Deep Learning Lab-2

immediate

October 3, 2024

Abstract

This project focuses on computing Gaze Fixation Density Maps (GFDM), also known as Wooding maps, from the gaze fixation points collected during viewer interactions with images. Using the MexCulture142 dataset, which features Mexican architectural styles, we analyze 284 images of 142 Mexican monuments categorized into three distinct styles: Prehispanic, Colonial, and Modern. By examining the saliency maps generated from these fixation points, we aim to understand viewer attention patterns and preferences associated with each architectural style. The generated saliency maps will be saved as grayscale images and overlaid on the original images with color mapping for enhanced visual clarity. This analysis not only deepens our understanding of human visual perception in architectural contexts but also contributes to the broader field of cultural heritage analysis through data-driven insights. The training process was conducted in a virtual environment at CREMI and on a PC.

1 GFDM Methodology

A Gaze Fixation Density Map (GFDM), also known as a Wooding map, is generated by analyzing gaze fixation data, which consists of (x, y) coordinates where observers focused on the image. For this project, images were resized and displayed on a screen with a vertical resolution of 1200 pixels and a height of 325 mm.

The GFDM highlights areas of interest in architectural images by utilizing gaze fixation data. By generating saliency maps from these fixation points, we can identify viewer preferences and cognitive responses to various architectural elements. To represent fixation density, we convolve the image with a two-dimensional isotropic Gaussian function centered on each gaze point, simulating the foveated nature of human vision. The Gaussian spread is controlled by a sigma value, ensuring uniform distribution in both x and y dimensions.

GFDMs are computed for each fixation point, resulting in partial saliency maps that are summed to create a global saliency map. This map is then normalized to a range of 0 to 1. In this project, the saliency maps will be saved as grayscale images and color-mapped overlays, enhancing visual understanding while aiding in the classification of architectural styles and deepening our insight into human visual perception within the context of architectural heritage.

2 Data Loading & Preprocessing

The MexCulture142 dataset consists of 284 images of Mexican monuments, along with corresponding gaze fixation points and ground truth Gaze Fixation Density Maps (GFDM). The dataset was divided into training and validation sets, with images labeled based on their filename prefixes.

Table 1: Fixation Points for Colonial_AcademiaDeBellasArtes_Queretaro_GazeFix_N_1

X	Y
816	584
730	654
705	763
974	824
1011	810
750	550
705	541
324	654
958	700
791	581
710	558
568	511

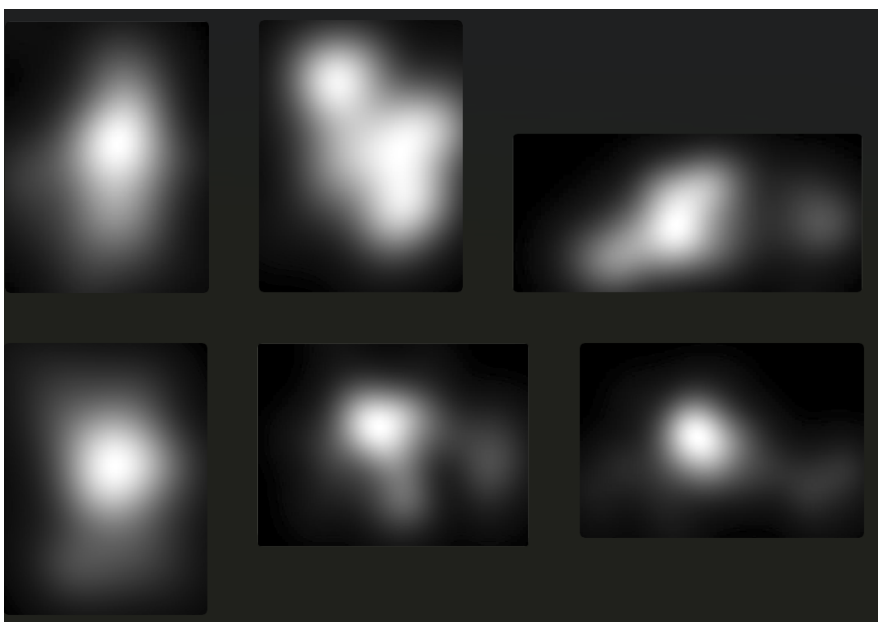

Fig. 1: Shows 6 images from gaze fixations density maps of MexCulture142

3 Implementation

The implementation of Gaze Fixation Density Maps (GFDM) involves several key steps, including saliency map generation, blending, and error metrics calculation.

3.1 Saliency Map Generation

The saliency map generation process begins with calculating the spread of the Gaussian function, denoted as σ. This is achieved using the formula:

$$\sigma = R \cdot D \cdot \tan(\alpha)$$

Here, R is computed as:

$$R = \frac{\text{VERTICAL\_RES}}{\text{SCREEN\_HEIGHT}} = \frac{1200\,\text{pixels}}{325\,\text{mm}} \approx 3.6923\,\text{pixels/mm}$$

The angle α is set at 2° (converted to radians for calculation), and D represents the distance from the viewer to the screen, set at 325 mm. Each gaze fixation point is then used to create partial saliency maps through the Gaussian function:

$$S(I, m) = Ae^{-\left(\frac{(x-x_{0m})^2}{2\sigma^2} + \frac{(y-y_{0m})^2}{2\sigma^2}\right)}$$

In this equation, (x_{0m}, y_{0m}) are the coordinates of the fixation point, while x and y are the variables that represent the pixel positions across the entire image and A is a constant set to 1.

3.2 Blending and Saving Results

The generated saliency maps are blended with the original images using alpha blending, where both images are combined based on a specified alpha value (set to 0.5 for equal transparency). This is represented by the equation:

$$I_{\text{blended}} = \alpha \cdot I_{\text{original}} + (1 - \alpha) \cdot I_{\text{saliency}}$$

The resulting blended images and grayscale saliency maps are saved in designated output folders for further analysis.

3.3 Error Metrics Calculation

To evaluate the accuracy of the generated saliency maps, the Mean Absolute Error (MAE) is calculated using the following formula:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

Here, y represents the ground truth values, and $\hat{y}$ denotes the predicted saliency values. The computed MAE provides insights into how closely the predicted saliency maps align with the actual fixation patterns.

This comprehensive implementation facilitates the analysis of viewer attention patterns and preferences associated with various architectural styles, contributing valuable insights to the field of cultural heritage analysis.

4 Results

The average Mean Absolute Error (MAE) obtained is 0.04376

4.1 Modern

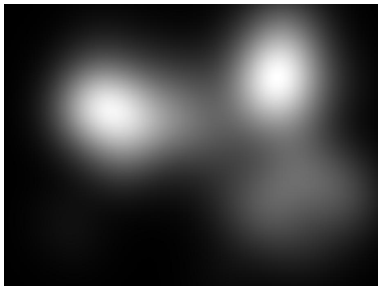
Fig. 2: Ground Truth

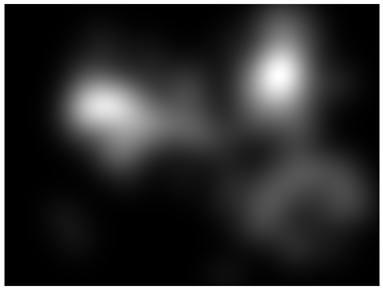
Fig. 3: GFDM


Fig. 4: Blended

4.2 Colonial

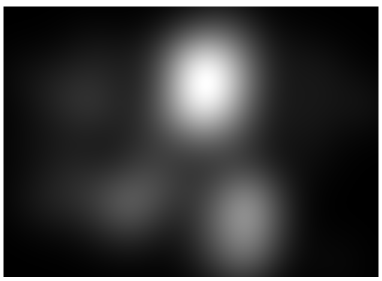
Fig. 5: Ground Truth

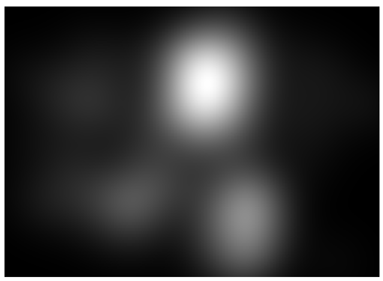
Fig. 6: GFDM


Fig. 7: Blended

4.3 Prehispanic

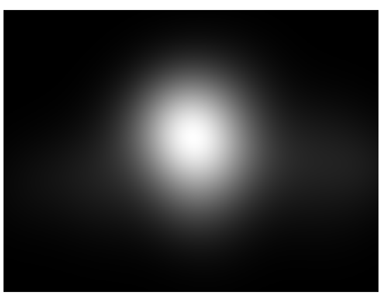
Fig. 8: Ground Truth

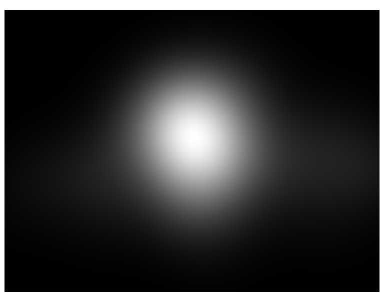
Fig. 9: GFDM


Fig. 10: Blended

5 Conclusion

The analysis of Gaze Fixation Density Maps (GFDM) for the MexCulture142 dataset revealed an average Mean Absolute Error (MAE) of approximately 0.0438, indicating a strong alignment between the generated saliency maps and the ground truth fixation points. This low MAE reflects the effectiveness of the methodology in capturing viewer attention across various architectural styles: Modern, Colonial, and Prehispanic. The findings enhance our understanding of how different architectural features engage viewers, contributing valuable insights to cultural heritage analysis and paving the way for further exploration of user experience in architecture.