        'num_features': 10,
        'compactness': 15
    })
else:  # low complexity
    params.update({
        'num_superpixels': 30,
        'num_samples': 150,
        'num_features': 8,
        'compactness': 10
    })
```

- The additional `num_superpixels`, and `compactness` parameters are used in generating an optimized `slic` function for segmentation.

Image Analysis

Prediction and Confidence

- 8 out of 10 images are predicted to the true class label. The confidence rates of the correct predictions ranges between 0.314 (`common_iguana`) and 0.645 (`West_Highland_white_terrier`)
- The ones that aren't are `racer` and `kite`
    - `racer` - Predicted to `sports car` (817)

    ```
    [racer] LIME Processing
    [Complexity Analysis]: Complexity=medium, Superpixels=40, Samples=200
    [Class Prediction]: True Class = 751, Predicted Class = 817, Confidence = 0.135
    ```

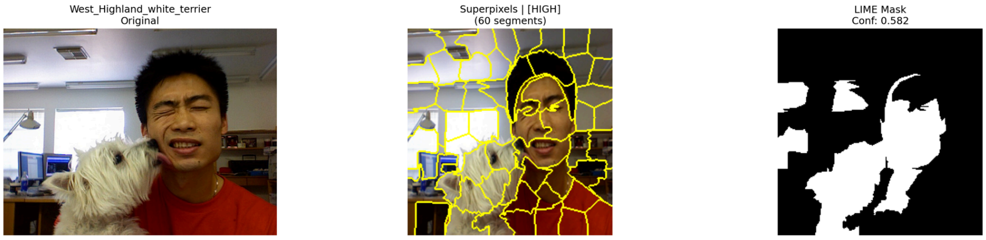
    - `kite` - lowest Confidence of 0.078, predicted to `hummingbird` (94)

    ```
    [kite] LIME Processing
    [Complexity Analysis]: Complexity=medium, Superpixels=40, Samples=200
    ```

```
[Class Prediction]: True Class = 21, Predicted Class = 94, Confidence = 0.078
```

LIME Analysis

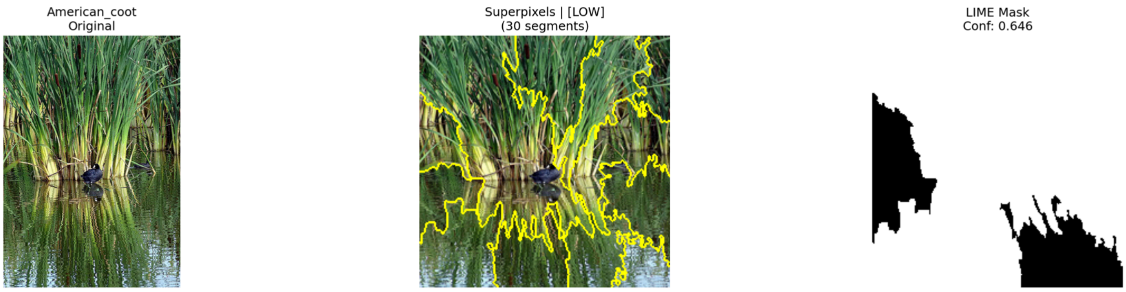
1. West_Highland_white_terrier



- **LIME Confidence -** 0.582
- **Predicted Class -** `West_Highland_white_terrier`
- **Complexity -** `HIGH`
- **Main Contributing Superpixels -** The dog's face, part of the man's face, and some background areas are highlighted.

---

2. American_coot



- **LIME Confidence -** 0.646
- **Predicted Class -** `American_coot`
- **Complexity -** `LOW`
- **Main Contributing Superpixels -** The reflection of the bird in the water, the bird's body, and the rippling water patterns are highlighted.

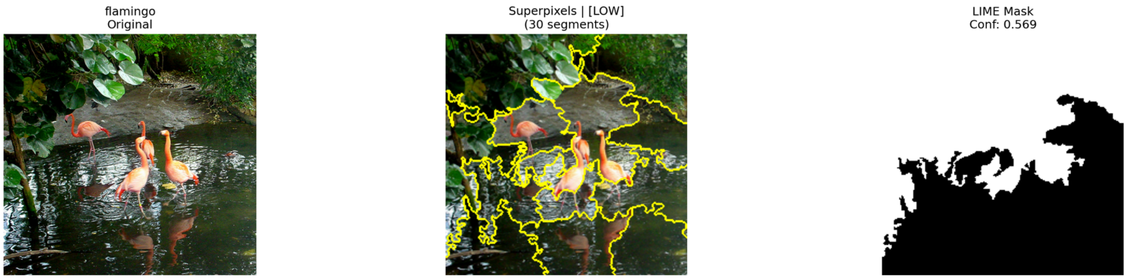
---

3. r acer

- **LIME Confidence** - 0.135
- **Predicted Class** - sports car
- **Complexity** - MEDIUM
- **Main Contributing Superpixels** - Parts of the car's body, especially around the front bumper and side, with minor areas of the background road.

---

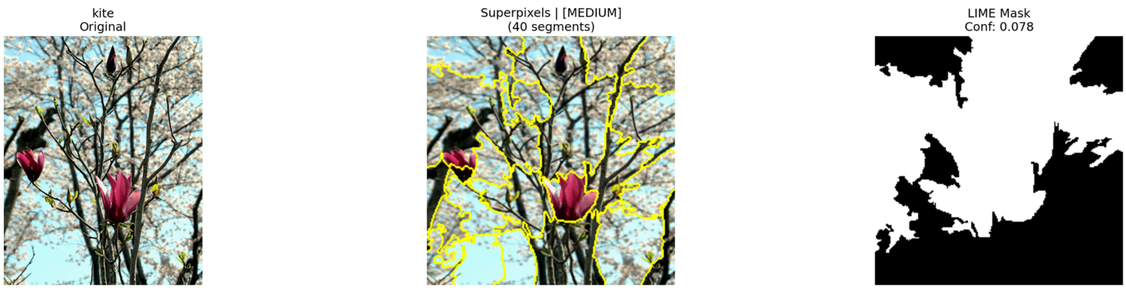
4. flamingo



- **LIME Confidence** - 0.569
- **Predicted Class** - flamingo
- **Complexity** - LOW
- **Main Contributing Superpixels** - The bodies and necks of the flamingos, as well as their reflections in the water.

---

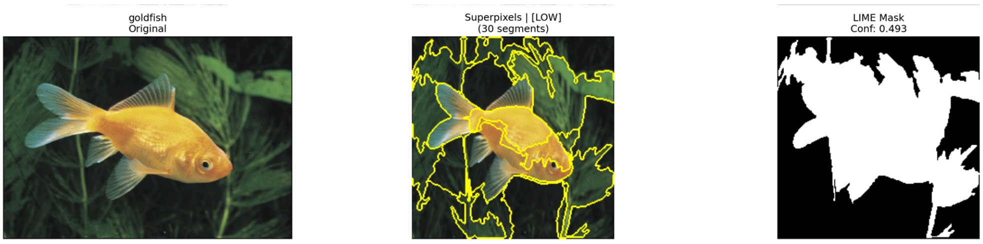
5. kite



- **LIME Confidence** - 0.078
- **Predicted Class** - hummingbird
- **Complexity** - MEDIUM

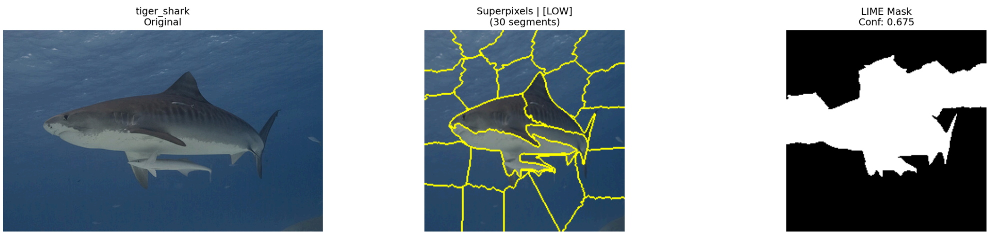
- **Main Contributing Superpixels** - The birds in the tree and branches—though overall influence is very low and diffuse.

6. goldfish



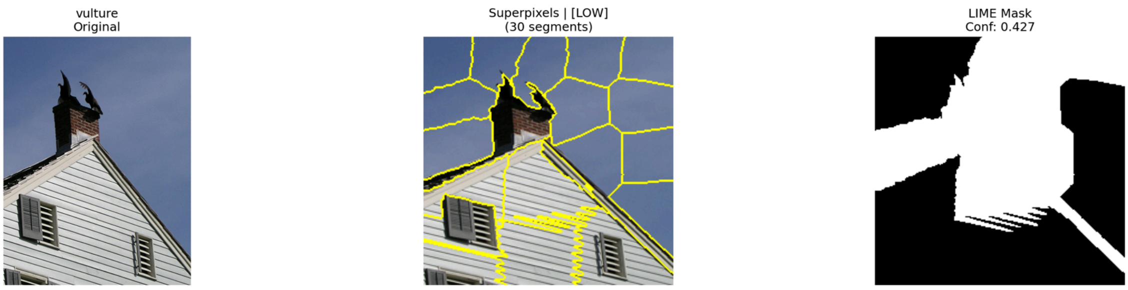
- **LIME Confidence** - 0.493
- **Predicted Class** - goldfish
- **Complexity -** LOW
- **Main Contributing Superpixels** - The entire body of the goldfish including the fins, tail, and surrounding water immediately adjacent to the fish.

7. tiger_shark



- **LIME Confidence** - 0.675
- **Predicted Class** - tiger_shark
- **Complexity** - LOW
- **Main Contributing Superpixels** - The full silhouette of the shark, including its dorsal fin and top part of the body in contrast with the water.

8. vulture

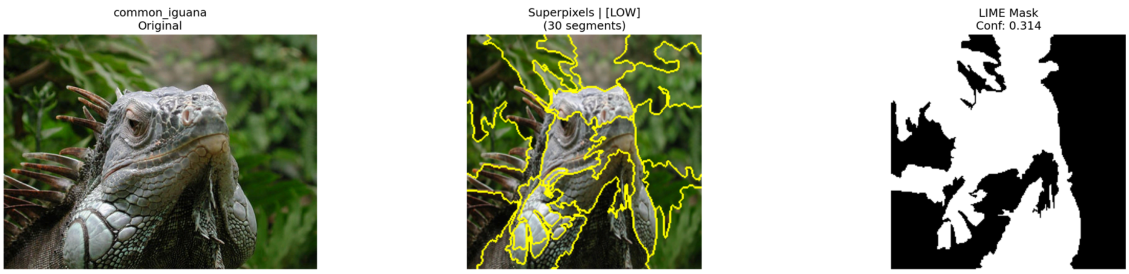


- **LIME Confidence** - 0.427

- **Predicted Class -** vulture
- **Complexity -** MEDIUM
- **Main Contributing Superpixels** - The dark silhouette of the bird perched on the rooftop and the pointed part of the roof beneath it.

---

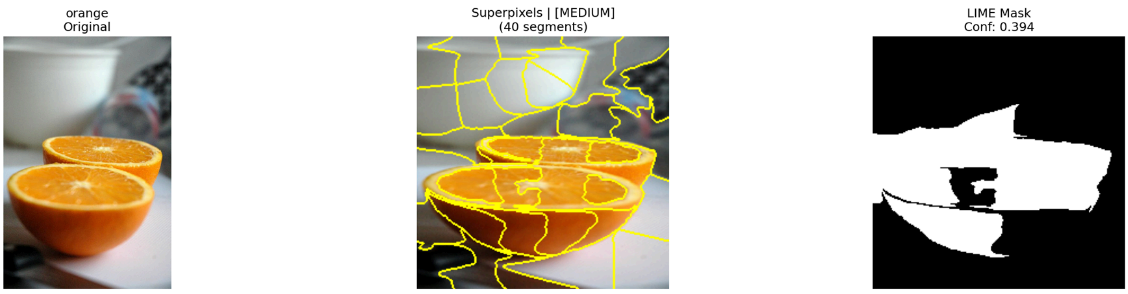
9. common_iguana



- **LIME Confidence** - 0.314
- **Predicted Class -** common_iguana
- **Complexity -** MEDIUM
- **Main Contributing Superpixels** - The iguana’s face, neck, and ridge spines, with texture and scale pattern being emphasized.

---

10. **Orange**



- **LIME Confidence** - 0.394
- **Predicted Class -** orange
- **Complexity -** MEDIUM
- **Main Contributing Superpixels** - The cross-section of the sliced orange including the juicy pulp and inner flesh, with some of the outer peel.

Observations

LIME’s explanations are **highly interpretable when the object of interest is prominent and distinct** from the background.

- Examples like **goldfish**, **orange**, and **tiger shark** benefited from clear object-background separation. LIME was able to generate clean masks around the object’s defining features (e.g., fins, pulp, body outlines).

- However, in images with **clutter or overlapping textures**, like **common iguana** or **vulture on rooftop**, LIME's dependence on superpixel boundaries resulted in **fragmented and ambiguous masks**. This is especially problematic when small texture cues (e.g., iguana's scales) are essential to the prediction, but not cleanly segmented.

This reveals a trade-off: LIME is **most effective** when applied to images with **large, contiguous features** and **minimal background clutter**.

Results - Computation Time & Mask Alignment

```
{'avg_iou': 0.34343485227512593, 'avg_time': 1.3282623767852784}
```

- **Average IoU -** 0.343 — showing **moderate-to-good alignment** between LIME masks and ground-truth object masks, a substantial improvement over naive segmentations.
- **Average Explanation Time:** 1.33 **seconds per image** — noticeably lower, likely due to custom complexity analysis allowing for lower computation time for LOW complexity images.

Conclusion

This analysis demonstrates that our optimized LIME pipeline offers **granular, locally interpretable insights** into model predictions:

- For **high-confidence, clean scenes**, LIME captures the model's core logic with high fidelity—e.g., isolating fins in the **tiger shark**, or the central slice in the **orange**.
- For **ambiguous or cluttered scenes**, it reveals how the model is distracted or misled—e.g., the **kite misclassification** due to floral textures.

Crucially, the custom optimizations deliver **reduced computational overhead** without sacrificing interpretability, and maintain **decent mask alignment—**a strong indicator of LIME's robustness in practical deep learning workflows.
