Image Segmentation without User Input - Report

Filip Łazowski, 448341

December 16, 2025

1 Grad-CAM Visualization

Below is an exemplary visualization of the pipeline components. It displays the original input image, the generated Grad-CAM heatmap (showing the model's focus on the target shape), the segmentation masks produced by both SAM approaches, and the ground-truth mask.

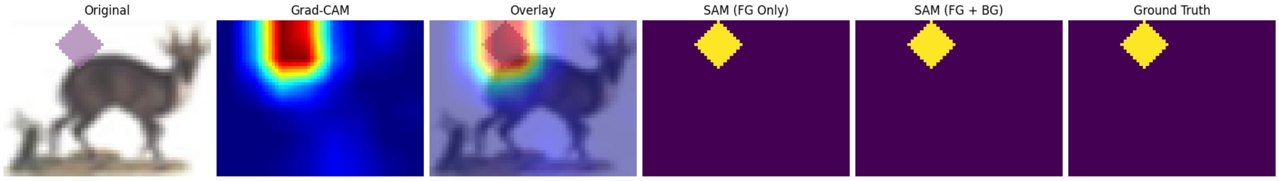

Figure 1: Visual comparison of the input image, Grad-CAM heatmap, Overlay, SAM predictions (Foreground Only and FG+BG), and the Ground Truth mask.

2 SAM Pipelines Description

Two approaches were implemented to automatically generate point prompts for the Segment Anything Model (SAM) based on Grad-CAM heatmaps:

1. **Foreground Only Pipeline:** This approach identifies the single pixel with the maximum value in the Grad-CAM heatmap (global maximum) and feeds its coordinates to SAM as a positive "foreground" point prompt.
2. **Foreground + Background Pipeline:** This approach extends the previous one by additionally identifying the pixel with the minimum value in the heatmap and providing it to SAM as a negative "background" point prompt, aiming to explicitly define irrelevant regions.

3 Evaluation Metrics

Both pipelines were evaluated on the test dataset containing 100 images. The metrics include Intersection over Union (IoU), Hit Rate (percentage of foreground points falling inside the ground-truth mask), and the mean distance of the point from the object's center of mass.

Table 1: Performance comparison of the two SAM pipelines.

Pipeline Approach	**Mean IoU**	**Hit Rate**	**Mean Distance**
Foreground Only	**72.56%**	68.40%	5.99 px
Foreground + Background	71.91%	68.40%	5.99 px

Both methods successfully exceeded the required threshold of 65% IoU.

4 Discussion

The results indicate that adding a background point did not improve the segmentation quality (IoU decreased slightly from 72.56% to 71.91%), suggesting that for this dataset of distinct synthetic shapes, a single well-placed foreground point is sufficient for SAM. To further improve performance, we could extract multiple foreground points (e.g., top-k local maxima) instead of a single global maximum for better cover of larger or non-convex objects. Additionally, using the thresholding Grad-CAM heatmap as a coarse mask prompt instead of point prompts might provide better spatial guidance to the model.