MSDS631
Deep Learning Project
Helped blind people

Group 18: Victor Zhang, Hsuan Lin

Task: Image Captioning

- **We train a model to help the blind people to know the world around**

Captioning Images Taken by People Who Are Blind

Danna Gurari, Yinan Zhao, Meng Zhang, Nilavra Bhattacharya

University of Texas at Austin

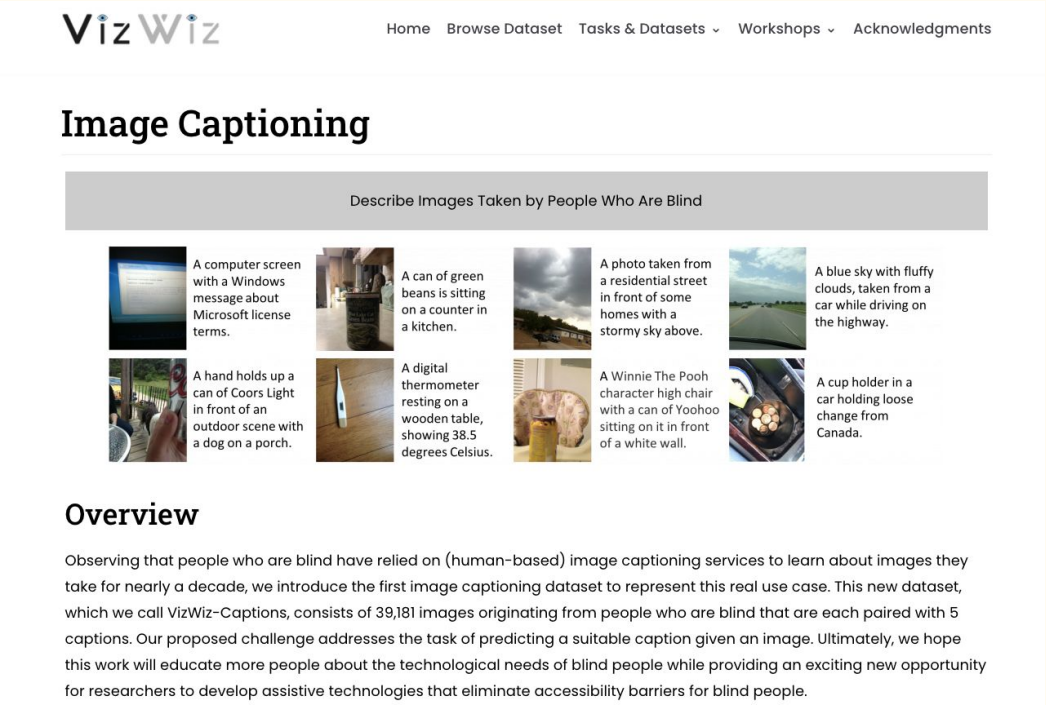
Abstract. While an important problem in the vision community is to design algorithms that can automatically caption images, few publicly-available datasets for algorithm development directly address the interests of real users. Observing that people who are blind have relied on (human-based) image captioning services to learn about images they take for nearly a decade, we introduce the first image captioning dataset to represent this real use case. This new dataset, which we call VizWiz-Captions, consists of over 39,000 images originating from people who are blind that are each paired with five captions. We analyze this dataset to (1) characterize the typical captions, (2) characterize the diversity of content found in the images, and (3) compare its content to that found in eight popular vision datasets. We also analyze modern image captioning algorithms to identify what makes this new dataset challenging for the vision community. We publicly-share the dataset with captioning challenge instructions at `https://vizwiz.org`.

Paper: https://arxiv.org/pdf/2002.08565.pdf

Dataset

The VizWiz-Captions dataset includes:

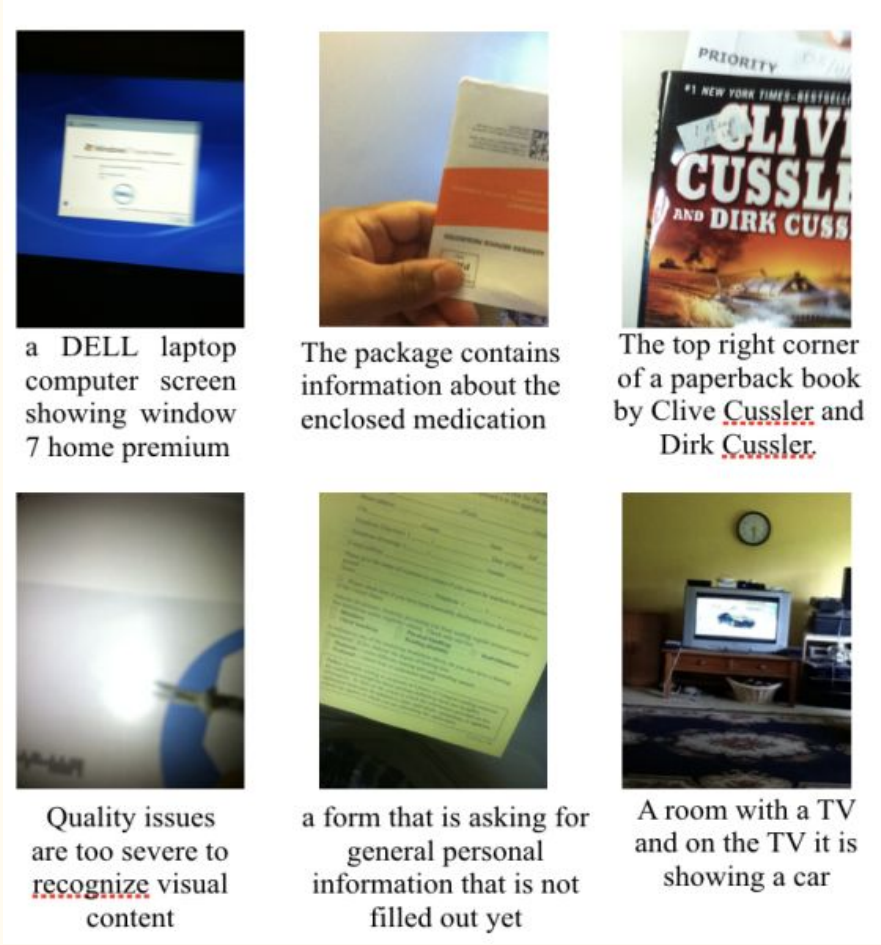
- 23,431 training images
- 117,155 training captions
- 7,750 validation images
- 38,750 validation captions
- 8,000 test images
- 40,000 test captions



https://vizwiz.org/tasks-and-datasets/image-captioning/

Plan

1. Implement different model architectures and compare the model performances
2. Train the model ourself (or use the pre-trained weight)
3. Collect image data from our daily life to see the prediction results
4. Put our code on GitHub



Code Reference: https://github.com/yahoo/object_relation_transformer

Thank you!