GOOD SYSTEMS
A UT Grand Challenge

Being Watched: Technical Challenges

Tong Wang[1], Zhangyang (Atlas) Wang[1]
[1]Chandra Family Department of Electrical and Computer Engineering

Introduction

Video analytics and software design research will probe the possibilities for engineering privacy levels into computer vision algorithms so that the video streams from public cameras can be 'tagged' at different levels of visual acuity, leading toward what we call differential access strategies in order to use camera-generated visual data in productive ways.

Differential access grapples with managing the utilities in visual data that enhance public culture and society more broadly while balancing the needs for privacy and the tenets of fairness, appropriate access, and transparency.

Objective

Create a versatile set of visual feature extractors, each of which has different trade-offs between:

- the extracted feature's utility expectation (e.g., how sophisticated a task it will be used for)
- the privacy leak risk incurred by sharing it (e.g, what privacy attributes a third party may be able to probe by accessing)

Challenges

To achieve the goal, we aim to solve multiple challenges:

- The first challenge is to identify and annotate identification-related attributes in smart city scenarios
- The second challenge is to provide an innovative automatic solution of learning-based image transformation for de-identification
- The last challenge is to perform evaluation of our algorithm. Specifically, we need to show that applying the transform will not hurt the utility task performance, but at the same time can remove the identification information

Project Overview

Challenge 1: Identify and annotate identification-related attributes

Question: Why not just faces? Many visual elements could result in unintended identification, such as hair, ears, jewelry, tattoos, or clothing with names or symbols

How?

1. Compile a general-purpose list of attributes akin to the privacy setting in daily life
2. Annotate and record a range of identification-related attributes that occur in daily scenarios

Why it is difficult?

1. The attribute list needs to be comprehensive to avoid unintended leakage of privacy
2. The annotation of attributes require significant efforts

Challenge 2: Automatic learning-based image transformation for de-identification

Goal: (1) to obscure the identity information; (2) and to preserving the utility-related (behavior and contextual) information

Ad-hoc baselines

Original image | HoG features | Optical flows

The problem of ad-hoc methods:
- privacy protection came as a "side product", making the privacy protection capability very limited

Learning-based methods

We can use the classification model's accuracy and/or confidence in correctly predicting an attribute (being identification- or utility-related), as a metric of evaluating how well this particular attribute has been preserved or removed.

Example

Learning objective:

$$f_A^*, f_T^* = \arg\min_{f_A, f_T}[L_T(f_T(f_A(X)), Y_T) + \gamma \max_{f_B} J_B(f_B(f_A(X)), Y_B)]$$

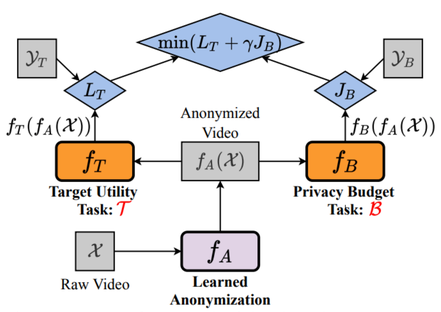
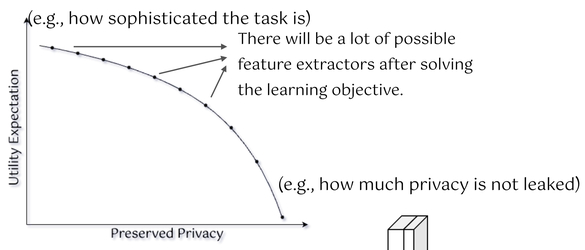


Fig: The optimization framework of the above learning objective

Trade-off between utility expectation and privacy leak risk

Different applications require different levels of features. For example:
- Anomaly detection may require only frame-wise differences or optical flow
- Detect an individual may require many raw appearance features

Fig: The trade-off between preserved privacy and utility expectation



Switching among multiple "differential access" level features

Rather than training multiple models, incorporating an in-situ learning framework can enable a seamless switch.

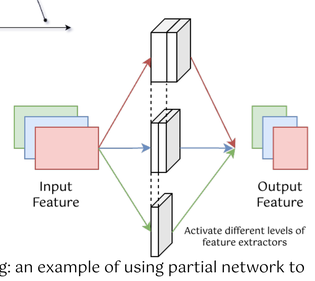


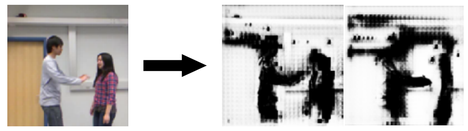
Fig: an example of using partial network to switch between multiple access level features

Expected Output

Annotated Videos

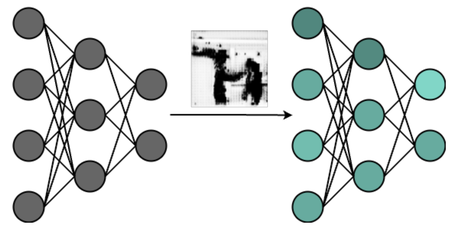
Gender: M & F
Faces: not visible
Colored Skin: Yes
Hair Colors: Black
Nudity: No
...

Learned de-identification transformation



Original image | Possible anonymized images

Trained machine learning models (challenge 3)



- The trained models still have high utility task performance (such as anomaly event detection, crowd counting, or flow estimation)
- The models learned reduced identification information (manifested by face verification/ recognition performance drop)

References

Wu et al., Privacy-Preserving Deep Action Recognition: An Adversarial Learning Framework and A New Dataset

Wang et al., Once-for-All Adversarial Training: In-Situ Tradeoff between Robustness and Accuracy for Free

The University of Texas at Austin
Bridging Barriers
Office of the Vice President for Research