



Privacy Issues with Visual Datasets

How to protect the privacy of incidental people in images?



Husky







Beer bottle

Problem

- 1. The ImageNet Challenge has only 3 people-centric categories (scuba diver, bridegroom, baseball player) among its 1000 categories. But people show up incidentally in the images
- 2. Similar issues for other visual datasets such as Open Images, Places, and COCO
- 3. People's privacy is at risk since these datasets are freely available and widely used ^[1]

Private information in images

- 1. Faces, addresses, credit cards, etc. ^[2]
- 2. We focus on faces

Contributions

- 1. Annotate faces in ImageNet, facilitating subsequent research on privacy protection
- 2. Empirically demonstrate that face obfuscation does not hurt large-scale visual recognition

Faces in ImageNet

- Annotated face bounding boxes on 1,431,093 images
- 243,198 images have at least one face (17%)
- 562,626 faces in total
- **106** categories have more than half images with faces







A Study of Face Obfuscation in MAGENET

Kaiyu Yang¹, Jacqueline Yau², Li Fei-Fei², Jia Deng¹, and Olga Russakovsky¹

Department of Computer Science, Princeton University Department of Computer Science, Stanford University



