Recapture Detection to Fight Deep Identity Theft

Anis Trabelsi

Digital Security Department, EURECOM, France, anis.trabelsi@eurecom.fr

Marc Michel Pic

Digital Labs, SURYS, France, m.pic@surys.com

Jean-Luc Dugelay

Digital Security Department, EURECOM, France, jean-luc.dugelay@eurecom.fr

The progress made in deep learning has allowed the deployment of more powerful biometric authentication systems instead of traditional ones based on passwords or PIN codes. Facial recognition is widely used on smartphones to grant user access. However, advances in deep learning also improve methods for doctoring images and videos. A fraudulent user can use these methods to steal the identity of another person. It is very easy for impostors to present to the smartphone an image or video of the victim's face displayed on another screen. In this paper, we describe the security risks when a facial recognition system is attacked by presenting an image, a video or an interactive deepfake displayed on a screen. We also present a deep learning-based method to detect this kind of attack.

CCS CONCEPTS • Computing methodologies → Artificial intelligence • Computer vision • Computer vision tasks • Scene anomaly detection

Additional Keywords and Phrases: recaptured image detection, digital image forensics, e-KYC, face anti-spoofing, identity theft.

ACM Reference Format:

Anis Trabelsi, Marc Michel Pic, and Jean-Luc Dugelay. 2022. Recapture Detection to Fight Deep Identity Theft. In 2022 4th International Conference on Video, Signal and Image Processing (VSIP '22), November 25–27, 2022, Shanghai, China, 9 pages.

1 INTRODUCTION

The identity of people on the Internet is becoming a major security issue. Since the Bale agreements, banking institutions have integrated the verification of the identity of persons or Know-your-Customer (KYC) into their registration process. With the dematerialization of banks, this procedure has become the e-KYC or remote KYC which operates remotely through the user's smartphone. Likewise, remote identity verification has become the standard for enrollment in electronic signature tools. New regulations are emerging to secure this approach, for example, in France, the PVID framework which regulates the remote acquisition of identity documents and of

people's face within the framework of eIDAS. This is necessary because a growing trend of new digital crime is emerging: The Deep Identity Theft.

Thanks to new deep learning tools, impostors can change their appearance to look like someone else in real time. Imposters can then accomplish all the common actions required during a remote enrolment without being detected by identity verification algorithms. Today, public application like Zao [1] or tools for a more limited audience [2] allow impostors to easily transform their appearance in real time.

There are even methods for stealing an identity with a single image of the victim's face [3]. Thanks to all these new methods for tampering images and videos, many authenticity verification algorithms included in facial recognition systems are no longer effective. When an impostor wants to steal an identity by tampering an image or a video of the face of someone else, there are two solutions.

The first solution consists in performing the falsification in the digital domain by replacing the standard stream video by a falsification (injection). The second approach consists in presenting to the camera of the facial recognition system a doctored image/video displayed on a high-resolution screen. This type of attack is known as "Presentation Attack" [4].

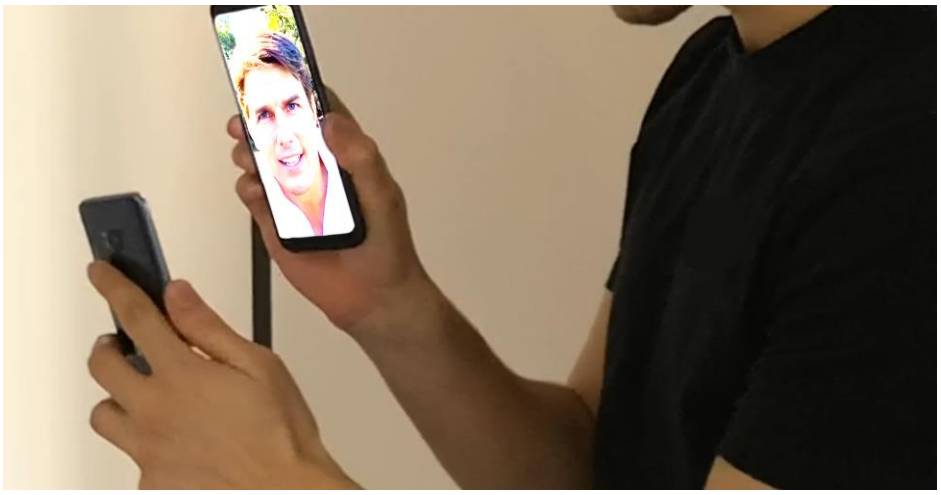

Figure 1: Presentation attack with a screen. One of the smartphones displays a face image to attack a remote KYC application running on the second smartphone.

Biometric authentication systems, in particular facial recognition systems, are highly vulnerable to presentation attacks. In the past, this type of attack was widely employed by falsifiers. They used a printed image of a face or 3D physical fake face by wearing masks. These could be simple cardboard masks and sometimes they could use quite expensive and very realistic latex or silicone masks. To prevent facial recognition systems from being fooled by presentation attacks, methods have been developed and introduced into these systems. Such methods are called liveness tests. A liveness test is an algorithm that asks the user to complete some actions in order to check that the user is a real person and not a printed image or wearing masks. The basic liveliness tests consist of blinking, smiling or rotating the head to the side.

Today, the most common scenario for a counterfeiter to attack a facial recognition system using a presentation attack is to use a smartphone (Figure 1). Smartphones are very widespread and are easier to manipulate than other types of screens such as computer or television screens. It is therefore necessary to take this type of attack into consideration and to propose methods to detect the presentation of replayed video on smartphone screens. By looking at the camera stream, we can analyze the nature of the image in detail to detect this type of attack. In this article we focus on presentation attacks that use a screen displaying an image/video. We currently consider this type of attack as the first source of Deep Identity Theft.

2 PRIOR WORK

Since 2014 and the work of Lu [5] and Taigman [6], the efficiency of facial recognition systems has exceeded human performance thanks to deep learning-based algorithms. Other more recent works have also led to important advances [7, 8, 9]. At the same time, social networks have become very popular. On social networks, people share photos or videos of their face (selfies) almost every day. As a result, it is very simple for an impostor to obtain an image/video of the person he wants to steal the identity in order to fool a facial recognition system.

These attacks represent a significant threat to facial recognition and are of great interest to researchers. Facial authentication systems are used in a variety of applications. These may include security applications such as airplane onboarding, access to bank accounts, or authentication to gain access to a building [10]. Indeed, the detection of counterfeit images and videos is an important research topic known as Digital Image Forensic. Many methods have been proposed to detect attacks on a facial recognition system. However, two technological advances are changing the current state of the art.

The first one concerns the technology integrated into smartphones. The performance and quality of smartphones have evolved significantly over the years. This is especially the case for the screen and camera technologies that are used in smartphones. Smartphone cameras are very powerful and manage to capture real-world scenes in high detail. Today, new smartphones even have multiple cameras. For instance, the latest smartphone, presented in September 2022 by Apple, will embed three image sensors (in its top-of-the-range version). In addition, the screens of smartphones have an increasingly detailed display quality. It is usual for a smartphone to have a resolution between 2560 x 1440 pixels and 3840 x 2160 pixels. Today, the two technologies most often used on smartphone screens are LCD (Liquid Crystal Display) and OLED (Organic Light-Emitting Diode). OLED is the most recent technology and provide a better level of brightness, a better level of contrast and better viewing angles than LCDs [11].

The second one is the technology for tampering images and videos. As said in the introduction, powerful and easy-to-use tools are available to digitally manipulate a face in an image or a video. At the end of 2017, a new method to forge a face in a video was released on the internet. This method is based on deep learning and the result of the falsification is very realistic. A video produced using this method is known as a deepfake.

Deepfakes technology can be used to swap one person's face with another's or to manipulate a person's facial expressions and movements in real time.

It is due to these two technological advances that an imposter will be most likely to use a screen to attack a facial recognition system. The attacker just has to use a second smartphone for the face transformation, display the result of its high-resolution screen and present it to the camera of the first smartphone where the remote KYC application is running (Figure 2). A recaptured image is an image that has been captured by a first photo

sensor, then displayed on a surface (paper print or digital screen) and finally captured by another photo sensor. It is very difficult to visually distinguish the recaptured image from the natural image for the human eye [12].

Therefore, a lot of work has been done in the literature to automatically detect recaptured images, especially images displayed from a screen. However, to the best of our knowledge, all this work has focused on detecting recaptured images from an LCD screen. According to a study by IHS Market, smartphone manufacturers will equip more than 50% of the displays of their devices with OLED technology by 2023. Images recaptured from an LCD screen can leave traces (aliasing, blurriness, noise). These traces are much less present on recaptured images from an OLED display (Figure 3). Many methods have emerged to detect recaptured images based on the traces left during the recapture process. As in many other fields of digital image forensics, new approaches based on deep learning have also appeared in recent years to detect recaptured images from screens.

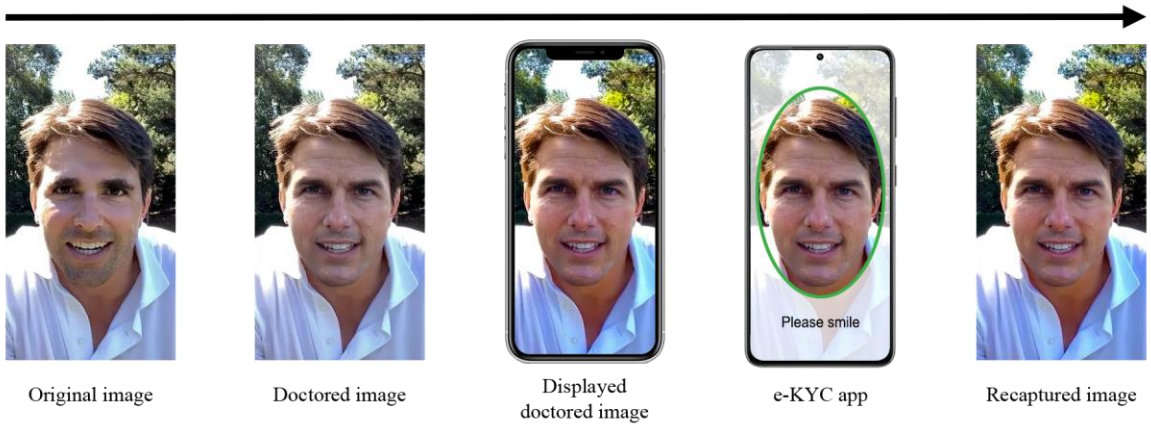

Figure 2: The usual process followed by impostors to attack a facial recognition system by presenting a doctored image displayed on a screen. Original and doctored image are from vfxchrisume.

2.1 Traditional recaptured images detection methods

Traditional techniques to detect recaptured images from LCD screens have been relying on a multitude of signal processing methods. Recaptured images from LCD screens often exhibit moiré patterns. A moiré pattern is an example of aliasing due to the overlapping of the digital grids of the sensor of the device which leads to high-frequency noise in the image. Muammar et al. proposed in [13] an investigation about aliasing and moiré patterns from recaptured images from LCD screens. They proposed an anti-forensic method that can detect recaptured images by analysis the presence of moiré patterns.

A similar idea has been proposed by Mahdian et al. [14]. The authors detect aliasing based on the presence of cyclostationary patterns. A cyclostationary pattern is a pattern that has a property of periodicity. This type of pattern is also present in images recaptured from an LCD screen. After a pre-processing step to enhance the image and obtain stronger features, the image is converted to the spectral domain. An image is considered recaptured if a strong correlation is found between the enhanced version and its spectral domain version.

The authors in [15] proposed to detect recaptured images using multiple features. The features they used include sensor pattern noise feature, texture feature and color information. In order to classify recaptured images, they trained a binary classification algorithm by using support vector machine (SVM).

Figure 3: Comparison between a recaptured image from an LCD screen (left) with a recaptured image from an OLED screen (right). We can see patterns in the sky on the image recaptured from an LCD screen.

2.2 Deep learning-based recaptured images detection methods

In [16], the authors propose a Laplacian Convolutional Neural Network (L-CNN) to detect recaptured images. The feature learning part (first layers) has integrated a Laplacian Filter in order to improve the noise signal ratio introduced by the recaptured process. What is challenging in this paper is that the recaptured images are small. The images have a size of 64x64 pixels.

Several other relevant articles describe deep learning methods to detect recaptured images from LCD screens as in [17] where the authors proposed to combine Convolutional Neural Networks (CNNs) with a Recurrent Neural Network (RNN). They use a preprocessing step composed of convolutional operation. After the training, the features extracted from the trained CNN model were fed into a recurrent neural network to classify the images.

In [18], Abraham proposed to detect moiré patterns using a convolutional neural network. After using Haar Wavelet Decomposition on the input image, he passes the three resulting bands (low-low, low-high and high-low) as input for the network. Then he uses the low-low band as the weight parameter for low-high and high-low band during training to estimate the spread of the moiré patterns in the image.

2.3 Existing datasets

Concerning, the existing datasets, there are four major databases of recaptured images from a screen. In [19], the authors present a recaptured image dataset acquired by only smartphone cameras called Astar database. Agarwal et al. proposed in [20], a large dataset of recaptured images composed of 14500 samples. They used various devices to both display and capture images. In NTU-ROSE [12] dataset, the recaptured images have been acquired only on an LCD screen and they vary the ambient light conditions. For ICL database [21], a total of twelve cameras were used.

These datasets are very useful for development of recaptured image detection methods. However, we note that among all these datasets, there are only LCD screens. None of the databases include OLED screens. As

said, it is very important to consider also OLED technology. Most of the state-of-the-art methods are based on the analysis of visual traces present in recaptured images from an LCD screen. As these artefacts are not often visible on images recaptured from OLED screens, these methods may no longer work on this type of display.

3 PROPOSED DETECTOR AND DATASET

3.1 Proposed dataset

In order to evaluate the performance of our detector, it is necessary to build a dataset of recaptured images from OLED screens. As previously mentioned, existing datasets only include recaptured images from LCD screens. This dataset will be used to train and test different neural networks. It is important that the recaptured image dataset does not include training bias. During the recapture process, we have therefore varied different parameters to avoid any training bias. These parameters are:

- The distance between the camera and the screen;
- The viewing angle between the camera and the screen;
- The content of the images displayed;
- The brightness levels of the screen;
- The model of camera that collects the images.

One of the most important parameters is the content of the image. The recaptured images must be sufficiently varied. The choice of the natural image database (i.e. the images used for the recapture) is therefore crucial. We have decided to use the MS-COCO [22] image dataset. The MS-COCO database is a large image dataset of real-world objects. It is a dataset that is adapted to our constraint because the content of the images is very varied (Figure 4).

Figure 4: Examples of images from our dataset. First row: The original images from MS-COCO, second row: The corresponding recaptured images.

The original images from the MS-COCO dataset were displayed on an OLED screen integrated in a 15-inch Gigabyte AERO laptop. The images are then captured with the front cameras of two different smartphones: a Samsung Galaxy S6 (5 megapixels) and a Samsung Galaxy S8 (8 megapixels). To remove the screen's edges

from the recaptured images, we have cropped all the images. The dataset is composed of 25,000 images (12,500 original images and 12,500 recaptured images).

3.2 Detector training

To be able to automatically detect a recaptured image, we decided to develop an approach based on deep learning by training a Convolutional Neural Network (CNN). A CNN is a neural network well suited for working with images, and this architecture has shown excellent results in image classification tasks. In theory, neural networks are able to find a function to classify two classes if a significant difference exists. For our problem it is appropriate because this distinction exists between an original image and a recaptured image.

We have chosen to use the architecture called EfficientNet [23]. We trained the EfficientNet-B0 variant on the recaptured image dataset we have built. We achieve the accuracy of 0.916 after a 50-epoch training and with a learning rate of 10-3. The sensitivity of the model is 0.860, and the specificity of the model is 0.9703.

3.3 Detector evaluation

To be able to evaluate our solution, we have built several different test sets. We need to assess if our model can detect recaptured image contents that are not present in the dataset. We, therefore, build a facial recaptured image test set by displaying images from the Face Recognition Grand Challenge (FRGC) dataset.

It is also necessary to test our model on recaptured images on screens that are not present in the database. To do this, we recaptured new images on two other OLED screens from a Samsung Galaxy S6 and a Samsung Galaxy S8. We also tested our detector on images recaptured from a Samsung S24D390H LCD monitor. Results are detailed in Table 1.

Table 1: Results of the detector applied to various test sets

Testing set	Number of images	Precision	Accuracy	F1-score
FRGC (OLED)	1000	0.996	0.948	0.950
MS-COCO (OLED)	1000	0.956	0.919	0.921
MS-COCO (LCD)	1000	0.604	0.743	0.701

We evaluated our method on the FRGC test set to compute the false acceptance rate (FAR) and the false rejection rate (FRR). The FAR is 0.14 and the FRR is 0.018. The results show that our method reduces the success rate of attacks by 86%. Moreover, without any recaptured images from LCD screens in the training dataset, our model is able to detect recaptured images with an accuracy of 74%. Our technology is therefore not dependent on the type of screens of the recaptured images.

3.4 Comparison with a state-of-the-art method

In this section, we will compare our results with a method that achieves a good accuracy on recaptured images from LCD screens. We will compare our method with the technique describes in [18]. We chose this method because it used both moiré pattern as features and convolutional network as classification tasks. We can, therefore verify our hypothesis by applying this method to images recaptured from an OLED screen. For this experiment, we implemented the method provided by the author. When the method is applied on a dataset composed of natural images and recaptured images from an LCD screen, the accuracy is about 95%. But, when

we use our dataset with OLED recaptured images, the method is not able to detect more than one third of the recaptured images.

4 CONCLUSION

The development of deepfake technologies in real time generates a significant risk for all remote-KYC processes and more. One of the keys to avoiding their too rapid diffusion is to detect recaptures from screens and in particular from smartphone screens. However, the algorithms previously designed were largely based on interference phenomena such as moiré, which are easier to detect at low resolutions. Screen technologies have also evolved in their nature and new detectors must take this into account and therefore, these algorithms are no longer relevant.

We redesigned a new algorithm able to take into account the new screens. We have compared them with previous algorithms on both old screens and new screens and shown a clear gap of performance on the detection of recaptured from OLED screens. But even better, as illustrated in this article, our algorithm is also able to adapt between generations of screens and we expect it will work on future screen technologies.

Video content will take a growing place in the future years in the share of trust between peoples and between companies. Detecting recapture, which is a clue of tampering, will be a major requirement to increase the level of confidence for our future society of remotely connected people.

REFERENCES

[1] Changsha, *ZAO*. 2019.

[2] I. Perov *et al.*, « DeepFaceLab: A simple, flexible and extensible face swapping framework », *ArXiv200505535 Cs Eess*, 2020.

[3] A. Siarohin *et al.*, « First Order Motion Model for Image Animation », 2019.

[4] J. Galbally *et al.*, « Biometric Antispoofing Methods: A Survey in Face Recognition », *IEEE Access*, vol. 2, 2014.

[5] C. Lu and X. Tang, « Surpassing Human-Level Face Verification Performance on LFW with GaussianFace », *ArXiv14043840*, 2014.

[6] Y. Taigman *et al.*, « DeepFace: Closing the Gap to Human-Level Performance in Face Verification », in *Conference on Computer Vision and Pattern Recognition*, USA, 2014.

[7] G.G. Patil and R.K. Banyal, « A Dynamic Unconstrained Feature Matching Algorithm for Face Recognition », in *Journal of Advances in Information Technology,* 2020.

[8] J.A. Moreano and N.L. Palomino, « Global Facial Recognition Using Gabor Wavelet, Support Vector Machines and 3D Face Models », in *Journal of Advances in Information Technology*, 2020.

[9] P. Easom-Mccaldin *et al.*, « Towards Building A Facial Identification System Using Quantum Machine Learning Techniques », 2022.

[10] Chandra *et al*., « Application of "Face Recognition" Technology for Class Room Electronic Attendance Management System », in *International Conference on ICT for Smart Society*, 2020.

[11] Z. Luo and S.-T. Wu, « OLED Versus LCD: Who Wins? », 2015.

[12] H. Cao and A. C. Kot, « Identification of recaptured photographs on LCD screens », in *International Conference on Acoustics, Speech and Signal Processing*, USA, 2010.

[13] H. Muammar and P. L. Dragotti, « An investigation into aliasing in images recaptured from an LCD monitor using a digital camera », in *International Conference on Acoustics, Speech and Signal Processing*, Canada, 2013.

[14] B. M. Adam Novoz, « Detecting Cyclostationarity in Re-Captured LCD Screens », *J. Forensic Res.*, vol. 06, n° 04, 2015.

[15] Y. Ke *et al.*, « Image Recapture Detection Using Multiple Features », *Int. J. Multimed. Ubiquitous Eng.*, vol. 8, n° 5, 2013.

[16] P. Yang *et al.*, « Recapture Image Forensics Based on Laplacian Convolutional Neural Networks », in *IWDW*, 2017.

[17] H. Li *et al.*, « Image Recapture Detection with Convolutional and Recurrent Neural Networks », *Electron. Imaging*, vol. 2017, n° 7, 2017.

[18] E. Abraham, « Moiré Pattern Detection using Wavelet Decomposition and Convolutional Neural Network », in *Symposium Series on Computational Intelligence (SSCI)*, India, 2018.

[19] X. Gao *et al.*, « A Smart Phone Image Database for Single Image Recapture Detection », in *Digital Watermarking*, vol. 6526, 2011.

[20] S. Agarwal *et al.*, « A Diverse Large-Scale Dataset for Evaluating Rebroadcast Attacks », in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, 2018.

[21] T. Thongkamwitoon *et al.*, « An Image Recapture Detection Algorithm Based on Learning Dictionaries of Edge Profiles », *IEEE Trans. Inf. Forensics Secur.*, vol. 10, n° 5, 2015.

[22] T.-Y. Lin *et al.*, « Microsoft COCO: Common Objects in Context », *ArXiv14050312 Cs*, 2015.

[23] M. Tan et Q. V. Le, « EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks », *ArXiv190511946 Cs Stat*, 2020.