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

- Supplementary Materials

Danna Gurari[1], Qing Li[2], Chi Lin[1], Yinan Zhao[1], Anhong Guo[3], Abigale Stangl[4], Jeffrey P. Bigham[3]

[1] University of Texas at Austin, [2] University of California, Los Angeles,
[3] Carnegie Mellon University, [4] University of Colorado Boulder

Supplementary Materials

This document supplements the main paper with the following details:

1. Algorithm benchmarking on the VISPR dataset (supplements **Section 5.1**).
2. Correlation of predictions on uncorrupted and corrupted versions of VizWiz-Priv images (supplements **Section 5.1**).
3. Most confident predictions for recognizing that private information is present and is not present by the top-performing model tested on the uncorrupted images in VizWiz-Priv (supplements **Section 5.1**).
4. Most confident predictions that the question is asking about and is not asking about private content in an image by the top-performing model tested on the uncorrupted images in VizWiz-Priv (supplements **Section 5.2**).
5. Analysis of algorithm performance with respect to each privacy type (supplements **Section 5.2**).

1. Cross-Dataset Algorithm Benchmarking

We benchmarked the performance of the ten privacy recognition algorithms described in the main paper in Section 5.1 to show how well they generalize to the VISPR dataset [1]. Figure 1 shows the detailed results, which correspond to the results that were summarized in the "Results on VISPR Images" section of the main paper. As shown, the algorithms trained on VizWiz-Priv generalize well to the VISPR dataset.

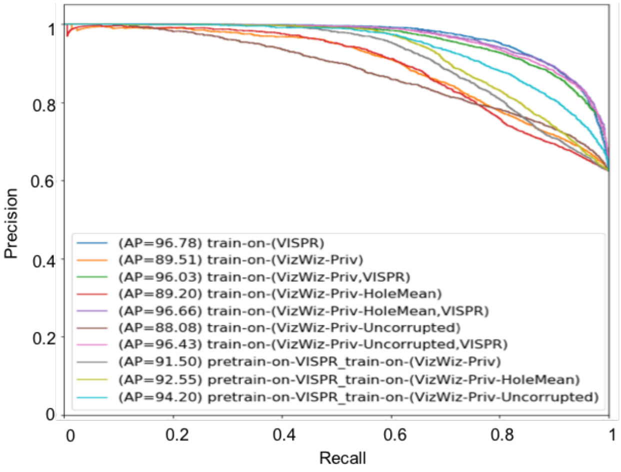


Figure 1. Precision-recall curves and average precision scores for privacy recognition algorithms evaluated on the VISPR dataset.

2. Prediction Correlations

We summarized in the “Results on VizWiz-Priv Images” section of the main paper the similarity of predicted scores from the ten benchmarked privacy recognition algorithms when tested on the original, uncorrupted VizWiz-Priv images versus the VizWiz-Priv images that have inpaintings to replace the private content. These results highlight the suitability of the publicly-available VizWiz-Priv images to offer a reasonable privacy-free substitute for benchmarking algorithm performance. Table 1 shows the detailed results, measured using Pearson’s correlation coefficient, for the similarity between the predictions on the two sets of images across all ten algorithms described in the main paper:

	Correlation
train-on-VISPR	0.75
train-on-VizWiz-Priv	0.70
train-on-{VizWiz-Priv,VISPR}	0.89
train-on-VizWiz-Priv-HoleMean	0.93
train-on-{VizWiz-Priv-HoleMean, VISPR}	0.72
train-on-VizWiz-Priv-uncorrupted	0.72
train-on-{VizWiz-Priv-uncorrupted, VISPR}	0.87
pretrain-on-VISPR,finetune-on-VizWiz-Priv	0.89
pretrain-on-VISPR, finetune-on-VizWiz-Priv-HoleMean	0.93
pretrain-on-VISPR, finetune-on-VizWiz-Priv-uncorrupted	0.78

Table 1. Comparison of predicted scores on the original, uncorrupted images in VizWiz-Priv and the hole-filled images in VizWiz-Priv.

We also illustrate in Figure 2 the correlation of the predicted scores for the top-performing privacy recognition from the main paper that uses the publicly-available VizWiz-Priv: train-on-{VizWiz-Priv-HoleMean,VISPR}:

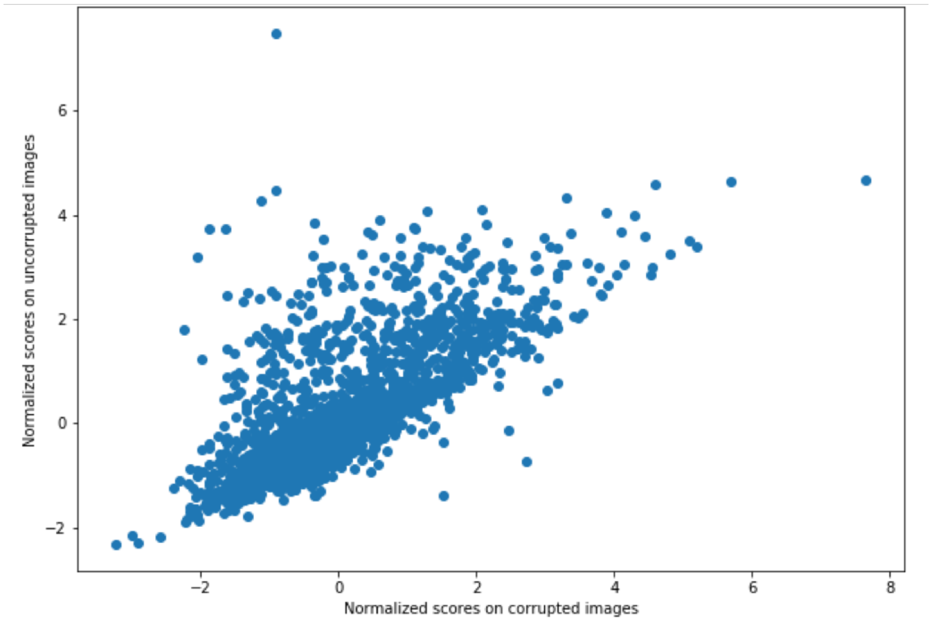


Figure 2. Scatter plot showing similarity of predicted scores from the top-performing algorithm (train-on-{VizWiz-Priv-HoleMean,VISPR}) when predictions were made on the original, uncorrupted images in VizWiz-Priv versus the hole-filled images in VizWiz-Priv.

3. Confident Predictions by the Privacy Recognition Algorithm

We also show the top 10 most confident predictions for when images show private and non-private information on the uncorrupted VizWiz-Priv test set in Figures 3a and 3b respectively. This is shown for the top prediction model from the

Category		Q	Q+I-hole-inpaint	Improvement
Overall		61.63%	64.74%	3.11%
Object	Face	1.22%	1.22%	0.01%
Object	Framed Photo	4.94%	12.26%	7.32%
Object	Other	1.39%	2.93%	1.54%
Object	Pregnancy Test Result	68.45%	68.51%	0.06%
Object	Tattoo	0.82%	1.43%	0.61%
Text	Business Card	14.23%	15.61%	1.38%
Text	Computer Screen	5.47%	4.12%	-1.35%
Text	Credit Card	9.90%	7.13%	-2.76%
Text	Letter	17.68%	22.60%	4.91%
Text	Miscellaneous Papers	27.11%	37.67%	10.56%
Text	Newspaper	4.66%	3.13%	-1.53%
Text	Other	10.16%	10.04%	-0.13%
Text	Pill Bottle/Box	68.17%	74.81%	6.64%
Text	Receipt	7.17%	5.81%	-1.36%
Text	Street Sign	4.01%	1.01%	-3.00%
Text	Suspicious	2.44%	3.72%	1.28%

Table 2. The average precision for `Q` and `Q+I-hole-inpaint` with respect to each privacy type.

main paper for which the training data will be publicly-available: `train-on-(VizWiz-Priv-HoleMean, VISPR)`. These findings suggest that the predictor is most confident about finding faces for private information, but is not yet able to distinguish when faces are on advertisements or products and so not private. These findings also highlight the predictor may be picking up on uniform textures across the image as predictive of a non-private image.

4. Confident Predictions by the (Un)necessary Privacy Leak Recognition Algorithm

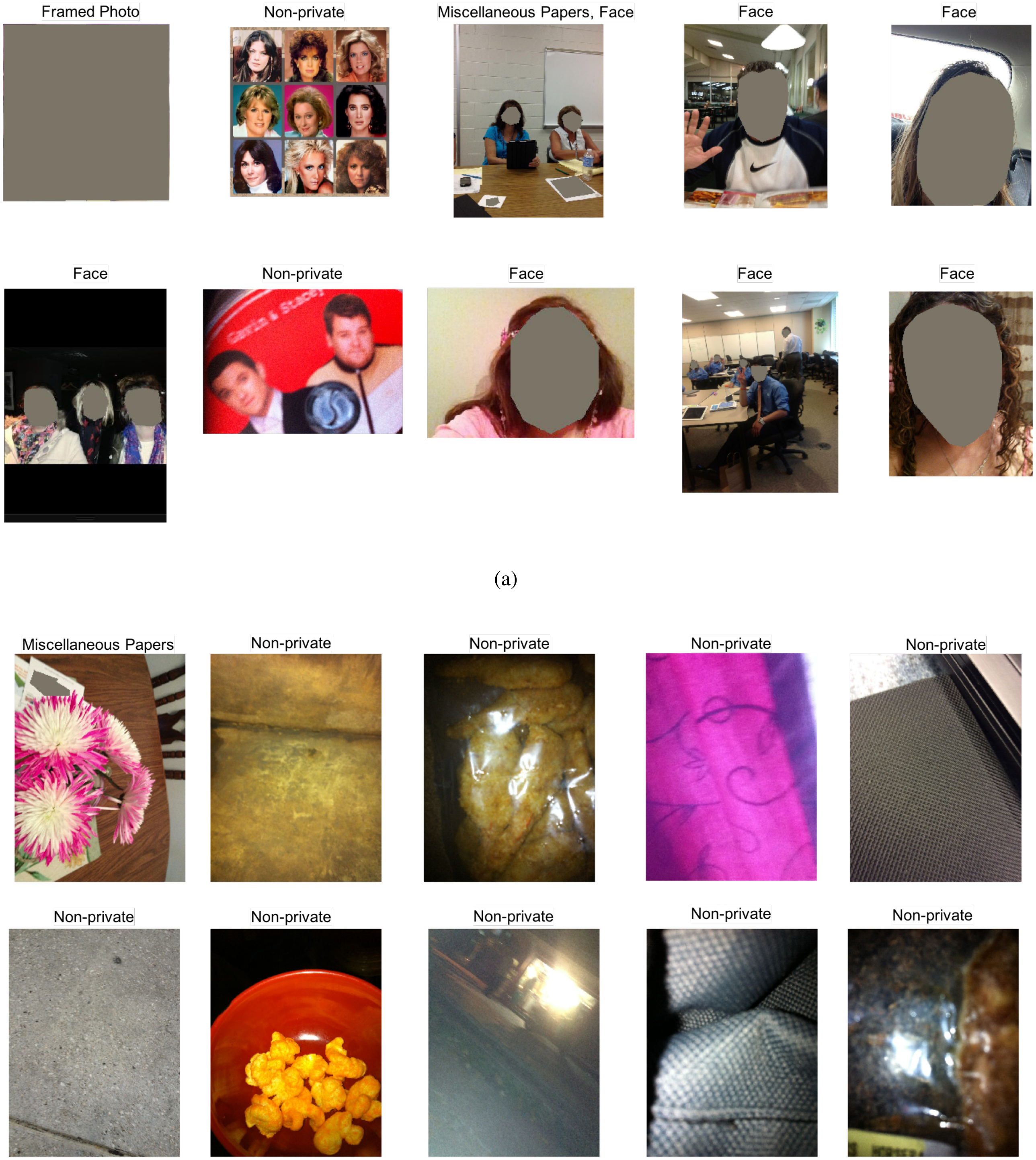
We next show the top 10 most confident predictions for when questions ask about private content in an image versus do not ask about private content in an image in Figures 4a and 4b respectively. This is shown for the top prediction model from the main paper for which the training data will be publicly-available: `Q+I-hole-inpaint`. As shown, the algorithm appears most confident for pregnancy tests and pill bottles.

5. Algorithm Analysis Per Privacy Type

Finally, we supplement the Section 5.2 experiments to illustrate the benefit of the image information. To do so, we compare the results for `Q` and `Q+I-hole-inpaint` with respect to each privacy type. Results are shown below in Table 2. The greatest gains for `Q+I-hole-inpaint` over `Q` alone are for miscellaneous papers (10.6%), framed photos (7.3%), pill bottles/boxes (6.6%), and letters (4.9%).

References

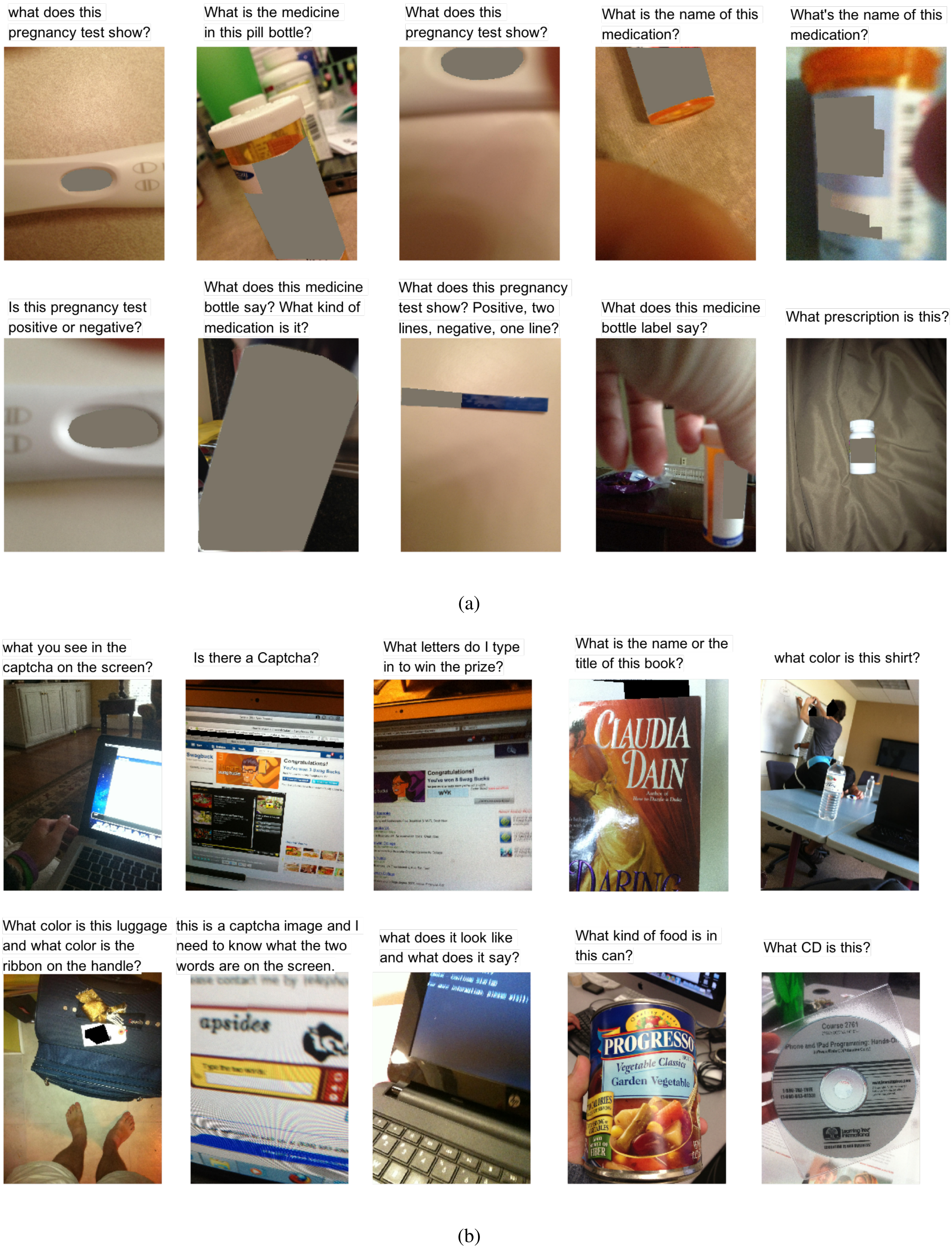
[1] T. Orekondy, B. Schiele, and M. Fritz. Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In *Computer Vision (ICCV), 2017 IEEE International Conference On*, pages 3706–3715. IEEE, 2017. 1



(a)

(b)

Figure 3. Top 10 most confident predictions by the top-performing `train-on-(VizWiz-Priv-HoleMean, VISPR)` model for predicting if private information is (a) present and (b) not present. We mask out the private information for these publicly-available images to preserve privacy.


what does this pregnancy test show?
What is the medicine in this pill bottle?
What does this pregnancy test show?
What is the name of this medication?
What's the name of this medication?
Is this pregnancy test positive or negative?
What does this medicine bottle say? What kind of medication is it?
What does this pregnancy test show? Positive, two lines, negative, one line?
What does this medicine bottle label say?
What prescription is this?
what you see in the captcha on the screen?
Is there a Captcha?
What letters do I type in to win the prize?
What is the name or the title of this book?
what color is this shirt?
What color is this luggage and what color is the ribbon on the handle?
this is a captcha image and I need to know what the two words are on the screen.
what does it look like and what does it say?
What kind of food is in this can?
What CD is this?

(a)

(b)

Figure 4. Top 10 most confident predictions by the top-performing `Q+I-hole-inpaint` model for predicting if a question is asking about (a) private information in an image and (b) non-private information in an image. We mask out the private information for these publicly-available images to preserve privacy.