Deep learning on Photo Privacy

Shanshan Yuan
Department of Computer Science

12/02/2024

Georgia State University

Outline

- Introduction
  - Background
  - Current Situation
  - Motivation
- Framework
- Implementation
- Evaluation
- Demo

Google Street View

- In May 2007, Google added its Street View feature to Google Maps, and it also provides panoramic views of streets gathered by webcams.

Leads to a series of problems !

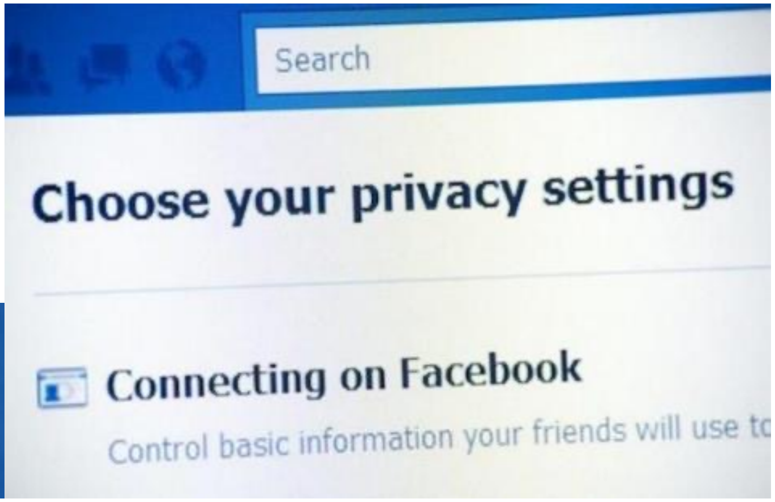
Background

- Photo sharing becomes extremely common with the development of technology and social media
- User can share their photos with each other on social networks by just one simply tapping or click.

Social Networking Platform Situation

- Some social networking websites also make their efforts to ensure the privacy of users’ photos.
- Facebook allows the user to specify specific viewing privileges for specific groups of people
- This action mostly relies on the users themselves to report privacy violation before any action is taken.

The websites do not analyze the photos before the photos are published and available to others!



Choose your privacy settings

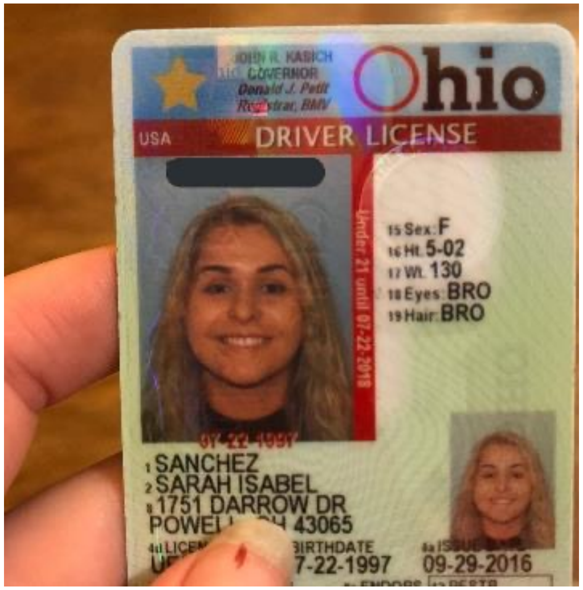
Connecting on Facebook

Control basic information your friends will use to

Social Networking Platform Situation (Cont.)

- Once the photos are posted on the social network to the public, it is nearly impossible to permanently delete the uploaded photos.
- Photos posted on these social networking websites release users' home location, contact information, family members, schedules and other sensitive information to the world before the users realize the privacy problem

This photo may be used by criminals for fraud activity or even assaulting this girl!



JOHN R. KASICH
GOVERNOR
Donald J. Petit
Registrar, BMV

Ohio

USA DRIVER LICENSE

[redacted]

Under 21 until 07-22-2018

15 Sex: F
16 Ht. 5-02
17 Wt. 130
18 Eyes: BRO
19 Hair: BRO

07-22-1997

1 SANCHEZ
2 SARAH ISABEL
8 1751 DARROW DR
POWEL H 43065

4d LICEN
UF

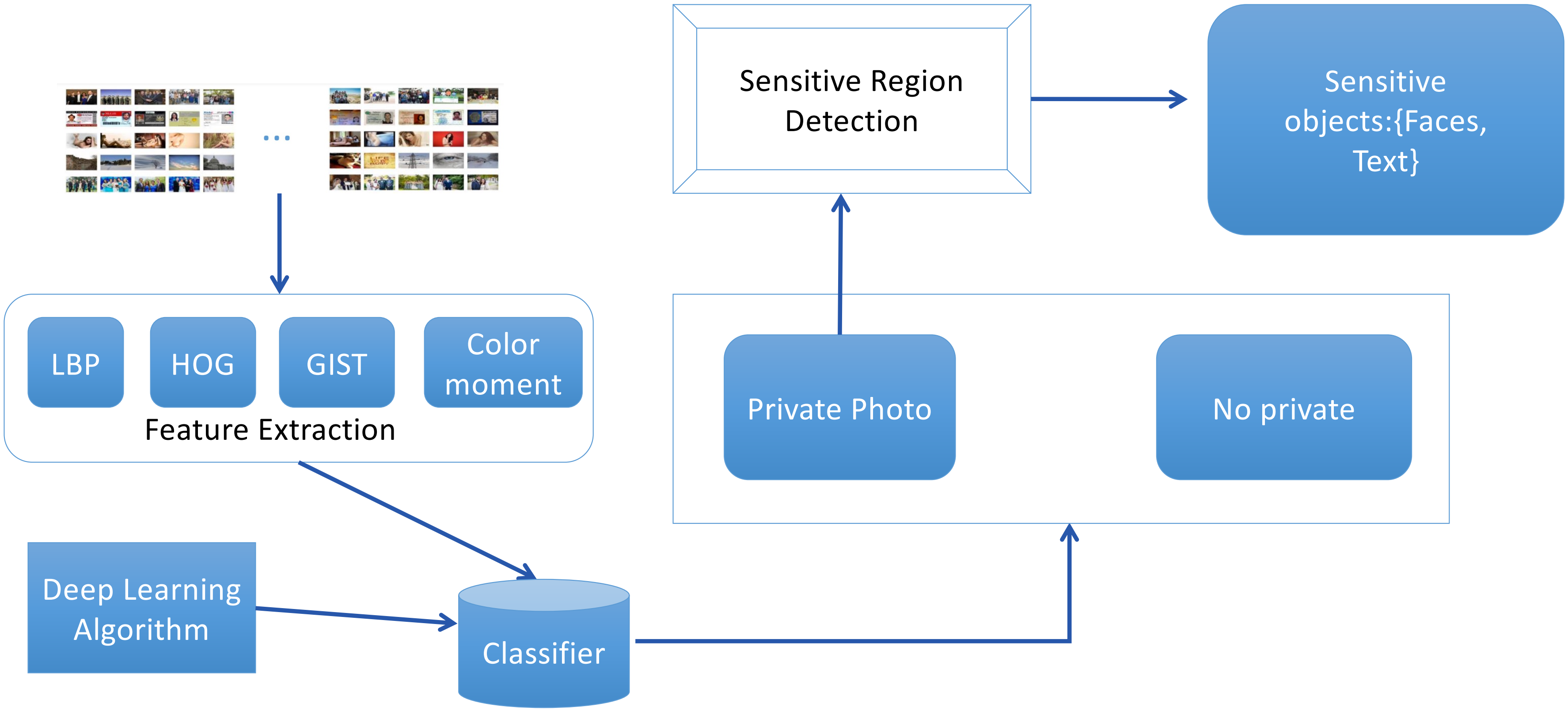
BIRTHDATE
7-22-1997

4a ISSUE DATE
09-29-2016

The photo downloaded directly from Twitter website

Motivation

- Build a model to automatically detect the privacy of a photo before it is published.
- When users want to share photos with others, they will be aware about the potential risks of information leakage.



...

LBP

HOG

GIST

Color moment

Feature Extraction

Deep Learning Algorithm

Classifier

Private Photo

No private

Sensitive Region Detection

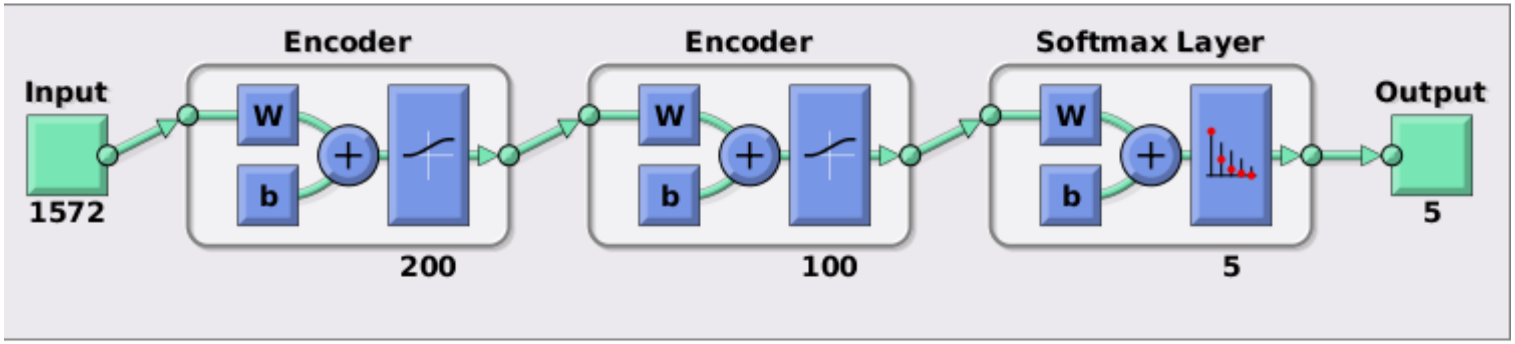
Sensitive objects:{Faces, Text}

Implementation--- Features extracted

- Local Binary Pattern(**LBP**) is a type of visual descriptor used for classification in computer vision. (http://www.ee.oulu.fi/mvg/page/lbp_matlab)
- The histogram of oriented gradients (**HOG**) is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image.(http://vision.ucsd.edu/~pdollar/toolbox/doc/)
- **GIST** a low dimensional representation of the scene, which does not require any form of segmentation(http://people.csail.mit.edu/torralba/code/spatialenvelope/)
- **Color moments** are measures that characterise color distribution in an image in the same way that central moments uniquely describe a probability distribution.(Implement by myself)

Implementation---Classifiers

- Classifier architecture



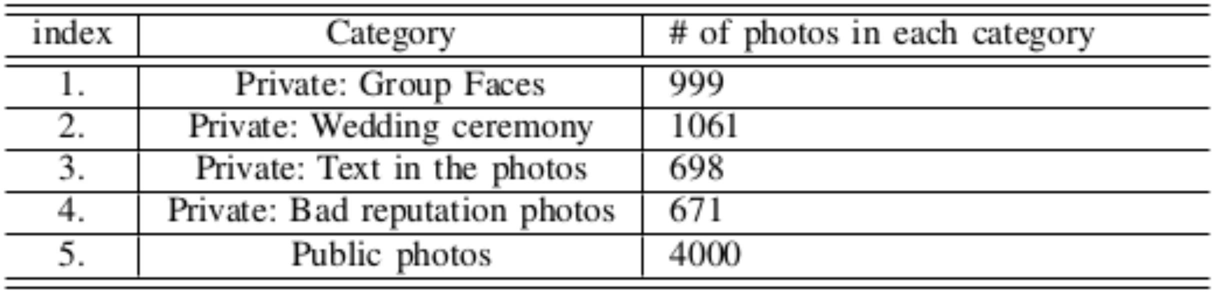
Implementation---Sensitive Objects Detection

- Face Detection --- Cascade object detector
  - Cascade object detector uses the Viola-Jones algorithm to detect people’s faces, noses, eyes, mouth, or upper body

- Text Detection --- OCR
  - Optical Character Recognition (OCR) is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo

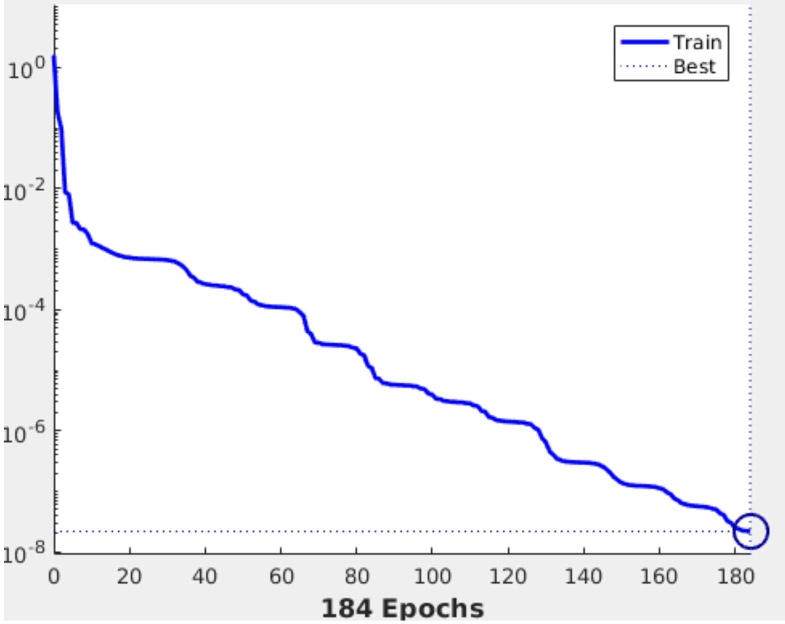
Evaluation---Dataset

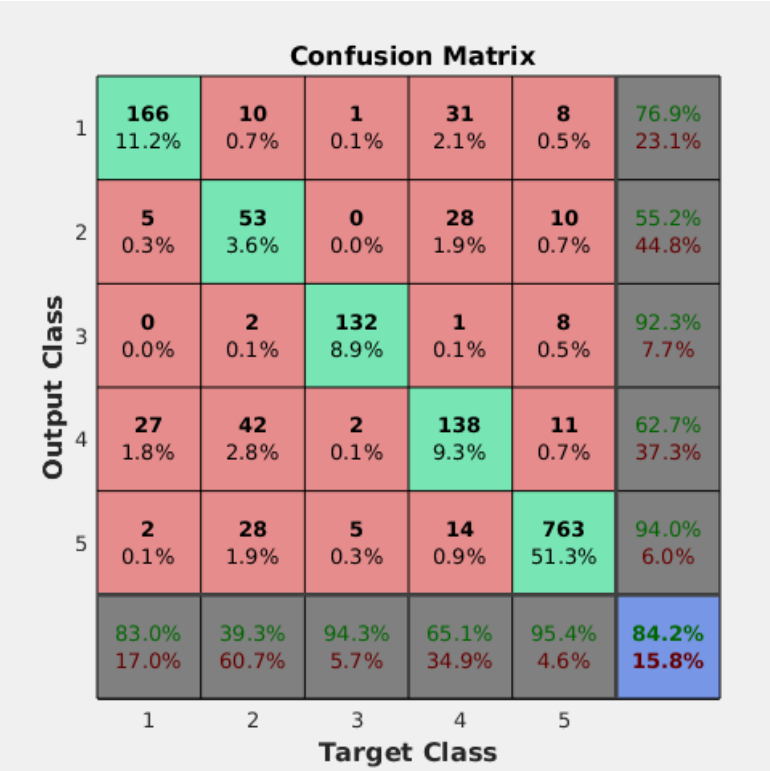
- We collected additional data to evaluate our algorithm from Flickr. Our dataset consists of about 8**000** photos of driver license, legal document, pornographic, and group/family portrait as private photos.



index	Category	# of photos in each category
1.	Private: Group Faces	999
2.	Private: Wedding ceremony	1061
3.	Private: Text in the photos	698
4.	Private: Bad reputation photos	671
5.	Public photos	4000

Evaluation – classification result





Confusion Matrix

Output Class	1	2	3	4	5	
1	166 11.2%	10 0.7%	1 0.1%	31 2.1%	8 0.5%	76.9% 23.1%
2	5 0.3%	53 3.6%	0 0.0%	28 1.9%	10 0.7%	55.2% 44.8%
3	0 0.0%	2 0.1%	132 8.9%	1 0.1%	8 0.5%	92.3% 7.7%
4	27 1.8%	42 2.8%	2 0.1%	138 9.3%	11 0.7%	62.7% 37.3%
5	2 0.1%	28 1.9%	5 0.3%	14 0.9%	763 51.3%	94.0% 6.0%
	83.0% 17.0%	39.3% 60.7%	94.3% 5.7%	65.1% 34.9%	95.4% 4.6%	84.2% 15.8%

Target Class

Performance

Label	Precision	Recall	Accuracy
Private Face	78%	77%	94%
Private Wedding	65%	64%	90%
Private Text in photo	90%	93%	98%
Private-bad reputation	60%	58%	92%
Public photo	94%	95%	94%

Thank you!