Chapter 19
Face Manipulation Detection in Remote Operational Systems

Marc Michel Pic, Gaël Mahfoudi, Anis Trabelsi, and Jean-Luc Dugelay

Abstract In this chapter, we present the various categories of Face Manipulation and their use within different remote operational systems. We then use the example of remote identity document onboarding systems to illustrate how each category can be used in practice to compromise such a system. After a definition of the different Face Manipulation categories and the common algorithms used to produce them, we go through the various manipulation detection algorithms and common image and video forgery datasets. We then introduce some known counter-forensics methods that can be used by an attacker to avoid detection. Knowing the detection methods and the counter-forensic, we present how we can build up a safer system by using the correct methods at the correct time. But also how knowledge about the tampering process could be used to design the user experience to make the systems harder to compromise. We complete this review by the standardisation effort and legal aspect on the matter. And we conclude by discussing the remaining challenges and perspectives for better use of nowadays detection methods in practical usage.

19.1 Introduction

The worldwide crisis of 2020 due to the COVID-19 pandemic changed our day-to-day interaction significantly. It confirms the global trend of generalising the use of remote operations, and we believe that this trend will continue in the coming years.

M. M. Pic (✉) · G. Mahfoudi · A. Trabelsi
SURYS, Bussy-Saint-Georges, France
e-mail: m.pic@surys.com

G. Mahfoudi
e-mail: g.mahfoudi@surys.com

A. Trabelsi
e-mail: a.trabelsi@surys.com

J.-L. Dugelay
EURECOM, Biot, France
e-mail: jean-luc.dugelay@eurecom.fr

C. Rathgeb et al. (eds.), *Handbook of Digital Face Manipulation and Detection*,
Advances in Computer Vision and Pattern Recognition,
https://doi.org/10.1007/978-3-030-87664-7_19

Many remote technologies are heavily based on facial recognition, but also on the more general behaviour and context analysis such as liveness challenges or even verifying if a person is wearing a medical face mask. For those operational systems, the ability to detect Face Manipulation is essential. We can classify those systems into three main categories and many subcategories.

The first category is the systems with a direct Face Recognition need such as Automated identity authentication like Automated Border Controls (ABCs) at Airports and Remote face authentication systems. The second category is systems using Indirect Face Recognition. Those would be used for tracking individuals across one or many acquisition devices or detecting an individual in a specific area. The last category would be the Face Behaviour Analysis Systems. Those aim at verifying that a specific attended action/behaviour is performed by an identified person andetecting an unexpected action/behaviour linked to a specific person or to detect actions/behaviours in the context of a group of people.

In this chapter, we will illustrate the common kind of face manipulation and the ways of securing an operational system against those. For the sake of clarity, we will look at those attacks within the context of a remote identity document and person acquisition scenario. Even though this does not fully embrace the many aspects of the different remote operational systems, it will allow us to give a practical example of all types of forgeries and means to secure such applications.

We will start by introducing a typical remote identity document onboarding system and explain which part is most likely to be attacked. We will discuss about the different types of Facial Manipulation attacks and how they fit into our particular system. We will then give a definition of each attack and present the common technologies and methods to create those forgeries. After, we will present common face manipulation detection methods and more general image manipulation detection algorithms. We will also introduce datasets used to study those attacks and to train the detection algorithms. Then, we will discuss some typical counter-forensic methods and how one can design his/her system to reduce the chances of forgeries. Finally, we will conclude with a discussion of the remaining challenges and perspectives for better use of nowadays detection methods in practical usage.

19.2 Remote Identity Document Onboarding

For the rest of this article, we will place ourselves within the framework of a generic remote identity document onboarding system. A brief overview of such a system is given in Fig. 19.1. We can see that such systems are made of two main steps. First, the user is asked to take a picture/video of his/her ID document. Then, he is asked to take a picture/video of himself/herself. The challenges for the system are then multiple. It must first authenticate the ID documents. Then, it must verify that the user is the owner of the document. Once all the verification steps are passed, the systems store user information such as name and age. But also a picture of the user that will later be used to authenticate him/her again when needed.

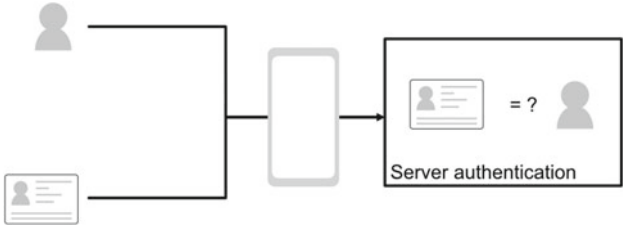

Fig. 19.1 Generic Remote identity document onboarding system

Table 19.1 Attacks and associated scenario

	Non-biometric	Face swapping	Face morphing	De-identification
Live attack	X	X		X
Portrait attack		X	X	X

In such a scenario, an attacker could have three main **strategies**. He can try to perform some kind of identity theft. Helped by someone else, they can create a common biometry so the two people can later on share their identity. Or lastly, he could try to create and use a completely fake identity for privacy concern or other.

Whatever his goal, he can only attack one or both parts of the system, i.e. the portrait during the document acquisition or his biometry during his self-portrait acquisition. We will refer to this as **support**. We will call an attack during the self-portrait acquisition a Living Person Attack and an attack during the document acquisition a Portrait Attack.

Then, depending on his objective, he will apply one or many of the four main **categories** of Face Manipulation. The Non-Biometric Face manipulation, Face Swapping, Face Morphing or Face De-Identification.

In the case of remote identity document onboarding, Non-Biometric Face manipulation would typically be used to fool liveness challenges. Face Swapping would serve in case of identity theft and might be used either for a Living Person Attack or a Portrait Attack. Face Morphing would be employed to create a shared biometry, which is typically used during a Portrait Attack. And finally, De-identification would be used to create a synthetic identity for both a Living Person Attack and a Portrait Attack. The manipulation and their associated supports are summed up in Table 19.1.

In the next section, we will give a more precise definition of each attack and the common algorithms used to perform them.

19.3 Face Manipulation Algorithms

Here, we will first give a description of each category of attacks and give general uses cases for each. Even though those attacks are conceptually different, nevertheless they all target the face area, they are inherently based on the same tampering algorithm. We will give a brief overview of the best-known tampering methods.

19.3.1 Categories of Attacks

Non-biometric manipulation

As stated, we observed an increasing used in systems such as the remote identity document onboarding. Those systems imply some face-related controls (e.g. face recognition behaviour). We define the non-biometric manipulations as any manipulation of the face that does not alter biometric traits.

The first application of such manipulations takes place during liveness detection. Liveness detection is often defined as the verification that the person in front of the camera is indeed alive and interacting voluntarily. We wish to detect attempts of fooling the systems with attacks such as photo presentation, screen presentation and mask presentation. But also that the person is not forced to perform actions by someone. Typical liveness challenges include eye-blinking, smile, head movements, etc. Recent examples have shown the importance of such detection. It is possible to create synthetic eye-blinking digitally without having to alter any biometric traits of an individual.

When a proof of action or inaction is needed, such manipulations can also be involved. For example, within the context of the COVID-19 crisis, verifying that a taxi driver wore a medical mask was necessary to allow him to drive a customer. Such verification is not at all related to the biometry, but can suffer from a non-biometric manipulation. When managing a large fleet, one might be the subject of attacks.

Those manipulations may not be as severe as identity theft or other. Though it is important to acknowledge those as they are easy to achieve and can lead to more problematic issues.

Face Swapping

Face swapping is a well-known technique that consists of the replacement of someone's face in an image or a video. There exist two main kinds of face swapping. It can either be applied on a portrait or to a live acquisition of a person [1].

The first case is what is usually called Face swapping and is typically used to perform an identity theft. Applying it to a portrait does not reduce its usage to images only. On Fig. 19.2, an example of a real-time face swapping is given for id document

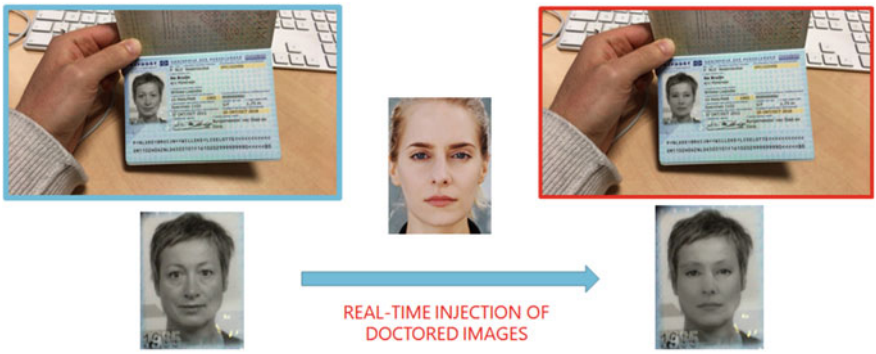

Fig. 19.2 Video Replacement of the portrait picture thanks to inverse fit swapping

Fig. 19.3 An example of deepfake by face reenactment. From left to right: target actor, pilot actor, reenacted actor

portrait within a video stream. Face swapping is often realised with classical methods [2] but can also use some more advanced deep learning techniques [3]. More details will be given in Sect. 19.3.2.

When applied on a live acquisition of a person, Face swapping is often referred to as Face Reenactment. The idea behind face reenactment is to animate a target face according to a video of a source actor or with a given set of expressions. In a sense, it can be used to perform non-biometric manipulation. Here, we are interested in the specific case where an attacker would reenact the face of someone else.

Face swapping on a still portrait or using Face reenactment is mostly used to perform identity theft. It is a very versatile manipulation as it can be applied at different stages of an operational system (Fig. 19.3).

Face Morphing

As described earlier, one can decide to perform a complete face replacement which is commonly called a Face Swap. It is important to consider a face swapping as a subset of a more general attack often called a face morphing.

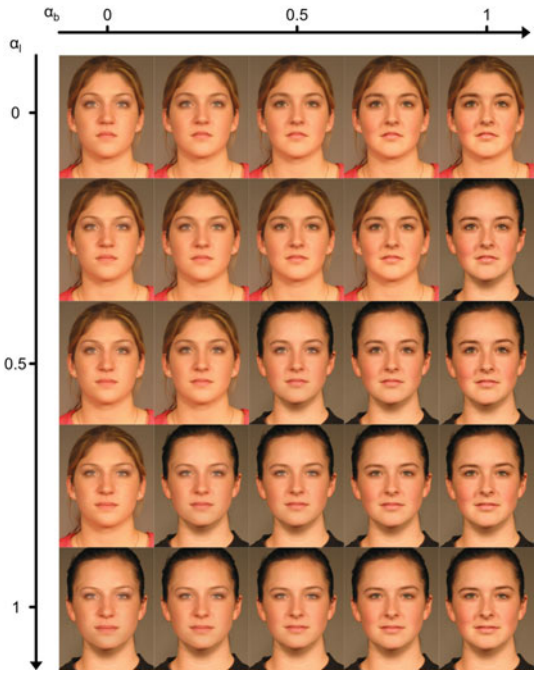

Fig. 19.4 Morphing with various α_b and α_l (genuine images from [6])

When performing a face swap, the attacker completely replaces the facial area. In a Face Morph Attack, two parameters are introduced to both control the blending factor α_b of the two faces but also a deformation factor α_l that aims at averaging the two face shapes. By adding control over α_b and α_l, it enables to exploit a flaw in common face recognition software. This attack was introduced in ref. [4] where it was shown that using $\alpha_b = 0.5, \alpha_l = 0.5$ allows producing a composite face that can later be used to authenticate two people using the same ID document. Later, it has been shown in ref. [5] that α_b seems to have a more significative impact on various face recognition systems. Figure 19.4 illustrates the effect of varying (α_b, α_l) couples.

In general, we assume that the face morphing can be produced before being printed and scanned back to a digital format to hide any traces of manipulation. In a typical remote onboarding scenario, it is important to keep in mind that those attacks exist. In particular, if we intended to use the photo to later authenticate the user.

Face De-identification

The rapid development of GANs (see Sect. 19.3.2) has led to more advanced face manipulation methods.

One good example is the advance in face de-identification methods which consist in removing some or all biometric traits of an individual. Those came from an

Fig. 19.5 From left to right, original identity, covering, blurring, pixelization, d-id

increasing concern about privacy as biometric authentication methods are becoming more common, but also from new regulations such as the General Data Protection Regulation (GDPR).

Three common methods were used before the introduction of the GAN-based face manipulation, i.e. masking, blurring and pixelization [7]. An example of each of those approach is given in Fig. 19.5.

Masking consists of covering the face with a graphic object (e.g. smiley) or a plain colour. Blurring and pixelization use simple filtering applied to the face area (e.g. Gaussian blur). These de-identification methods are simple and effective but come with some limitations. Firstly, the destructive approach produces an unpleasing result. Secondly, it is possible to partially reverse some of those methods [8] (e.g. de-blurring, de-pixelization and de-noising). And finally, these techniques do not allow suppressing specific characteristics (e.g. age and ethnicity). More sophisticated methods, using GANs, solve these problems. In ref. [9], the authors have been able to completely suppress a source face biometry while preserving the visual aspect of a face. The de-identified face is neither identifiable by a human nor by facial recognition algorithms. Their method allows to automatically de-identify a face in a video in real time. Using this technique, it is also possible to modify or remove biometric characteristics such as the age [10], gender or ethnicity.

19.3.2 Common Face Manipulation Algorithms

Landmark-based face manipulation

Even though deep learning-based face manipulation algorithm performs extremely well, there is still use cases where a more classical method is appropriate. One advantage of classical methods is that they do not need training data and can most of the time produce convincing results in real time.

Landmark-based methods usually come down to three simple steps. First, a face detection algorithm is applied. Then salient face features, often called landmarks, are extracted. And finally, the manipulation is performed.

In general, classical face manipulation algorithms used common landmark detection techniques like [11]. In such cases, those methods would only be used in simple

2D cases, i.e. fixed portrait tampering. More advanced methods exist to perform 3D face alignment. In general, classical face manipulation algorithm used common landmark detection techniques like [11]. In such cases, those methods would only be used in simple 2D cases, i.e. fixed portrait tampering. More advanced methods exist to perform 3D face alignment. They rely on a 3D mask synthesis which offers very good results. They are particularly well suited to be applied to images and also videos. One of the earliest methods [12] proposed to replace the face with a 3D model. This 3D face is then edited to modify facial expressions. However, it was not possible to make a real-time facial reenactment at that time. Newer methods yield dense 3D alignment in real time such as [13–15]. In [16], the authors have successfully developed a facial reenactment system that allows editing a face in the video in real time with a simple camera. They first detect facial expressions in a source and the target video. Then, they generate a 3D model of the face of both the target and the source video. Next, they transfer the facial expressions from the source to the target 3D model. Finally, they blend the 3D model on the target video. This produces very convincing results. Today, it is also possible to give life to a still image [17] by transferring facial expressions from a pilot video and optionally the voice.

At the time of writing, those more advanced methods tend to be much more difficult to implement. Because of that, AI-based methods are usually preferred for their ease of use. And thanks to the numerous existing face databases, many powerful methods have been proposed.

AI-Based Face Manipulation

With recent advances in generative models (Variational Auto Encoders (VAEs) and Generative Adversarial Networks)GANs)), Deep learning-based face manipulation has received a lot of attention. In particular, with the apparition of the popular deepfakes. The term deepfake is a portmanteau word composed of the “deep” to refer to “deep learning” and the word fake. The inventor of deepfake is an Internet user under the pseudonym “deepfake”. He was inspired by a paper that proposed a method to modify the environment of a video [18], and he applied it to faces. In this paper, Liu et al. built a framework that uses VAEs and GANs to apply modifications on each frame of a video.

The DeepFakes allow exchanging in a fast, automatic and realistic way a face in a video. Nowadays, the term deepfake is also used to designate more generally a “hyper-realistic” falsification of a video or audio signal.

The biggest danger is that it does not require special technical skills to make a deepfake unlike the more complex landmark-based method. Nor is it necessary to master complicated software. Today, anyone can make a deepfake.

The first deepfake method only used VAEs to replace a face in a video. A deepfake based on auto-encoders consists of using two auto-encoders and crossing the decoders. An auto-encoder is a type of neural network used to reconstruct an image from compressed information (called latent space) of the same image. In order to build a deepfake, it is necessary to train one of the auto-encoders with images of faces of a first individual and to train the other auto-encoder with images of the second

individual. Then, once the training phase is complete, the decoders are swapped to force the reconstruction of another face from the latent space. One of the important points in this method of deepfake generation is that both auto-encoders need to share the same encoder during the training stage.

But thanks to the various advances in GANs, in particular since the well-known SyleGAN [19] introduced by NVDIA research team, many deep facial forgery methods are now based on this technology rather than on VAEs.

A deepfake based on GANs used the neural networks introduced in 2014 by Ian Goodfellow [20]. In the same spirit as auto-encoders, a GAN is made of two distinct parts, a generator G and a discriminator D. In the case of deepfake generations, the role of the generator is to synthesise a video capable of deceiving the discriminator, and the role of the discriminator is to determine whether the content proposed by the generator is authentic or not.

Most recently, the authors of [21] were even able to remove the needs of the GAN to be trained on a specific individual. Rather, they trained their network to animate a single photo according to a source video. This last advance simplified even further the sequence of processing attached to the creation of deepfakes for non-experts.

19.4 Detecting Face Manipulation

In many practical applications, it is possible to control some parts of the acquisition process. Whether we can directly control the acquisition device or only access the incoming streaming for the said device, it is important to look at the problematic of Face Manipulation detection as a subset of image manipulation detection.

This does not mean that specialised detection should not be considered but rather that they should be completed with additional methods. In this section, we will introduce the common Face Manipulation detection method along with some more general Image Manipulation detection algorithms.

We will also talk about how much the control of the acquisition device and the definition of the User Interface can play a significant role in the detection of image tampering.

19.4.1 Face Specific Methods

When we do know that the acquired media will contain a human face, ignoring face-specific tampering detection method would be rather unwise.

Knowing that the acquired face cannot be considered as genuine, there exist many methods that aim at exposing digital tapering. We will consider two main categories, first the methods looking for weaknesses in the manipulation process and secondly the methods that try to expose physical inconsistencies.

Deepfake Detection Methods

With the increasing creation of deepfakes, many detection methods have been proposed. In the literature, there are mainly three categories of deepfake detection methods: based on physiological analysis, based on image texture analysis and based on automatic detection with artificial intelligence. As part of the physiological analysis, Li et al. [22] observed some inconsistencies in the eye blinking in a deepfake video. Using a Long-term Recurrent Convolutional Network (LCRN), they successfully detected deepfake videos. In ref. [23], the authors determine whether a video is a deepfake or not by analysing inconsistencies in head position. For detection methods based on image or texture analysis, the authors mainly look for inconsistencies in the optical flow [24] or the presence of artefacts [25]. Finally, approaches purely based on a detection using artificial intelligence pass the frames of a video through neural networks. The neural networks can be recurrent neural networks [26], 3D convolutional networks [27], recurrent networks or an ensemble of them. Kumar et al. [28] proposed a method dedicated to detect videos to which face reenactment has been applied using [16]. They proposed an ensemble of 5 ResNets trained to identify noise patterns or artefacts. Despite very good results and robustness at different levels of video compression, their method cannot be used in real time. Megahed et al. have described a method [29] for detecting face reenactment manipulations based on Histogram of Oriented Gradient and SVM. Unfortunately, because of the significant diversity of the different ways to generate a deepfake, it is very difficult to develop a method suitable to detect all possible types of deepfake videos. Some methods have been proposed to detect both deepfakes by face swapping and deepfakes by face reenactment. In ref. [30], the authors trained two convolutional neural networks to detect both swapped face videos and reenacted face videos. Despite encouraging results, their methods are dependent on the training database and, therefore, do not work very well on unseen types of deepfakes. In ref. [31], the authors also propose a method based on deep learning to automatically detect deepfakes by looking at the facial regions. Their method is robust to different levels of compression. However, it is also very dependent on the training database. A model able to detect with a high level of accuracy a type of deepfake can easily be fooled by a deepfake that has been generated with another method. In addition to the variety of deepfakes generation methods that make it difficult to generalise detection techniques, it is also important to consider that the models must be robust to adversarial attacks. Indeed, in ref. [32], it has been shown that it is possible to easily deceive a detector by injecting an adverse noise into a video.

19.4.2 Face Agnostic Methods

Image forensic can be divided into two main categories, i.e. active and passive image forgery detection. In passive image forgery detection, we have to authenticate an image without any knowledge of the digital media. Whereas active forgery detection

has access to at least a partial information about the image provenance and leverage that knowledge to authenticate the media.

One can intuitively understand how active forgery detection is preferred as we can pinpoint very specific properties of the image to assess its authenticity. As mentioned earlier, in practical cases, we often have some controls on the acquisition device. This can be used to our advantage as we can somewhat impose some constraints on the acquired media to go from passive forgery detection to active forgery detection.

This allows us to add many layers of authentication on top of the previously mentioned face-specific detection methods and build a more confident prediction about the image integrity.

Camera-based methods

The first category of image manipulation detection algorithm is based on the characteristics and steps involved in the creation process of a digital image. A brief overview of the acquisition pipeline is given in Fig. 19.6. Light passes through the lenses before reaching the camera sensor. Sensors are not perfectly manufactured and small defects can be used as fingerprints of a particular sensor. A widely known fingerprint for camera model identification is the Photo Response Non Uniformity (PRNU) which was first introduced in ref. [33]. A sufficient number of images from one camera (about 50) allows us to produce a fingerprint which can then be used to assess if an image comes from a specific camera. In some cases, the PRNU has further been used to locate the digital tampering by searching for partial mismatch of the fingerprint [34, 35].

Common camera sensors first separate the lights using what is called a Colour Filter Array (CFA). This filter generally separates the information into three channels (typically red, green and blue) which are later on interpolated to produce the full-size image. This whole operation is known as image demosaicing. This operation tends to leave correlation between the final pixels which can be used to detect digital forgeries by locating the CFA grid and type. Because one sensor can only use one specific CFA, the detection of portions of an image with either a misaligned CFA or a mismatching type can inform of the presence of a digital tampering [36, 37].

Once the acquisition pipeline is complete, a final image pixel z can be roughly modelled as

$$z = x + \eta \tag{19.1}$$

where x was the true pixel value corrupted by some noise η. In reality, more elaborate models than Eq. 19.1 are used such as [38] when x is corrupted by various noises and processes. Typical noise sources are the so-called shot noise (due to the way light reaches the sensor) and read noise due to amplification, quantification, etc. On top of that, different post-processing such as gamma correction or compression would further corrupt the observed pixel x.

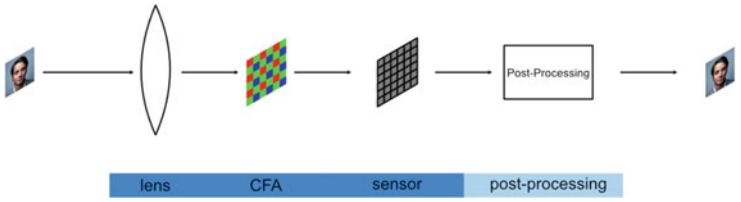

Fig. 19.6 Simplified acquisition pipeline

Whatever the noise model used, one can make the assumption that the counterfeit will not be able (or will not try) to reproduce a consistent noise across the image. Many methods aim at exposing those inconsistencies [39].

Few other methods focus on other artefacts such as inconsistencies in chromatic aberrations [40] and inconsistent camera response function [41, 42].

Pixel-based methods

Another class of image manipulation detection methods is the pixel-based method. Instead of targeting parts of the acquisition pipeline, those methods rather try to target some parts of the tampering process.

In fact, there exist a few common tampering categories: Image splicing where an element from one image is inserted into another image; copy-Move where elements are duplicated within a single image; object-removal where objects are removed from an image (typically using inpainting algorithms).

While producing a forged image, a digital artist often uses a combination of those types of forgery and produce what we usually call a compositing.

Because the digital artist most likely wants his/her compositing to be convincing, he would use various techniques and post-processing to reach his goal. Think of the production of a deepfake. As we mentioned, CNNs tend to produce blurry results. A counterfeit might then want to enhance the quality of his deepfake by applying a sharpening filter. Such post-processing is performed by the widely used DeepFaceLab software for instance. Whatever the processing use, it will leave traces.

Most of the time, the detection of splicing forgeries relies on camera-based methods. In fact, in such forgeries, a large portion of the image may carry a significant difference with respect to intrinsic camera properties. It is not unreasonable to suppose that the spliced part might have gone through several post-processing such as sharpening and Gaussian blurring. In those cases, some methods [43, 44] aim at exposing those post-processing as a trace of image forgery.

For copy-move forgery, camera-based methods tend to be ill-suited as the duplicated element will share most of the initial image properties. For this reason, there exist many algorithms targeting copy-move forgery directly [45]. One challenge of copy-move forgery detection is the presence of Similar but Genuine Object (SGO). In the case of SGO, most algorithms tend to produce a high false positive rate [46]. This becomes a strong issue in the case of remote ID onboarding as pictures of ID

documents contain many SGO. This has been addressed by a few previous papers [47, 48] and has yet to be further developed with more work and more datasets.

Object-removal is a more specific attack and is less studied in the state of the art. One reason is that most object-removal forgeries are performed either with direct copy move or with inpainting algorithms, because current professional software tends to use exemplar-based inpainting algorithms. Thus, object-removal can often be detected with copy-move detection methods, but there exists research focussing especially on Object-Removal [49, 50].

Format-Based Methods

The last category of image forgery detection algorithm is the format-based method. Those algorithms make neither particular assumption on the acquisition pipeline nor on the tampering process. Rather they aim at authenticating the format of a given media.

In a system such as Fig. 19.7 (see Sect. 19.5.2), the acquisition pipeline is controlled.

By assuming the acquisition process is secure, the authentication of the media can be reduced to specify a check on the format properties. For instance, if the media was a JPEG image sent with a quality factor of 95, one could verify that this property did not change. One advantage of format-based methods is that they have proved to be extremely effective and have been well studied.

For images, several approaches have been proposed for JPEG images. It is possible with state-of-the-art methods to verify various properties such as the JPEG quality factor [51] or the presence of a double compression [52]. While less studied, methods for analog video formats exist too.

19.4.3 Datasets

In this section, we will introduce the most common image and video tampering datasets for both Face Manipulation and general tampering detection.

Images Datasets

A list of common image tampering datasets is given in Table 19.2. This table is far from being exhaustive and thus illustrates the wide availability of datasets. In particular since 2019, three extremely large datasets (PS-Battles [53], DEFACTO [54] and IMD2020 [55]) have been publicly released containing all kinds of forgeries that should allow researchers to properly evaluate, train or test their forensic algorithms.

On the other hand, there exist only a few datasets that are face specific. In 2017, the FaceSwap dataset was released containing only face swap manipulations. They used two different methods to develop the dataset. The Biometix datasets, released in 2017, contained 1082 Face Morphing based on the FERET face dataset. In the

Table 19.2 Image and face manipulation datasets and information

Dataset	Manipulation	Size
Columbia gray and colour [56]	Splicing	1092
MICC F220 and F2000 [57]	Copy-move	810
Casia v1 and v2 [58]	Splicing, copy-move	6044
COVERAGE [46]	Copy-move	100
Biometix [59]	Face morphing	1082
FaceSwap [60]	Face swapping	1,927
PS-Battles [53]	All	102,028
DEFACTO [54]	All	229,000
IMD2020 [55]	All	72000
OpenMFC2020 [61]	All	16,000

DEFACTO dataset, about eighty thousand face morphing are available. Apart from those, we could cite the DSI-1 dataset for completeness, but it only contains about 25 tampered images.

Even though most splicing detection methods are applicable to Face Manipulation detection, we believe that more specific datasets would help in the development of more specific methods.

Video Datasets

In contrary to image forgery datasets, there exist only a few general video tampering datasets. Some common datasets are given in Table 19.3. We believe that this is due to the extremely large degrees of freedom when compressing a video that make it difficult to produce a dataset that would suit everyone's needs. Often, researcher ends up crafting their own dataset depending on the feature they use.

Thanks to the initiative of the NIST, researchers now have access to a dataset of about 1500 tampered video.

If we lack general-purpose video forgery datasets, many video datasets dedicated to the detection of deepfake forgeries have been proposed. This imbalanced could be explained by the dangers they represent against biometric authentication systems but also by the fact that deepfake detection algorithms often operate blindly. Thus,

Table 19.3 General video tampering datasets

Dataset	Manipulation	Size
SULFA [62]	Copy-move	10
MTVFD [63]	Copy-move, splicing, frame swapping	30
OpenMFC2020 [61]	Various	1,500

Table 19.4 Deepfakes Datasets and informations

Dataset	Fake videos	Real videos	Identity	Methods	Augmentation
UADFV	49	49	49	1	–
DeepfakeTIMIT	640	320	43	2	–
FaceForensics++	4000	1000	–	4	2
Google DFD	3000	363	28	5	–
Celeb-DFD	5639	890	59	1	–
Deeper	1000	59000	100	1	7
DFDC	104500	23654	960	5	19

the large degree of freedom when creating the forgeries is considered as part of the problem and the detection method should be able to work in all scenarios.

To the best of our knowledge, we count seven large datasets of deepfakes (Table 19.4). The most important and recent database is the one attached to the Deepfake Detection Challenge (DFDC) [64]. The Deepfake Detection Challenge is an international competition, launched in December 2019, to help the scientific community to develop new techniques to detect deepfakes. The competition closes at the end of March 2020, and the winning solution achieved an accuracy score of 82%.

19.5 Counter-Forensics and Countermeasures

19.5.1 Counter-Forensics

We described many different tampering detection approaches. In a way, all those methods implicitly assume that the attacker will perform a naive tampering which in turn will leave many traces.

Though it is not unreasonable to assume so, it is important to also consider cases where the attacker will try his/her best to hide traces of his forgery. This can go from simply correcting incorrect EXIF metadata to directly target detection methods described in Sect. 19.4.2.

For example, in ref. [65], the authors propose a method to suppress the PRNU of an image and to replace it with the one of another camera. In ref. [66], the authors showed that it was possible to mislead CFA-based algorithms. Other approaches try to hide the forgery at the compression level such as [67], and some methods even target specific forgeries such as [68].

Regarding face-specific methods, rapid advances in the realistic rendering of deepfakes and GAN-based facial forgeries are strongly connected to the progress of methods for detecting such contents. But as each new detection method reveals a “weakness” in the synthesised content, it is then rapidly fixed in order to hide the forged image or video from detection.

First counter-forensic has corrected physical or physiological inconsistencies by adding, in the case of deepfakes, a natural eye blink that was missing and in the general case of face swapping, correcting the orientation of the face in relation to the head. They also adjusted inconsistencies in colour, brightness or artefacts that can randomly appear on a GAN-generated image.

Another powerful counter-forensic is the use of adversarial attacks. An adversarial attack consists of adding a computed noise imperceptible to the human eye into the image. This noise has a big impact on deep learning-based detectors [69]. Based on this approach, several methods have been proposed to make a deepfake video recognised as an original video by deep learning-based deepfake detectors [70, 71]. The authors of [32] have successfully generated a single adversarial attack that misled three different deepfake detectors.

19.5.2 Countermeasures

Here, we will take a more macroscopic look at typical remote identity document verification systems and discuss how one would use the methods described in Sect. 19.4.2 to ensure the integrity of the end media.

We will consider two main scenarios. In the first case, we have access to the acquisition device and can somewhat control it as in Fig. 19.7. The captured media is then sent over the network to some servers that will later have to authenticate it.

In the second scenario Fig. 19.8, we have no knowledge about the acquisition device. A server simply receives a media that needs to be authenticated. In that case, it is still possible by design to impose some constraints on the media such as the format and size.

Controlled acquisition device

Whenever possible, it is always preferred to have some sort of control over the acquisition process. The reason is that active detection methods are arguably more effective than completely passive approaches.

In particular, within the scenario of Fig. 19.7, it is possible to use every single category of detection algorithms described earlier which allows for a more reliable decision.

Fig. 19.7 Controlled acquisition pipeline

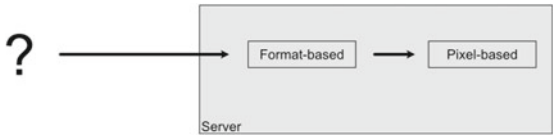

Fig. 19.8 Uncontrolled acquisition pipeline

In such a case, we can assume that the attacker can perform his/her tampering during three different stages. The first option is to attack the stream at the earliest stage of the acquisition pipeline, i.e. at the driver level. Any electronic component of a system is controlled by a piece of software called a driver. In the case of a camera, the role of the driver would be to directly control the sensor to retrieve the raw image data. It would then apply every basic image processing needed to pass forward a readable RGB image, i.e. demosaicing, camera response function and some basic noise filtering.

If the image is altered at the driver level and unless there is a known watermarking algorithm used at a hardware level, camera-based detection algorithms are the way to go. In fact, at this stage, the image is supposed to directly come from the sensor and thus must fulfil some models such as the uniqueness of the CFA, a precise model of the sensor noise and so on. As for every method, false positives are possible. But repeated tampering detection by multiple methods at this stage must imply a deeper analysis at a later stage of the system.

After passing the drivers, we receive an RGB-like image. The client is now in charge of sending the media to a distant server. Because we should not assume the network channel to be safe, as it could be subject to a man-in-the-middle attack, for instance, it is then the last opportunity to inject knowledge on the media before sending it on the network. Think of applying a specific JPEG/MPEG compression adding a specific watermark, etc. which will later be used to enhance the format-based detection method for instance.

Once the media has reached the server and after applying format-based method to verify the last known properties of the image. Pixel-based and more face-specific methods can now be used to further confirm any previous detection that might have occurred. It is fine to use such methods at the very end of the pipeline as they often assume nothing about the properties of the acquisition device nor the format used in case of a compression.

One major advantage of having control over the acquisition device is also the possibility to interact with the user in some cases such as remote KYC. In that scenario, it is possible to ask the user to perform actions in order to make the tampering process harder. Suppose we ask the user to capture his ID document. We know that the ID photo is a potential target for a counterfeit. We also know that to alter the photo in real time, a precise detection of the face and some landmarks is needed. One challenge of both face and landmark detection is the presence of an occluding element in the face. Asking the user to hide part of the photo at some point might make the automatic tampering algorithm to fail which could cause visible artefacts.

Another strategy consists in making difficult the manipulation creation. This can be applied only in some contexts. Ruiz et al. [72] proposed a method based on adversarial attacks as a defense. By adding a specific noise in the image, they are able to make that image unusable by a deepfake generator. In that sense, having a good knowledge of the common tampering methods is necessary to develop more challenging user experiences for a counterfeit.

Uncontrolled acquisition device

As already mentioned, having control over the acquisition stage is preferable. It is not always possible though and one might have to accept media from an unknown source as in Fig. 19.8. But it does not mean that such a system must accept anything as an input.

First of all, imposing a specific format for the incoming media is mandatory as this already allows the use of specific format-based methods. Also, if the quality of the media is sufficient enough, the blind camera-based method can still be applied. As already pointed out, pixel-based methods are always a good option as they are most of the time blind detections.

Unlike the controlled case, one should keep in mind that the counterfeit has all the needed time to try and hide his/her manipulation. It is, thus, reasonable to assume that he/she will try to apply as many counter-forensic methods as he/she can make every decision algorithm less reliable.

As for the controlled case, asking the user to send pictures or video with specific constraints (e.g. hiding part of the face) is a good idea, even though in this scenario the counterfeit will have the time to correct visible artefacts.

A general rule of thumb for both scenarios would be to impose as many constraints as possible to ensure that only really experienced counterfeit will be able to fool the system.

Preventive Measures

When the context allows conceiving the initial ID document in order to prevent specifically digital attacks on the portrait, different countermeasures can be taken to facilitate the detection of manipulations. First strategies can be related to the addition of semi-fragile watermarks [73] in the content, which will disappear during an image manipulation and thus trigger a detection. The second category of strategies is to secure the image thanks to cryptographic seal, based on perceptual hashes [74, 75].

19.6 Reference Framework, Standardisation and Legal Aspects

Growing usage of remote identification generates growing types of frauds and growing needs for a safer environment. New regulations are emerging.

In France, a new reference framework, named PVID, describes the recommendations to remotely acquire the identity of someone [76]. In its preliminary version, it forbids limiting the control only to still images to authenticate a person. Rather, the authentication must be performed based on a video stream. It also requires a hybrid approach with both human and machine checks. This reference framework is planned to be included in the French implementation of the E-IDAS, giving it a European impact. PVID is supposed to be published in April 2021. It would probably impose constraints on resolution, frame rate and perhaps bandwidth of the transmitted video and will forbid any disruption of the video stream. It also imposes a double verification process involving parallel humans and machines.

In Europe, the security Agency ENISA has published in February 2021 an analysis of the Remote ID Proofing [77], in which they describe remote verification in several cases with or without a human operator and in the context of a video acquisition to prevent manipulation. They based their work on the 2018 document of German BSI [78] which precise clearly the threats and attacks, but also the condition of success of attacks.

In order to qualify the solutions proposed by vendors, some dedicated laboratories begin to appear to test the capacities of manipulation detectors, with private large sets of software or hybrid attacks. Comparable methodologies, tools and datasets are clearly required in the next years to insure a global level of security against face manipulations.

19.7 Conclusions

With the worldwide crisis of 2020 due to the COVID-19, we observed significant adoption of remote technologies. We believe that this trend will continue and that many systems will need to be able to confidently authenticate their end-users which will come with an increasing use of face and behaviour recognition software.

In the meantime, the deepfakes visual quality has Improved, and in the future, the deepfakes will be even more realistic by correcting all the imperfections they still include. Nowadays, deepfakes are mainly used to create adult content or entertainment videos. However, we believe that in the next few years, deepfakes will be the most used way for attacking facial recognition systems. With those evolving technologies, the creation of a forged media stream is becoming easier every day.

We must anticipate the use of those powerful technologies against any operational system based on face and behaviour recognition software. Thus, the detection of those tampered media is becoming increasingly important.

In this chapter, we introduced the four common Face Manipulation categories. We described how each of these methods could be used against an operational system. We focussed our attention on the example of a remote system for identity document verification. It is an interesting use case as it gives many opportunities for an attacker to fool the system. We also gave a brief overview of many detectors specific to faces, general image manipulation detection methods but also some datasets for image and video tampering detection. Then we introduced a few common counter-forensics methods, but also countermeasures to increase the overall system reliability whether the acquisition device is controlled or not. Finally, we discussed standardisation and legal aspects covering such systems.

As of today, the detection of deepfakes and other attacks is far from being solved. Looking at the bright side, we saw that many detection algorithms already exist. Even though none of those can fully operate in a completely blind manner, most of them can be used in a more active approach. This requires the overall system to be carefully designed. A proper combination of all those tools may give powerful insights about a given media integrity. We believe that most of the time, some constraints over the media properties can be imposed. Completely blind detection algorithms are welcomed but should not prevail. We encourage researchers to develop a more formal definition of operational systems and to imagine precise lifecycles of digital media. This would allow the development of more specific datasets, detection methods and interpretability of such methods. This in turn would be extremely beneficial as it would help to close the gap between theoretical research and operational applications.

We believe that Face Manipulation Detection must be included in the pipeline of required processing for any official document. It could be organised in a way similar to the ICAO Picture Compliance test for instance. We also recommend the standardisation of the tests and the methodologies to measure the efficiency of the Face Manipulation Detection algorithms in order to obtain fair and efficient comparison between vendors. This standardisation effort needs to take into account the various strategies, contexts and goals described here, in order to define meaningful metrics.

References

1. Tolosana Ruben, Vera-Rodriguez Ruben, Fierrez Julian, Morales Aythami, Ortega-Garcia Javier (2020) Deepfakes and beyond: a survey of face manipulation and fake detection. Inf Fusion 64:131–148
2. Bitouk D, Kumar N, Dhillon S, Belhumeur P, Nayar S (2008) Face swapping: automatically replacing faces in photographs. In: SIGGRAPH 2008
3. Korshunova I, Shi W, Dambre J, Theis L (2017) Fast face-swap using convolutional neural networks. In: Proceedings of the IEEE international conference on computer vision, pp 3677–3685

4. Ferrara M, Franco A, Maltoni D (2014) The magic passport. In: IEEE international joint conference on biometrics. IEEE, pp 1–7
5. Ferrara M, Franco A, Maltoni D (2019) Decoupling texture blending and shape warping in face morphing. In: 2019 international conference of the biometrics special interest group (BIOSIG)
6. Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of the face recognition grand challenge. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1. IEEE, pp 947–954
7. Korshunov P, Araimo C, De Simone F, Velardo C, Dugelay JL, Ebrahimi T (2012) Subjective study of privacy filters in video surveillance. In: 2012 IEEE 14th international workshop on multimedia signal processing (MMSP). IEEE, pp 378–382
8. Ruchaud Natacha, Dugelay Jean-Luc (2016) Automatic face anonymization in visual data: are we really well protected? Algorithms and Systems, In Image Processing
9. Gafni O, Wolf L, Taigman Y (2019) Live face de-identification in video. In: 2019 IEEE/CVF international conference on computer vision (ICCV). IEEE, pp 9377–9386
10. Antipov G, Baccouche M, Dugelay JL (2017) Face aging with conditional generative adversarial networks. In: 2017 IEEE international conference on image processing (ICIP). IEEE, pp 2089–2093
11. Kazemi V, Sullivan J (2014) One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp 1867–1874
12. Vlasic D, Brand M, Pfister H, Popovic J (2005) Face transfer with multilinear models. In: SIGGRAPH 2005
13. Cao Chen, Hou Qiming, Zhou Kun (2014) Displaced dynamic expression regression for real-time facial tracking and animation. ACM Trans Graph (TOG) 33(4):1–10
14. Jeni LA, Cohn JF, Kanade T (2015) Dense 3d face alignment from 2d videos in real-time. In: 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG), vol 1. IEEE, pp 1–8
15. Thies Justus, Zollhöfer Michael, Nießner Matthias, Valgaerts Levi, Stamminger Marc, Theobalt Christian (2015) Real-time expression transfer for facial reenactment. ACM Trans. Graph. 34(6):183–1
16. Thies J, Zollhöfer M, Stamminger M, Theobalt C, Nießner M (2016) Face2face: real-time face capture and reenactment of rgb videos. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 2387–2395
17. Averbuch-Elor H, Cohen-Or D, Kopf J, Cohen MF (2017) Bringing portraits to life. ACM Trans Graph (TOG) 36:1–13
18. Liu MY, Breuel T, Kautz J (2017) Unsupervised image-to-image translation networks. In: NIPS
19. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. IEEE, pp 4401–4410
20. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial networks. arXiv:1406.2661
21. Siarohin A, Lathuilière S, Tulyakov S, Ricci E, Sebe N (2020) First order motion model for image animation. arXiv:2003.00196
22. Li Y, Chang MC, Lyu S (2018) In ictu oculi: exposing ai created fake videos by detecting eye blinking. In: 2018 IEEE international workshop on information forensics and security (WIFS). IEEE, pp 1–7
23. Yang X, Li Y, Lyu S (2019) Exposing deep fakes using inconsistent head poses. In: ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 8261–8265
24. Amerini I, Galteri L, Caldelli R, Del Bimbo A (2019) Deepfake video detection through optical flow based cnn. In: 2019 IEEE/CVF international conference on computer vision workshop (ICCVW). IEEE, pp 1205–1207
25. Li Y, Lyu S (2019) Exposing deepfake videos by detecting face warping artifacts. arXiv:1811.00656

26. Guera D, Delp EJ (2018) Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS). IEEE, pp 1–6
27. Lima O, Franklin S, Basu S, Karwoski B, George A (2020) Deepfake detection using spatiotemporal convolutional networks. arXiv:2006.14749
28. Kumar P, Vatsa M, Singh R (2020) Detecting face2face facial reenactment in videos. In: 2020 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 2578–2586
29. Megahed A, Han Q (2020) Face2face manipulation detection based on histogram of oriented gradients. In: 2020 IEEE 19th international conference on trust, security and privacy in computing and communications (TrustCom). IEEE, pp 1260–1267
30. Afchar D, Nozick V, Yamagishi J, Echizen I (2018) Mesonet: a compact facial video forgery detection network. In: 2018 IEEE international workshop on information forensics and security (WIFS). IEEE, pp 1–7
31. Rossler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M (2019) Faceforensics++: learning to detect manipulated facial images. In: Proceedings of the IEEE/CVF international conference on computer vision. IEEE, pp 1–11
32. Neekhara P, Dolhansky B, Bitton J, Ferrer CC (2020) Adversarial threats to deepfake detection: a practical perspective. arXiv:2011.09957
33. Lukas J, Fridrich J, Goljan M (2006) Digital camera identification from sensor pattern noise. IEEE Trans Inf Forensics Secur 1(2):205–214
34. Chierchia Giovanni, Poggi Giovanni, Sansone Carlo, Verdoliva Luisa (2014) A bayesian-mrf approach for prnu-based image forgery detection. IEEE Trans Inf Forensics Secur 9(4):554–567
35. Korus Paweł, Huang Jiwu (2016) Multi-scale analysis strategies in prnu-based tampering localization. IEEE Trans Inf Forensics Secur 12(4):809–824
36. Ferrara Pasquale, Bianchi Tiziano, De Rosa Alessia, Piva Alessandro (2012) Image forgery localization via fine-grained analysis of cfa artifacts. IEEE Trans Inf Forensics Secur 7(5):1566–1577
37. Le N, Retraint F (2019) An improved algorithm for digital image authentication and forgery localization using demosaicing artifacts. IEEE Access 7:125038–125053
38. Thai TH, Retraint F, Cogranne R (2015) Generalized signal-dependent noise model and parameter estimation for natural images. Signal Process 114:164–170
39. Pan X, Zhang X, Lyu S (2011) Exposing image forgery with blind noise estimation. In: Proceedings of the thirteenth ACM multimedia workshop on multimedia and security. IEEE, pp 15–20
40. Johnson MK, Farid H (2006) Exposing digital forgeries through chromatic aberration. In: Proceedings of the 8th workshop on multimedia and security. IEEE, pp 48–55
41. Hsu Yu-Feng, Chang Shih-Fu (2010) Camera response functions for image forensics: an automatic algorithm for splicing detection. IEEE Trans Inf Forensics Secur 5(4):816–825
42. Chen C, McCloskey S, Yu J (2017) Image splicing detection via camera response function analysis. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp 5087–5096
43. Cao G, Zhao Y, Ni R, Yu L, Tian H (2010) Forensic detection of median filtering in digital images. In: 2010 IEEE international conference on multimedia and expo. IEEE, pp 89–94
44. Cao G, Zhao Y, Ni R, Kot AC (2011) Unsharp masking sharpening detection via overshoot artifacts analysis. IEEE Signal Process Lett 18(10):603–606
45. Christlein Vincent, Riess Christian, Jordan Johannes, Riess Corinna, Angelopoulou Elli (2012) An evaluation of popular copy-move forgery detection approaches. IEEE Trans Inf Forensics Secur 7(6):1841–1854
46. Wen B, Zhu Y, Subramanian R, Ng TT, Shen X, Winkler S (2016) COVERAGE-A novel database for copy-move forgery detection. In: 2016 IEEE international conference on image processing (ICIP). IEEE, pp 161–165
47. Li Yuanman, Zhou Jiantao (2018) Fast and effective image copy-move forgery detection via hierarchical feature point matching. IEEE Trans Inf Forensics Secur 14(5):1307–1322

48. Mahfoudi G, Morain-Nicollier F, Retraint F, Pic MM (2019) Copy and move forgery detection using sift and local color dissimilarity maps. In: 2019 IEEE global conference on signal and information processing (GlobalSIP). IEEE, pp 1–5
49. Zhang Dengyong, Liang Zaoshan, Yang Gaobo, Li Qingguo, Li Leida, Sun Xingming (2018) A robust forgery detection algorithm for object removal by exemplar-based image inpainting. Multimed Tools Appl 77(10):11823–11842
50. Mahfoudi G, Morain-Nicolier F, Retraint F, Pic MM (2020) Object-removal forgery detection through reflectance analysis. In: 2020 IEEE international symposium on signal processing and information technology (ISSPIT) (ISSPIT-2020), virtual
51. Retraint F, Zitzmann C (2020) Quality factor estimation of jpeg images using a statistical model. Digital Signal Process 103:102759
52. Farid H (2009) Exposing digital forgeries from jpeg ghosts. IEEE Trans Inf Forensics Secur 4(1):154–160
53. Heller S, Rossetto L, Schuldt H (2018) The ps-battles dataset-an image collection for image manipulation detection. arXiv:1804.04866
54. Mahfoudi G, Tajini B, Retraint F, Morain-Nicolier F, Dugelay JL, Pic MM (2019) DEFACTO: image and face manipulation dataset. In: 2019 27th European signal processing conference (EUSIPCO). IEEE, pp 1–5
55. Novozamsky A, Mahdian B, Saic S (2020) IMD2020: a large-scale annotated dataset tailored for detecting manipulated images. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision workshops. IEEE, pp 71–80
56. Hsu Y-F, Chang SF (2006) Detecting image splicing using geometry invariants and camera characteristics consistency. In: 2006 IEEE international conference on multimedia and expo. IEEE, pp 549–552
57. Amerini Irene, Ballan Lamberto, Caldelli Roberto, Del Bimbo Alberto, Serra Giuseppe (2011) A sift-based forensic method for copy-move attack detection and transformation recovery. IEEE Trans Inf Forensics Secur 6(3):1099–1110
58. Dong J, Wang W, Tan T (2013) Casia image tampering detection evaluation database. In: 2013 IEEE China summit and international conference on signal and information processing. IEEE, pp 422–426
59. Biometix dataset. https://www.linkedin.com/pulse/new-face-morphing-dataset-vulnerability-research-ted-dunstone. Last Accessed 22 Feb 2021
60. Zhou P, Han X, Morariu VI, Davis LS (2017) Two-stream neural networks for tampered face detection. In: 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW). IEEE, pp 1831–1839
61. Guan H, Kozak M, Robertson E, Lee Y, Yates AN, Delgado A, Zhou D, Kheyrkhah T, Smith J, Fiscus J (2019) MFC datasets: large-scale benchmark datasets for media forensic challenge evaluation. In: 2019 IEEE winter applications of computer vision workshops (WACVW). IEEE, pp 63–72
62. Qadir G, Yahaya S, Ho AT (2012) Surrey university library for forensic analysis (sulfa) of video content. In: IET conference on image processing (IPR 2012). IEEE, pp 1–6
63. Al-Sanjary OI, Ahmed AA, Sulong G (2016) Development of a video tampering dataset for forensic investigation. Forensic Sci Int 266:565–572
64. Dolhansky B, Bitton J, Pflaum B, Lu J, Howes R, Wang M, Canton Ferrer C (2020) The deepfake detection challenge dataset. arXiv:2006.07397
65. Villalba LJ, Orozco AL, Corripio JR, Hernandez-Castro J (2017) A PNRU-based counter-forensic method to manipulate smartphone image source identification techniques. Future Gener Comput Syst 76:418–427
66. Chuang WH, Wu M (2012) Robustness of color interpolation identification against anti-forensic operations. In: International workshop on information hiding. Springer, pp 16–30
67. Stamm MC, Tjoa SK, Lin WS, Liu KR (2010) Undetectable image tampering through jpeg compression anti-forensics. In: 2010 IEEE international conference on image processing. IEEE, pp 2109–2112

68. Amerini Irene, Barni Mauro, Caldelli Roberto, Costanzo Andrea (2013) Counter-forensics of sift-based copy-move detection by means of keypoint classification. EURASIP J Image Video Process 2013(1):1–17
69. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R (2014) Intriguing properties of neural networks. CoRR. arXiv:1312.6199
70. Hussain S, Neekhara P, Jere M, Koushanfar F, McAuley J (2020) Adversarial deepfakes: evaluating vulnerability of deepfake detectors to adversarial examples. arXiv:2002.12749
71. Carlini N, Farid H (2020) Evading deepfake-image detectors with white- and black-box attacks. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). IEEE, pp 2804–2813
72. Ruiz N, Bargal SA, Sclaroff S (2020) Disrupting deepfakes: adversarial attacks against conditional image translation networks and facial manipulation systems. In: ECCV workshops 2020
73. Lee CF, Shen JJ, Hsu FW (2019) A survey of semi-fragile watermarking authentication. In: Jeng-Shyang P, Akinori I, Pei-Wei T, Lakhmi CJ (eds) Recent advances in intelligent information hiding and multimedia signal processing. Springer International Publishing, Cham, pp 264–271
74. Pic MM, Ouddan A (2017) Photometrix^{TM}: a digital seal for offline identity picture authentication. In: European intelligence and security informatics conference (EISIC). IEEE
75. Pic MM, Mahfoudi G, Trabelsi A (2019) A phygital vector for identity, robust to morphing. In: Digital document security (DDS2019)
76. Référentiel d'exigences pvid (v1.0). https://www.ssi.gouv.fr/uploads/2020/11/anssi_pvid_referentiel_exigences-v1.0.pdf. Last Accessed 22 Feb 2021
77. Remote id proofing. https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing/at_download/fullReport. Last Accessed 22 Feb 2021
78. Technical guideline tr-03147 assurance level assessment of procedures for identity verification of natural persons. https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/TechGuidelines/TR03147/TR03147.pdf. Last Accessed 22 Feb 2021