The Visual Computer (2023) 39:813–833
https://doi.org/10.1007/s00371-021-02347-4

SURVEY

A detailed analysis of image and video forgery detection techniques

Shobhit Tyagi[1] · **Divakar Yadav**[1]

Accepted: 25 October 2021 / Published online: 13 January 2022

Abstract
With the recent advancement in modern technology, one can easily manipulate a digital image or video using computer software or a mobile application. The purpose of editing visual media could be as simple as to look good before sharing to the social networking site's or can be as malicious as to defame or hurt one's reputation in the real world through such morphed visual imagery. Identity theft is one of the examples where one's identity get stolen by some impersonator who can access the personal and financial information of an innocent person. To avoid such drastic situations, law enforcement authorities must use some automatic tools and techniques to find out whether a person is innocent or the culprit. One major question that arises here is how and what parts of visual imagery can be manipulated or edited. The answer to this question is important to distinguish the authentic images/videos from the doctored multimedia. This survey provides a detailed analysis of image and video manipulation types, popular visual imagery manipulation methods, and state-of-the-art image and video forgery detection techniques. It also surveys different fake image and video datasets used in tampering. The goal is to develop a sense of privacy and security in the research community. Finally, it focuses to motivate researchers to develop generalized methods to capture artificial visual imagery which is capable of detecting any type of manipulation in given visual imagery.

Keywords Visual imagery forgery detection · Image and video manipulation and forensics · Deep learning

1 Introduction

There is exponential rise of image generation in recent few years. An estimate suggests that around 1.4 trillion digital photographs are being generated in the year 2020 alone [1]. Nowadays, digital images are necessary for our daily lives, by not only storing images of family and friends but also present on the cover of every major information source like magazines, newspapers, journals, etc. From teaching aids to pieces of evidence in courts, images are used everywhere for both office purposes and personal memories. The approximate projection of these generated digital images is shown in Table 1. Even though most people edit images for fun but if one's intention is to hurt some innocent person through facial manipulation, then that person can face some consequences in the real world. This explosion of digital images creates a hike for the development of image editing tools. In this digital era, there are several image processing softwares and mobile applications that can produce doctored images with high sophistication. The manipulated images are very hard to differentiate from the authentic ones to the naked eyes. Image editing softwares like Adobe Photoshop [2], GIMP [3], BeautyCam [4], etc., are very powerful and simple to use tools, available in the market. These softwares require little or no knowledge related to image processing with few instructions set for the user. This is the reason for their popularity among the users, particularly teenagers. They create many slow-motion videos by morphing individual frames to generate special effects for applications like TikTok, Instagram, etc. In image morphing, one image is transformed into another image using special effects. Earlier, image or video morphing was considered to be tough and performed by professional or graphic designers, but nowadays anyone can perform morphing using many softwares that are available in the market. Some of the popular image morphing software's are FotoMorph, SqirlzMorph, FreeMorphing, etc. Needless to say that they have a very simple and easy-to-use interface to perform morphing without any prior knowledge or experience. Image segmentation is used as a preprocessing phase to extract necessary and meaningful information for analyzing the objects within an image [5]. Recently, sophisticated

✉ Divakar Yadav
dsy99@rediffmail.com

Shobhit Tyagi
shobhit.tya@gmail.com

[1] National Institute of Technology, Hamirpur, India

Table 1 Estimated number of images captured and stored per year

Year	Photographs taken (in trillions)	Photographs stored (in trillions)
2019	1.4249	6.4583
2020	1.4363	7.3754
2021	1.4397	8.2956
2022	1.5593	9.2944

high-level editing and manipulation have been automated by the developments in computer vision technology. Localizing a particular area of an image and switching them with other images can be achieved with ease in few seconds using desktop softwares [6] and mobile applications like [7]. The advancement of image manipulation techniques can be interpreted in two usually contradictory ways. The advantages include the beautification of images, better artistic and creative methods for graphic designers to visualize the arts of photo-editing. Disadvantages include easier forgery methods for a given image without any traces making it harder to track such perpetrators thus helping forgers to produce and deliver fakes IDs and information. According to a report released by NortonLifeLock in their annual Cyber Safety Insights Report, nearly four out of ten Indians (around 39%) have encountered identity theft [8]. Some common types of forgeries are morphing, copy and move, retouching, swapping, etc. For instance, generative adversarial networks (GANs) proposed by [9] have demonstrated that fake images can be easily generated with high realism. Super-resolution GAN (SRGAN) [10] proposed a method to generated artificial images with super-resolution. Such methods can be used to generate fake identities for malicious purposes or to create fake accounts on social networking sites like Facebook to cheat some innocent person [11]. Using these artificial images, the harm can be extended in the political and commercial affairs such as these doctored images can affect our behavior, childhood memories as well as who we vote [12].

The computer vision algorithms [13] made visual effects that are very hard to distinguish for common people from the real images. The automation of such a method not only makes forgers easily generate realistic images but also promotes the common population to misuse such software. The number of fake images or videos on the internet keeps on increasing day by day. According to a report by Deeptrace Labs, there were 7964 deep fake videos, available on the internet. Nine months later, the number increased to 14,678 and keeps on increasing [14]. It is important to focus and evolve new image tampering techniques for efficiently tracking and detecting tampered images or videos on the Internet. In this paper, we will discuss and inspect some of the popular image and video manipulation techniques. We also analyze their working methodology and datasets used by them which help them to generate high-quality realistic fake images and videos. We first introduce general image manipulation operations used by these methods.

The architecture of this paper is as follows. In Sect. 2, prerequisites for image and video manipulation are discussed, along with common forgery types. In Sect. 3, we discuss the popular image and video tampering datasets. The datasets are also compared based on various parameters. In Sect. 4, we analyze the deep learning-based image forgery detection methods along with traditional detection methods. Section 5 examine video manipulation and detection techniques and compare various techniques based on different parameters. In Sect. 6, we discuss the findings, along with research challenges. Finally, in Sect. 7, we discuss the conclusions based on our observations.

2 Prerequisites for visual imagery manipulation

The term "Visual Imagery" can be used to describe both images and videos. Imagery manipulation (IM) is a process of editing an image or video using some operation via computer software or other digital devices like mobile and tablets. IM is also known as image editing. Nowadays, cameras are substituted by mobiles which can generate high-quality digital photographs and are also equipped with some editing applications like movie, videos, etc. Digital photographs are becoming an integral part of our daily life. From birthdays to marriage ceremonies, photographs are used to remember and store ones personal memory. Image editing applications are used to add some visual effects to images and videos that make them look beautiful in such images. In this paper, the term image manipulation is used instead of digital image editing for clarification purposes. Image manipulation and its types are shown in Fig. 1. The image manipulation can include operations like resampling which is a pixel-level operation to remove and add objects within an image. Table 2

Table 2 Definitions of image manipulation, image forgery, image tampering, and image generation

Terminology	Definition
Image manipulation	21 Editing or altering digital images using computing methods.
Image forgery	Creating fake graphic contents which deliver false information using image manipulation
Image tampering	A specific kind of image forgery in which the manipulator alters several scenes of a given image
Image generation	Generating non-existing images that contain real-life objects using deep learning techniques

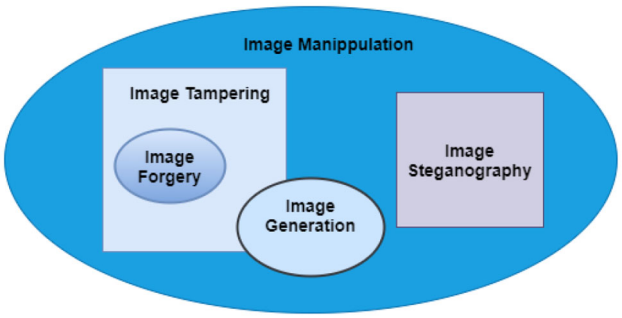

Fig. 1 Image manipulation and its types

shows the differences between image manipulation, image forgery, image tampering, and image generation.

2.1 Image forgery (IF)

It is one of the types of image manipulation that includes delivering false information using graphic content. For example, a forged driving license gives fake information regarding ones driving capabilities. There are six types of image forgery [15], morphing, compositing, retouching, enhancing, computer painting, and generating. The first three methods tamper with the appearance and semantic information of an image. In the image-enhancing approach, the image details are manipulated by changing the color map or contrast. The last two approaches generate a cipher or non-existing image which is also known as a fake image. The generated image can contain any real-life objects or non-existence people without an original image. So, image forgery basically promotes the warped or falsified portion of an image to further promoting false information. The image painting and generating approaches are based on the creativity of the manipulator's experience and need. There are several image forgeries examples apart from Fig. 1 can be found on the Georgia Tech College article named "Photo Tampering Throughout History" [16].

2.2 Image tampering (IT)

Both real life, as well as computer-generated images, can be manipulated and transformed into doctored images. Even though, we already rule out image generation to be a part of an image forgery, manipulating computer-generated images also generates a tampered image. The resulting fake image can be used on social media outlets, to cause catastrophic problems. Generative deep learning algorithms such as variational autoencoders (VAEs) and generative adversarial networks (GANs) have been mainly used to generate a portion of pseudo images or videos. The image translation tasks [17] applied by GANs are hard to detect by the human eye in a limited time. The problems created by GANs are increased when fake videos of famous personalities embedded with speech and facial content surfaced, causing severe problems in social and political activities. Wang et al. [18] defined image tampering as concealing or altering an object within the image. The doctored or tampered image contains newly added objects as well as real objects from the original image. The tampered areas can be easily visually isolated when compared to the original image [19]. However, without the original image as a reference, it would be hard to locate the tampered areas and to decide the intention of the manipulator whether being good or evil. The process of detecting tampered images with the naked eye without an original image as reference is called blind detection. Blind detection can be a tedious task if the manipulator has good creativity and art experience.

Image tampering aims to substitute or modify the content within a particular region of an actual picture with new content. The type of tampering is determined by the content source and composition. Table 3 represents types of image tampering along with their features and uses in tampering. The most common is the copy and paste [20] in which an entire object is added to the original image. This method is generally used for removing undesired objects from the scenes. Similar to that, if new content is cut from one image and then added to an image, it is referred to as cut and paste [21]. The erase and fill [22] is a tiresome process as the content/object being added was scrubbed from various patches of different images or from the original image itself. This method can also be used for object removal by covering it with neighboring textures. The goal of image in-painting is to convert an input image with holes and gaps into a complete image as output. Some examples of image in-painting techniques are context encoder [23], multiscale method [24], consistent completion [25], etc. Both object removal and addition are additive operations. In object removal, a part of the image background or neighboring textures is added to fill the space created while inserting an object is removing a part of the image background. Usually, copy–move is used for object removal; erase and fill are used for object addition. In 2009, Farid [26] categorized image tampering into five different categories based on their detection approaches, i.e., pixel, camera, format, geometric and physically based techniques.

The performance and complexity of a proposed tampering algorithm are based on the dataset containing both authentic and manipulated images. Most of the datasets contain labeled ground truth images compiled by human experts. The general approach to achieve this purpose is to simply mark the images with binary labels. Generally, "**0**" is used for authentic, whereas "**1**" is used for manipulated images. For localizing the tampered regions within an image, it is necessary to provide a black and white mask, showing the region of modified pixels [20,27]. This mask is compared with the

Table 3 Types of image tampering and their use in tampering

Type	Splicing		Inpainting		
	CM	CP	EF	Context encoder	Consistent completion
Visible difference	Yes	Yes	Yes	Yes	Yes
Single Image source	Yes	No	Yes	No	No
Region duplication	Yes	Yes	No	No	Yes
Uses	OR	OA	OR	OR and OA	OR and OA

The (OR) stands for object removal and (OA) for object addition. The (CP), (CM) and (EF) stands for copy and move, cut and paste, and erase and fill, respectively

predicted mask, generated by the algorithm of a test image to assess the performance of the localization task [28,29].

Generally, the dataset is categorized into three sets: training, validation, and test set to generalize the model better. The training set is used to train the model to learn or recognize irregularities in an image using the corresponding mask and ground truth. The validation set is used to validate the model after the training is completed. This step is crucial to check whether the model suffers from overfitting or other machine learning problems. The test set is used to check the accuracy of the model and whether the model generalizes well or not. There is various evaluation criterion to predict the result of the model with the corresponding ground truth, e.g., precision and recall (PR) [27], prediction accuracy (PA) [30] and receiver operating characteristic (ROC) curve [31]. The PA is a standard measure representing the percentage of the correct labeled images. The precision measures the proportion of how many predictions are correct. Mathematically, precision is defined by Eq. 1. Recall measures the proportion of how many true positives are identified correctly. Mathematically, recall is defined by Eq. 2. The T_P, F_N and F_P denotes the true positive, false negative and false positive. If there are no F_P exists in the model, the precision becomes 1.0. Similarly, if there is no F_N exists in the model, the recall becomes 1.0. The recall is also referred to as the sensitivity of a model.

$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{1}$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \tag{2}$$

The ROC curve is a graphical representation of a classification model in different threshold settings. The curve shows the true positive value and the false positive value. The ROC curve helps to evaluate the outcomes of binary classification. The PR values can also be calculated at the different thresholds to plot a PR curve. In contrast to these three, the accuracy of a classification model on a dataset can also be measured by F-Score, commonly known as the F1-score [32]. It classifies the result into "positive" or "negative." The F1-Score combines the precision and recall of a model to simply evaluate binary classification models. The formula of F-score is defined in Eq. 3.

$$\text{F-Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{3}$$

2.3 Image generation (IG)

The process of IG is not considered to be a part of image forgery. The reason is that generating a fake image does not necessarily prove forgery in it. IG is an image creating method which uses machine and deep learning-based techniques to produce an artificial image containing real-world objects and scenes from an existing dataset. There are two types of image generation: (a) **conditional image generation (CIG)** generates the samples conditionally from the dataset, based on label, i.e., $p(y|x)$. Examples of CIG are MIX-MHinge GAN [33], U-net GAN [34]. (b) **unconditional image generation (UIG)** which refers to the process of generating samples unconditionally form the dataset, i.e., $p(y)$. Examples of UIG are StyleGAN [35], StyleGAN2 [36], Fine-GAN [37].

2.4 Image warping and morphing

The process of manipulating a digital image such that the shape of the objects present in the image is/are transformed or distorted. Warping is the geometric deformation of a single object in a given image. Warping can be used for correcting image distortion as well as for creative purposes. On the other hand, image morphing interpolates two or more graphical objects. It means morphing is a combination of image warping and blending techniques to interpolate objects to create a novel object. The goal is to take an input image and gradually distort it while generating the target image. In practice, morphing is commonly used in the entertainment industry for instance, in movies, animations, and TV shows (Fig. 2).

3 Image and video manipulation datasets

This section investigates the popular image and video tampering datasets according to their year of publications.

Fig. 2 Image morphing

3.1 Image tampering datasets

Table 4 represents the most relevant image tampering datasets, used for forgery detection. Furthermore, these datasets are discussed and evaluated based on various parameters such as tampering type involved, image size, format, and mask availability.

3.1.1 The Columbia gray dataset

It is one of the earliest image datasets for analyzing image tampering that was made publicly available online. This image splicing detection dataset[1] was created by DVMM laboratory of Columbia University in 2004 [38]. It has 933 authentic and 912 spliced images with a block size of 128×128 pixels and is saved in BMP file format. The image blocks are extracted from the CalPhotos image dataset.[2] This dataset uses copy–paste (CP) tampering to generate tampered images. The two types of operations used are (a) CP along object boundaries, (b) CP of horizontal (or vertical) strips without any post-processing using Adobe Photoshop [8]. The authentic and spliced categories are divided into five subcategories, which are further divided into three subsubcategories, according to the orientation of the object boundary. The limitation of this dataset is that it only contains grayscale (i.e., black and white) images, and also tampering mask is not available so it cannot be used for localization tasks.

3.1.2 The Columbia color dataset

The Columbia gray dataset does not provide color images; to overcome this limitation, the Columbia community launched another dataset named Columbia Color in 2006 [39]. There are 183 original color and 180 cut and paste images in this dataset.[3] The images are available in high resolution and uncompressed. The exchangeable image file format (EXIF) information is also retained in authentic images to make the dataset not only suitable for splicing detection but other computer vision algorithms as well, since the ground truth settings are accessible. The image sizes are between 757×568 to 1152×768 in either BMP or TIFF formats. Four digital cameras were used to capture the original images, whereas Adobe Photoshop was used to produce the tampered images [2] from the authentic images with no post-processing. To check the tampered region boundaries, edge masks are also provided.

Both the Columbia datasets are not effective as the *cut* operation was applied randomly making the tampering too obvious in the pasted regions. Even though most of the images in the Columbia dataset can be easily identified by human ability, both Columbia datasets provide an important benchmark for evaluating forgery detection algorithms.

3.1.3 The CASIA datasets

In 2009, to fulfill the demand for large and realistic tampered image tampering datasets, the CASIA team [40] launched two datasets[4] to the public. There are two versions of CASIA datasets named v1.0 and v2.0. The datasets contain both copy–paste and cut-paste images, unlike the Columbia datasets, which only include cut-paste images. Post-processing is not included in v1.0, but it was added in v2.0 to improve the realism of tampered images.

There are 1721 color images in the v1.0 dataset, each with a fixed image size of 384×256 pixels. There are 800 genuine images and 921 tampered images among the 1721 total images. The images were all saved in JPEG format. The CASIA v2.0 dataset contains 12,614 color images with variable image sizes ranges from 240×160 to 900×600. Crop and paste operations were used on authentic images to create the tampered images using Photoshop software [2].

Recently, another version named modified CASIA dataset[5] is introduced by Zheng et al. [41] in 2020. This dataset includes 400 real and 400 tampered images from eight different categories. It also includes 3600 images created by applying a single content-preserving manipulation to an authentic image and 3600 tampered images created by applying the same manipulation to tampered images. There are a total of nine manipulations used to generate such images, e.g., speckle noise, Gaussian noise, salt & pepper noise, Gaussian filter, JPEG compression, motion blur, gamma correction, scaling, and rotation. An example from CASIA [40] dataset is shown in Fig. 3.

[1] https://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/AuthSplicedDataSet.html.

[2] http://elib.cs.berkeley.edu/photos/use.html.

[3] https://www.ee.columbia.edu/ln/dvmm/downloads/authsplcuncmp/.

[4] https://github.com/namtpham/casia1groundtruth.

[5] https://ieee-dataport.org/open-access/modified-casia#files.

Table 4 Popular image tampering datasets used in image detection

Dataset	Year	Tampering type	#Authentic / #Manipulated images	Image size	Format	Mask availability	Post-processing	Color
Columbia Grey	2004	CP	933/912	128×128	BMP	No	No	No
Columbia Color	2006	CP	183/180	757×568, 1152×768	TIFF	Yes	No	No
CASIA v 2.0	2009	CP and CM	7491/5123	240×160, 900×600	TIFF, JPEG	No	Yes	Yes
MICC-F2000	2011	CM	1300/700	2048×1536	JPEG	No	No	Yes
DRESDEN	2011	CM	14000/0	1024×1024	JPEG	No	No	Yes
BOSS [49]	2011	Steganalysis	800	2000×3008, 5212×3468	JPEG	No	Yes	Yes
IMD	2012	CM	48/48	3000×2300	JPEG, PNG	Yes	Optional	Yes
Manip	2012	CM	48/48	2305×3020	JPEG	Yes	No	Yes
MICC-F600	2013	CM	440/160	800×553, 3888×2592	JPEG, PNG	Yes	Yes	Yes
CoMoFoD	2013	CM	5200/5200	512×512	JPEG, PNG	Yes	Yes	Yes
IEEE IFS-TC	2013	CM and Splicing	1050/450	1024×768	PNG	Yes	No	Yes
Wild Web	2015	CP, CM and EF	0/10646	Multiple	Multiple	Yes	Yes	Yes
SCUT-FBP [50]	2015	–	500	384×512	Multiple	No	No	Yes
COVERAGE	2016	CM	100/100	Multiple	TIFF	Yes	No	Yes
Realistic Tampering	2016	CM	0/220	1920×1080	TIFF	Yes	No	Yes
FaceForensics++	2019	Face2Face, DeepFake, etc.	0/1.8 Million	Multiple	Multiple	Yes	Yes	Yes

The CP, CM, and EF stand for copy–paste, copy–move, and erase–fill tampering. The BMP, TIFF, JPEG, and PNG stand for BitMap, Tag Image File Format, Joint Photographic Experts Group, and Portable Network Graphics

Fig. 3 An example of CASIA dataset [40], the original Image **A** is on the left and its spliced image **B** is on the right

Fig. 4 An example of MICC dataset [42], Images **A** and **C** are original and their corresponding spliced Images are **B** and **D**

3.1.4 The MICC datasets

Two datasets[6] named MICC-F220 and MICC-F2000 were released by MICC team in their paper on copy–move detection and localization [42]. As the name suggests, the F220 is an image copy–move manipulation dataset that contains 220 color images among which half are authentic and half tampered images. The image size ranges from 722 × 480 to 800 × 600 pixels. The size of the tampered region represents 1.2% of the whole image area. In contrast, the F-2000 dataset contains 2000 JPEG images with an image size of 2048 × 1536. There are 1300 authentic and 700 tampered images, respectively. The tampered region represents 1.12% of the whole image area. Similar to Columbia dataset, the MICC datasets take no effort in selecting realistic tampering regions. The tampered regions did not accord well with neighboring regions making localization of tampered regions easier to detect.

Later in 2013, Amerini et al. [43] published another dataset named MICC-F600. To improve the tampering effect, some images are post-processed. The F600 dataset contains 600 images of which 440 are authentic and 160 tampered images. The image size varies from 800 × 553 to 3888 × 2592 and is available in either JPEG or PNG format. For better localization, the ground truth of tampered images was also given. The sizes of the tampered region vary from one image to another. An example from MICC [42] dataset is shown in Fig. 4.

3.1.5 The DRESDEN image dataset

Gloe et al. [44] introduces a novel image database for the development and testing of camera-based digital forensic techniques. This dataset contains over 14 k images taken by 73 digital cameras under controlled conditions from various indoor and outdoor scenes. The "Dresden Image Database" was made publicly available in 2011 for researchers and forensic investigators. The images are of sizes 1024 × 1024 pixels and are available in JPEG format. The dataset only contains real images, and it is up to the users to synthesize their own tampered images from these original images. For example, [45] uses IEEE forensics Challenge [46] and DRESDEN [44] datasets to synthesize a new dataset of 65k images for image tamper localization task. Examples from the DRESDEN [44] dataset are shown in Fig. 5.

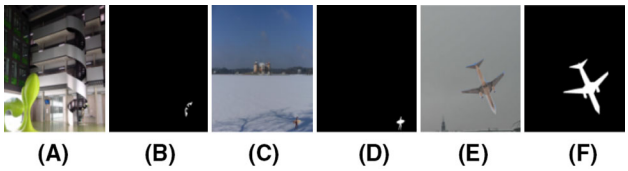

Fig. 5 An example of DRESDEN image dataset [44], images **A**, **C**, and **D** are original images, and images **B**, **D**, and **F** are their corresponding ground truth masks

3.1.6 The IMD dataset

The IMD dataset[7] was released in 2012 to study and understand the visually imperceptible copy–move tampering [20]. Unlike the other datasets mentioned above, the IMD dataset used skilled graphic experts (i.e., human photo-shoppers) for manipulating images to semantically select tampered regions for a better tampering effect. These meaningful tampered regions are called "snippets" making the spliced images harder to detect for human ability than previous datasets.

The image manipulation dataset (IMD) contains a total of 48 image pairs, i.e., the authentic and their spliced versions. The image size is around 3000 × 2300 pixels. The dataset is available in two formats; uncompressed images are saved in PNG, while compressed images are saved in JPEG format. Apart from post-processing, other optional operations like JPEG compression and Gaussian noise are available for use. The software used for generating tampered images was also released and can be used to generate many other user-designed spliced images.

3.1.7 The CoFoMo dataset

The CoFoMo dataset[8] was released in 2013 for a better understanding of snippets transformations and manipulation operations to forgery detection algorithms [47]. Translation, rotation, scaling, combination, and distortion are the five

[6] http://lci.micc.unifi.it/labd/2015/01/copy-move-forgery-detection-and-localization/.

[7] https://www5.cs.fau.de/research/data/image-manipulation/.

[8] https://www.vcl.fer.hr/comofod/.

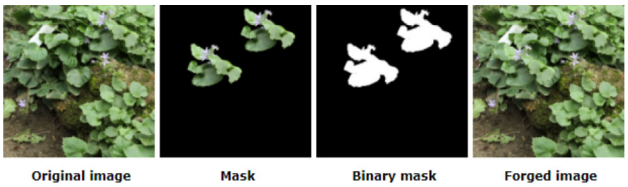

Fig. 6 An example of CoMoFo forgery detection dataset [47]

types of manipulations applied to the images. Furthermore, all forged and initial images are subjected to six different forms of post-processing techniques, including JPEG compression, noise addition, blurring, color reduction, contrast changes, and brightness improvement. The dataset contains 200 original images. The 200 tampered images were generated from original images by applying one snippet transformation along with a copy–move operation. Later, 25 post-processing operations were applied to each of the 200 original and 200 tampered images to generate 5200 original and 5200 manipulated images. The image is of size 512×512 and are available in both JPEG and PNG formats. Some examples from the CoFoMo [47] dataset are shown in Fig. 6.

3.1.8 The IEEE IFS-TC dataset

The IEEE Information Forensics and Security Technical Committee (IFS-TC) organized an international Image Forensics Challenge[9] and released a public open-source dataset [46] in 2013. The objectives were (a) to support and provide an open-source dataset for tampering detection, (b) create a specific set of rules and protocols for evaluating recent forensics techniques that can be used to identify different types of forgeries in a digital image, and (c) to evaluate other state-of-the-art detection techniques and provide a standardization protocol as common ground truth for new techniques comparison. The dataset contains a total of 1500 images among which 1050 are authentic and 450 tampered images. The images size is 1028×768 pixels and is available in PNG format with no post-processing operations.

The images are categorized into two groups: "pristine" or "never manipulated" and "forged" or "fakes" images. Training and testing are provided for both pristine and fake images. Only the training set contains the corresponding class and masks. Some examples from the IEEE [46] dataset are shown in Fig. 7.

3.1.9 The wild web dataset

As the name suggests, all the images in the Wild Web dataset [48] are collected from the Web (i.e., Internet) making them harder for tampering localization because the tampered

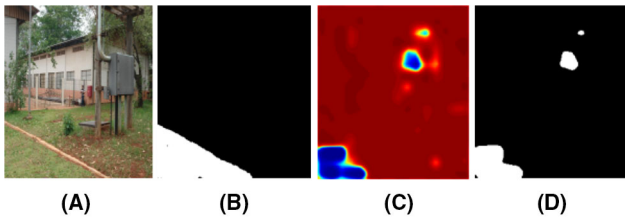

Fig. 7 An example of IEEE forensics challenge [46] dataset, Images **A** and **B** represent the input images and ground-truth masks for manipulated regions. Images **C** and **D** represent the probability heat-map and predicted binary masks

images had gone through many post-processing operations than in tampered mentioned above. In contrast to other publicly available datasets, the wild web can be used for research purposes[10] upon request from the authors. This dataset includes a large number of forgeries gathered from multiple online and social media outlets. The dataset contains 13,577 unique fake images in multiple formats. The dataset contains 80 cases of forgeries such as cut–paste, copy–move, and erase–fill, confirmed by multiple sources and experts photographers. Due to the lack of original images, this dataset is not suitable for forgery detection but can be used by researchers by adding their authentic images to the dataset. Few examples from the wild web [48] dataset are shown in Fig. 5.

3.1.10 Retouching forgery datasets

In 2011, Bas et al. [49] introduced the first international challenge on Steganalysis. The authors explain the BOSS (Break Our Steganographic System) dataset used for the challenge and also presented several results submitted to the competition. The main challenges of the dataset are discrepancies in the training and testing sources of images. The authors called this problem the cover-source mismatch which makes the designer's job difficult as now they had to design their steganalyzers specific to the source of images. The datasets contain a total of 800 images in different resolutions which are unaltered in raw format. A similar type of dataset named SCUT-FBP was introduced by Xie et al. [50] in 2015. The dataset contains 500 images of Asian females along with their corresponding score based on attractiveness. The ratings were given by 70 different observers which have been verified by rating distribution, self-consistency, and standard deviation. The images have a resolution of 384×512 and a CNN model achieved the best Pearson correlation 0.8187. The results also indicated by the authors shows that the dataset provides a reliable benchmark for facial beauty perception (FBP).

[9] https://signalprocessingsociety.org/newsletter/2013/06/ifs-tc-image-forensics-challenge.

[10] https://mklab.iti.gr/results/the-wild-web-tampered-image-dataset/.

3.1.11 Other dataset resources

Many popular forensics Challenges and competitions also released image datasets online for forgery detection. Any forensics expert or interested researchers group can use the datasets to evaluate their algorithms on the test data and participate in such challenges. Generally, the dataset is divided into two subsets of data named, testing and test dataset. The ground truth labels of training dataset images (i.e., authentic and tampered) are also available for better predictions and tampering effects. The ground truth labels of the test dataset were hidden by the organizing committees. The participants can submit their findings (i.e., predicted labels and masks) through a submission system and the organizers publish the report based on their prediction results.

The Open Media Forensics challenge[11] evaluation (OpenMFC) is organized by NIST that release an image and video tampering dataset to assess and measure the capability of media forensic algorithms and systems. The objective of OpenMFC is to help researchers to study media forensics and help advance the state-of-the-art media forensics algorithms. OpenMFC focuses on the performance measurements of the automated imagery (image and video) manipulation detection and localization technologies. It also supports GAN manipulation detection tasks. Every year, the challenge was organized by NIST in collaboration with Defense Advanced Research Projects Agency (DARPA) to evaluate current algorithms and techniques for media forensics in which any interested group can participate and submit their findings on their official website[12]. The COVERAGE [51] dataset[13] is another copy–move dataset which contains 100 original images and 100 manipulated images. The images are saved in TIFF format.

Other image forensics datasets except NIST are RAISE[14] and Uncompressed Colour Image Database (UCID) dataset.[15] The raw images dataset (RAISE) contains 8156 high-resolution raw and uncompressed images. All images are unprocessed and camera-native. The images are captured by four photographers in different indoor and outdoor places in Europe. The UCID [52] dataset has more than 1300 images with their ground truth images. The objective of the authors was that the dataset be used for evaluating image retrieval techniques.

[11] https://www.nist.gov/itl/iad/mig/open-media-forensics-challenge.

[12] https://mfc.nist.gov.

[13] https://github.com/wenbihan/coverage.

[14] http://loki.disi.unitn.it/RAISE/index.php.

[15] https://qualinet.github.io/databases/image/uncompressed_colour_image_database_ucid/.

3.2 Video tampering datasets

Video forgery also plays a crucial role in media forensics and has even more severe implications for society than fake images. The recent progress of synthetic video generation and manipulation can cause significant problems like spreading false information, loss of trust in digital content, and fake news. In this section, we discuss some popular video manipulation and tampering datasets used to generate synthetic videos and how hard is to detect such videos, either by using deep learning techniques or by humans. The most relevant DeepFake video datasets are discussed and compared in Table 5.

3.2.1 First-generation datasets

UADFV [54] dataset contains 49 real as well Deepfake videos generated using FakeApp [55]. Later, Korshunova et al. [56] introduced a dataset named **DeepFake-TIMIT**, which contains 640 fake videos generated using faceswap-GAN.[16] Rossler et al. [57] introduced **Faceforensics** which is one of the largest facial forgeries dataset containing roughly 1.8 million manipulated images. The tampered images are created with four state-of-the-art methods, namely Face2Face [58], DeepFake [59], FaceSwap [60], and NeuralTextures [61]. The aim is to help researchers to use deep learning approaches in forgery detection. Apart from the novel dataset, the authors also provide a benchmark for facial manipulation detection. The dataset[17] is easily accessible after filling a google form for research purposes.

3.2.2 Second-generation datasets

The databases consisting of second generation are released in late 2019. These datasets goals are to provide high-quality synthetic videos that are available on the Internet. [18] **Deepfake Detection** [62] dataset contains around 3K videos generated from 28 consented individuals from different ages, sex, and ethnic groups. The DFD dataset is available in three levels of video quality: (i) Original quality (RAW), (ii) Low quality (LQ), and (iii) High quality (HQ). The **DFDC** [53] dataset is from the Facebook Deepfake detection challenge which contains around 4K fake videos generated from 66 consented individuals of various ages, sex, and ethnic groups. In 2020, Jiang et al. [63] introduced a real-world face forgery detection dataset. **The Deeper Forensics-1.0** dataset[19] contains both images and videos from 100 paid actors giving

[16] https://github.com/shaoanlu/faceswap-GAN.

[17] https://github.com/ondyari/FaceForensics.

[18] https://www.youtube.com/channel/UCKpH0CKltc73e4wh0_pgL3g.

[19] GitHub: https://github.com/EndlessSora/DeeperForensics-1.0.

Table 5 DeepFake video datasets

Database, year and authors	Mask-SSIM score	#Real Videos/Frames (Source)	#Fake Videos/Frames (Source)	Generation method
First generation datasets				
UADFV (2018) [54]	0.82	49/17.3K (Youtube)	49/17.3K (FakeApp)	Videos are generated using DNN model for Identity Swap
Deepfake TIMIT-HQ (2018) [56]	0.80	320/34.0K	320/34.0K (FaceSwap-GAN)	Uses Faceswap-GAN and vid -TIMIT dataset for video synthesis
FaceForensics++ (2019) [57]	0.81	1000/509.9K (Youtube)	1000/509.9K (FaceSwap) and 1000/509.9K (Deepfake)	Uses faceswap to synthesize deepfake videos
Second generation datasets				
DeepFake Detection (2019) [62]	0.88	363/315.4K (Actors)	3068/2242.7K (DeepFake) from 28 consented individuals	Method of generation is not disclosed
Celeb-DF (2019) [107]	0.92	590/225.4K (Youtube)	5639/2116.8K (DeepFake) from 59 Celebrities in 5K	Celebrities videos generated by improved synthetic process
DFDC Preview (2019) [53]	0.84	1,131/488.4K (Youtube)	4119/2116.8K (Unknown)	Uses two facial modification algorithms for video generation but the synthesis algorithms are not disclosed
DeeperForensics-1.0 (2020) [63]	–	50K (Collected by authors)	10K (Face Swap) from 100 consented actors	Fake Videos are generated using a new end-to-end face swapping framework

The generations are based on the release time and synthesis algorithm used for synthesizing videos. However, the second-generation datasets are bigger and have better quality videos as compared to the first-generation datasets

consent to use and manipulate their faces. It also includes a hidden test set of tampered videos that consistently improve deceptive scores in human evaluations. A total of 35 perturbations are employed on the dataset to create a more challenging and high diversity dataset. The dataset contains up to 60k high-quality videos of which 50k are original and 10k are fake.

4 Image manipulation and detection

Before going further, we need to understand some state-of-the-art visual manipulations methods that have emerged in recent years. The imagery (i.e., image and video) content has become one of the most relevant topics in digital forensics.

4.1 Image manipulation methods

4.1.1 Graphics-based methods

In this subsection, we discuss two popular graphics-based methods for facial manipulation named *FaceSwap and Face2Face*. The FaceSwap [60] approach transfers the face region from the source to a target video. The face regions from the source videos are extracted using sparse detected facial landmarks. These landmarks are used by the FaceSwap algorithms to generate a 3D template model which fits the face to the projected target video. To minimize the difference between the target shape and the localized landmarks, the textures of input images are used. Then, the rendered model is mixed in with the desired video flow. Finally, the color cor-

rections are applied to make the synthetic video look more natural.

The Face2Face [58] approach re-enacts the source video facial expressions in the target video. Needless to say, the identity of the target person remains unchanged. This method uses a dense reconstruction of target facial expressions under different illuminations and expressions. The objective is to use a source video to animate the target video's facial expressions and produce a photo-realistic manipulated output video. The original paper contains information about the re-enactment and illumination process [58].

4.1.2 Learning-based methods

The learning-based methods use deep learning techniques to replace face or other scenes of a source to target video. The popular learning-based methods are *DeepFakes and NaturalTextures*. The term DeepFakes is commonly used as a synonym for face replacement in images and videos by the general public or digital media networks (i.e., News). To differentiate the term from the published paper, we denote the published approach by DeepFakes [59] in the following paper.

The DeepFakes [59] method replaces a face from the source image sequence collection to a target sequence. There are several real-world uses of DeepFakes available such as FakeApp [55] and the Faceswap[20] Github. Faceswap is a software that utilizes deep learning to recognize and swap faces in images and videos. A shared encoder along with two autoencoders is required in this method. These encoders use source and target videos to reconstruct images. Faces are then cropped and aligned from such images using a face detector. Finally, a trained encoder and decoder are used to generate a fake image by replacing the source with the target face.

Theis et al. [61] introduce another deep learning approach for facial re-enactment, named as NaturalTextures. This method uses original video and photometric reconstruction in combination with adversarial loss for training to learn natural textures data of the target person. This method can be used in texture synthesis and image in-painting as well.

4.2 Image forgery detection methods

Image manipulation is a set of one or more operations that alters the authenticity of digital images using any software or mobile app. The process of manipulating the content stored in an image to spread false information is known as image forgery. Image tampering is a process in which some of the content of an image is replaced with false information to misguide someone (Fig. 8).

[20] https://github.com/deepfakes/faceswap.

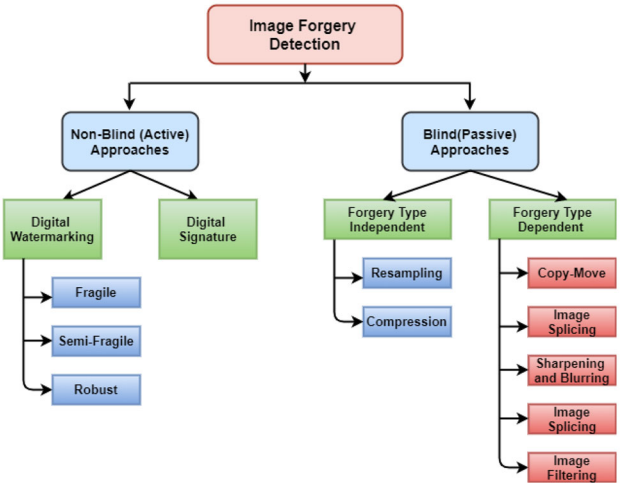

Fig. 8 Categorization of image forgery detection approaches

The IFD methods can be divided into two categories: (a) **Active** and, (b) **Passive**. In the active approach, during the image acquisition stage, additional information is added to the image. This information was later authorized by some person to detect the manipulation. If the additional information was found incorrect, the image can be identified as tampered image; otherwise, the image is authentic.

The passive approach does not depend on any additional data to detect a forgery in an image. Instead of additional information, these approaches extract features from an image to detect forgery. Hence, these methods are also known as "**Blind Approaches**." The passive approaches are subdivided into two categories: (a) **Forgery type independent** which is used to detect forgeries such as sampling and compression, and (b) **Forgery type dependent** which is used to detect specific types of copy–move and splicing forgeries in an image.

4.2.1 Traditional methods

Copy–Move (CM) Detection Methods In this tampering, a scene or object of any size is copied from one image and pasted into a different portion of the same image. CM tampering aims to detect the presence of duplicated regions in a given image. The image is classified as authentic if no such regions are found; otherwise, the image is classified as tampered image. To locate similar regions, it is essential to divide the image into smaller parts such as blocks or extract features using keypoints [64]. The keypoints are identified in the entire image, and then features are extracted from each key point.

The CM detection is divided into two groups by Christlein et al. [20]: (a) block-based [65], (b) keypoint-based [66]. By comparing features extracted from each block or keypoint using a certain algorithm, it is easy to find matched pairs.

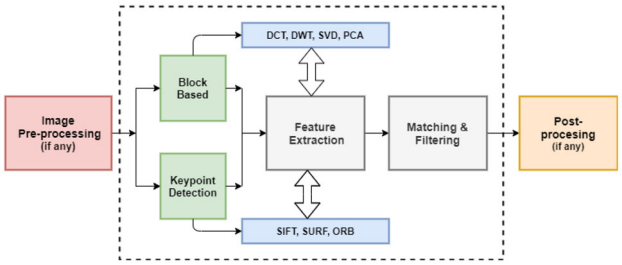

Fig. 9 A flowchart for identifying copy–move tampering in images

If the matched pairs are found in two separate regions, the image is considered tampered. Both methods assume that the tampered area is wide enough to include several blocks or key points and that the duplicated regions contain some matched components. A flowchart for identifying copy–move tampering in images is given in Fig. 9. The dashed rectangle shows the important components of detection. The features are extracted either from block or keypoint-based methods, and identical feature points are matched and filtered as pairs. Finally, if the two different regions are densely matched pairs, they are classified as duplicates. The process of feature extraction is different for both approaches. For block method, one can use DCT [65], DWT [67], SVD [68], and PCA [69], etc., while the keypoint method uses SIFT [70] [71], SURF [70], and ORB [72] etc., to extract features from an image. The use of post-processing and preprocessing operations to improve detection and localization efficiency is optional.

Image Splicing (IS): IS is the process of compositing two or more images to generate a spliced image. In contrast to copy–move tampering, there is no region duplication in image splicing. This makes the localization task even harder to detect as compared to copy–move. To detect tampering in image splicing, one has to check for the clues that are left behind after the tampering process has been completed. Some common clues are inconsistency in lighting, discontinuity in edges, geometric situation, and camera. Edge discontinuity occurs when a copied area in an image leaves an unnatural irregularity on the sides of the manipulated region [73]. Another clue rises from the camera using which the image is taken. Since various cameras have different properties, the images taken are often different, and tampering can be proven using this evidence. The lighting conditions of the two photographs, for example, are not the same. There must be a slight difference in optics between the two images. There may be some lighting differences between the tampered region and the rest of the image, which has consistent lighting. Johnson and Farid [26,74] proposed a method that computes the lighting path of different scenes in an image. If there is a conflict between the measured paths, it indicates the existence of tampering. The different formats of images in which they are stored also have different compression and properties. The image data in JPEG images are lossy and compressed. Despite the significant file size reduction, JPEG images have reasonable image quality. When a forger creates a tampered JPEG image from a genuine JPEG, the forged JPEG image has double quantization (DQ) effect due to two compression operations. Popescu et al. [75] proposed a tool to identify the DQ effect developed from the double JPEG compression. The existence of a DQ effect in an image indicates that it has been tampered with.

4.2.2 Deep learning-based approaches

The earlier approaches (both block and keypoint based) uses crafted features (such as DCT [65], DWT [67], PCA [69], SIFT [70], SURF [70], etc.) for forgery detection. The advancement of deep learning (DL) techniques in computer graphics fields motivated researchers to use DL models for image forgery detection. These models can learn abstract as well as complex features from the tampered image given a sufficient amount of data. DL combines phases like feature extraction and classification of tampered regions which saves time and improves the efficiency needed to find hand-crafted features from manipulated images. However, training a deep neural network (DNN) is difficult and requires a large amount of data and high computational power. For computer vision purposes, CNN is popular against the other DL models. The convolutional layers are capable of creating the feature map by applying filters to a given image (input). These feature maps are responsible for detecting features in the input. A comparison of various deep learning-based forgery detection methods is shown in Table 6.

5 Video manipulation and detection

Video manipulation aims to manipulate the content of digital video using video editing techniques and video processing. The purpose of these approaches ranges from well such as entertainment purposes, movies, and educational lectures to bad such as fake news, photo-manipulation, and DeepFakes. Figure 10 shows various types of video manipulation.

5.1 Types of video manipulation

With the increasing demand for digital content, images and videos are becoming an integral part of our day-to-day lives. Video manipulation can be categorized into three types.

5.1.1 Missing context

In this manipulation, a part of the original video is pasted into another video to misguide the audience. The context of the video is changed so that it presents a different narrative than

Table 6 Comparison of deep-learning-based approaches for image tampering detection in chronological order

Authors and year	Tampering type	Features	# Layers	Model	Datasets	Accuracy
Chen et al. [108] (2015)	Median filtering, CP	Median filter Residuals	9 (1 FL, 5 CL, 3 FCL)	CNN	DRESDEN, UCID	85.14%
Zhang et al. [109] (2016)	CM and CP	2D Wavelet decomposition	–	SAE	CASIA and Columbia	91.09%
Bayar et al. [110] (2016)	Gaussian blurring, resampling, filtering	Prediction error filters	8 (1 new proposed CL, 2 CL,2 MPL, 3 FCL)	CNN	Images collected by authors	99.1%
Cozzolino et al. [111] (2016)	CP	Noise residual filters	–	Auto-encoder	Images collected from several devices	F-measure 0.41(basic) 0.37(with post-proc.)
Rao et al. [112] (2016)	CM and CP	Hierarchical representation from color images	10 (8 CL, 2 pooling layers, and 1 FCL)	CNN	CASIA, Columbia gray, DVMM	98.04% (CASIA)
Amerini et al. [113] (2017)	JPEG compression, CP	RGB features and DCT histogram	–	Multi-domain CNN	UCID	95%
Bondi et al. [114] (2017)	CP	Camera model features	11 (4 CL, 3 MPL, 2 FCL, 1 ReLU and 1 Softmax)	CNN	DRESDEN	Detection-8% Localization-82%
Salloum et al. [115] (2018)	Image Splicing	Surface and Edge probability map	–	MCFN	CASIA, Columbia, & Carvalho	F1-Score 0.61 (Columbia)
Wu et al. [116] (2018)	CM	Features from VGG16	16	BusterNet DNN	CASIA and CoFoMo	–
Bi et al. [117] (2018)	CP	Image residuals	–	Ringed residual U-Net (RRU-Net)	CASIA, Columbia	Accuracy-76% F-measure-0.91
Wang et al. [118] (2019)	CM and CP	Residual CNN ResNet-101	101	Mask-RCNN	Cover, Columbia	AP-93% (Cover), 97% (Colum.)

Table 6 continued

Authors and year	Tampering type	Features	# Layers	Model	Datasets	Accuracy
Wang et al. [119] (2019)	Photoshopped Images	DesNet for local prediction	Same as in ResNet	Dilated ResNet (DRN-C-26) [124]	Open Images, Flickr	AP-50.0% (Face-Forensics++), 2AFC-98
Gul et al. [120] (2019)	Copy–move Detection and localization	256-dimensional feature vectors and PCA	8 (5 CL fol- lowed by MPL and 3 FCL)	AlexNet CNN	GRIP	F-measure 0.93
Kuznetsov et al. [121] (2019)	Splicing	Border distortions	16	VGG-16 CNN	CASIA v2	96.4%–97.8%
Marra et al. [122] (2020)	Localized forgeries	CNN-based network	–	Resnet-101 with Xcep-tionNet as feature extractor	UCID, DRESDEN, MGC2019, etc.	AUC-0.846
Nejad et al. [123] (2021)	Image mosaicing (Image stitching)	Blending algorithms based on Gaussian-weighted function	–	Clustered Redundanat keypoint elimination	Tree, Wall images, bough, building and waterfall images	MEE-3.72 RMSE-4.46 MAE-6.12

The FL, CL, MPL, and FCL stand for filtering layers, convolutional layers, max-pooling layers, and fully connected layers. The CM and CP represent the copy–move and copy–paste tampering. The P, R, AP, AFC, MEE, MAE, and RMSE represent the precision, repeatability, average precision, alternative force choice, median error, maximum error and root-mean-square error

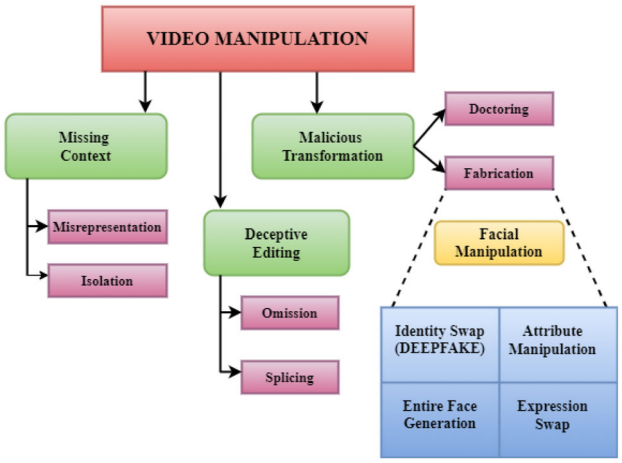

Fig. 10 Different types of video manipulation

the original video. The goal is to misrepresent the original video to mislead others.

5.1.2 Deceptive editing

It is used for deceiving the general public by removing a part of the original video so that it presents a different story. In this type, a new video can also be generated using different pieces of different videos and combining these segments to form a video not created by the original individual.

5.1.3 Malicious transformation

This is the most dangerous types of video manipulation. In this type, doctored and fabricated videos are generated using computer software and video editing techniques. The doctored videos are the ones that are directly modified frame by frame to create a different video. On the other hand, the fabricated video (or Deepfakes) are bogus/artificial videos created using computer software such as Photoshop and Face2Face to harm an individual or a group.

5.2 Video forgery detection methods

Video manipulation can be of different kinds depending upon the need and requirement of the manipulator. In this paper, we only focus on manipulations related to the face (also known as facial Manipulations). Facial manipulations can be categorized into four types based on the level and extent of the manipulation: (a) Artificial face generation, (b) Identity swap (also known as Deepfakes), (c) Attribute manipulation, and (d) Expression swap. The process of face generation cannot be considered as a forgery because the manipulation is done on image level and the generated face is random with no resemblance to real people. Therefore, we do not discuss this manipulation but other three facial manipulations in this paper. The recent advancement of deep learning methods helps researchers to develop novel methods to detect forgeries in videos. The most common type of forgeries in videos is facial manipulations. Table 7 represents various state-of-the-art facial manipulation detection approaches published in recent years.

5.2.1 Identity swap

It is one of the most popular facial manipulation areas of interest in recent years. This manipulation consists of replacing the face of one person in a video with another person using computer software. These videos are commonly known as DeepFakes. The aim is to generate realistic synthetic videos. For detecting identity swap, Matern et al. [76] proposed a detection system based on visual features. These features consist of eyes, teeth, nose, missing eye details, and reflections. Classifiers used in the systems are (i) the logistic regression, and (ii) a multilayer perceptron (MLP) [77]. The performance is calculated to be 85.1% AUC using a private database. In this line, Yang et al. [78] proposed another fake detection model based on facial expressions and head movements. The authors extracted 68 different facial features from DeepFakes videos and used them to normalize their model. At the final stage, an SVM is used for classification. The best AUC achieved is 89.0% on UADFV database. Another method is proposed by Jung et al. [79] in which an algorithm called DeepVision is used to detect and analyze blinking patterns in videos. The blinking features are counted and used to detect whether the video is artificial or not. The approach obtained a final 87.5% accuracy on confidential databases. The deep-learning-based approach of 3DCNN is proposed by Wang et al. [80] which uses spatial and motion information for forgery detection. The authors used a combination of CNN and RNN models to obtain promising results on the FaceForensics++ dataset. Finally, facial regions features are used by Tolosana et al. [81]. The authors used XceptionNet to create a fake detection model. Datasets from both generations were used in experimentation. The AUC achieved by the model in first- and second-generation datasets is 83.6% and 91.0%. The literature given in Table 7 shows poor generalization performance on wild or unseen datasets.

5.2.2 Attribute manipulation

In this manipulation, the modification is done on some specific attributes of the face. These attributes are color of skin, eye, and hair, gender, age, adding glasses, etc. To detect attribute manipulation, Wang et al. [82] proposed a method based on neuron behavior. The authors capture neuron activation patterns to detect facial manipulations. An SVM is used for classification and their approach is named FakeSpotter. Later, Marra et al. [83] proposed their approach based on

Table 7 Comparison of different identity swap and attribute manipulation detection approaches

Author and year	Approach	Classifiers	Databases	Best performance
Identity swap				
Matern et al. [76] (2019)	Visual features	Logistic	Own	AUC-85.1%
		Regression	UADFV	AUC-70.2%
		MLP	Deepfake-TIMIT (HQ)	AUC-77.3%
			Celeb-DF	AUC-55.1%
Yang et al. [78] (2019)	Head pose features	SVM	UADFV	AUC-89.0%
			Deepfake-TIMIT (HQ)	AUC-53.2%
			Celeb-DF	AUC-54.6%
Jung et al. [79] (2020)	Eye blinking	Distance	Created by authors	Acc.-96.%
Wang et al. [80] (2020)	Deep learning features	3DCNN	FaceF++ (LQ)	TCR-95.13%
			FaceF++ (HQ)	TCR-95.13%
Tolosana et al. [81] (2020)	Facial regions features	CNN	UADFV	AUC-100.0%
			DFDC preview	AUC-91.0%
			Celeb-DF	AUC-83.6%
Attribute manipulation				
Wang et al. [82] (2019)	GAN-pipeline features	SVM	Created by authors from (StyleGAN and others)	Acc.-84.7%
Marra et al. [83] (2019)	Deep learning features	CNN	Created by authors from (Glow/StarGAN)	Acc.-99.3%
Zhang et al. [84] (2019)	Spectrum domain features	GAN discriminaor	Created by authors from (Star-GAN/CycleGAN)	Acc.-100%
Pathgeb et al. [85] (2020)	Photoresponse non-uniformity features	Score level fusion	Created by authors from several apps	EER-13.7%
Expression swap				
Nguyen et al. [86] (2019)	Deep learning features	Autoencoder	FaceF++ (LQ)	EER-7.1%
			FaceF++ (HQ)	EER-7.8%
Amerini et al. [87] (2019)	Image and Temporal features	CNN + Optical flow	FaceF++	Acc.-81.6%
Sabir et al. [89] (2019)	Image and Temporal features	CNN + RNN	FaceF++ (Face2Face, LQ)	Acc.-94.3%
Dang et al. [90] (2020)	Deep learning features	CNN + attention mechanism	FaceF++	AUC-99.4%
				EER-3.4%

FaceF++—FaceForensics++, TCR—True Classification Rate, EER—Equal Error rate, Acc.—Accuracy, AUC—Area Under Curve, SVM—Support Vector Machine, CNN—Convolutional Neural Network, MLP—Multilayer Perceptron

multitasking incremental learning to classify images generated from GANs. The model was based on XceptionNet and able to achieve an accuracy of 99.3% to correctly detecting on new GAN-generated images. Another approach is proposed by Zhang et al. [84] which is based on spectrum domain features. They used AutoGAN as the classifier, which is a GAN simulator to generate an image without any need for pre-training. The method is tested on various GAN models to obtain promising results. Finally, Rathgeb et al. [85] proposed photoresponse non-uniformity (PRNU). The spatial and spectral features are extracted from PRNU patterns. The EER achieved by the model is 13.7% on a private database.

5.2.3 Expression swap

In this manipulation, the aim is to modify the facial expressions of the target person (in a video or image). The most common techniques for expression swap are Face2Face and NaturalTextures. They are capable of replacing the original person's facial expression with the target facial expression. The available database for this manipulation is Faceforensics++ [57], as it contains several facial expression replacement videos for training and testing. Nguyen et al. [86] proposed a method based on multi-task learning. They used the Faceforensics++ database and achieved an EER of 7.1% for the Face2Face method on HQ videos. Another interesting approach is used by Amerini et al. [87]. They used CNN along with optical flow fields to detect possible facial dissimilarities, using PWC-nets [88]. The motivation of their approach is that fake videos must have unnatural optical flow because of the unnatural movement of the eye, lips, etc. The accuracy achieved is 81.6% using both VGG16 and ResNet50 architecture. A similar approach was used by Sabir et al. [89], in which recurrent CNN is used instead of VGG16. The AUC achieved on the Faceforensics++ database is 94.3% on the Face2Face method. They only used low-quality videos in their analysis. Finally, Dang et al. [90] proposed an approach based on an attention mechanism to further enhance the training process. These mechanisms help the CNN models to process and extract features maps more efficiently. The approach uses the DFDC database and obtained an AUC of 99.4% and EER of 3.1%. Vinolin et al. [91] used deep CNN for video forgery detection. The DCNN is trained using a dual-adaptive Taylor-rider optimization algorithm (DA-Taylor-based ROA). For detection, the concept creates a 3D model of the video frame to generate light coefficients. Later, the Viola–Jones algorithm is used for the detection of face objects. The research achieves better optimal convergence to find the best solution. The accuracy achieved by the authors is 94.031% (in the absence of noise), 93.86% (in case of salt and pepper noise), and 92.376% (in case of speckle noise).

6 Evaluation and findings

The subsequent points were discovered after a detailed evaluation of different published papers:

- The tampering detection is more difficult to detect than localization task. The tampering detection centers around coarse-grained examination, while the localization task centers around the fine-grained examination of an image. There are several methods for tampering detection but only a few of these are capable of localizing the tampered region.
- The traditional methods (both keypoint and block-based) uses hand-crafted feature extraction approaches for tampering detection. The DL-based methods are more self-sustaining and capable of learning complex features from tampered regions without any human intervention. In terms of accuracy, the DL-based approaches have the upper hand in both classifications of tampered regions and localization tasks. However, training DNN is difficult and requires a large dataset and high computational power.
- Each technique is specific for some type of forgery detection performed on the authentic image. Therefore, the use of a specific detection technique might not work when the person performing the examination is unaware of the forgery type. Hence, there is a need for forgery detection techniques that can be used to detect any forgery type.
- Different Researchers used different metrics (such as Accuracy, F1-Score, ROC Curve, MCC, IoU, etc.) to measure the performance of tampering detection. Therefore, there is a requirement for a common benchmark for comparing different tampering detection algorithms. Some new benchmarks are introduced by Rossler et al. [57] in their published work but it only covers facial forgery.
- The datasets used to evaluate the performance of tampering detection algorithms must be standard. For better evaluation of algorithms, the datasets must contains tampered images with a variety of different tampering attacks to increase the diversity of the test set. Recent facial manipulation datasets such as DepperForensics-1.0 [63] contain challenging hidden test set to better simulate the real-world distributions with high deceptive scores.

6.1 Research challenges and future scopes

After analyzing several research papers, it has been noted that this field requires a lot of time and work. Based on these reviews, future work can be proposed. The following are some of the research challenges:

- Computational complexity and efficient feature extraction models [92,93].
- Accurate localization of tampered regions in a tampered image [94,95].
- Slow learning rate for DL models [96,97].
- Robust forgery detection techniques to solve issues like pre- and post-processing procedures, multiple forgery attacks, low learning rate (LR) problem [94,98,99].
- The absence of global dataset that contains all types of possible forgery attacks and high deception score [100, 101].
- Human interpretation required to cross-check the results of localization task predictions of DL models [102–104].

- There are several problems in computer vision like pattern recognition, feature extraction, image analysis, video surveillance and analytics, geometrical transformations, etc. Such problems are addressed by techniques such as deep learning and Weber local descriptor (WLD) [105].

To solve the problem of feature extraction, advanced DL models are needed to be developed that can maintain features characteristics even when the dimensionality reduction (DR) operations are applied. The leaning can be improved using momentum or by applying varying learning rates instead of fixed learning rate [106]. To speed up, the detection process, GPU acceleration, and parallel programming may be considered. At last, to improve the localization of tempered regions, multi-scale analysis can be applied.

7 Conclusion

This paper reviews techniques on image and video tampering and its detection. Although the intent of a person cannot be predicted without evaluating the tampered image, one must be cautious about their videos and images before posting them on any social media networks. In nutshell, image tampering is defined as changing the content of a real-world image (not computer-generated) with some new content. As newer altering software's being developed every day, it is difficult for humans to distinguish tampered images from real images. This problem becomes worse when someone tries to localize the tampered region. The best approach to prove fraudulence is to find the real image from which the corresponding altered image is created. If the real image is destroyed by the attackers or not to be found, the next natural step is to inspect the invisible artifacts left by the image manipulation operations.

In conclusion, this research contributes is to the area of digital media forgery detection. A detailed analysis of high-quality published papers is compared to provide various image forgeries detection methods. A taxonomy is provided to represent popular image tampering datasets. Several deep learning-based image forgery detection techniques are also compared in this paper. The techniques are categorized based on the modeling and approach used for detection. Furthermore, various video datasets are explained and compared to the field of digital media forensics. The objective is to provide a brief overview of various fundamental topics related to digital image detection. Moreover, by observing different techniques, research challenges and potential scopes are identified, to provide researchers with possible future directions.

Declarations

Conflict of interest On behalf of all authors, I (corresponding author) state that there is no conflict of interest.

References

1. Carrington, D.: (2020 Website: Mylio.com) How Many Photos in 2020? Detailed report here (2020).https://focus.mylio.com/tech-today/how-many-photos-will-be-taken-in-2020
2. Gyncild, A.C., Team, B., Team: Adobe Photoshop CC Classroom in a Book. Pearson Education, London (2013)
3. The GIMP Development Team: GIMP. Retrieved from (2019)
4. Zhang, W.: Smartphone photography in urban China. World Acad. Sci. Eng. Technol., Int. J. 960 Soc. Behav. Educ. Econ. Bus. Ind. Eng. **11**(1), 231–239 (2017)
5. Chouhan, S.S., Kaul, A., Singh, U.P.: Image segmentation using computational intelligence techniques. Arch. Comput. Methods Eng. **26**(3), 533–596 (2019)
6. Sneumueller.: Auto face swap (2016). https://www.microsoft.com/en-us/store/p/auto-face-swap/9nblggh3m5nq
7. LTD RTP.: Face swap booth—photo faceswap and face changer (2017). https://itunes.apple.com/us/app/face-swap-booth-photo-faceswap-face-changer/id826921329?mt=8
8. The Economic Times Bureau.: Detailed report at: https://economictimes.indiatimes.com/tech/internet/4-in-10-indians-have-experienced-identity-theft-report/articleshow/75029916.cms?from=mdr (2020)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative adversarial nets. Adv. Neural Inf. Process. Syst. pp. 2672–2680 (2014)
10. Ledig, C., Theis, L., F. Husz ar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al.: Photo-realistic singleimage super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690 (2017)
11. AI can now create fake porn, making revenge porn even more complicated. The conversation media group (2018). http://theconversation.com/ai-can-now-create-fake-pornmaking-revengeporn-even-more-complicated-92267
12. Rose Eveleth (BBC News 13th December 2012), Detailed article here: https://www.bbc.com/future/article/20121213-fake-pictures-make-real-memories (2012)
13. Thyagharajan, K.K., Kalaiarasi, G.: A review on near-duplicate detection of images using computer vision techniques. Arch. Comput. Methods Eng. pp 1–20 (2020)
14. Rob Toews.: (Forbes 25May 2020). Detailed article here: https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=c2edef574940 (2020)
15. Farid H.: Creating and detecting doctored and virtual images: implications to the child pornography prevention act. Department of Computer Science, Dartmouth College, TR2004-518, 13 (2004)
16. Farid H.: Photo tampering throughout history. Read more at: https://www.cc.gatech.edu/~beki/cs4001/history.pdf

17. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29, pp. 2242–2251 (2017)
18. Wang, W., Dong, J., Tanm T.: A survey of passive image tampering detection. In: IWDW, 9, pp. 308–322 . Springer (2009)
19. He, J., Lin, Z., Wang, L., Tang, X.: Detecting doctored JPEG images via DCT coefficient analysis. In: Proceedings of ECCV., pp 423–435 (2006)
20. Christlein, V., Riess, C., Jordan, J., Riess, C., Angelopoulou, E.: An evaluation of popular copy-move forgery detection approaches. IEEE Trans. Inf. Forensics Secur. **7**(6), 1841–54 (2012)
21. Ng, T.T., Chang, S.F., Sun, Q.: Blind detection of photomontage using higher order statistics. In: Proceedings of ISCAS, vol. 5. IEEE (2004)
22. Schetinger, V., Oliveira, M.M., da Silva, R., Carvalho, T.J.: Humans are easily fooled by digital images. Comput. Gr. (2017)
23. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., and Efros, A.: Context encoders: feature learning by inpainting. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2536–2544 (2016)
24. Yang, C., Lu, X., Lin, Z., Shechtman, E., Wang, O., Li, H.: High-resolution image inpainting using multi-scale neural patch synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6721–6729 (2017)
25. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Gr. (ToG) **36**(4), 1–14 (2017)
26. Johnson, M.K., Farid, H.: Exposing digital forgeries by detecting inconsistencies in lighting. In: Proceedings of the 7th Workshop on Multimedia and Security, pp. 1–10, ACM (2005)
27. Wen, B., Zhu, Y., et al.: COVERAGE: A novel database for copy-move forgery detection. In: Proceedings of ICIP. IEEE, 161–165 (2016)
28. Zhang, Y., Goh, J., Win, L.L., Thing, V.L.: Image region forgery detection: a deep learning approach. In: Proceedings of SG-CRC., pp. 1–11 (2016)
29. Bunk, J., Bappy, J.H., Mohammed, T.M., Nataraj, L., Flenner, A., Manjunath, B., Chandrasekaran, S., Roy-Chowdhury, A.K., Peterson, L.: Detection and localization of image forgeries using resampling features and deep learning. In: Proceedings of CVPRW. IEEE, pp. 1881–1889 (2017)
30. Rao, Y., Ni, J.: A deep learning approach to detection of splicing and copy-move forgeries in images. In: Proceedings of WIFS (2016)
31. Shi, Y.Q., Chen, C., Chen, W.: A natural image model approach to splicing detection. In: Proceedings of MM& Sec. ACM, 2007:51–62 (2007)
32. Ng, A.: Machine learning yearning (2018). (http://www.mlyearning.org/)
33. Tang, S.: Lessons learned from the training of GANs on artificial datasets. IEEE Access **8**, 165044–165055 (2020)
34. Schonfeld, E., Schiele, B., Khoreva, A.: A u-net based discriminator for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020)
35. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2019)
36. Karras, T. et al.: Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020)
37. Singh, K.K., Ojha, U., Lee, Y.J.: Finegan: Unsupervised hierarchical disentanglement for fine-grained object generation and discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2019)
38. Ng, T.T., Chang, S.F., Sun, Q.: A data set of authentic and spliced image blocks. Columbia University, ADVENT Technical Report (2004)
39. Hsu, Y.F., Chang, S.F.: Detecting image splicing using geometry invariants and camera characteristics consistency. In: Proceedings of ICME (2006)
40. Dong, J, Wang, W., and Tan, T.: CASIA image tampering detection evaluation database. In: 2013 IEEE China Summit and International Conference on Signal and Information Processing, Beijing, pp. 422–426 (2013)
41. Zheng, Y., Cao, Y., Chang, C.: A PUF-based data-device hash for tampered image detection and source camera identification. IEEE Trans. Inf. Forensics Secur. **15**, 620–634 (2020)
42. Amerini, I., Ballan, L., Caldelli, R., Del Bimbo, A., Serra, G.: A SIFT-based forensic method for copy-move attack detection and transformation recovery. IEEE Trans. Inf. Forensics Secur. **6**(3), 1099–110 (2011)
43. Amerini, I., Ballan, L., Caldelli, R., Del Bimbo, A., Del Tongo, L., Serra, G.: Copy-move forgery detection and localization by means of robust clustering with J-Linkage. Signal Process. Image Commun. **28**(6), 659–69 (2013)
44. Gloe, T., Bohme, R.: The dresden image database for benchmarking digital image forensics. J. Digit. Forensic Pract. **3**(2–4), 150–159 (2010)
45. Bappy, J.H., Simons, C., Nataraj, L., Manjunath, B.S., Roy-Chowdhury, A.K.: Hybrid LSTM and encoder–decoder architecture for detection of image forgeries. IEEE Trans. Image Process. **28**(7), 3286–3300 (2019). https://doi.org/10.1109/TIP.2019.2895466
46. IEEE IFS-TC Image Forensics Challenge Dataset. http://ifc.recod.ic.unicamp.br/fc.website/index.py
47. Tralic, D., Zupancic, I., Grgic, S., Grgic, M.: CoMoFoD—new database for copy-move forgery detection. In: Proceedings of ELMAR (2013)
48. Zampoglou, M., Papadopoulos, S., Kompatsiaris, Y.: Detecting image splicing in the wild (Web). In: Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on, pp. 1–6. IEEE (2015)
49. Bas, P., Filler, T., Pevný, T.: Break our steganographic system: the ins and outs of organizing BOSS. In: International Workshop on Information Hiding. Springer, Berlin, Heidelberg (2011)
50. Xie, D., et al.: Scut-fbp: A benchmark dataset for facial beauty perception. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics. IEEE (2015)
51. Wen, B., Zhu, Y., Subramanian, R., Ng, T., Shen, X., Winkler, S.: COVERAGE—A novel database for copy-move forgery detection. In: Proc. IEEE Int. Conf. Image Processing (ICIP) (2016)
52. Schaefer, G., Stich, M.: UCID—An Uncompressed Colour Image Database. In: Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia 2004, pp. 472–480, San Jose, USA (2004)
53. Dolhansky, Brian, H., Russ, P., Ben, B., Nicole, F., Cristian C. (eds.): The deepfake detection challenge (dfdc). Preview dataset (2019). arXiv preprint, arXiv:1910.08854
54. Xin Y., Yuezun L., and Siwei L.: Exposing deep fakes using inconsistent head poses. In: ICASSP, pp 8261–8265 (2019)
55. Fakeapp. https://www.fakeapp.com/. Accessed: 2018-09-01
56. Korshunov, P., Marcel, S.: Deepfakes: a new threat to face recognition? Assessment and detection (2018). arXiv preprint arXiv:1812.08685
57. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., Nießner, M.: Faceforensics: A large-scale video dataset for forgery detection in human faces (2018). arXiv preprint arXiv:1803.09179

58. Justus T., Michael Z., Marc S., Christian T., and Matthias N.: Face2Face: Real-time face capture and reenactment of RGB videos. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 2387–2395 (2016)
59. Deepfakes github.: https://github.com/deepfakes/faceswap. Accessed: 2018-10-29
60. Faceswap. https://github.com/MarekKowalski/FaceSwap/. Accessed: 2018-10-29
61. Thies, J., Zollhöfer, M., Nießner, M.: Deferred neural rendering: image synthesis using neural textures. ACM Trans. Gr. (TOG) **38**(4), 1–12 (2019)
62. Google AI Blog. Contributing data to deepfake detection research. https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html. Accessed: 2019-09-25
63. Jiang, L., Li, R., Wu, W., Qian, C., Loy, C.C.: Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2889–2898 (2020)
64. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceeding of ICCV, vol. 2. IEEE, 1150–1157 (1999)
65. Fridrich, A.J., Soukal, B.D., and Lukáš, A.J.: Detection of copymove forgery in digital images. In: Proceedings of Digital Forensic Research Workshop, Citeseer (2003)
66. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
67. Bashar, M., Noda, K., Ohnishi, N., Mori, K.: Exploring duplicated regions in natural images. In: IEEE Trans. Image Process. (2010)
68. Wall, M.E., Rechtsteiner. A., Rocha, L.M.: Singular value decomposition and principal component analysis. In: A Practical Approach to Microarray Data Analysis, pp. 91–109. Springer (2003)
69. Shlens, J.: A tutorial on principal component analysis (2014). arXiv preprint arXiv:1404.1100
70. Bay, H., Tuytelaars, T., Van Gool, L.S.U.R.F.: Speeded up robust features. In: Proceeding of ECCV, pp 404–417 (2006)
71. Huang, H., Guo, W., Zhang, Y.: Detection of copy-move forgery in digital images using sift algorithm. In: 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, vol. 2, pp. 272–276. IEEE (2008)
72. Zhu, Y., Shen, X., Chen, H.: Copy-move forgery detection based on scaled ORB. Multimed. Tools Appl. **75**(6), 3221–3233 (2016)
73. Chen, W., Shi, Y.Q., Su, W.: Image splicing detection using 2D phase congruency and statistical moments of characteristic function. In Security, Steganography, and Watermarking of Multimedia Contents IX, vol. 6505, p. 65050R. International Society for Optics and Photonics (2007)
74. Johnson, M.K., Farid, H.: Exposing digital forgeries in complex lighting environments. IEEE Trans. Inf. Forensics Secur. **2**(3), 450–461 (2007)
75. Popescu, A.C., Farid, H.: Statistical tools for digital forensics. In: International Workshop on Information Hiding, pp. 128–147. Springer (2004)
76. Matern, F., Riess, C., and Stamminger, M.: Exploiting visual artifacts to expose DeepFakes and face manipulations. In: Proc. IEEE Winter Applications of Computer Vision Workshops (2019)
77. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning (2016)
78. Yang, X., Li, Y., Lyu, S.: Exposing deep fakes using inconsistent head poses. In: Proc. International Conference on Acoustics, Speech and Signal Processing (2019)
79. Jung, T., Kim, S., and Kim, K.: DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern, IEEE Access, vol. 8, pp. 83 144–83 154 (2020)
80. Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., Natarajan, P.: Recurrent convolutional strategies for face manipulation detection in videos. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2019)
81. Tolosana, R., Romero-Tapiador, S., Fierrez, J., and Vera-Rodriguez, R.: DeepFakes evolution: analysis of facial regions and fake detection performance (2020). arXiv preprint arXiv:2004.07532
82. Wang, R., Ma, L., Juefei-Xu, F., Xie, X., Wang, J., and Liu, Y.: FakeSpotter: a simple baseline for spotting AI-synthesized fake faces (2019). arXiv preprint arXiv:1909.06122
83. Marra, F., Saltori, C., Boato, G., and Verdoliva, L.: Incremental learning for the detection and classification of GAN-generated images. In: Proc. IEEE International Workshop on Information Forensics and Security (2019)
84. Zhang, X., Karaman, S., and Chang, S.: Detecting and simulating artifacts in GAN fake images. In: Proc. IEEE International Workshop on Information Forensics and Security (2019)
85. Rathgeb, C., Botaljov, A., Stockhardt, F., Isadskiy, S., Debiasi, L., Uhl, A., Busch, C.: PRNU-based Detection of Facial Retouching. IET Biometrics (2020)
86. Nguyen, H., Fang, F., Yamagishi, J., and Echizen, I.: Multi-task learning for detecting and segmenting manipulated facial images and videos (2019). arXiv preprint arXiv:1906.06876
87. Amerini, I., Galteri, L., Caldelli, R., and Bimbo, A.: Deepfake video detection through optical flow based CNN
88. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: PWC-Net: CNNs for optical flow using pyramid, warping, and cost Volume. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018)
89. Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., Natarajan, P.: Recurrent convolutional strategies for face manipulation detection in videos. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2019)
90. Dang, H., Liu, F., Stehouwer, J., Liu, X., Jain, A.: On the detection of digital face manipulation. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020)
91. Vinolin, V., Sucharitha, M.: Dual adaptive deep convolutional neural network for video forgery detection in 3D lighting environment. Vis. Comput. pp. 1–22 (2020)
92. Ardizzone, E., Bruno, A., Mazzola, G.: Copy-move forgery detection by matching triangles of keypoints. IEEE Trans. Inf. Forensics Secur. **10**(10), 2084–2094 (2015)
93. Zhao, X., Wang, S., Li, S., Li, J.: Passive image splicing detection by a 2-D noncausal Markov model. IEEE Trans. Circuits Syst. Video Technol. **25**(2), 185–199 (2015)
94. Li, B., Ng, T., Li, X., Tan, S.: Revealing the trace of high-quality JPEG compression through quantization noise analysis. IEEE Trans. Inf. Forensics Secur. **10**(3), 558–573 (2015)
95. Yin, T., Yang, G., Li, L., Zhang, D.: Detecting seam carving based image resizing. Comput. Secur. **55**, 130–141 (2015)
96. Zhang, Y., Goh, J., Lei, L., Thing, V.: Image region forgery detection: a deep learning approach. Singap. Cyber-Secur. Conf. **14**, 1–11 (2016)
97. Yu, J., Zhan, Y., Xiangui, Yang J., KB.: A multi-purpose image counter-anti-forensic method using convolutional neural networks. In: International Workshop on Digital Watermarking, pp. 3–15 (2017)
98. Lee, J.-C., Chang, C.-P., Chen, W.-K.: Detection of copy-move image forgery using histogram of orientated gradients. Inf Sci (NY) **321**, 250–262 (2015)
99. Cozzolino, D., Poggi, G., Verdoliva, L.: Efficient dense-field copy-move forgery detection. IEEE Trans. Inf. Forensics Secur. **10**(11), 2284–2297 (2015)
100. Chen, J., Kang, X., Liu, Y., Wang, Z.J.: Median filtering forensics based on convolutional neural networks. IEEE Signal Process. Lett. **22**(11), 1849–1853 (2015)
101. Tuba, V., Jovanovic, R., Tuba, M.: Digital image forgery detection based on shadow HSV inconsistency. In: 5th International Symposium on Digital Forensic and Security (ISDFS) (2017)

102. Bahrami, K., Kot, A.C., Li, L., Li, H.: Blurred image splicing localization by exposing blur type inconsistency. IEEE Trans. Inf. Forensics Secur. **10**(5), 999–1009 (2015)
103. Li, H., Luo, W., Qiu, X., Huang, J.: Image forgery localization via integrating tampering possibility maps. IEEE Trans. Inf. Forensics Secur. **12**(5), 1240–1252 (2017)
104. Abdul Warif, N.B., Abdul Wahab, A.W., Idna Idris, M.Y., Fazidah Othman, R.S.: SIFT-Symmetry: a robust detection method for copy-move forgery with reflection attack. J. Vis. Commun. Image Represent. **46**, 219–232 (2017)
105. Banerjee, A., Das, N., Santosh, K.C.: Weber local descriptor for image analysis and recognition: a survey. Vis. Comput. pp. 1–23 (2020)
106. Bowling, M., Veloso, M.: Multiagent learning using a variable learning rate. Artif. Intell. **136**(2), 215–250 (2002)
107. Li, Y., Xin Y., Pu, S., Honggang, Q., and Siwei, L.: A new dataset for deepfake forensics. arXiv preprint, Celeb-df (2019)
108. Chen, J., Kang, X., Liu, Y., Wang, Z.J.: Median filtering forensics based on convolutional neural networks. IEEE Signal Process. Lett. **22**(11), 1849–1853 (2015)
109. Zhang, Y., Goh, J., Win, L.L., Thing, V.L.: Image region forgery detection: a deep learning approach. In: SG-CRC, pp. 1–11 (2016)
110. Bayar, B., Stamm, M.C.: A deep learning approach to universal image manipulation detection using a new convolutional layer. In: Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, pp. 5–10, ACM (2016)
111. Cozzolino, D., Verdoliva, L.: Single-image splicing localization through autoencoder-based anomaly detection. In: 2016 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–6, IEEE (2016)
112. Rao, Y., Ni, J.: A deep learning approach to detection of splicing and copy-move forgeries in images. In: 2016 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–6, IEEE (2016)
113. Amerini, I., Uricchio, T., Ballan, L., Caldelli, R.: Localization of jpeg double compression through multi-domain convolutional neural networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1865–1871, IEEE (2017)
114. Bondi, L., Lameri, S., D. Güera, P. Bestagini, E. J. Delp, and S. Tubaro,: Tampering detection and localization through clustering of camera-based cnn features, in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1855–1864. IEEE (2017)
115. Salloum, R., Ren, Y., Kuo, C.-C.J.: Image splicing localization using a multi-task fully convolutional network (mfcn). J. Vis. Commun. Image Represent. **51**, 201–209 (2018)
116. Wu, Y., Abd-Almageed, W., Natarajan, P.: Busternet: Detecting copy-move image forgery with source/target localization. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 168–184 (2018)
117. Bi, X., Wei, Y., Xiao, B., Li, W.: Rru-net.: The ringed residual u-net for image splicing forgery detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2019)
118. Wang, X., Wang, H., Niu, S., and Zhang, J.: Detection and localization of image forgeries using improved mask regional convolutional neural network (2019)
119. Wang, Sheng-Yu., et al.: Detecting photoshopped faces by scripting photoshop. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2019)
120. Muzaffer, G., Ulutas, G.: A new deep learning-based method to detection of copy-move forgery in digital images. In: 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). IEEE (2019)
121. Kuznetsov, A.: Digital image forgery detection using deep learning approach. J. Phys. Conf. Ser. Vol. 1368. No. 3. IOP Publishing (2019)
122. Marra, F., Gragnaniello, D., Verdoliva, L., Poggi, G.: A full-image full-resolution end-to-end-trainable CNN framework for image forgery detection. IEEE Access **8**, 133488–133502 (2020). https://doi.org/10.1109/ACCESS.2020.3009877
123. Hossein-Nejad, Z., Nasri, M.: Clustered redundant keypoint elimination method for image mosaicing using a new Gaussian-weighted blending algorithm. Vis. Comput. 1–17 (2021)
124. Yu, F., Koltun, V., Funkhouser, T.: Dilated residual networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (2017)

Mr. Shobhit Tyagi is PhD Scholar in the Department of Computer Science & Engineering at National Institute of Technology, Hamirpur (HP). He did his undergraduate and Post Graduate in Computer Science & Engineering. His area of interest includes Machine Learning, Image Processing.

Dr. Divakar Yadav is currently working as Associate Professor in the Department of Computer Science & Engineering at National Institute of Technology, Hamirpur (HP), India. Prior to join this institute, he worked at Madan Mohan Malaviya University of Technology, Gorakhpur (UP) as Associate Professor and Jaypee Institute of Information Technology, Noida as Assistant as well as Associate Professor. He did his undergraduate (B. Tech.) in Computer Science & Engineering in 1999 from IET, Lucknow, Postgraduate (M.Tech.) in Information Technology in 2005 from Indian Institute of Information Technology, Allahabad and PhD in Computer Science & Engineering in 2010 from Jaypee Institute of Information Technology, Noida. He also worked as Post Doctoral Fellow at University of Carlos-III, Madrid (Spain) between 2011–2012. He supervised 4 PhD theses, 22 M.Tech. dissertations and many undergraduate projects. He published two books, 12 book chapters and more than 100 research articles in reputed International Journals and Conference Proceedings. His area of research includes information Retrieval and Machine Learning. He is senior member of IEEE.