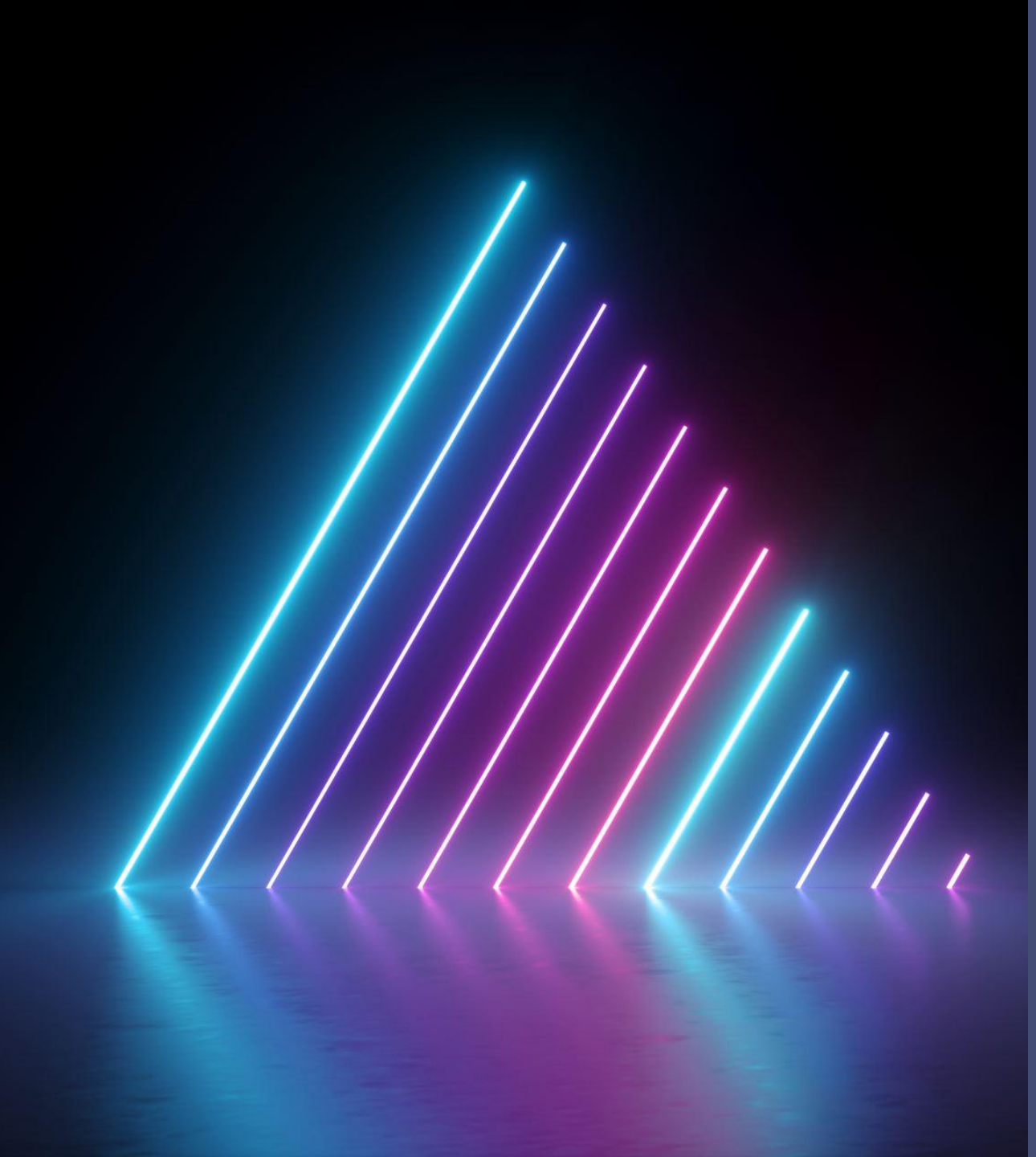


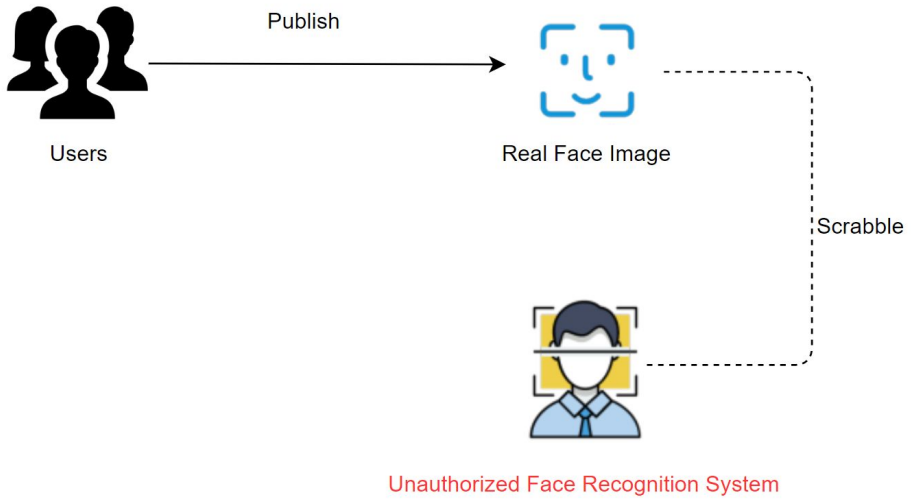
FACE ENCRYPTION

Now you don't.

Team TensorOverflow
Jiaxun Gao 300063462
Bowen Zeng 300115382
Xiang Li 300056427

MOTIVATION

- Privacy leakage on social media
- Personal identity stealing by 3rd parties

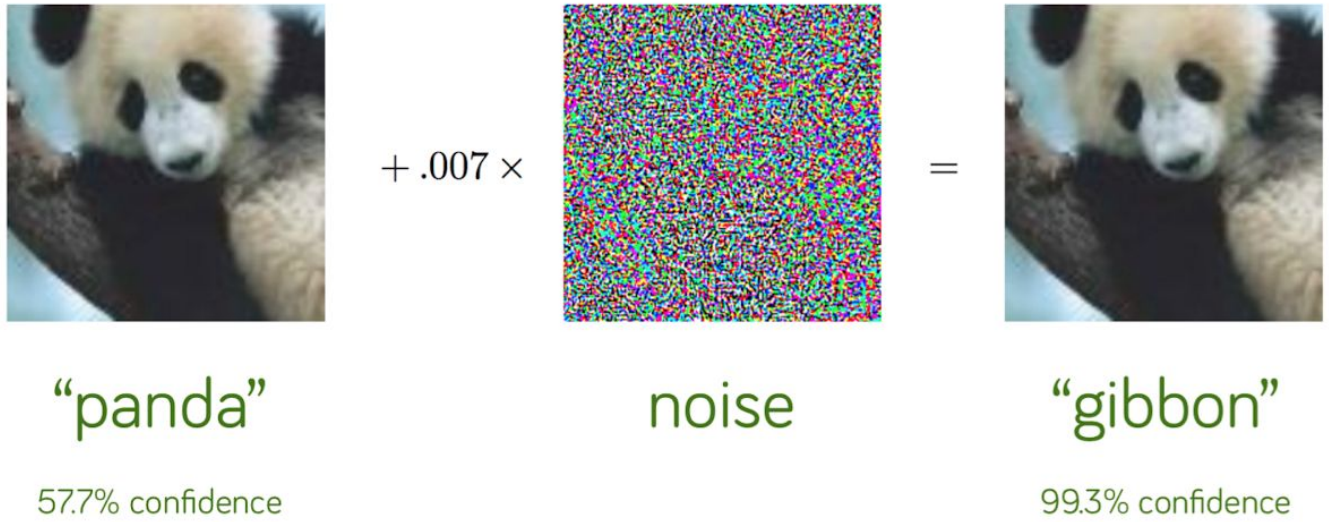


How can we protect our personal image from being abused?

A face encryption system that can:

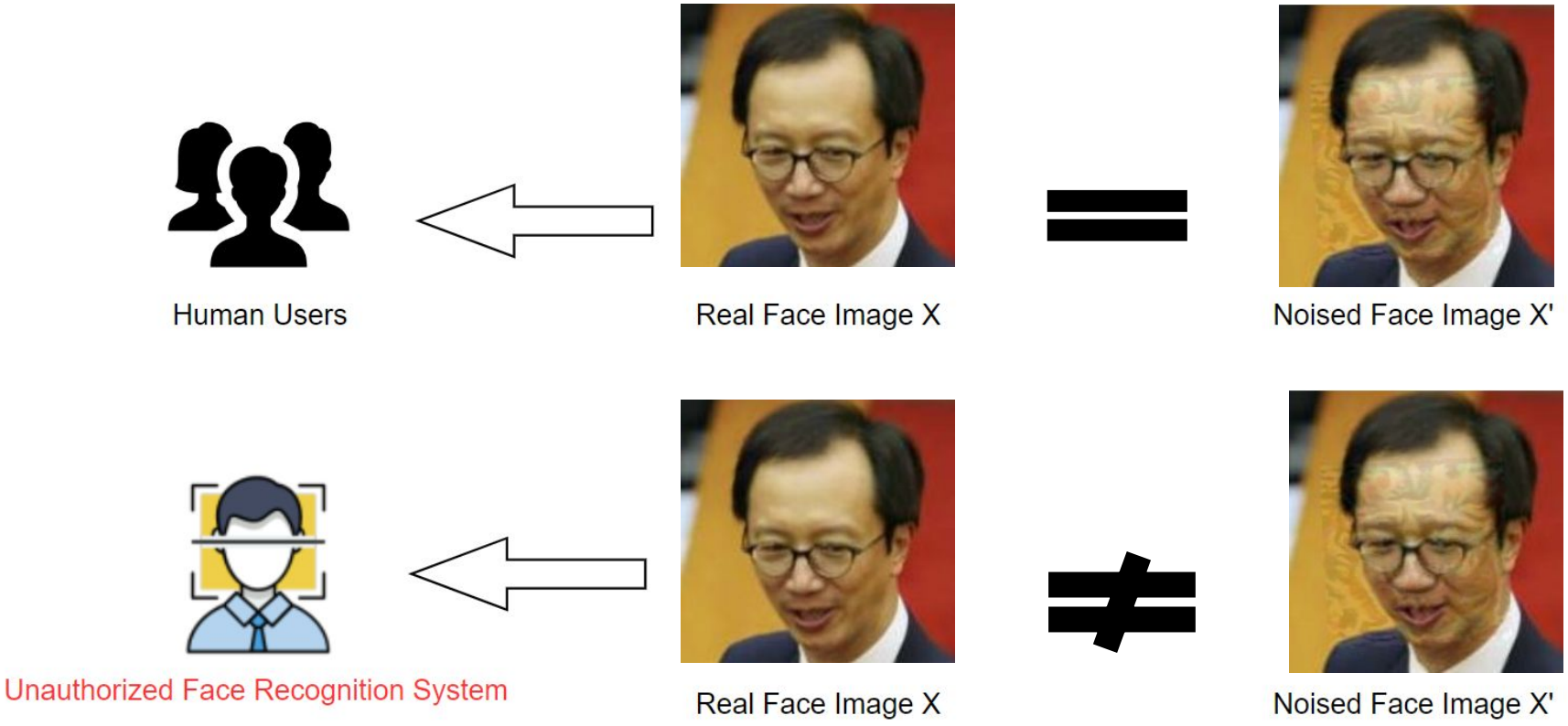
- Fool the SOTA facial recognition system to protect personal privacy
- Doesn’t harm UX
- Lightweight enough to be able to run on phone/laptop

ADVERSARIAL ATTACKING



adversarial example

Figure: the picture is taken from (Goodfellow et al).


Human Users
Real Face Image X
Noised Face Image X'
Unauthorized Face Recognition System
Real Face Image X
Noised Face Image X'

Methods to generate Adversarial Examples

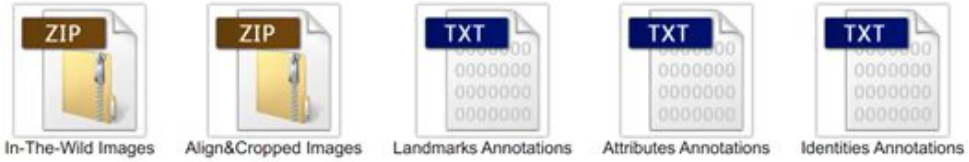
- Projected Gradient Descent (PGD)
- Boundary Attack

DATASET: CelebA

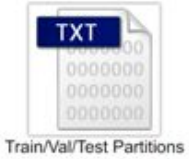
- 

dataset [1]
  - More than 200K high resolution celebrity images
  - More than 10K of identities

Downloads



Evaluation



DATASET: CelebA

- 

dataset [1]
  - Reflect real day scenario on social media
  - Widely used in facial recognition projects

DATASET: Custom

- Custom dataset (for testing)
  - Add another 10-20 personal identities
  - bzeng087@uottawa.ca
  - Title: [SDS3386] Custom dataset

All data will be deleted after the end of the project!

TASKS

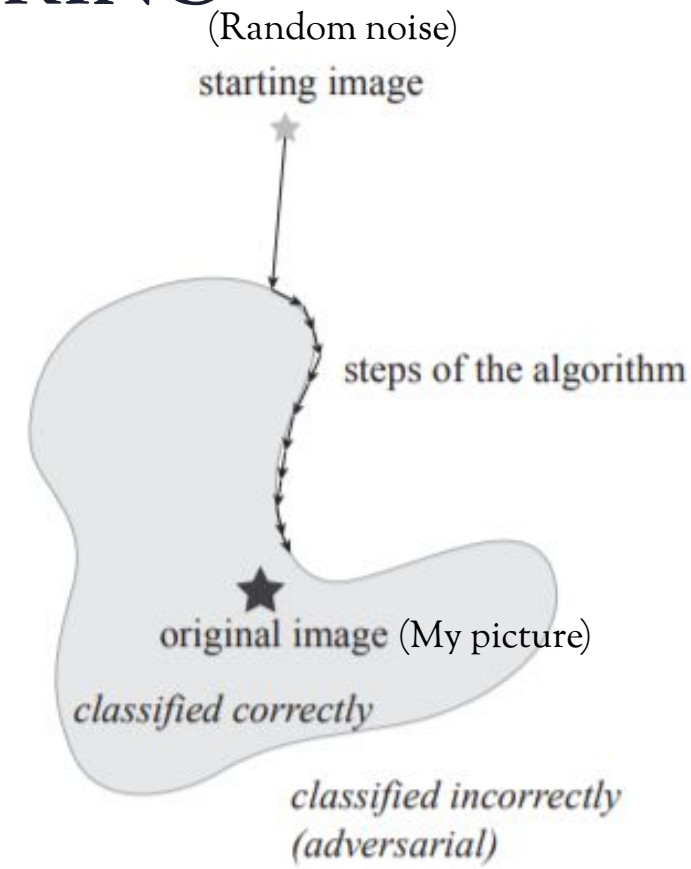
- Data cleaning & augmentation
- Implement a facial identity recognition system
  - Approach 1: Multiclass classification with Neural network (e.g. ResNet)
  - Approach 2: Compute embedding + clustering (e.g. KNN)
- Develop and compare various light-weight attacking strategies
  - Obfuscation-based method (baseline)
  - Adversarial method

METHODOLOGY

- Obfuscation-based method [2]
  - Blurring
  - Darkening
- Adversarial method
  - Whitebox
  - Blackbox [3]

ADVERSARIAL ATTACKING

- Whitebox attacking:
  - PGD attacking
  - Know the implementation details about the FR sys
- Blackbox attacking:
  - Boundary attack
  - Only query to the FR system is allowed



REFERENCE

[1] https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

[2] Slobodan Ribaric, Aladdin Ariyaeeinia, and Nikola Pavesic. De-identification for privacy protection in multimedia content: A survey. Signal Processing: Image Communication, 47:131–151, 2016.

[3] https://openaccess.thecvf.com/content/ICCV2021/papers/Yang_Towards_Face_Encryption_by_Generating_Adversarial_Identity_Masks_ICCV_2021_paper.pdf

[4] https://towardsdatascience.com/adversarial-eyeglasses-to-trick-facial-recognition-887c9f90930f

THANKS!