Towards a Visual Privacy Advisor: Understanding and Predicting Privacy Risks in Images

Tribhuvanesh Orekondy, Bernt Schiele, Mario Fritz

Max Planck Institute for Informatics, Saarbrücken, Germany

ICCV'17

goo.gl/euCv7N

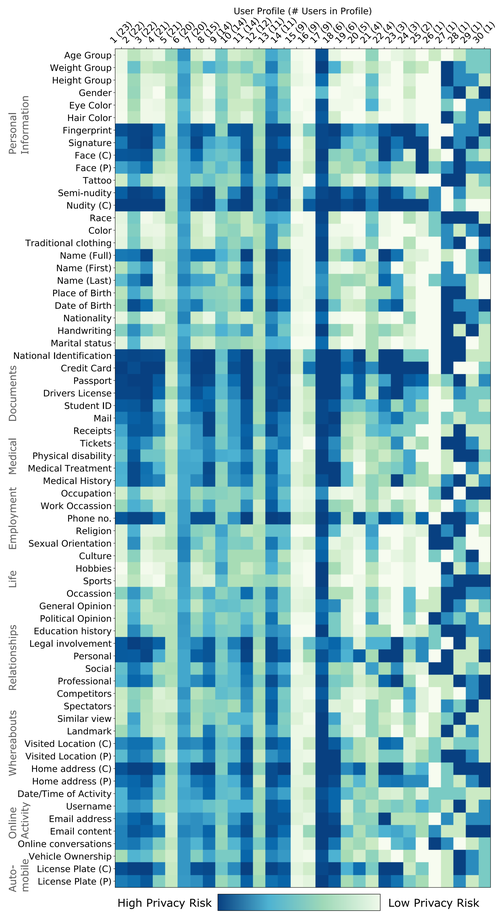
Motivation

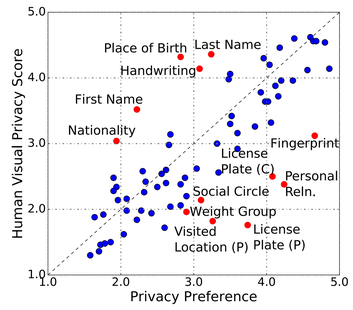
- Images contain a broad range of private information
- Users unintentionally expose such information when sharing images online (e.g, Twitter, Flickr, Facebook)
- Can we extend the concept of "privacy settings" to visual content?

The Visual Privacy (VISPR) Dataset

- ~22k publicly available Flickr images
- Challenges from natural everyday scenes: foreground and background clutter, text content
- 68 Privacy Attributes across 9 groups
- Attributes gathered from: EU Data Protection Directive, US Privacy Act of 1974, Social network website rules
- ~116k assigned attributes, 5.22 per image

User Studies





Study 2: Users' Visual Privacy Assessment

- For each attribute, users:
  a) assess privacy risk based on a group of images (y-axis) and then
  b) provide explicit privacy preference for the attribute (x-axis)
- Users are inconsistent in enforcing their own privacy preferences
- With everyday images (relationships, landmarks, etc.), users severely underestimate privacy risk

Study 1: Understanding User Preferences

- Diverse Preferences ⇒ Same image, different privacy risks
- Some users especially sensitive to some attributes

Acknowledgements

This research was partially supported by the German Research Foundation (DFG CRC 1223)

Approach

Privacy Attribute Prediction

- User-independent multilabel classification task: given an image, predict multiple attributes
- We compare multiple baseline methods for this task
- ResNet-based model achieves an MAP of 47.45 %

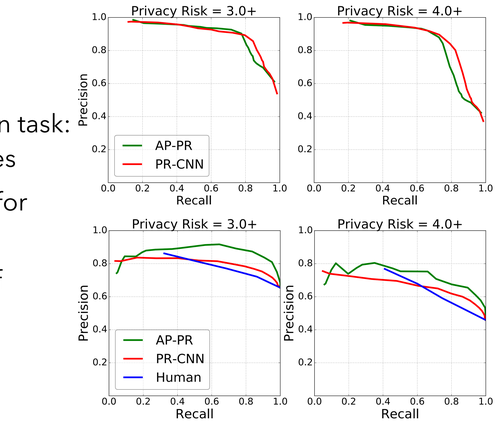
Personalizing Privacy Risk

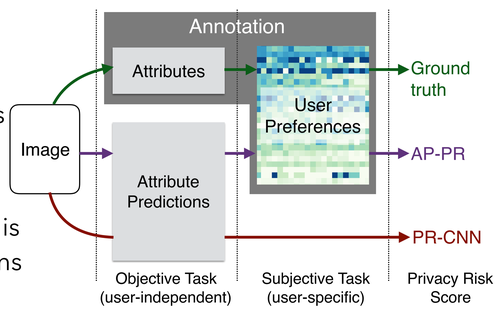
Let Privacy Risk = $\max_a$ (privacy rating of attribute a in image)

Two proposed approaches:

1. AP-PR: Directly uses attribute predictions (AP) to estimate privacy risk (PR)

2. PR-CNN: End-to-end learning to predict user-specific privacy risk from images. This is better at handling noisy attribute predictions





Humans vs. Machine

- We compare these approaches to users' visual privacy assessment (Study 2)
- Our approach achieves better Precision-Recall and L1 scores for the same images, when compared to users themselves