Supplemental Material for A Benchmark of Computational Models of Saliency to Predict Human Fixations

Tilke Judd, Frédo Durand, and Antonio Torralba

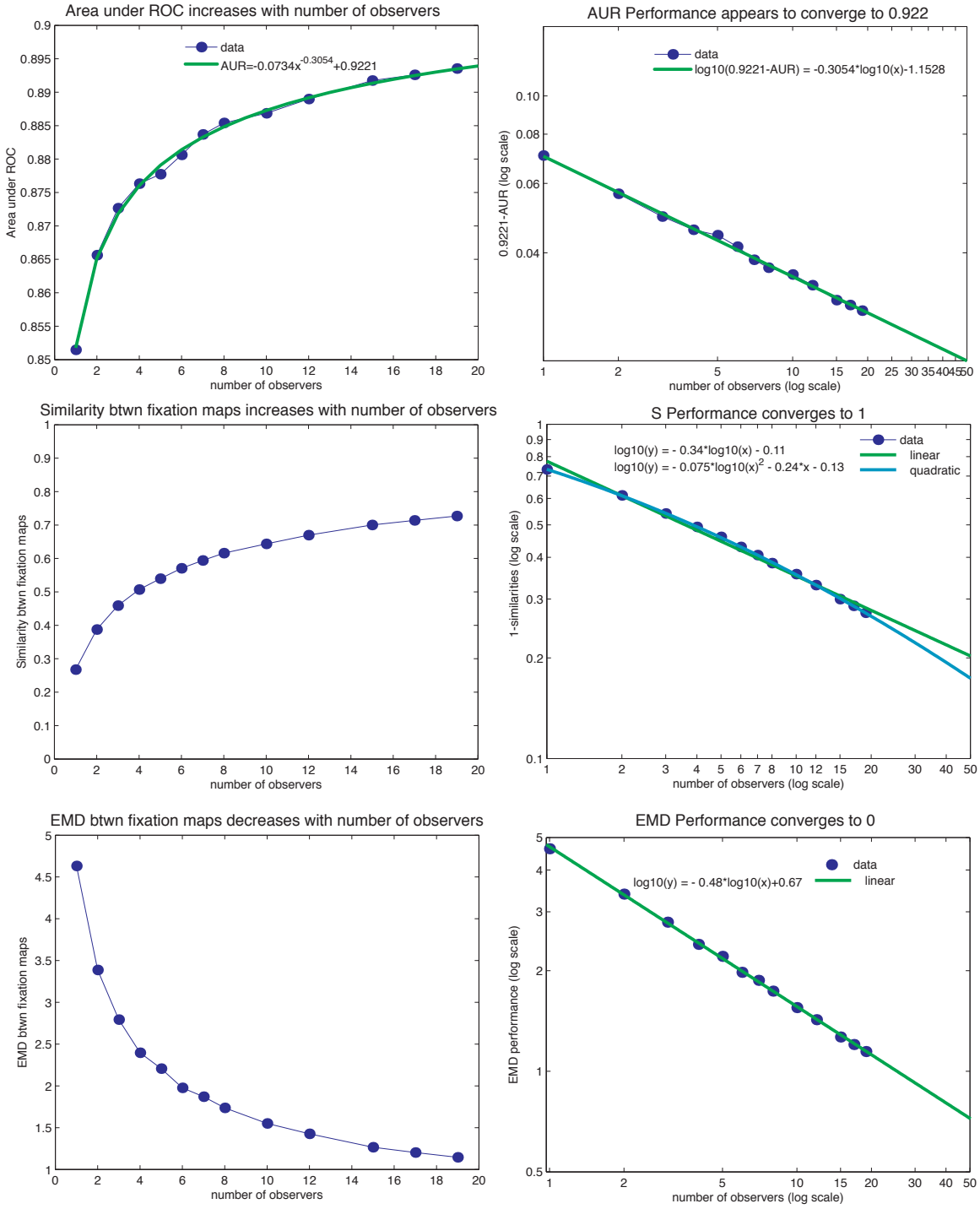

Fig. 1. Measuring human performance. In these plots we see the performance of n observers to predict fixations from n observers (for the ROC metric), and the similarity and earth movers distance between two fixation maps, each of n independent observers. As the number of human observers used to create a human "ground-truth" fixation map increases, performance to predict other humans increases. Therefore the human baseline performance depends on the number of humans used. The limit of the human performance is 0.907 for ROC, 1 for Similarity and 0 for EMD. The plots on the right with fit curves indicate the rate at which performance reaches that limit.

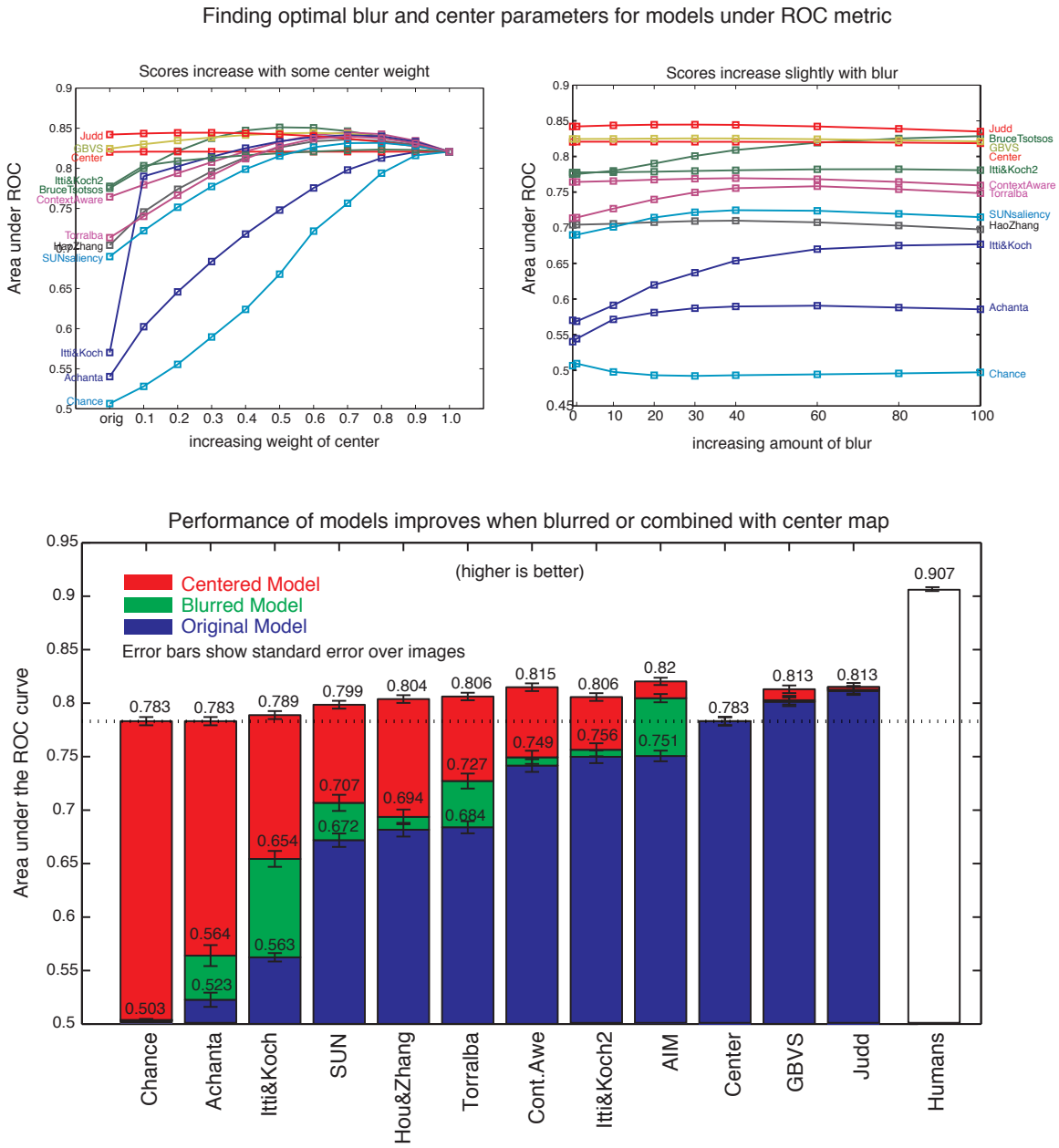

Fig. 2. We optimize the blur and the center parameters for each model under the area under the ROC metric. Top left shows the change in performance with increasing weight of the center bias as tested on 100 images from the MIT ICCV 2009 data set. Top right shows the change in performance with increased blur on the same tests images. While blurring makes only a slight improvement per model, adding a center bias greatly improves many models. The best center bias and blurring parameter for each model was chosen and then used to measure the updated performance on the benchmark data set. Updated performances are shown in the bottom chart.

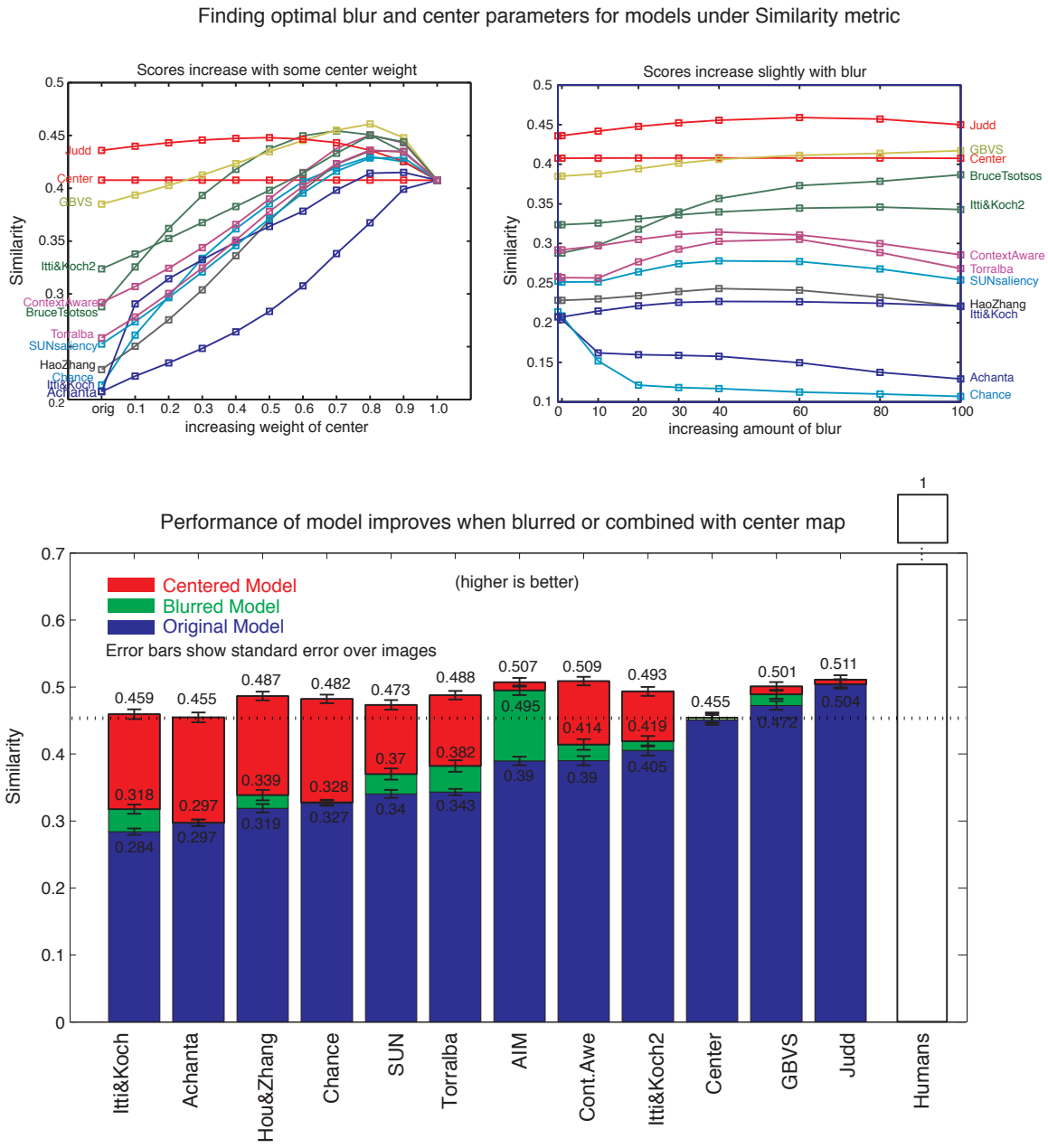

Fig. 3. We optimize the blur and the center parameters for each model under Similarity metric. Further description in Figure 2

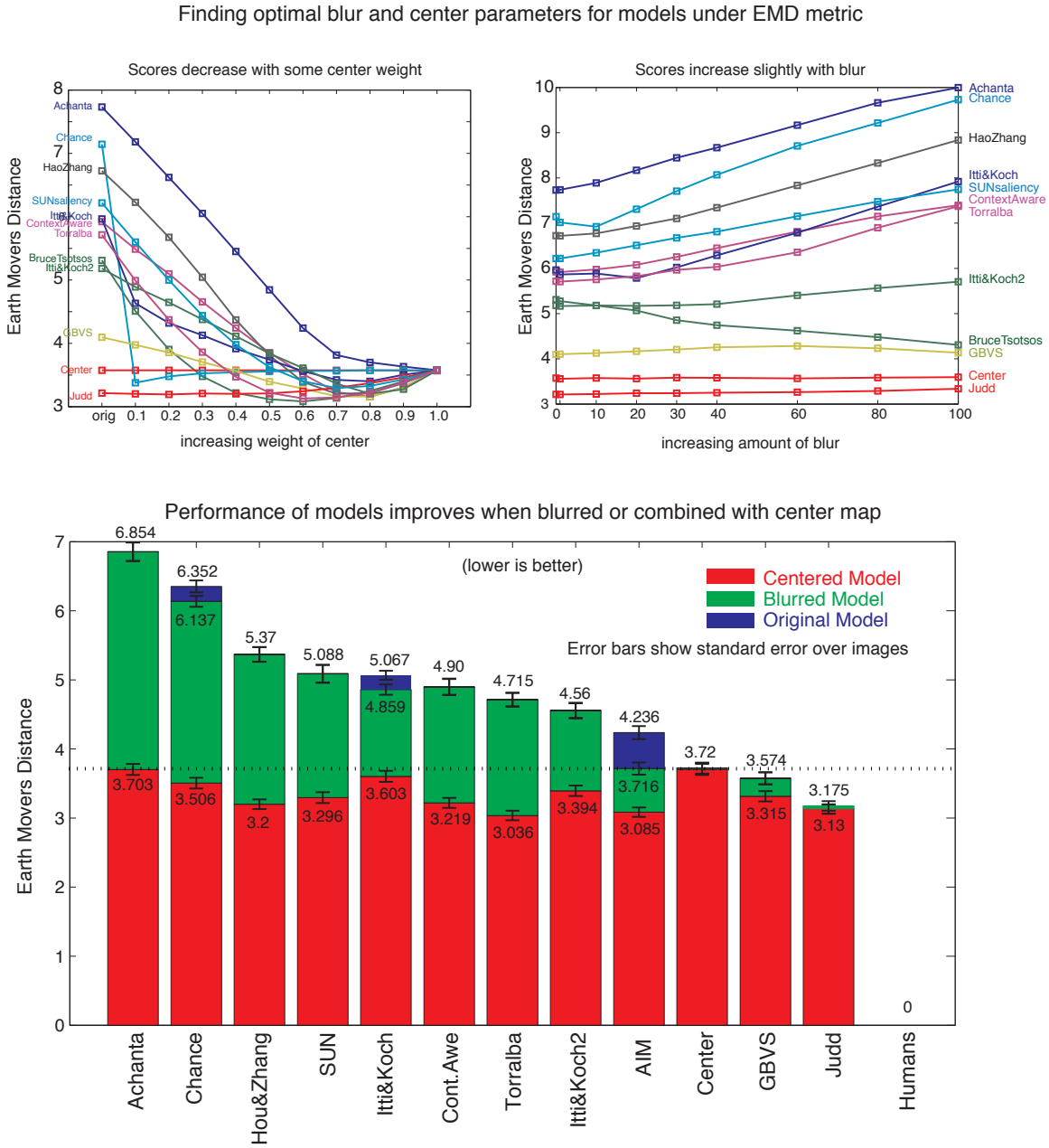

Fig. 4. We optimize the blur and the center parameters for each model under the EMD metric. Further description in Figure 2

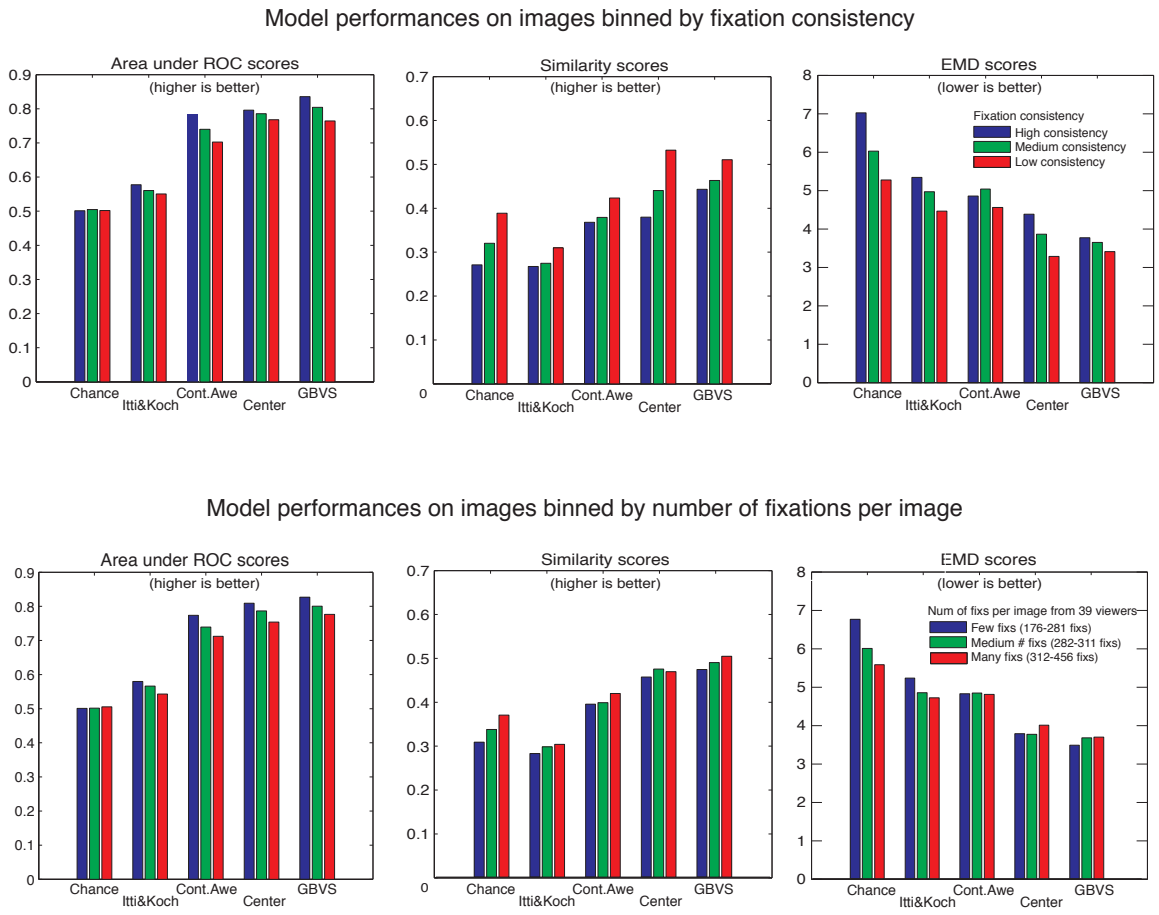

Fig. 5. **Top row:** Performance of 5 models for all three metrics on images binned by fixation consistency. As images become less consistent, performance scores get slightly worse under the ROC metric, but gets slightly better under the Similarity and EMD metrics. On images with low consistency, saliency maps often predict low saliency everywhere which produces higher Similarity and EMD scores but lower ROC scores. **Bottom row:** the same trend is seen when images are binned by number of fixations per image.

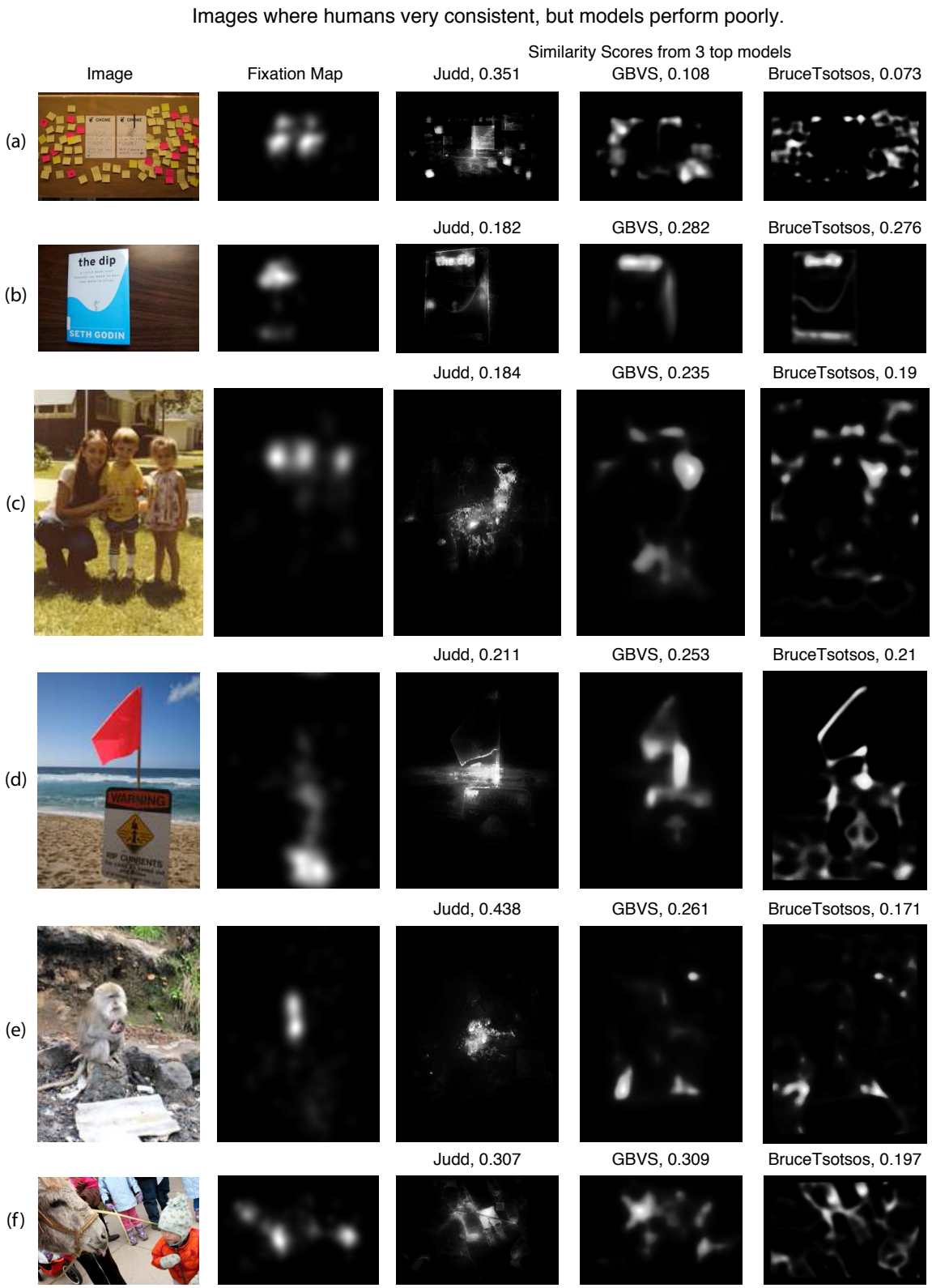

Fig. 6. Current models have a very hard time predicting fixations of some images even though human fixations are very consistent. Here are images for which ratio of the average performance over 3 of the top performing to the consistency of humans models was lowest. For example, in image (d) observers read the text about rip currents rather than look at the bright orange colors. The saliency models do not locate the text at all. In image (e) observers fixate on the monkey's faces while saliency models pick up the bright white paper and orange mushroom in the background.