

An exemplary visualization of the output of the Grad-CAM

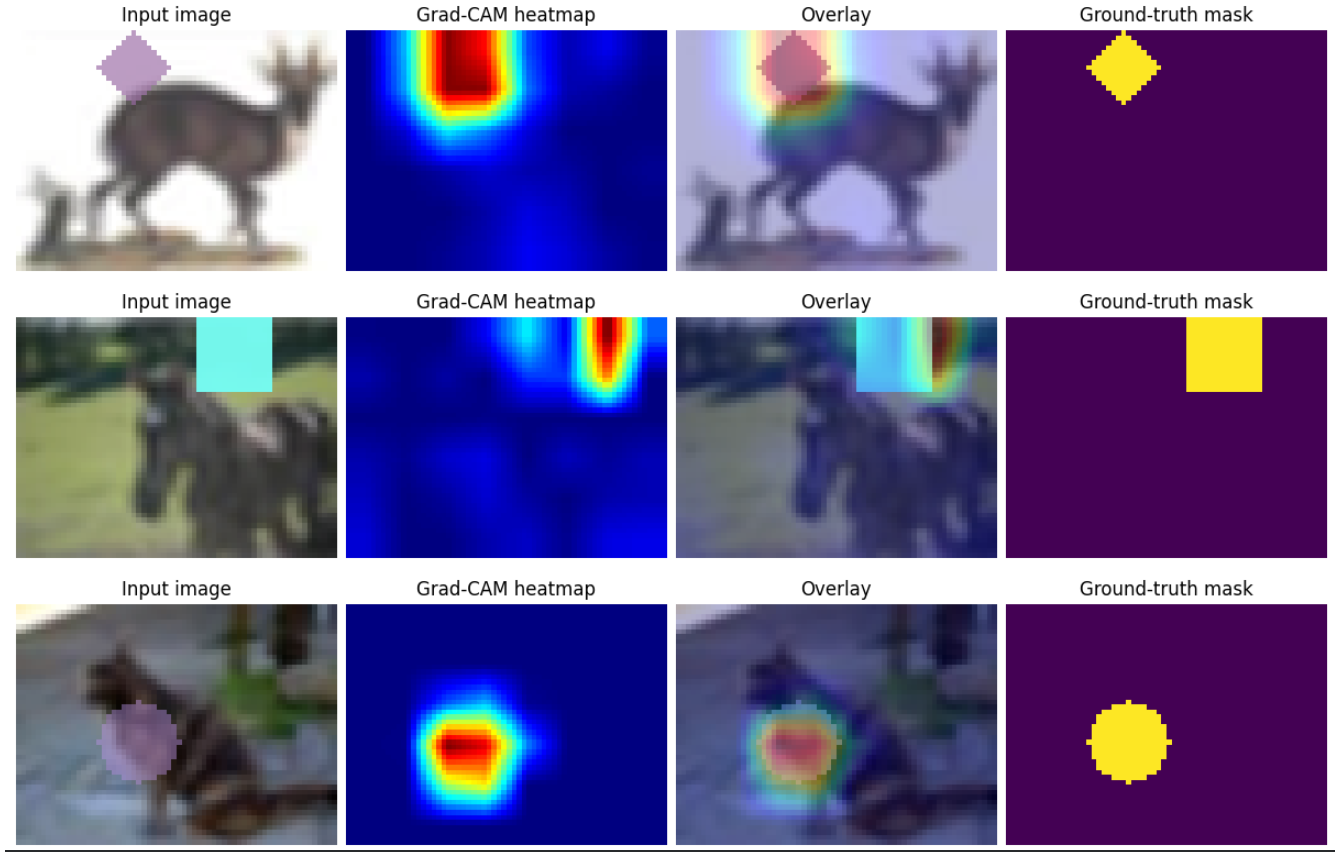


Figure 1: Visualization of the output of the Grad-CAM

A concise description of each approach for the SAM pipelines (1-3 sentences each)

ForegroundOnlySamPipeline

Firstly, we generate an approximate mask from the Grad-CAM heatmap by applying binarization using a threshold value.

This class generates only one point by finding centroid of generated mask. It is analogous to the center of gravity in physics. Assuming every point of our mask has the same weight, the centroid represents the point of balance, allowing the entire mask to be supported by a single finger placed underneath.

ForegroundBackgroundSamPipeline

Here, our mask is generated in the same manner as above.

This class generates only one point for foreground (the same as described above) and one point for background. We define four points, each with a distance of $5\sqrt{2}$ from a respective corner of the image. The background point is the one furthest from the foreground centroid, chosen from those four corner-proximate points.

A presentation of all metrics

ForegroundOnlySamPipeline

IoU	Hit rate	Distance
0.8729	0.2260	4.39

ForegroundBackgroundSamPipeline

IoU	Hit rate	Distance
0.8635	0.2260	4.39

Discussion (up to 5 sentences) on potential areas for improvements

The current approach utilizing only single points for the foreground and background can be enhanced by providing more points or bounding boxes derived from the hottest regions of the Grad-CAM map. The selection of the convolutional layer used by Grad-CAM should also be optimized, as utilizing an earlier layer might yield more detailed activation maps for better object localization. Furthermore, similar object and background colors occasionally lead to the lowest IoU scores, which represents an area for further research. A low IoU is also sometimes exclusively caused by the classification network itself (e.g., Grad-CAM selecting an irrelevant area). This strongly suggests that a solution involving a more robust classification network could significantly improve the initial guidance for the SAM segmentation model.

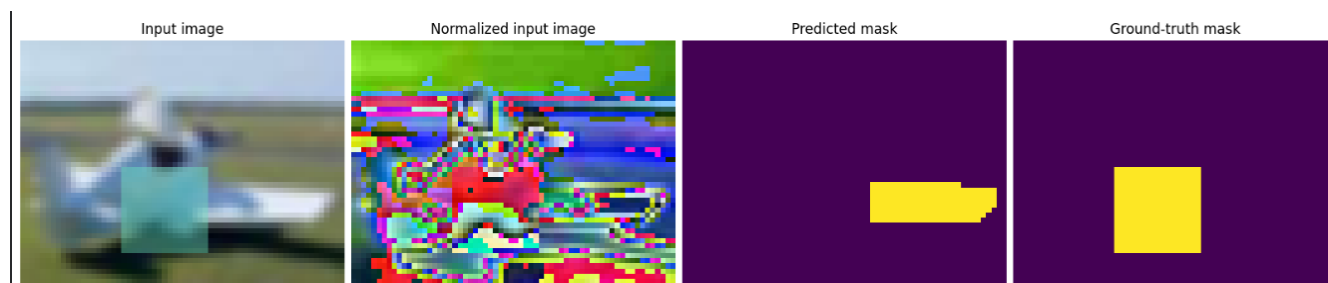


Figure 2: Shape and image have similar colours