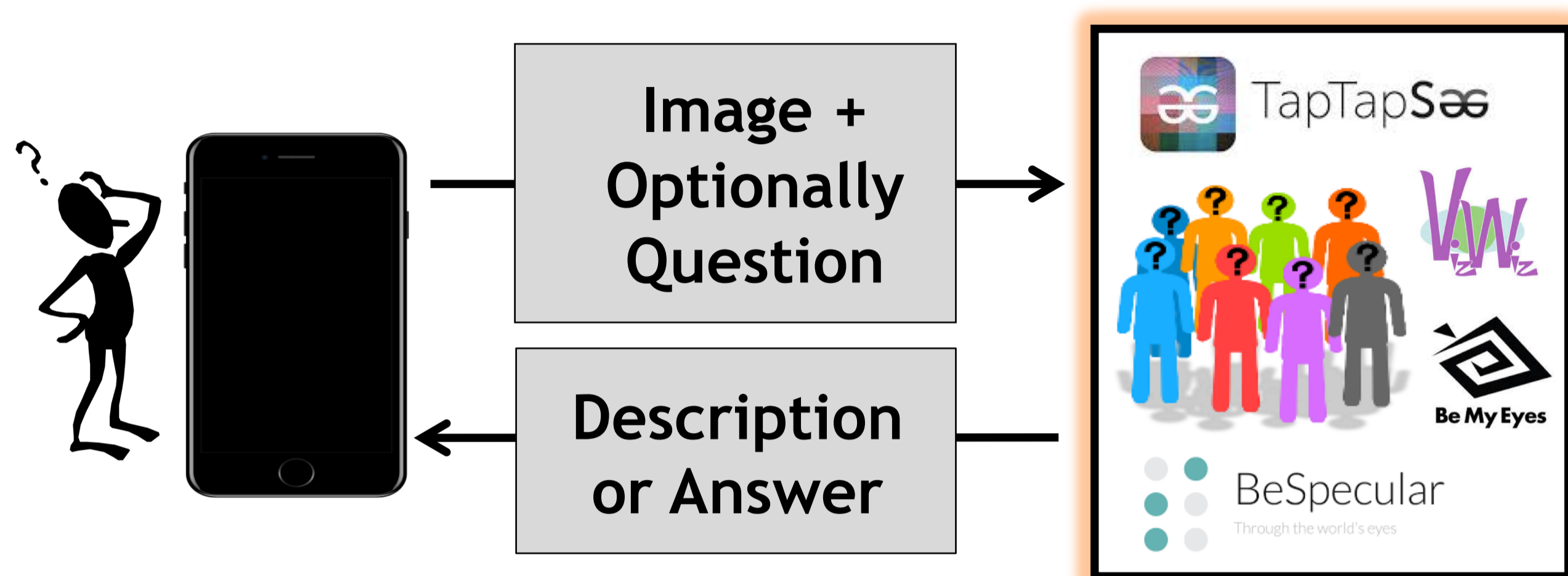


VizWiz-Priv: A Dataset for Recognizing the Presence and Purpose of Private Visual Information in Images Taken by Blind People

Danna Gurari¹, Qing Li², Chi Lin¹, Yinan Zhao¹, Anhong Guo³, Abigale J. Stangl¹, Jeffrey P. Bigham³

¹ The University of Texas at Austin, ² University of California, Los Angeles, ³ Carnegie Mellon University

Situation: Visual Assistance Applications



Observation 1: Private Content is in Images Taken by People Who are Blind

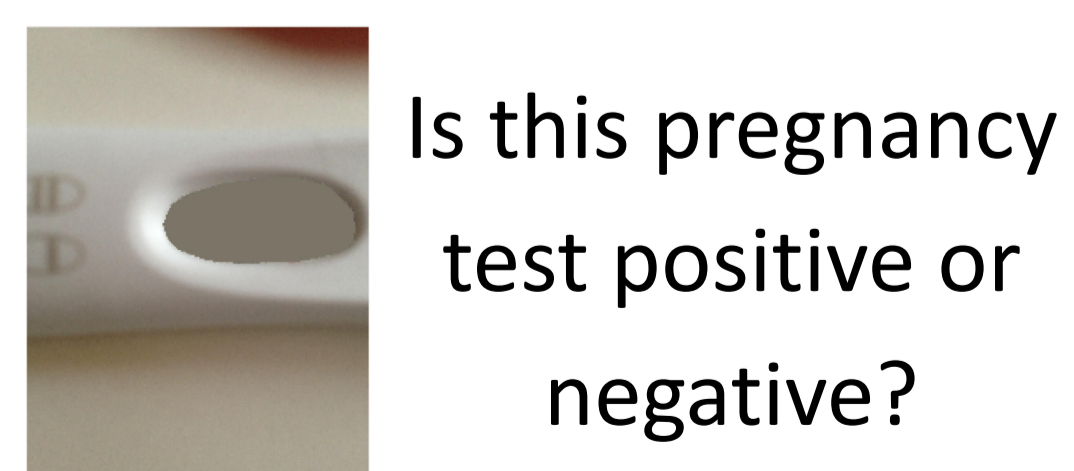
- 12% of their 44,799 pictures show private content.



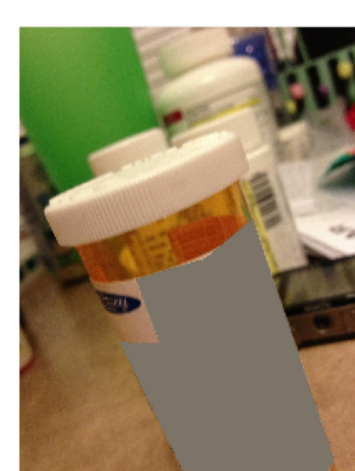
- Recognizing whether private information is in an image is important for supporting people who are blind.

Observation 2: Visual Questions Can Ask About Private Content in Images

- 1 of every 40 VQs (for 33,858) ask for private content.



Is this pregnancy test positive or negative?

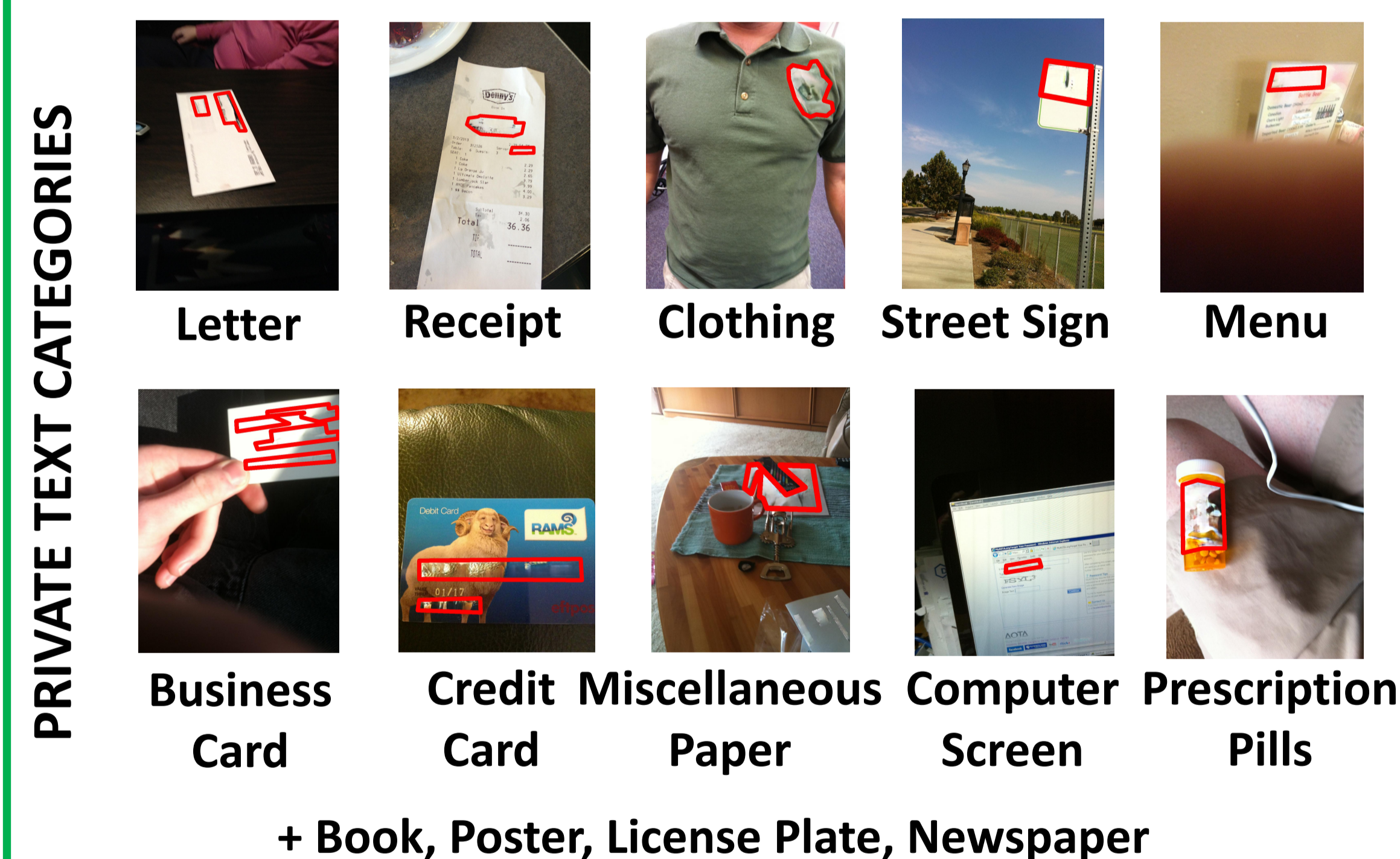
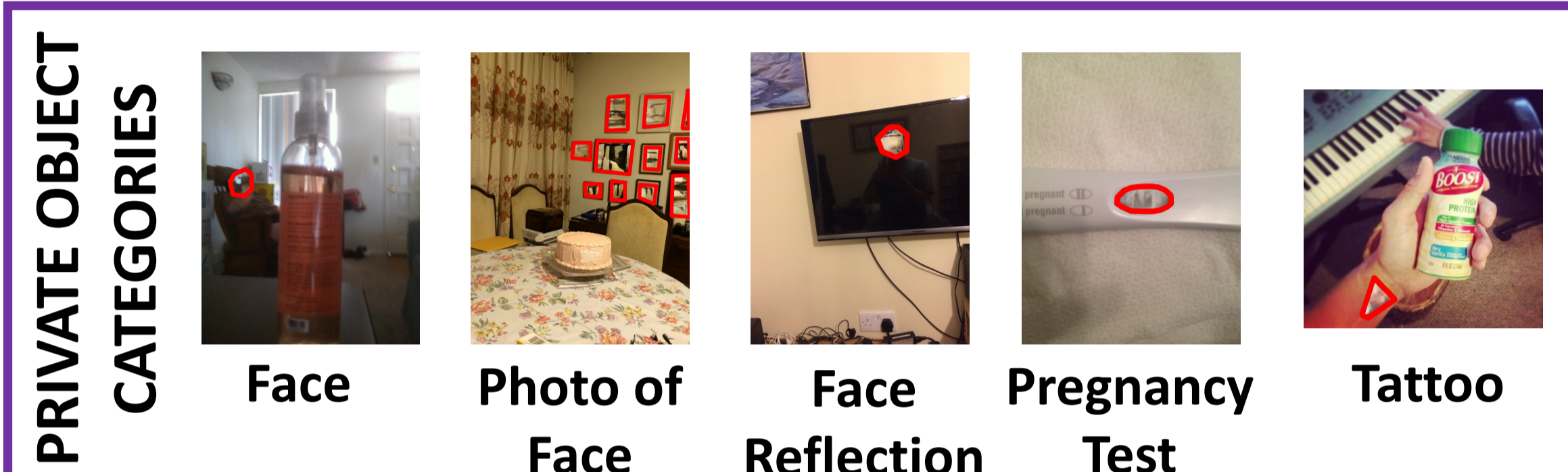


What is the medicine in this pill bottle?

- Knowing if private content is needed to answer a VQ is important for supporting people who are blind.

Visual Privacy Taxonomy & Dataset

- 5,537 private images with 8,862 private regions
- Replaced private regions with (see **red regions**):
 - 1) Inpainting algorithm results [Yu et al. 2018]
 - 2) ImageNet mean



- Of these images, 1,403 come with questions.
- 62% (875) of the VQs ask about private content.

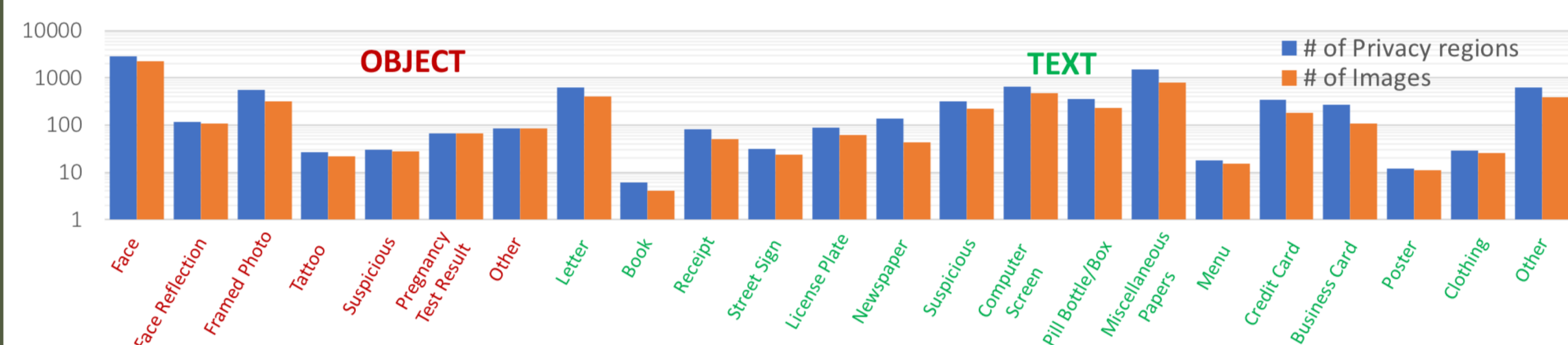
Our dataset originates from people who are blind and supports two tasks: recognizing private information and its purpose from their images and visual questions respectively.

Dataset: <http://www.vizwiz.org/>

Task 1: Is Private Content in the Image?

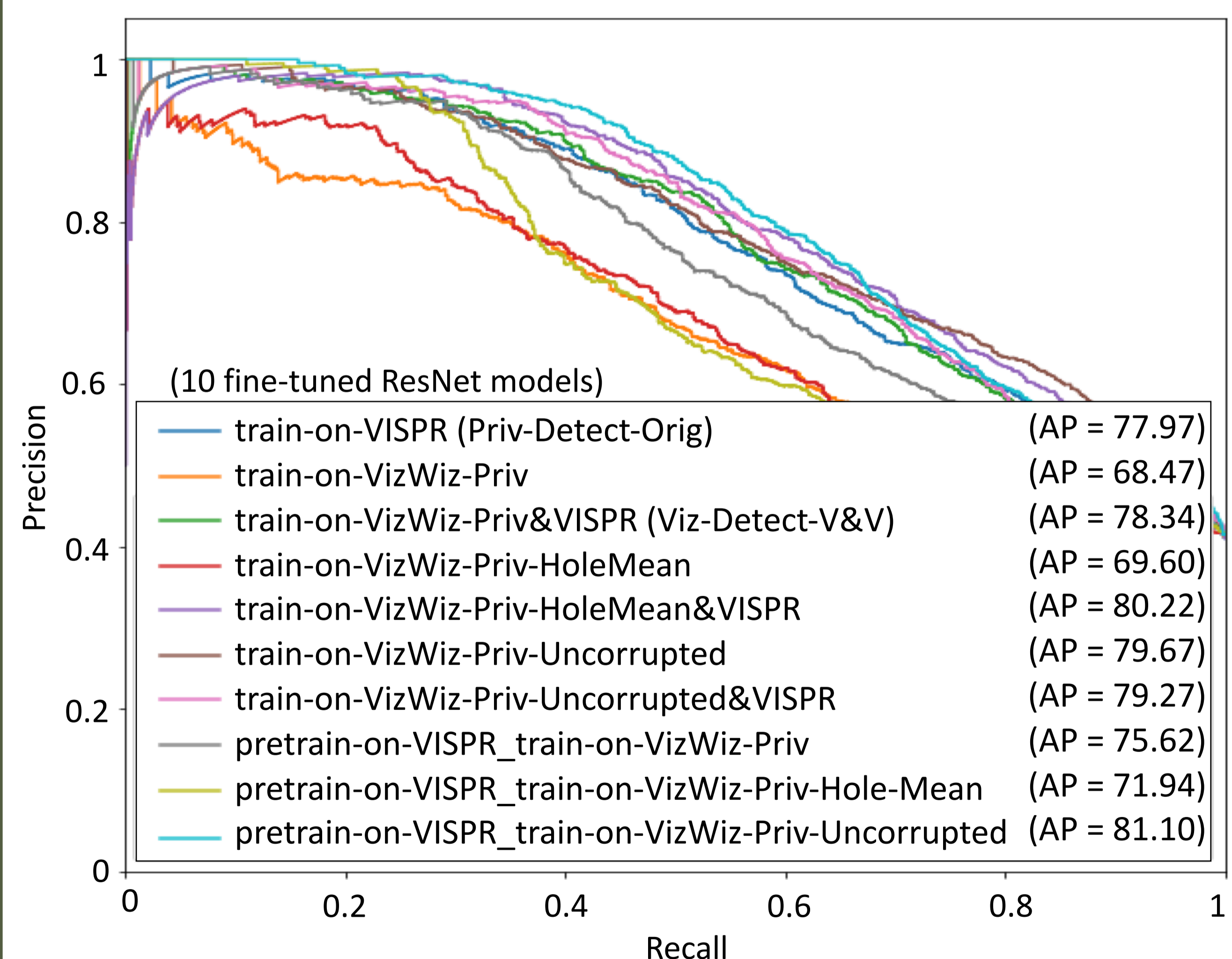
Dataset Analysis:

- 42% of private regions show objects, and the rest text.



Algorithm Benchmarking:

- 5,537 private images and 8,089 non-private images

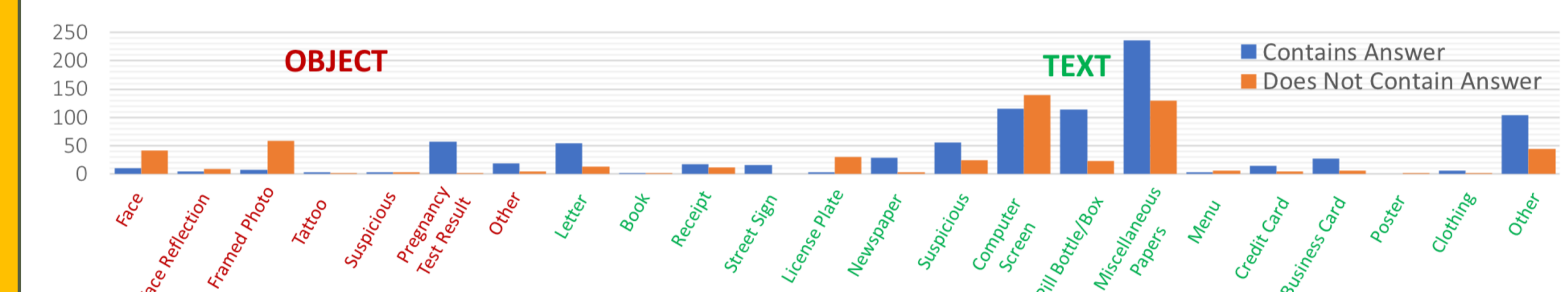


Training algorithms with our corrupted dataset where private content is masked out of the images yields similar performance to training algorithms on the original, uncorrupted images.

Task 2: Is Private Content in the Image Needed to Answer the Visual Question?

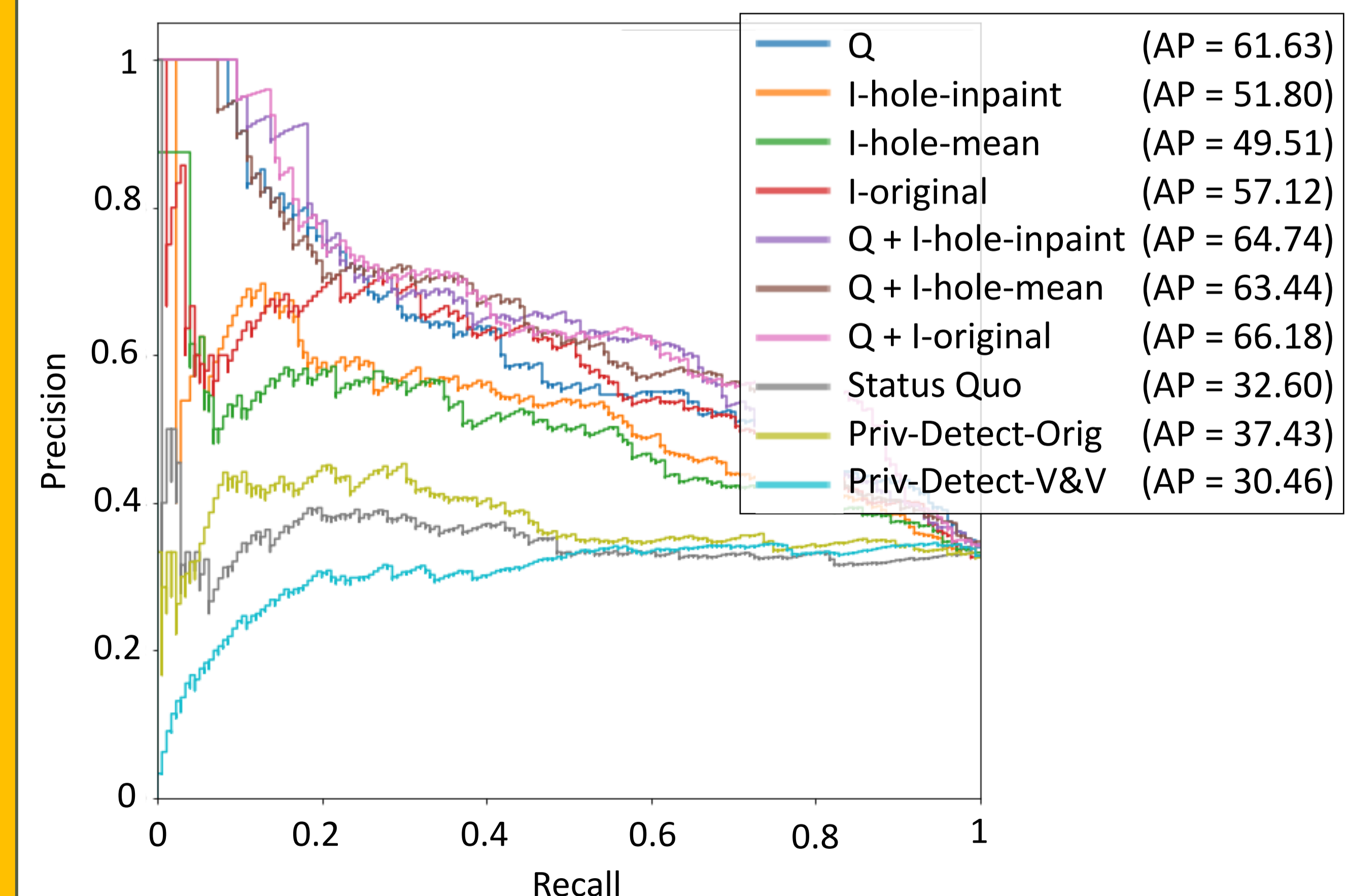
Dataset Analysis:

- Private content is needed for images of pregnancy tests, pill bottles, letters, street signs, & credit cards but not for faces, framed pictures, & license plates.



Algorithm Benchmarking:

- Q: GRU trained on GloVe embedding
- I: 2-layer NN trained on ResNet-50 features



Algorithms can predict whether private content in an image is needed to answer the visual question.