Deep Learning Lab-3

immediate

October 10, 2024

Abstract

This project focuses on computing various representations of gaze fixation points, including Isoline representation (isoline heatmaps and blended images), Soft Selection, and Hard Selection representations. These methods provide alternative visualizations to traditional saliency heatmaps, offering deeper insights into visual attention patterns. Comparison with ground-truth fixation data can be done using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Pearson Correlation Coefficient (PCC), and Structural Similarity Index (SSIM). These metrics provide a comprehensive assessment of both pixel-wise accuracy and perceptual similarity. The training process was conducted on a PC.

1 Methodology

This section describes the techniques used to generate alternative visual representations of gaze fixation points, beyond the standard saliency heatmap. These methods aim to provide different perspectives on attention patterns within an image, allowing for a more nuanced understanding of where and how viewers focus their gaze. The two primary approaches used in this project are the Isoline representation and Selection representation, each offering a distinct way to highlight areas of interest

1.1 Isoline Representation

The Isoline representation visualizes a saliency map using contour lines to connect points with the same saliency level. Dense contour lines indicate areas with high saliency, while sparse lines represent lower saliency regions. These contour lines can be color-mapped based on the saliency value, and overlaid onto the original image to highlight areas with the most gaze points.

1.2 Selection Representation

Selection representation visualizes attention by prioritizing certain regions in an image and comes in two types:

- **Hard Selection**: A binary mask is applied to the image, spotlighting only regions where saliency exceeds a threshold, while the rest is blacked out.
- **Soft Selection**: A weighted approach that superimposes the saliency map onto the image, smoothly blending the regions based on varying fixation densities.

2 Data Loading

We have been provided with the images `img` and `test_saliency_img`, and in addition, we will use the MexCulture142 dataset, which consists of 284 images of Mexican monuments, along with corresponding gaze fixation points and ground truth Gaze Fixation Density Maps (GFDM). The dataset is divided into training and validation sets, with images labeled according to their filename prefixes.

3 Implementation & Code Structure Overview

The project consists of two main files: `main.py` for execution flow and `utils.py` for utility functions. `main.py` handles loading images, generating saliency map representations, and displaying results. `utils.py` provides core functions for saliency map processing and image display.

4 Key Functions in `main.py`

- `get_user_inputs()`: Parses command-line arguments for image paths and display options, making the script flexible.
- `load_images()`: Loads the test image and saliency map, converting them to usable formats for further processing.
- `generate_representations()`: Generates different visualizations of the saliency map, such as heatmaps, isolines, and blended images.
- `display_images()`: Displays the images either in a grid or individually based on user input.

5 Key Functions in `Representations.py`

- `Saliency` and `RGBImage` Classes: Simple classes to wrap around image objects for easier access and handling.
- `represent_heatmap()` and `represent_heatmap_overlaid()`: Generate heatmaps and overlay them on the original image for visualizing saliency regions.
- `represent_isolines()` and `represent_isolines_superimposed()`: Create isolines (contour lines) from the saliency map and blend them with the image.
- `represent_hard_selection()` and `represent_soft_selection()`: Apply hard or soft masks to the image based on the saliency map to highlight specific regions.
- `grid_layout()`: Arranges the images into a 2x3 grid, adds titles, and saves the grid layout as an image.

6 Results

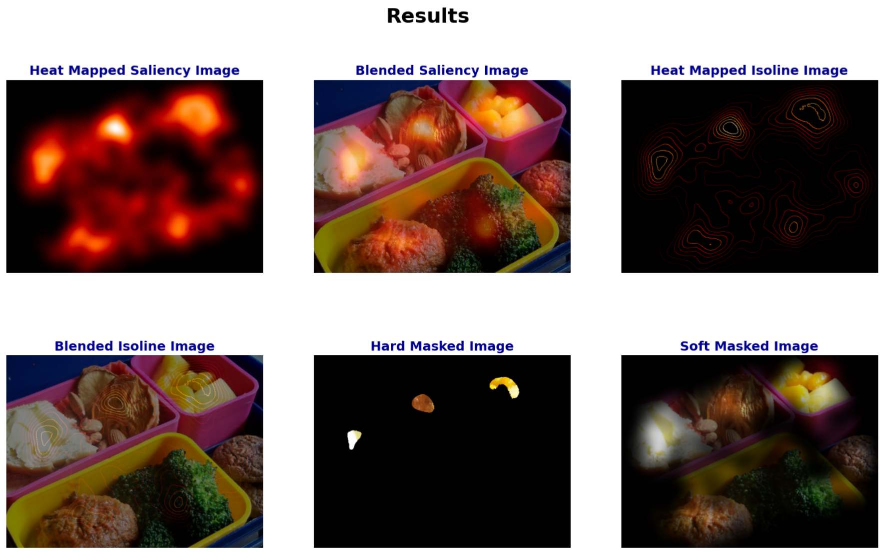

Fig. 1: illustrates various visual representations of saliency maps to depict gaze fixation patterns.

6.1 Explanation of Saliency Map Representations

The saliency map representations are described as follows:

- **Heat-mapped saliency**: This representation highlights regions of high attention using warm colors, while darker areas indicate lower attention. It provides a clear indication of the areas that attract the most focus.
- **Blended saliency map**: This representation combines the heatmap with the original image, making it easier to relate fixation points to specific visual elements. It provides a more contextualized view of attention distribution.
- **Isoline representation**: Contour lines are used to connect areas with similar saliency values. Denser lines indicate regions with higher attention. The isoline map is also blended with the original image to maintain clarity while illustrating concentration of fixation points.
- **Hard mask representation**: Only the most salient areas are highlighted, with the rest of the image masked out. This method provides a clear focus on regions of maximum attention.
- **Soft selection representation**: This representation offers a smoother overlay that gradually highlights regions based on varying attention levels. It provides a more continuous and less discrete visualization of attention distribution compared to the hard mask.

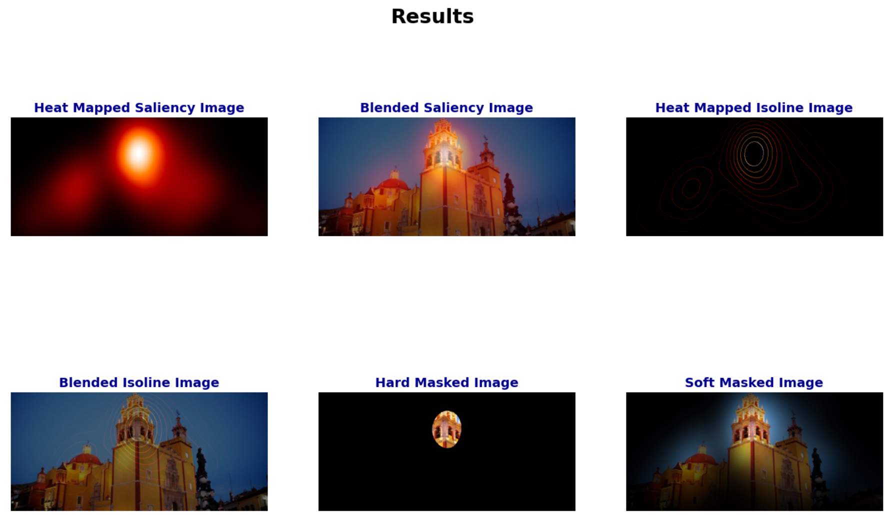

Fig. 2: illustrates various visuals of (Colonial BasilicaDenue)

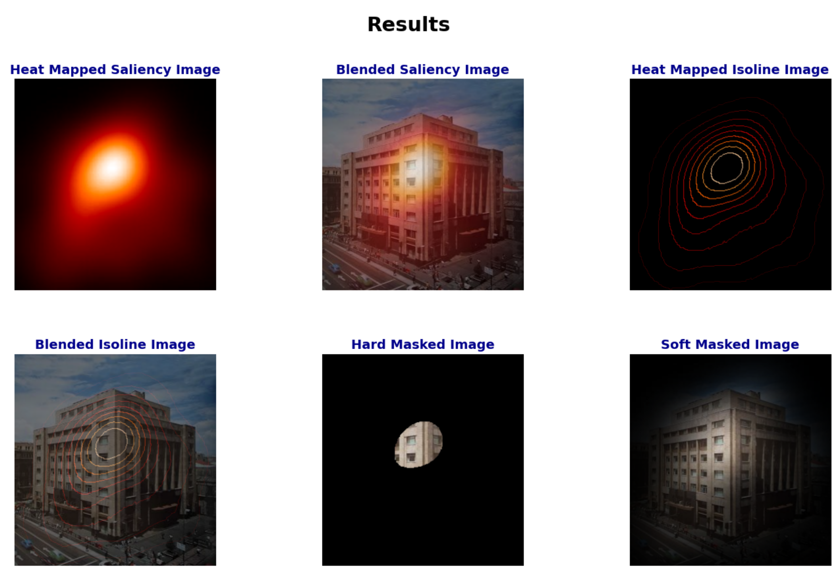

Fig. 3: illustrates various visuals of (Modern PlazaGuardiola).

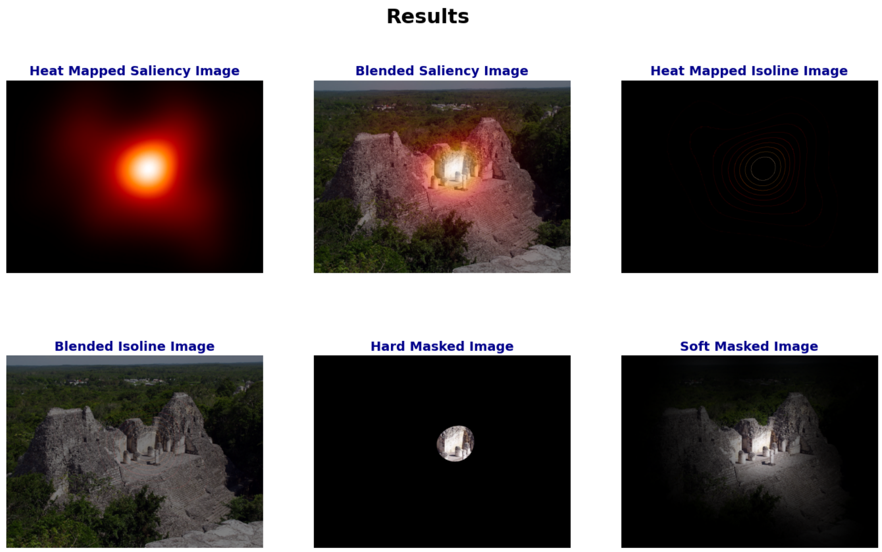

Fig. 4: illustrates various visuals of (Prehispanic EstructuraVIII).

7 Extra Work: Evaluation of the Saliency Map with ground truth

The saliency map created in LAB-2 was evaluated against the ground truth gaze fixation density map using four standard metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Pearson Correlation Coefficient (PCC), and Structural Similarity Index (SSIM). These metrics provide a comprehensive assessment of both pixel-wise accuracy and perceptual similarity.

- **Mean Absolute Error (MAE)**: This measures the average magnitude of absolute differences between the predicted saliency map and the ground truth. A lower MAE indicates that the predicted values are close to the actual values. The result of **0.0438** shows a low error, reflecting good alignment with the ground truth.
- **Mean Squared Error (MSE)**: This metric quantifies the average squared differences between the saliency map and the ground truth. The lower the MSE, the better the match. The result of **0.0069** indicates small errors, emphasizing the high similarity between the two maps.
- **Pearson Correlation Coefficient (PCC)**: This measures the linear correlation between the predicted and ground truth maps. A PCC of **0.9691** indicates a very strong positive correlation, meaning that the distribution of gaze points in the saliency map closely follows the ground truth.
- **Structural Similarity Index (SSIM)**: This assesses the perceptual similarity between two images by considering luminance, contrast, and structure. An SSIM value of **0.8628** reflects a high degree of structural similarity, confirming that the saliency map captures important visual features effectively.

Metric	Result
Mean Absolute Error (MAE)	0.0438
Mean Squared Error (MSE)	0.0069
Pearson Correlation Coefficient (PCC)	0.9691
Structural Similarity Index (SSIM)	0.8628

Table 1: Evaluation of Saliency Map against Ground Truth

8 Conclusion

The implementation of the functions successfully generated six different representations of gaze point density, including saliency heatmap, blended saliency map, isoline heatmap, image-isolines blend, soft selection, and hard selection representations. The use of `skimage`'s `measure` and `find_contours` provided better control and insight into the contour generation process, allowing for more precise manipulation compared to `matplotlib`. Each representation was visually accurate and aligned with the task requirements.