B04902103 資工四 蔡昀達

Reading critique #1

Paper : Sharif, Mahmood, et al. "Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition." *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016.

1. Summary

This paper aims to show that facial biometric systems using state-of-the-art DNNs are vulnerable to attacks in real world senarios. These facial systems are widely used for various sensitive purposes. Thus, attackers who mislead them can cause severe ramifications. This paper uses gradient descent for while-box and Particle Swarm for black-box as optimization to craft attacking pertubations on eyeglass. Their expermiments shows that most of their attempts succeed in fooling the FRS.

2. Strength of the paper

1. The experiments carefully consider physical realizability and inconspicuous to simulate real world attacking senarios.
2. It designes a complete process and expermiment to demonstrate the vulerbility in each perspective of the facial biometric systems.

3. Weakness of the paper

1. It is not suprising that FRS using DNNs can be attacked sucessfully since adversarial samples are already discovered in both white and black box early ago.
2. Attacking algorithms lack novelty. All experiments uses exisiting algorithms and are exactly the same as previous published adversarial attacking methods.
3. The novelty of this paper is that it simulates real world senarios. However, experiments cannot show that the methods for ensuring physical realizability (smoothness, printability, robustness) are useful or necessary.
4. The evaluation of the experiments is confusing.
    1. Very few samples are tested (20 out of 2622). It cannot show how effective the attack method really is.
    2. The evaluation images are from training data. It should not be.
5. Experiments except DNN_A are trained only with ten class and 40 images. It can only be considered as a toy example and is far from a commercial FRS.

4. My reflection

Adversarial attack happens to be my research area and thus most content in this paper is a known to me. This paper can be improved by using more advanced attacking methods such as GAN, RL, C&W etc (exist and published). The problem this paper trying to show is critical while more and more commercial services uses facial systems as authentication (Customs, transactions, mobile). In my opinion, since adversarial attack is still an unsoled problem, defence strategy and experiments for enhancing and controling physical realizability are more important and needs further investigation.