

# SIGN: A Statistically-Informed Gaze Network for Gaze Time Prediction and Inference

## Supplementary Material

### 6. More Sample Fixation Maps

#### 6.1. AdGaze-3500

In figures 4 and 5 we show two more sets of 4 examples of fixation maps inferred by SIGN on the images of AdGaze-3500. The inferred maps highlight the texts, products and human faces, which are known to be elements attracting human attention.

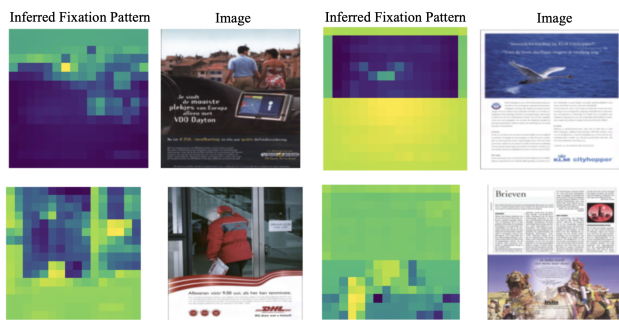


Figure 4. Inferred fixation maps for four samples from AdGaze-3500. For each sample, the left presents the inferred fixation map of the ad image on the right. Brighter regions are predicted to be more conspicuous.

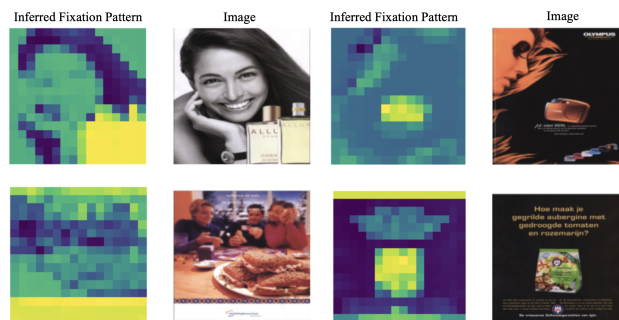


Figure 5. Inferred fixation maps for four more samples from AdGaze-3500.

#### 6.2. COCO-Search18

In figures 6 and 7 we show two more sets of 4 examples of fixation maps inferred by SIGN on the images of COCO-Search18, along with the ground-truth fixation maps. Large overlaps between the inferred fixation maps and the ground-truth ones are observed, especially for the Target Present cases. For Target Absent scenarios, centerbias has a domi-

nant influence on the fixation map, as many scattered fixations are involved in TA tasks.

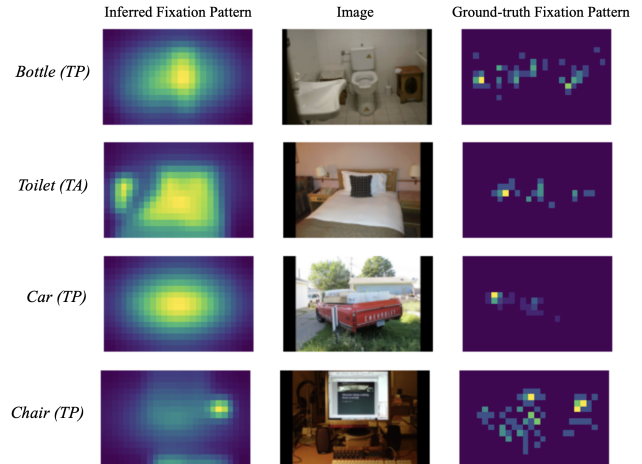


Figure 6. Inferred fixation maps for four samples from COCO-Search18. The inferred fixation maps are on the left, the original image is in the middle, and the ground-truth fixation maps are on the right (blurred by a Gaussian filter with a standard deviation of 35 pixels [10] and resized). Row labels indicate the target object, and the task type (TA: Target Absent; TP: Target Present).

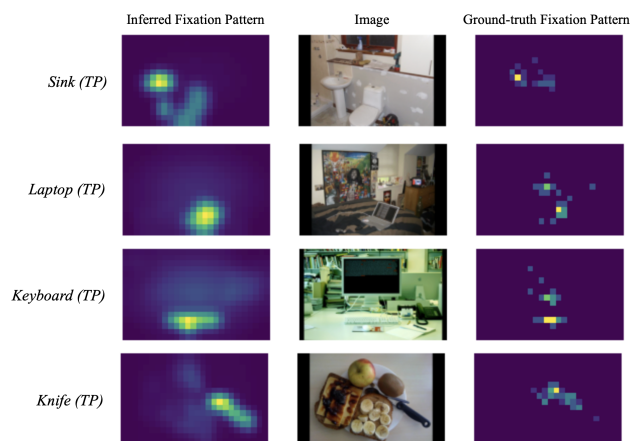


Figure 7. Inferred fixation maps for four more samples from COCO-Search18.