IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 November 2022, accepted 21 December 2022, date of publication 16 January 2023, date of current version 23 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3236993

SURVEY

Visual User-Generated Content Verification in Journalism: An Overview

SOHAIL AHMED KHAN[1], GHAZAAL SHEIKHI[1], ANDREAS L. OPDAHL[1], FAZLE RABBI[1], SERGEJ STOPPEL[2], CHRISTOPH TRATTNER[1], AND DUC-TIEN DANG-NGUYEN[1], (Member, IEEE)

[1]MediaFutures, Department of Information Science and Media Studies, University of Bergen, 5007 Bergen, Norway
[2]Wolftech, 5008 Bergen, Norway

Corresponding authors: Sohail Ahmed Khan (sohail.khan@uib.no) and Duc-Tien Dang-Nguyen (ductien.dangnguyen@uib.no)

This work was supported by the Industry Partners and the Research Council of Norway through MediaFutures: Research Centre for Responsible Media Technology and Innovation through the Centres for Research-Based Innovation Scheme under Project 309339.

ABSTRACT Over the past few years, social media has become an indispensable part of the news generation and dissemination cycle on the global stage. These digital channels along with the easy-to-use editing tools have unfortunately created a medium for spreading mis-/disinformation containing visual content. Media practitioners and fact-checkers continue to struggle with scrutinising and debunking visual user-generated content (UGC) quickly and thoroughly as verification of visual content requires a high level of expertise and could be exceedingly complex amid the existing computational tools employed in newsrooms. The aim of this study is to present a forward-looking perspective on how visual UGC verification in journalism can be transformed by multimedia forensics research. We elaborate on a comprehensive overview of the five elements of the UGC verification and propose multimedia forensics as the sixth element. In addition, different types of visual content forgeries and detection approaches proposed by the computer science research community are explained. Finally, a mapping of the available verification tools media practitioners rely on is created along with their limitations and future research directions to gain the confidence of media professionals in using multimedia forensics tools in their day-to-day routine.

INDEX TERMS Visual misinformation, multimedia forensics, journalistic verification, misinformation detection, disinformation detection.

I. INTRODUCTION

In today's digital society where everyone has a voice on the multitude of social media platforms, harnessing user-generated content (UGC) has become a daily routine in newsrooms. With prevalence of smartphones with high quality cameras and access to the Web, visual and textual information relating to trending and breaking news events on social media platforms such as, Facebook, Twitter and YouTube has grown in scale and scope. The large volume of textual and visual content shared on social media enables journalists to gather information for publishing timely breaking news and rapid updates on trending events. In the context of journalism, UGC is associated with the term citizen journalism, i.e., citizens contributing to the process of collecting, reporting and distributing news-related information in the time of crisis. When reporters cannot reach to the ground efficiently such as in the countries with limited press freedom or in cases where events unfold quickly such as during a natural disaster, UGC becomes a key element in media coverage. Some well-known examples of events that were hugely covered by street journalism are the Arab Spring in 2010s, the Hurricane Sandy in 2012, the 2019-2020 Iranian protests [1].

Social media, and other similar web based platforms are not only used by journalists and reporters to obtain updated information about the latest trending news stories around the globe [2], [3], but also prepare a ground for newsrooms and media outlets to increase audience reach [1], [4]. According to a study by the International Centre For Journalists (ICFJ) in 2019 [5], two-thirds of news organisations distribute content in at least four formats, e.g., (1) social media post, (2) news item on the website, (3) video, and (4) messaging

The associate editor coordinating the review of this manuscript and approving it for publication was Ziyan Wu.

apps, with social media being the most widely-used platform. Almost 80% of the news organisations which were surveyed found to be using social media platforms to distribute their content [5]. With the presence of media outlets as well as ordinary people creating content on these digital channels, the number of ordinary people relying on social media for news consumption is increasing [6]. According to Pew Research Center's report on ''News Consumption Across Social Media in 2021'', more than half of Twitter users regularly used the platform to consume news in 2021 [6]. The report also found that social media users often rely on the platforms such as Facebook, Twitter, YouTube, Reddit, LinkedIn, and TikTok to obtain up to date information about the trending stories online.

Unfortunately, social media becoming an integral part of news distribution and consumption is a double edged sword. On one hand, the public and the media community get easy and instant access to local and global news in almost real time. On the other hand, these digital channels are being used to spread misinformation (when someone unintentionally share misleading content) and disinformation (when someone knowingly share misleading content) [7]. The fight against mis-/disinformation has been an ongoing move by news outlets, fact-checkers, and social media companies. For journalists and news editors who have to leverage UGC, it is even more essential to effectively monitor, verify, and debunk fake/manipulated UGC shared on social media platforms especially during breaking news events. Because of the importance of verification, major newsrooms such as The Associated Press, or BBC, have dedicated teams focused on verifying UGC. It is evident from the ICFJ's 2019 report [5] that about one-third of the surveyed news organisations have dedicated fact-checking teams to verify and fact-check content. Moreover, non-profit fact-checking organisations such as FullFact[1] and Faktisk[2] have been established and are quickly growing. Some independent organisations offer UGC verification as a service. For instance, Storyful,[3] a social media intelligence agency, offers ''verification'' as one of its services, and they collaborate with major newsrooms such as The New York Times and Reuters.

The significance of visual UGC verification becomes more evident when we look back on the incidences in the past when newsrooms and professional journalists failed in identifying misleading photos/videos and shared them as reliable content related to a newsworthy event. For instance, during the devastating flooding in Queensland, Australia, in early 2019, photos of crocodiles on the streets of the flood-affected region were uploaded and shared on social media platforms. The well-known Australian news outlet, Nine News published those photos as if they were captured on the flood scene [8]. It was discovered later that the photos were originally taken in 2014 showing American alligators in Florida, USA. But it was too late for the news agency to undo the damage to its reputation. Many other images alleging to be of the 2019 flooding, shared sometimes even by professional journalists, actually belonged to other events from different time periods and geographic contexts. These kind of incidents where professional journalists fail in UGC verification and share unauthentic or out-of-context information promotes the spread of mis-/disinformation which journalists aim to fight against, profoundly affects the level of trust in news, and severely dents the reputation of the journalist as well as the media outlet [3], [7].

Besides the increasing amounts of visual UGC shared online every day, it is becoming even more effortless to produce false and misleading content using inexpensive and user-friendly photo editing software tools such as Adobe Photoshop and Gimp. Along with the classical image manipulation techniques, a contemporary form of visual content forgery known as ''Deepfake'' media (fake multimedia produced using deep neural networks) has emerged in recent years. These technological advances at everyone's fingertips pose more challenges for the newsrooms and media practitioners to verify visual UGC.

Manual UGC verification is an extremely time consuming task because this procedure typically entails interviewing eye witnesses, checking the digital footprint of the source who shared the UGC item online, gathering more details about the events being portrayed in the video/image (e.g., identifying location within the image, date and time), thus results in spending considerable amount of time before coming to conclusions [4]. Obviously, this is in stark contrast to the race against the clock in newsrooms and the necessity of debunking viral mis-/disinformation online. Therefore, digital verification using (semi) automated tools is crucial in reducing the time burden of visual UGC verification. According to the 2019 ICFJ study mentioned before, media practitioners have attracted to these tools in recent years and the trend of utilising computational tools to verify/fact-check UGC by the journalists and fact-checkers is rising [5]. However, only around 33% of the journalists have been found to use such tools to assist them during the verification procedure [5]. The obstacles on the path to encourage the usage of technology for visual UGC verification could be grouped into two categories. The first category is related to the lack of knowledge about manual verification procedure among computer scientists who develop these tools. It is essential for the technology experts to fully comprehend the process of the manual verification in order to align/improve the tools with the requirements in newsrooms. The second category of barriers concern the journalists' points of view on these digital verification tools. Media practitioners ought to recognise the capabilities of forensics techniques proposed by the computer science research community in detecting image/video forgeries. Moreover, mapping of the verification tools employed in newsrooms and by journalists along with their use cases is fruitful in picturing the assets

[1] https://fullfact.org/

[2] A non-profit organisation and independent editorial office for fact-checking of the public debate and the public discourse in Norway. https://www.faktisk.no/

[3] https://storyful.com/

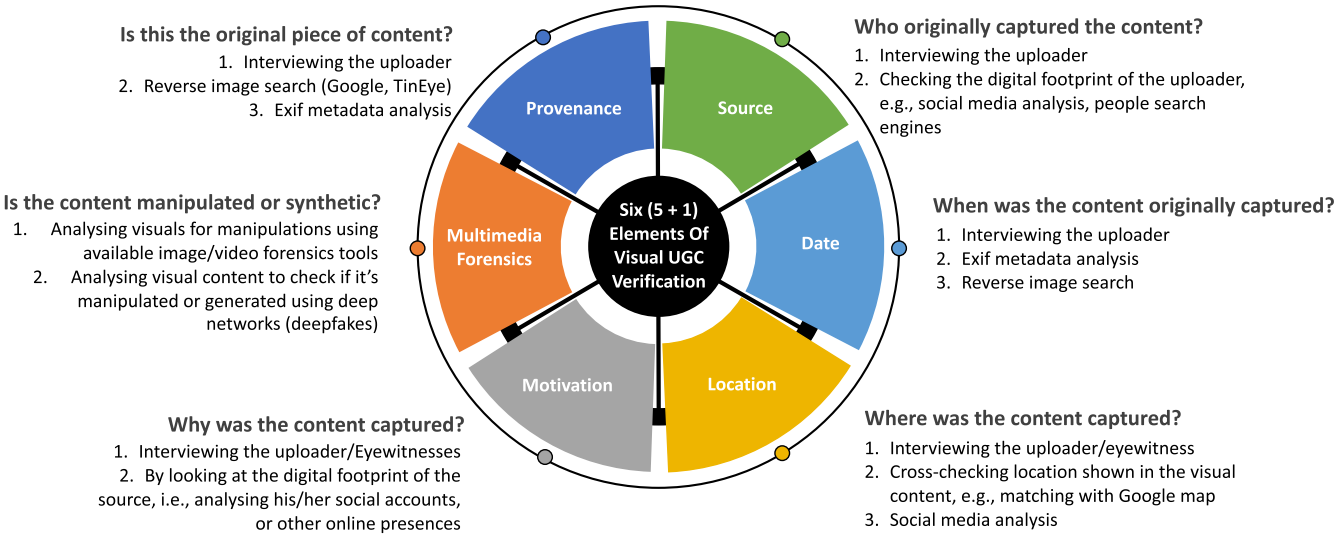


FIGURE 1. The proposed Six (5+1) elements of UGC verification. The first five elements are inspired from [9]: **Provenance:** checking if the same content has been shared, **Source:** person who captured the content initially, **Date** of capturing the content, **Location** where the content was captured, **Motivation** of capturing/sharing the content; and we proposed the sixth element "**Multimedia Forensics**" to help in identifying whether the visual UGC item is manipulated or synthetic.

and liabilities of the existing technology for visual UGC verification.

The main aim of this study is to highlight the merits of (semi) automated visual content verification. We contribute:

- A comprehensive overview on visual UGC verification in journalism. The five basic elements of visual UGC verification in journalism, are described in detail. We seek to understand how state-of-the-art in multimedia forensics could enhance existing tools to facilitate the procedure journalists follow when they verify visual UGC.
- In addition to the five elements of UGC verification, we also propose a sixth element which we call "Multimedia Forensics" and describe why we think it is necessary.
- An extensive study on visual content forensics from a technical perspective and present a number of different image/video forgery techniques and detection strategies along with examples of manipulated imagery from the news domain.
- A map of tools frequently employed by journalists and media practitioners for visual content verification, their use cases, and the limitations associated with them.

This paper is organised as follows. In section II, a detailed analysis of journalistic process for visual content verification is presented. Section III, discusses various classes of visual content forgeries and the technologies developed to detect these forgeries. In section IV, we present a comprehensive mapping of the tools and technologies available for visual content verification and the technologies being used in the media industry for visual content verification. Section V concludes the findings and proposes future research directions.

II. THE 5+1 ELEMENTS OF VISUAL UGC VERIFICATION IN JOURNALISM

Major newsrooms principally have their own verification guidelines. The Associated Press (AP), for instance, has well-established standards that haven't changed for years. These standards made it possible for the organisation to successfully deal with social media content [4].

At BBC (British Broadcasting Corporation), after all the essential measures to verify the content are completed, the journalists disseminate their findings across all of the BBC's platforms using a system called Electronic News Production System (ENPS) [10].

First Draft News[4] [9] proposed five elements constituting the investigative UGC verification process: (1) *Provenance*, (2) *Source*, (3) *Date*, (4) *Location* and (5) *Motivation*. In this section, we describe these five elements in detail, and present how journalists, fact-checkers base their investigations on these elements to verify UGC. In addition to these five elements, we describe our proposed sixth verification element i.e., *Multimedia Forensics*. We illustrate these six elements in Figure 1. This figure does not belong to *First Draft News*, and is created by us for the sake of this study.

Some real world examples of media practitioners performing visual UGC verification by following the five steps mentioned above are described in [4]. We have also conducted a set of discussions with journalists and fact-checkers from three media outlets including Bellingcat,[5] Faktisk, Verdens Gang (VG),[6] and Bergens Tidende (BT)[7] to discern how

[4]First Draft News is an independent non-profit organisation focused on fighting mis-/disinformation online. More information at: https://firstdraftnews.org/about/

[5]A Netherlands-based investigative journalism group.

[6]A Norwegian tabloid newspaper.

[7]Norway's largest newspaper outside Oslo.

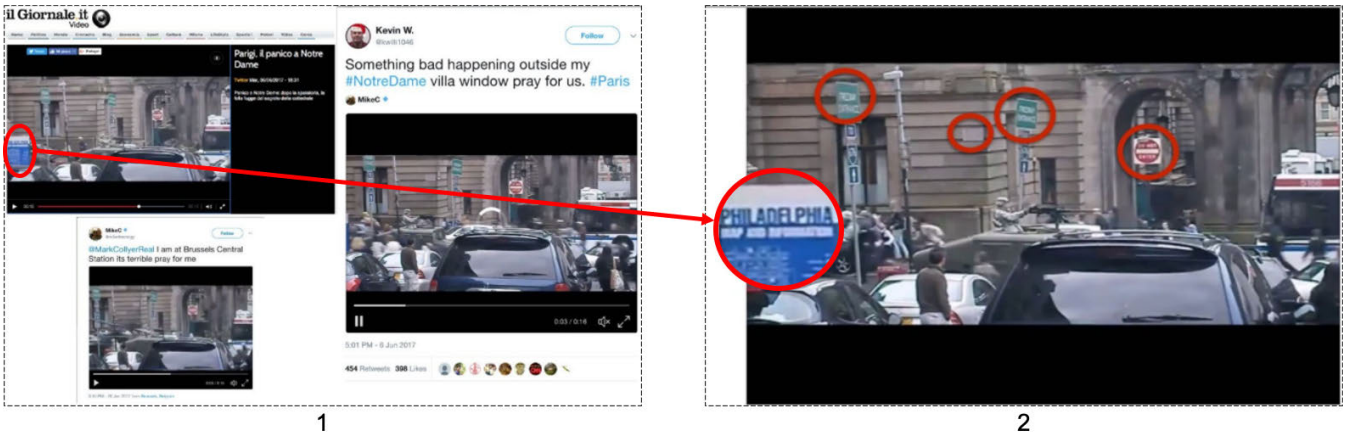


FIGURE 2. An example of verifying the location of an image by reading the signboards present inside the image [11].

their way of UGC verification aligns with the procedure in Figure 1. It reveals that visual UGC verification workflow in practice is fairly similar to the specified steps, but following every single step in the given order is not always the case.

Besides these journalistic workflows and recommended UGC verification strategies, there are other comprehensive conceptual frameworks available in order to analyse and mitigate manipulative content, or propaganda influence campaigns [12]. Some examples are, (1) Carnegie Mellon BEND Framework, (2) The ABCDE Framework [13], [14], [15], (3) The AMITT Framework [16], and (4) The Scotch Framework [17]. Analysts who attempt to interpret and mitigate mis-/disinformation employ these frameworks in real-world scenarios [12]. All of them propose somewhat similar fundamental steps which can be followed in order to uncover organised propaganda campaigns behind spreading mis-/disinformation online. We therefore suggest that following at least one of these frameworks, along with the six elements we are presenting in this paper, will result in a more comprehensive visual UGC verification.

A. PROVENANCE

Provenance is considered as the most important step in UGC verification process [9]. Through provenance, it is established whether the piece of visual content is indeed the original one or has been shared online in the past. It is also worth checking if it is a manipulated version of an image/video shared in the past. Sometimes, images are downloaded from the internet (e.g., social media platforms, websites) and then uploaded again, maybe on a different social media platform or website at a later time. These are called scrapes [9] and makes the provenance even more difficult.

A well known technique journalists and fact-checkers employ to establish provenance of the visual UGC is by carrying out a reverse image search. Reverse image search is the process of using search engines, such as, Google, TinEye, Yandex and others to find similar looking images to the one which is being queried. Browser extensions, for example RevEye, are also useful in finding similar looking images online. Reverse image search is an extremely powerful tool used to find out if a given image/video has been shared online before or not. If an older version of the queried image is found online from an earlier timestamp, this is an instant indication that the image may be re-purposed, presented out-of-context, or misleading [9]. Typically, the image with the highest resolution/size is considered to be the original image, which can help lead the journalists to its source [4].

To carry out video provenance, a similar strategy as reverse image search can be adopted. For example, an individual frame from the video is extracted and then a reverse image search is carried out for that specific frame. The InVID verification plugin [18] can be used to establish video provenance. The plugin is available freely in the form of a Chrome or Firefox browser extension. It makes video provenance easy by offering functionalities such as, breaking down videos into individual frames, extracting video metadata, by using natural language processing algorithms to show any associated comments which can be helpful in verifying the video. The InVID verification plugin also has a magnifier which can be used to read any small text within a video frame, or to analyse other smaller details within the frame.

Besides these strategies, to look for any other relevant information, media practitioners sometimes also look into anonymous platforms, for example, Reddit, 4chansearch.com, Gab.ai, Discord channels, Facebook groups and other similar websites. Looking into these sources is helpful as a variety of UGC, including memes, misinformation sometimes originate from these places [9].

B. SOURCE

Verifying source refers to finding out who originally captured the content (image/video), whereas provenance refers to finding out who uploaded/shared the content for the first time online. This is important because sometimes the content creator and the uploader maybe different, for example, if a person captures a video in Istanbul and send to another person

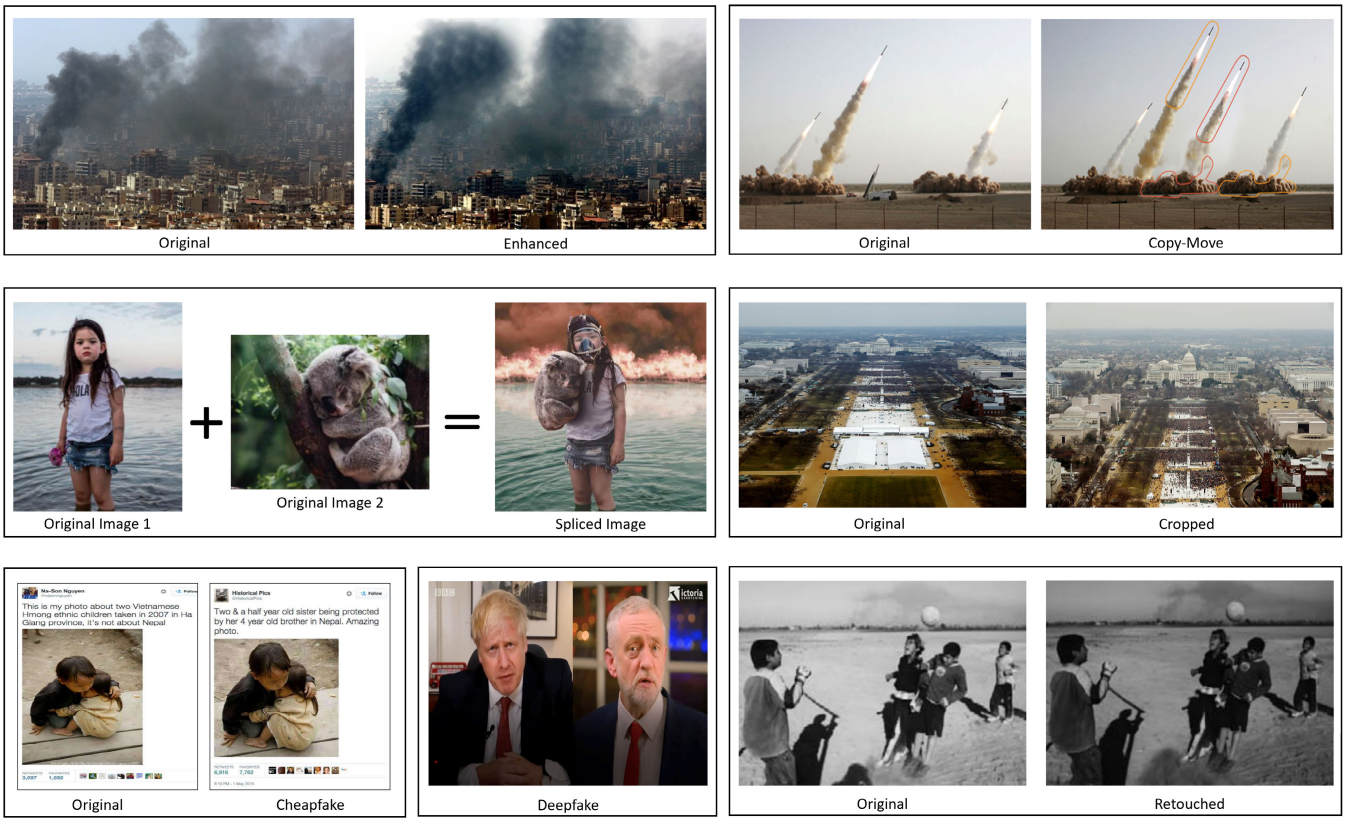


FIGURE 3. Examples of image manipulations employed in the news domain. On top left we have an example of image enhancement [19]. On top right we show an example of manipulated image shared by Iran's Defence Ministry using Copy-Move forgery [20]. On left of the middle row we present an example of image splicing forgery [8]. On right of the middle row we show an example of image cropping manipulation [21]. On bottom left we show an example of cheapfake media [22], where, on the right we have an example of an image which was shared along with an out-of-context caption on twitter, on the left side we show a tweet from the photographer who captured the image originally, stating that the image has been miscaptioned. In the middle of bottom row we have an example of deepfake media [23]. On bottom right we present an example of image retouching [3]. More details on each of the presented example in this figure can be found in upcoming sections.

(e.g., friend, family, colleague) in London who then uploads the video online. The primary source in this case is the person in Istanbul, who initially captured the video.

During the verification process, it is thus crucial for the journalists and fact-checkers to identify the primary source of the content by checking if the uploader is also the source of the content or not. Interviewing the uploader and establishing provenance can help in swift and reliable verification of the UGC item, and can lead to the primary source of the image/video being verified. Typical questions journalists might ask to confirm the identity of source might include (1) when was the image/footage captured, (2) the acquisition device, (3) what they saw on scene of the event, (4) what the source has been doing on scene of the incident, (5) if the source lives nearby etc [4]. In some cases, journalists request the source person to send the image/footage via email, since email services do not compress, strip metadata headers from the file, and thus can help journalists in verifying the source. Journalists might also ask for additional supporting evidence e.g., images or footage if any to confirm whether the person has been actually on the scene [4].

When interviews are not possible, journalists inspect the digital footprint of the uploader (to find out if he/she is the original source) by analysing the associated social media profiles, the kind of posts the person has created/shared in the past, checking if the person has other social media accounts (LinkedIn, Twitter, Facebook, Skype etc), search the web for any other relevant information about the account (email addresses, phone numbers, web-pages) [4], [9]. Investigating associated activities on the web could help in harvesting more details about the individual and reaching the genuine source of the image/video. If the person has a profile photo available, a reverse image search is conducted to retrieve more details.

There are also several tools for gathering more information about individuals on the web using person search engines such as "Pipl" or "Spokeo" [10]. For investigating a specific website, rather than a social media account, look-up tools such "WHOIS", "ViewDNS" or other related domain name search engines are utilised. Tools like "BotSentinel" or "Hoaxy" are employed to detect social media bots. Another useful tool is Twitonomy, a Twitter analytics tool to acquire detailed information about an account, for example, when the account was created, the associated tweet history, the percentage of retweeted tweets, the most used hashtags, to whom they reply the most, average tweet count per day and other

similar statistics. More information about the mentioned tools can be found in Table 3.

C. DATE

Although every social media post has an associated timestamp which tells when the post was created, that timestamp does not tell when the content was actually captured. Besides this, some visuals (scrapes) are uploaded multiple times on different social media platforms and have different timestamps. Finding the true date of creation of a visual item is thus no easy task. Journalists are aware of this and therefore share the date and time of the capture along with the content while publishing. The InVID verification plugin [18] can be used to get the exact upload time in Coordinated Universal Time (UTC) format (if the associated Exif header data is still intact).

Exif data headers are informative in finding out the date and time of acquisition. However, if a piece of content is downloaded from social media platforms the Exif header information might not be available. That is because these platforms drop the information in the Exif header when content is uploaded to save storage space [24]. Journalists might ask the eyewitness or the person who uploaded/shared the content to email the original image/footage in order to verify the Exif information. However, the information in Exif header can be easily modified and therefore it needs to be handled with care.

Journalists also make use of weather tools like ''Wolfram Alpha'' to check weather, or ''SunCalc'' to find the angle of sun on a specific date/time at a particular location. Further information about the mentioned tools can be found in Table 3.

D. LOCATION

When it is possible to interview eyewitnesses, in order to verify the date, time and location of an event, journalists ask direct questions [4]. For further confirmation, journalists and fact-checkers sometimes request more pictures/videos from the witnesses from the scene during the interview or right after it. Having multiple pictures/videos from the scene of incident provide additional details about the location. When interviews are not possible, journalists use computational tools to infer the location by analysing the associated metadata headers. Exif headers can provide vital information for the verification task, for example, the brand and the model of the capturing device, timestamp at which the image/footage was captured, GPS coordinates etc. Tools such as Photoshop or websites like ''Fotoforensics'' can be used to generate Exif reports [4].

UGC posted on social media platforms is often geotagged. However, the geotagged location might not be the same as the location in which the content was captured [9]. Journalists and fact-checkers obtain more information about the location within the image/video using available online software, e.g., Google Maps, Bing Maps, Apple Maps, Wikimapia, Google Earth, and others. Online maps are employed to identify surroundings, specific notable buildings, or other structures present in a shared image/video on an interactive map. The identification task becomes difficult when the buildings or surroundings in an image/video are damaged or destroyed in incidents such as airstrikes, bombings or natural disasters.

Location services like ''Geofeedia'' are also utilised by the media professionals to establish location from which a certain image was shared. To automatically extract text from signboards in images, journalists make use of optical character readers (OCRs) such as Tesseract.[8] If there are shops present in the scene, their names can be searched on online maps e.g., Google maps, Bing Maps to acquire further information. Google Translate or other similar translation services are used when the text on the signboards present in the image/video are in a different language. Other tools and services are also sometimes used in order to verify the location being presented in the image/video, for example, weather services similar to Wolfram Alpha, shadow information (SunCalc), temperature information.

Figure 2 shows an example of how the signboards present in the images can help journalists estimate the location. The images in Figure 2 were extracted from a viral video shared on Twitter. Image 1 on the left, contains two tweets shared on Twitter claiming that the video was captured in (1) Belgium and (2) France. However, when the fact-checkers investigated the video by extracting individual frames and focusing on the signboards as shown in image 2, they found out that the video was in fact captured in Philadelphia, United States. The journalistic process to find the location of an incidence captured in a video is the same as for an image. In addition, the audios associated with any given video also provide valuable information about the location, for instance, by analysing the language or the specific accent/dialect being spoken in the video. The BBC Monitoring Service helps its staff on analysing accents [3].

E. MOTIVATION

Finding out the motivation behind capturing the content and sharing it online is virtually impossible [9]. Journalists can ask basic questions about (1) the reason for being at the site of the incident, whether intentional or unintentional, (2) the person's social media footprint, (3) the person being an activist or not, (4) working for the government or a political organisation [9]. By figuring out at least some these questions, journalists and fact-checkers might end up having a sense of the motives.

F. MULTIMEDIA FORENSICS

Until now, we have described the 5 basic elements of UGC verification that journalists and fact-checkers typically employ, and the computational tools they use to carry out verification. However, we feel that the verification workflow can be further strengthened by adding an additional element into the UGC verification task. Thus in this study, we propose

[8]https://github.com/tesseract-ocr/

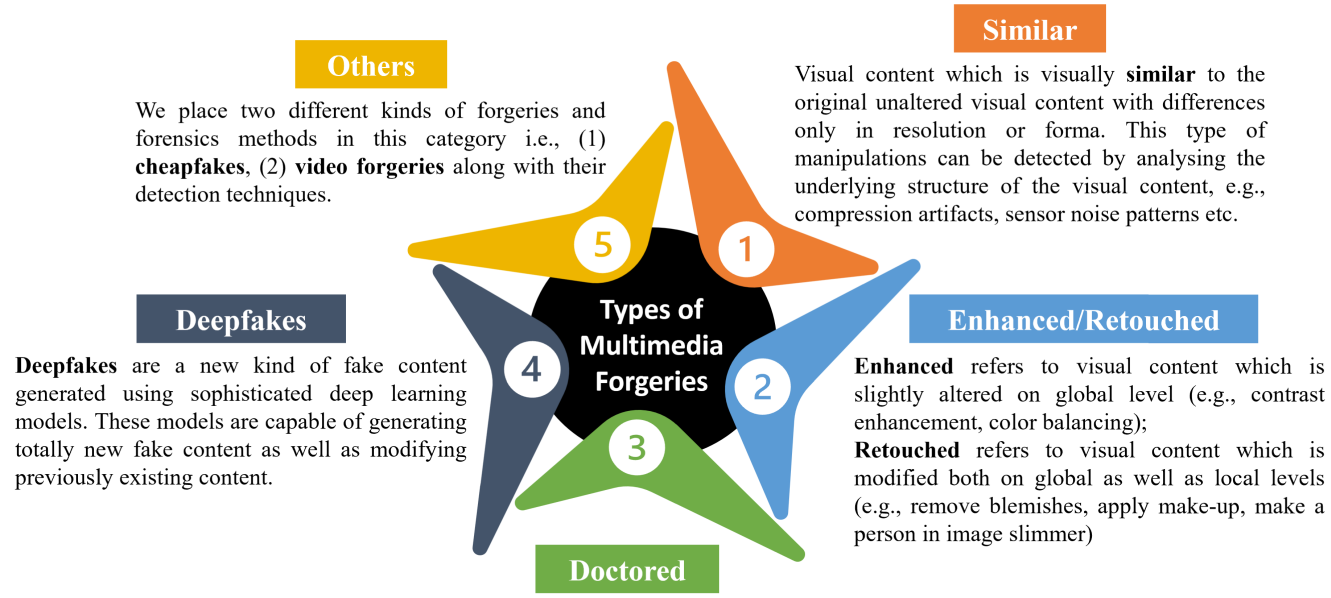


FIGURE 4. This figure categorises 5 different families of multimedia forgeries based on the degree of applied manipulation. In section III we discuss these forgeries and also present forensics techniques aimed at detecting these forgeries. Some content of this figure is adopted from [25].

the sixth element: *"Multimedia Forensics"* to strengthen the verification process.

The first five elements can help verify visual UGC which has been scraped from the web, manipulated and then shared again online. However, the tools (e.g., reverse image search, online maps or geo-location tools) used in the first five elements are not designed to verify manipulated content surfacing online for the very first time (until it is debunked, which will of course take some time). Through the sixth element, we suggest the use of image/video forensics tools which can help with detecting multimedia forgeries, for example copy-move or splicing.

Fact-checkers sometimes find themselves in trouble while verifying newly surfaced visual UGC. According to the interviews we conducted, even after successfully localising the location being depicted in the image/video using digital tools, to verify if the image/video is genuine or manipulated, is not a simple task. It's true that there are multimedia forensics tools available for verifying visual UGC, however at present, their widespread use within the news media organisations is not evident.

There are a number of image/video forensics tools available online which can help uncover manipulated visual UGC for example, FotoForensics, Forensically, Ghiro, DeDigi, WeVerify, InVID, MeVer.[9] For deepfake media detection web-based tools such as, "Deepware.ai", "DuckDuckGoose.ai" are available which can be used to debunk newly surfaced deepfake media. Context based visual UGC verification tools such as, Journalistic Decision Support System (JDSS) [26], Context Aggregation and Analysis Tool [27] are also available which are able to provide contextual information about a given UGC item at one place.

In the next section, we present an insight on some categories of multimedia forgeries and the available forensics solutions. Our aim is to present an insight on where the computer science research community stands in the fight against visual mis-/disinformation, what kind of tools/solutions are available and what is needed in the future.

III. STATE OF THE ART IN MULTIMEDIA FORENSICS

In this section, different visual content forgeries and forensics techniques proposed by the community of computer science researchers are presented. We also present some examples from the past where manipulated visuals were employed to spread mis-/disinformation online. See Figure 3 and Table 1 for reference. We categories visual content forgeries into five categories as shown in Figure 4 based on the degree of applied manipulations as proposed in [25]. These categories are:

- Similar: Images or videos visually similar to the original unaltered visual content with variations only in resolution or format are placed in this category;
- Enhanced/Retouched: Image enhancement operation is typically carried out globally on an entire image, for example boosting the color of image in order to make it look more pleasing, or enhancing the

[9] A list of these tools can be found in Table 3

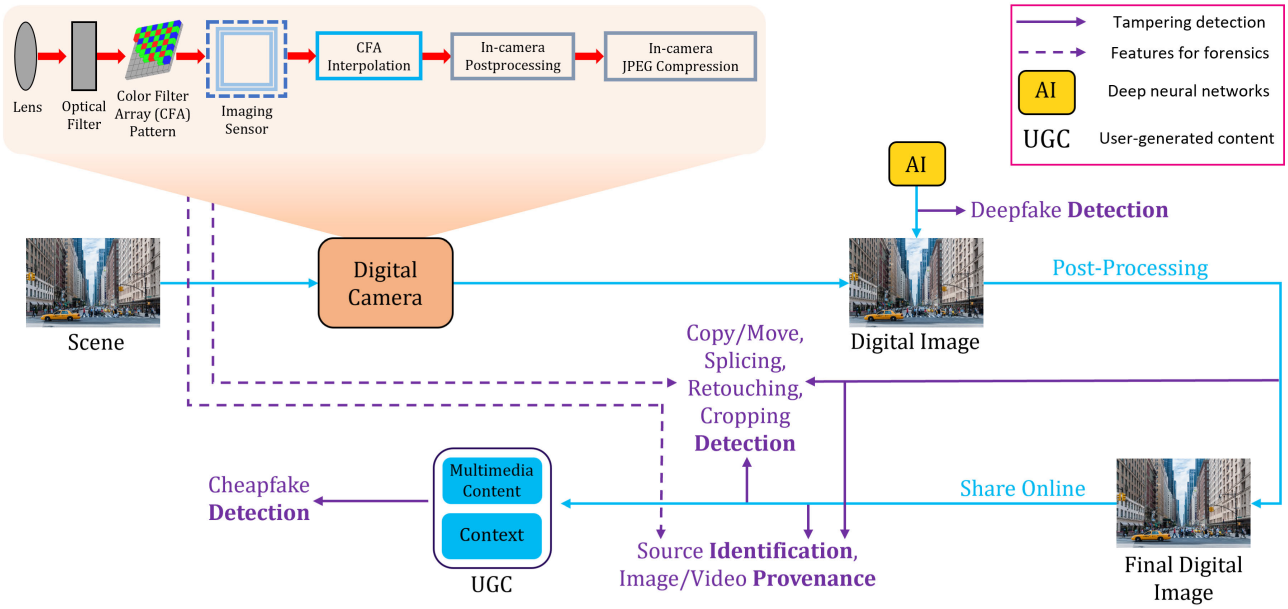


FIGURE 5. This figure illustrates the digital life-cycle of a visual content item (image, video) and the stages at which forensics operations can be applied to detect tampering the image/video might have undergone i.e., represented using solid purple lines. Dotted purple lines show the stages from which helpful features can be acquired to detect different forgeries, e.g., CFA interpolation patterns are used to identify the make/model of the capturing device [28], or in [29] sensor noise patterns were employed to detect image splicing forgeries.

contrast/brightness/saturation etc to make it look more attractive to the eyes. Image enhancement operation can also be employed to make minor corrections in order to highlight or suppress certain artifacts within an image, or to make an image look more dramatic (as can be seen on the top left corner of Figure 3). Generally, the enhancement operation is not performed with a malicious intent, e.g., it is not employed to change the semantics of the image, however, we do have some examples (in the upcoming sections) where this operation is used with a somewhat malicious intent;

Retouched: Similar to image enhancement operation, image retouching is also usually employed without having any malicious intent behind it. For example, image retouching operation is often used to eliminate imperfections from an image, such as removing blemishes, under-eye circles from a face. The idea behind using this operation is also to make the photos look better, however, although less frequently this operation can be employed with a malicious intent to hide, or misrepresent information being conveyed in the image. One difference between image enhancement operation and the image retouching operation is that, image retouching operation can alter local as well as global details within an image, whereas, image enhancement operation only alters global details of the image. Because of the fact that these two operations are not very much different, in this study we present Enhanced and Retouched as a single category;

- Doctored: Visual media altered using sophisticated editing techniques (e.g., copy-move, splicing) that change the semantics of the original visual content item and/or produce something different from the original data belong to this group;
- Deepfakes: A new class of fake media which is generated using deep neural networks is called deepfake media. The deepfake generation models can generate totally new fake content, as well as, they are capable of manipulating already existing content. The deepfake generation models are not only capable of generating visual content, but they can also generate audio, textual content as well. However, in this study we will mainly focus on visual deepfakes; and
- Others: In this category we present two different types of visual content forgeries i.e., (1) cheapfakes and (2) video forgeries. Cheapfakes refer to multimedia content produced using ''cheaper'', and more user-friendly tools (or in some cases, no tool is required at all) such as, Photoshop, Gimp, Final Cut Pro etc [30].

The following sections describe these categories. See [12] for an in-depth understanding of some of the concepts presented in this section.

A. SIMILAR

In visual UGC verification, variations in compression or scaling images/videos seem unaltered in human eyes. Multimedia forensics tools analyse the underlying structure of the visual content by analysing compression artifacts, sensor related artifacts of the capturing device and available metadata information.

TABLE 1. A summary of multimedia problems presented in Section III. We also list suitable forensics techniques, as well as available tools to detect/debunk these forgeries. Some of content in this table is inspired from [8]. An analysis of the tools can be found in Table 3.

Modification Category	Problem	Examples from the News Domain	Forensics Techniques	Tools
Similar	Source Identification	A video game clip was mis-captioned and shared on social media platforms in the context of Russian invasion of Ukraine. The computer generated clip claimed to show "Ghost of Kyiv", a fictitious Ukrainian fighter pilot shooting down a Russian fighter jet [31].	Source identification is carried out by analysing metadata information, CFA interpolation patterns, sensor noise fingerprints, JPEG compression artifacts. Deep CNN models have also been employed for the source identification task.	MeVer Image, InVID, Ghiro, FotoForensics, Forensically, DeDigi, Online Exif Viewer, exifdata, YouTube Data-Viewer
	Image/Video Provenance	An image went viral on social media in 2021 claiming to show a heart-shaped sunset over a beach. The image was found to be mis-captioned, and the original image (digital artwork) was actually posted on Instagram by a user in 2020 [32].	For provenance analysis metadata information, noise fingerprints, DCT features are used to train statistical models. Deep learning models are also proposed for provenance analysis.	MeVer Image, InVID, Ghiro, FotoForensics, Forensically, DeDigi, Google/TinEye Image Search
Enhanced/ Retouched	Retouching	US President Donald Trump's official Facebook and Instagram handles shared his edited photos to show him with a tightened waistline, elongated fingers, a slimmed neck and shoulder, higher crotch and tightened hair [8].	Retouching forgeries are typically detected using noise patterns, histogram analysis. Deep CNN models are also used to detect these forgeries.	MeVer Image, InVID, Ghiro, FotoForensics, Forensically, DeDigi, Google/TinEye Image Search
	Cropping	During the inauguration ceremony of US President Donald Trump, the White House cropped official photos in a way that made the crowd seem larger. For reference, see Figure 3.	Cropped images are normally multiple compressed, they can be detected by analysing the image compression qualities, image histogram, or blocking artifacts. Deep learning models are also proposed to detect image cropping.	MeVer Image, InVID, Ghiro, FotoForensics, Forensically, DeDigi, Google/TinEye Image Search
Doctored	Copy-Move	Sepah News, owned by Iran's Revolutionary Guards posted forged images using copy/move forgery to show four missiles, instead of the original 3. The image was edited by copying and pasting one of the missiles from the original image itself [20].	Two widely used detection methods are, (1) Block matching based method exploiting DCT and DWT features; and (2) Key-point matching based methods exploiting SIFT, SURF features to detect manipulated images. Some approaches use deep learning models as well.	MeVer Image, InVID, Forensically, Google/TinEye Image Search
	Splicing	A popular photo from G20 summit held in Hamburg, Germany in 2020 was a result of the image splicing forgery. The photo showed Donald Trump and other prominent world leaders surrounding Putin, looking towards him as if they were all listening to something important from him [33].	Diverse range of features are used to detect image splicing forgeries, for example, CFA interpolation artifacts, JPEG compression artifacts, noise patterns to detect and localise spliced image regions. Deep learning models are also used to detect and localise splicing forgeries in images.	MeVer Image, InVID, Ghiro, FotoForensics, Forensically, DeDigi, Google/TinEye Image Search
Deepfakes		In 2019, a deepfake video produced by two artists Bill Posters and Daniel Howe along with an advertising firm showing Mark Zuckerberg, founder of Facebook saying things, in reality he never said [34].	Deepfakes are typically detected using deep CNN models trained on large amounts of image data. Recurrent neural networks, and transformer models capable of learning temporal associations are also used to detect deepfakes. Diverse set of features such as, 3D decomposition, biological features, or optical flow are used to train these models.	Deepware, MeVer Video, Google/TinEye Image Search
Other	Cheapfakes	In the aftermath of the 2015 earthquake in Nepal, an image was circulated on the internet under the guise that it was an image captured in Nepal . The picture was actually captured in Vietnam in 2007 [35].	Deep networks trained separately on images and associated text captions are proposed to detect cheapfakes/multimodal fake news. Transformer models, for example, ViLBERT, VL-BERT can also be used to detect cheapfakes	MeVer Image, InVID, Ghiro, FotoForensics, Forensically, DeDigi, Google/TinEye Image Search, Snopes
	Video Forensics	A manipulated video of the House Speaker Nancy Pelosi was widely shared across different social media platforms including, TikTok, Facebook. The video was manipulated by slowing down the frame rate which made Nancy Pelosi's appear drunk [36].	Using diverse set of features, for example, Gray Level Co-occurrence Matrix (GLCM), Peak-signal-to-noise ratio (PSNR), Histogram of Oriented Gradients (HOG), optical flow. New approaches propose to use deep CNN models.	MeVer Video, InVID, DeDigi, Google/TinEye Image Search, YouTube Data-Viewer

1) SOURCE/CAMERA IDENTIFICATION

Source/camera identification relates to finding out information e.g., make, model of the capturing device. Sometimes this can be achieved by simply analysing the associated metadata information, however, for images/videos shared online such information is often stripped to save storage. To cope with this, researchers propose to employ features inherent to the underlying properties of the image/video. Such features result from different phases of the digital image acquisition process which takes place inside digital cameras/capturing devices. A simplistic overview of the digital image acquisition process is shown in Figure 5. For the source camera identification task along with metadata information, features such as sensor noise patterns, CFA interpolation artifacts, and

compression artifacts are employed by the experts to analyze image/video under question, as depicted in Figure 5.

A straightforward technique to identify the source/camera of an image is to analyse it's Exif (Exchangeable Image File) header. Some useful details about the image and acquisition device are saved in the Exif headers, for example, make and model of the device, image resolution, exposure settings, date/time of acquisition, and some other relevant details [37]. However, typically when an image is uploaded online or shared on a social media platform, the platform strips out the Exif header data to save memory [24]. Besides, the information present within the Exif header cannot be trusted in critical cases (e.g., police investigations, court proceedings) since it can be easily modified.

To address this problem, researchers proposed several innovative solutions to infer information about the source/camera properties of a given image. A diverse set of features inherent to a capturing device based on the artifacts produced during image acquisition process including Sensor Pattern Noise, CFA (Color Filter Array) interpolation, JPEG compression artifacts etc. are employed [37], [38], [39].

In [39], JPEG compression statistics are employed for source camera identification. Since different camera manufacturers employ different compression strategies considering the trade-off between the image size and quality, the authors argue that it is possible to classify images based on JPEG compression artifacts.

Machine learning approaches like Support Vector Machines (SVMs), Expectation Maximisation (EM), and Clustering algorithms are trained on these features extracted from images to identify the acquisition device [28], [40], [41], [42], [43], [44], [45].

In [28], authors proposed to employ CFA configuration and the associated demosaicing algorithm for source camera identification. Altogether, authors proposed 34 different features and trained a SVM classifier to classify camera make and model. In [45], authors proposed to cluster images from same capturing device together using on PRNU noise residuals using correlation clustering approach. Authors argue that since noise residuals of images coming from the same device possess a somewhat larger correlation as compared to the noise residuals of images coming from unrelated devices. This property can be leveraged for source camera identification task.

Below we provide a mathematical formulation of how sensor noise can help distinguish between images captured from different devices. To start the process, the images are denoised using any available denoising filter. The denoised version of the image is then subtracted from the original image as follows [46]:

$$W_k(x, y) = I_k(x, y) - \hat{I}_k(x, y) \tag{1}$$

In equation 1 above, $I_k(x, y)$ refer to the original images, $\hat{I}_k(x, y)$ refer to denoised version of the original images, where $k = 1 \ldots N$. The term $W_k(x, y)$ helps suppress the underlying content of the images and makes the PRNU noise estimation more effective [46]. The PRNU noise is then estimated as given in equation 2.

$$K(x, y) = \frac{\sum_{k=1}^{n} W_k(x, y) I_k(x, y)}{\sum_{k=1}^{n} I_K^2(x, y)} \tag{2}$$

The PRNU $K(x, y)$ can then be used to determine the specific device used to capture the image $I(x, y)$, i.e., by comparing the estmiated PRNU of the image $I(x, y)$ with available PRNU estimates from a dataset of images captured using a number of different devices. The PRNUs having a correlation more than a certain pre-defined threshold can be considered as resulting from the same device. The following equation 3 presents the correlation ρ as given in [46].

$$\rho = I(x, y)K(x, y) \otimes W(x, y) \tag{3}$$

In [47], a deep convolutional neural network (CNN) model was employed to carry out the source identification task for images captured using mobile devices. Study [48] proposed content-adaptive fusion residual networks for source camera identification on small-sized images. An efficient source camera identification method based on modified deep CNN (VGG[10]) network was adapted in [49].

The source/camera identification methods can be divided into two categories: (1) Perfect Knowledge Methods and (2) Limited/Zero Knowledge Methods [24]. These methods are briefly described below:

1) **Perfect Knowledge Methods:** Perfect knowledge methods carry out the source identification task while having a closed dataset containing reference camera fingerprint from a number of different camera makes and models.
2) **Limited/Zero Knowledge Methods:** Limited/Zero knowledge methods consider limited prior information about camera properties, or use small datasets having less details about the capturing devices.

2) IMAGE/VIDEO PROVENANCE

Image/video provenance concerns determining the last web/social media platform where the visual content was shared. Platform provenance analysis is an important step in visual content verification because it can help establish the full life cycle of the UGC item of interest.

Various research studies have been conducted in the past for both image, and video provenance analysis, using forensics techniques. Researchers rely on features obtained by signal processing methods i.e., noise residuals, DCT coefficients, or by using metadata information [50], [51], [52], [53]. A diverse set of machine learning and deep learning classifiers such as SVM, Logistic Regression, Decision Trees, Random Forests, and CNNs have been proposed in the literature for platform provenance analysis. In study [54] it was shown that for smartphones, the JPEG headers are to a certain extent useful in identifying the operating system, and sharing application.

[10]Visual Geometry Group (VGG) is a standard deep CNN architecture.

Study [52] proposed a social media platform provenance technique using ensembled convolutional neural network (CNN) architectures called, FusionNET. Authors employed diverse features for the provenance task, such as, (1) histogram of DCT coefficients, (2) noise residuals. Appending multiple features such as PRNU (Photo Response Non-Uniformity) to the DCT features improves classification accuracy. A video provenance network (VPN) which utilises both video and audio features is proposed in [55]. In study [56] a novel multi-branch CNN architecture called MultiFrame-Net was proposed to find the social network from which the video under analysis originated.

B. ENHANCED & RETOUCHED

Image enhancement and retouching operations manipulate the visual content in subtle ways. Contrast enhancement, sharpening and cropping operations fall into this category. In most cases these operations are not carried out with a malicious intent to deceive the audience, however, in some cases these operations can be employed to deceive by altering the semantics of the visual content.

1) ENHANCEMENT/RETOUCHING DETECTION

Image enhancement makes minor corrections to highlight or suppress certain artifacts within an image, often without any malicious intent. An example of image enhancement with a rather malicious intent is presented at the top left corner of Figure 3. The original image is on the left, and the colour enhanced image is on the right. The photographer darkened smoke to make destruction from an airstrike look more catastrophic [19]. After discovering the manipulation, Reuters news agency refused to work with the photographer who captured and enhanced this image.

Image retouching is similar to image enhancement to some degree. However, the retouching operation may be used to alter subtle global as well as local details within an image. In case of facial images, retouching operation might be employed to remove acne, blemishes, or scars. Normally, the retouching operation is harmless, as it does not conceal or misrepresent the information within an image [8]. A somewhat problematic image retouching operation is given at the bottom right of Figure 3 where the photographer removed his shadow from the photo [3]. The photographer was consequently dismissed for editing the photo [3].

In [57], a blind image forensic method to detect global contrast enhancement operations used to modify images by analysing their histograms was proposed. Study [58] proposed a facial image retouching detection technique by using spatial and spectral features obtained from PRNU noise fingerprints. The same author have suggested to detect facial retouched images using a differential detection system in [59]. The proposed system compares a (suspected) retouched image with a genuine reference image by using a number of different features such as texture descriptors, deep face representations, and face landmark data. In [60], a deep CNN model was employed to automatically detect warping (retouching) operation applied to human faces using Adobe Photoshop. In [61] a deep CNN model was proposed to detect GAN-based synthetic image alterations. In addition to the CNN model, authors employed two different algorithms for classification namely, (1) SVM and (2) thresholding.

2) CROPPING DETECTION

Cropping operation is typically carried out in order to remove unnecessary parts around the corners of an image or a video frame. Cropping forgeries are not as common as other kinds of image forgeries (copy-move, splicing) and are mostly considered harmless. However, they can be employed to spread mis-/disinformation [62], [63] when there is a clear intent to deceive audience by concealing information or distorting facts presented within an image, i.e., to shroud objects or conceal the wider perspective [8]. For example, in 2017 during the US president Donald Trump's inauguration ceremony, the White House pressurised the US National Park Services to crop out empty spaces from images and publish the cropped version of images where crowd was present [63]. For reference see Figure 3.

Cropping operation in JPEG images can be identified by detecting artifacts resulting from multiple JPEG compressions. Study [65], proposed to detect JPEG re-compression by using histogram discontinuities, periodic artifacts resulting from image re-quantisation process. In [66], a method to detect cropping and re-compression operations within JPEG images using blocking artifact characteristics matrix (BACM) was proposed. Cropping operation disturbs the symmetry of the BACM, and thus it can be employed as evidence to detect double compressed JPEG cropped images [67]. Study [68] devised a fully automated cropping forgery detector for images cropped asymmetrically by estimating the camera principal point. Analysing block artifact grids (BAG) which result from block processing during JPEG compression is another approach to investigate image cropping forgeries [69]. In [70], a method to detect upscale cropping operation in surveillance videos using sensor pattern noise (SPN) features was proposed. Mellin radial harmonic (MACE-MRH) correlation filter was used to unveil indications of upscaling. By omitting the high-frequency components of the video under investigation, and deciding the size of the local search window, this technique localizes partially tampered regions in an effective manner.

C. DOCTORED

Major of image forensics techniques developed in the last decades are dealing with revealing the sophisticated image modification techniques (e.g., copy-move, splicing) that change the semantics and/or produce something different from the original visual content.

1) COPY-MOVE DETECTION

Copy-Move image forgery is carried out by copying a specific region from an image and pasting the copied segment elsewhere in the same image [71]. Copy-Move forgery is

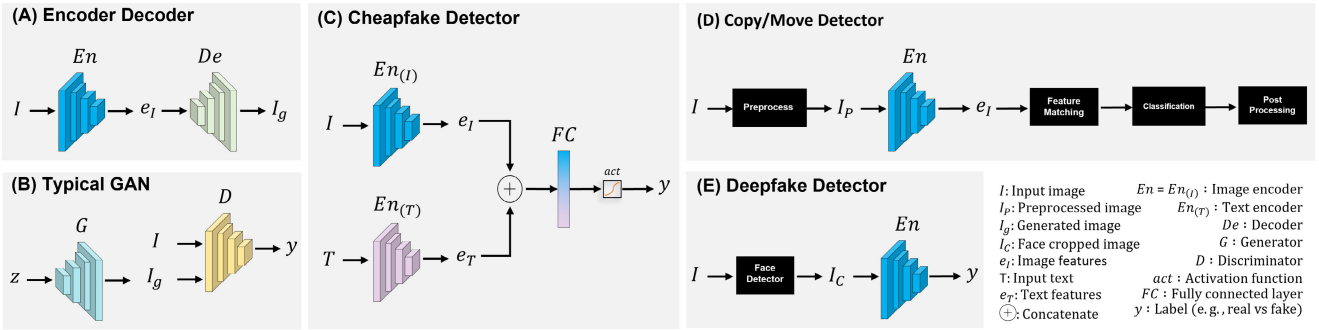


FIGURE 6. In this figure, on the left side we present two different methods employed to create deepfakes, e.g., (A) Encoder-Decoder netwroks, and (B) Generative adversarial networks or simply, GANs [64]. We also illustrate basic pipelines of forgery detectors employing deep networks for feature extraction and classification, e.g., (C) shows a basic multimodal cheapfake media detector, (D) copy/move forgery detector and localizer, and (E) a deepfake detection system.

carried out to hide something within an image or to increase the number of objects present in an image. For instance, take the popular fake Iranian missiles photo, in which copy-move forgery was used to hide a miss-fired missile with a fired missile. For reference, see Figure 3.

There are two different families of copy-move forgery detection techniques including (1) block matching-based techniques and (2) keypoint matching-based techniques. Block matching-based techniques divide the image into smaller overlapping blocks. Features are extracted from the resulting blocks and matched in order to identify duplicated regions within an image [37]. Block matching based copy-move forgery detection techniques employ Discrete Cosine Transform (DCT) features among others [72]. To apply DCT on an image, the image is first divided into N x N blocks (typically $N = 8$). Equation 4 shows how the DCT for the i^{th} j^{th} entry of an image is computed [73].

$$D(i,j) = \frac{1}{\sqrt{2N}} C(i)C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} p(x,y) cos\left[\frac{(2x+1)i\pi}{2N}\right] cos\left[\frac{(2y+1)j\pi}{2N}\right] \tag{4}$$

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}} \text{if } u = 0 \\ 1 \text{if } u > 0 \end{cases} \tag{5}$$

where $p(x, y)$ represent the x^{th}, y^{th} elements of the image as given in matrix p. For copy-move detection, these block are sorted and matched with other blocks of the image to detect any matching blocks i.e., having a correlation greater than the specified threshold. Correlation for a pair of sorted blocks can be calculated as below:

$$Corr = \frac{\sum_{i=1}^{n}(px_i - px_{mean})(py_i - py_{mean})}{\sqrt{\sum_{i=1}^{n}(px_i - px_{mean})^2 \cdot (py_i - py_{mean})^2}} \tag{6}$$

In the equation 6 above, px and py represent the two blocks, whereas n represent the number of coefficients within the block [72].

However, these techniques are computationally expensive [38]. Some studies proposed to employ dimensionality reduction techniques such as Principal Component Analysis (PCA) to reduce the feature space resulting in lower computational complexity [71]. In [74], PCA was carried out on DCT features to detect copy-move forgeries, reducing the computational complexity while achieving a higher robustness against noise and compression. A similar strategy was followed in [75] which used Singular Value Decomposition (SVD) for dimensionality reduction. To detect copy-move forgeries using block matching methods, researchers exploit a diverse set of features including Discrete Cosine Transform (DCT) [74], [76], Discrete Wavelet Transform (DWT) [75], [77], and Fourier-Mellin Transform (FMT) [78].

Unlike the block-based techniques, keypoint-based techniques extract features from certain regions in the image having high entropy rather than the whole image [79], thus reducing the computational complexity. Keypoint-based techniques are typically robust against geometric transformations and rely on features such as Scale Invariant Feature Transform (SIFT) [80], [81], [82] and Speeded-Up Robust Features (SURF) [83], [84] for the detection task.

More recently, researchers have started to employ deep learning models capable of automatically extracting useful features from images in order to detect copy-move forgeries [85], [86], [87], [88], [89], [90]. Figure 6 D shows simple copy-move detection pipeline.

2) SPLICING DETECTION

In image splicing forgery, a segment/block from a given *source* image is copied and pasted inside the *target* image [71]. Spliced images possess various artifacts for example, differing noise patterns, multiple colour distributions, abnormal dynamic range, lightning inconsistencies etc. This happens because the image is spliced using segments from the source image having different noise, dynamic range, and colour distribution as compared to the target image, thus introducing irregularities within the target image's statistics [38]. An example of image splicing is given at left side of the middle row in Figure 3. The photo shown on the right was spread far and wide on social media during the Australian bushfire crisis in 2019-2020. It was later found that the image

was a spliced version of multiple other images, as we show in the Figure 3.

To detect image splicing forgeries, diverse set of features such as noise residuals, CFA interpolation artifacts, and JPEG compression artifacts are employed. The same formulation of PRNU noise as given in the equation 2 presented above can be used to analyze images for tampering. However, in this case instead of estimating PRNU $K(x, y)$ for images captured using different devices, PRNU $K(x, y)$ is estimated only for the set of images known to have captured using the same device as $I(x, y)$. The correlation ρ given in equation 3 can then be used to estimate authenticity of the image [46].

Sensor pattern noise (PRNU) fingerprints to detect image forgeries including image splicing and copy-move forgeries was proposed in [29]. Popescue and Farid analyse Color Filter Array (CFA) interpolation inconsistencies emanated by the tampering operations, to detect image splicing and copy-move forgeries [91]. In [92], an approach to detect image splicing forgeries by detecting JPEG ghost, which appears when the two images (source, target) are compressed using different quantisation amounts was proposed. In [93], Markov features acquired from DCT and DWT coefficients are used to train an SVM classifier to detect image splicing forgeries. In [94], a technique to detect image splicing forgeries by analysing lighting inconsistencies within the images was proposed.

Some of the notable image splicing detection studies are presented in [95], [96], [97], [98], [99], [100], and [101]. Since recently researchers have started to employ deep learning based models for image splicing detection [102], [103], [104], [105], [106].

D. DEEPFAKE DETECTION

In the previous sections, we have outlined classical image forgeries. Nowadays, with the availability of enormous compute power at low cost and with the development of sophisticated deep learning models, producing realistic fake multimedia content known as *deepfakes* is becoming prevalent. Deepfakes are not (until now) a widely popular form of UGC mis-/disinformation at present, however, according to the journalists and fact-checkers we have consulted, it can be said that they have the potential to become problematic in the future. While we can say that the deepfakes as mis-/disinformation are not popular, they are still around us in the form of TikTok, Instagram filters which people use to add different types of effects (makeup etc) to their faces. These filters are also driven by the deep neural networks.

According to [64], deepfakes can be defined as, *"Believable media generated by a deep neural network"*. The term Deepfakes is a combination of two different words, *"deep learning"* and *"fake"*, referring to manipulating/producing fake realistic multimedia content including, images, videos, text and audios. Deep learning models such as Autoencoders [107] and Generative Adversarial Networks (GANs) [108] are typically used to generate realistic deepfakes.

Contemporary deepfake generation methods usually employ GANs. The generative adversarial network, or simply GAN, is comprised of two different networks i.e., (1) a generator, and (2) a discriminator [108]. As evident from the name, the GAN is trained in an adversarial manner, where the generator tries to fool the discriminator by generating plausible (fake) data samples similar to the training data. The discriminator on the other hand tries to differentiate between the (fake) samples produced by the generator network from the ones in the training set (real samples).

Simply put, the generator and the discriminator networks play the so called min-max game [108], which is defined by the following equations 7 and 8. The discriminator is trained so that it tends to maximize the function given in equation 7. Alternatively, the generator is trained in a way so that it tries to minimize the function in equation 8, i.e., by generating more plausible data samples similar to the data distribution in the training set.

$$\mathcal{L}_{adv}(D) = \max \log D(x) + \log(1 - D(G(z))) \quad (7)$$

$$\mathcal{L}_{adv}(G) = \min \log(1 - D(G(z))) \quad (8)$$

In equations above, x refers to real data sample, z is the latent vector, $G(z)$ refers to the fake data produced by the generator G, $D(x)$ is the prediction of discriminator D for real sample, $D(G(z))$ is the prediction of the discriminator of fake data [108]. After being trained for a large number of epochs, the generator is able to fool the discriminator by generating extremely plausible fake data, as can be seen in Figure 7.

Deepfakes extend further than just the visual content (images/videos), for example, in [111] it was shown that how generative networks can be employed to tamper medical evidence such as, MRI and CT scans. In 2019, a UK based energy firm's CEO was scammed for $250k [112], by using a voice cloning deepfake algorithm similar to the one proposed in [113]. Besides this, it has been shown that the generative models are capable of generating synthetic news articles and tweets [114], [115].

Deepfakes have the potential to be used to spread mis-/disinformation online and disrupt peace. In 2019, a video went viral on social media in which Boris Johnson and Jeremy Corbyn where seen endorsing each other for Prime Minister [23]. For reference, see Figure 3. Recently, amidst the Russian invasion of Ukraine, a deepfake video of Ukrainian president went viral on social media platforms [116].

Deepfake detection is a challenging task and a lot of studies have been proposed in the past to detect deepfake media employing diverse set of features for training deep learning models. Some examples include, biological signals, behavioural features, 3D face decomposition features, and optical flow [117], [118], [119], [120], [121], [122], [123],

FIGURE 7. This figure presents different types of GAN generated high quality synthetic content including human faces [109], birds [110], buses [109], indoor and outdoor scenes [109].

[124], [125]. In most cases the proposed systems consider the deepfake detection task as an n-class classification (typically $n = 2$, e.g., fake or real) problem. To train the classification model, majority of the proposed systems mentioned above employ cross-entropy loss as defined in equation 9.

$$\mathcal{L}_{CE} = -\sum_{i=1}^{|X|}\sum_{c=1}^{n} y_i[c]\log(y_i'[c]) \tag{9}$$

where X represents the training set, $y'[c]$ refers to the predicted probability for a given sample x_i of class c. Figure 6 E presents a simple overview of deepfake detection pipeline.

For more general deepfake media detection, in [126] Zhang et al. proposed to exploit unique artifacts which result from the up-sampling operation present in most of the common GAN pipelines. In [127], several different classical as well as deep learning based fake content detectors [128], [129], [130], [131] were employed to detect GAN generated images found on social media platforms. In [132], a techniques employing co-occurrence matrices extracted from the pixel domain for all of the three colour channels to train deep convolutional neural network to detect GAN generated images was proposed.

E. OTHERS

1) CHEAPFAKE DETECTION

The term ''Cheap Fakes'' was initially coined in 2019 [30]. Cheapfakes are manipulated media produced to spread fake news and misinformation/disinformation. Examples of cheapfakes can be, (1) photoshopped images, (2) slowing down, speeding up, and/or cutting video frames, and (3) re-contextualising genuine visual content by presenting it along with falsified textual captions etc.

Cheapfakes are created/manipulated using freely accessible editing tools such as, Photoshop or GIMP, unlike the deepfakes which are produced using sophisticated deep learning tools, and require technical expertise which makes them more prevalent online [133]. In case of re-contextualised cheapfakes, sometimes genuine images are presented along with false/out-of-context textual captions, thus requiring no editing tool to generate this type mis-/disinformation. For

example, shortly after the 2015 earthquake in Nepal, an image with two children, a brother and a sister went viral on the internet claiming to be captured in Nepal. The picture was originally captured in Vietnam in 2007. The image itself was not manipulated, but presented out-of-context [35]. The mentioned picture is given at the bottom left corner of Figure 3.

Typical deep learning based cheapfake detection systems usually comprise of two different deep neural networks, i.e., (1) an image CNN to extract image features, and (2) a text CNN for textual feature extraction. The extracted multimodal features are then fused togather in order to get final classification score. In [134], a self-supervised learning strategy to train neural network models to detect out-of-context captions associated with images was proposed. Authors also open-sourced a considerably large dataset comprising of around 200K images and 450K captions for further research in the domain. A neural network based system for multi-modal (image and text) fake news detection was proposed in [135]. ''FauxWard'', a novel framework based on graph convolutional neural network able to learn heterogeneous information extracted from a social media post's user comment network in order to effectively detect misleading information shared online was proposed in [136]. In [137], Khattar et al. proposed an autoencoder based fake news detection model, relying on both textual and visual content.

2) VIDEO FORENSICS

Video forensics is somewhat different than image forensics because unlike images, videos also carry temporal information along with spatial information. Video forensics techniques are divided into two categories, (1) inter-frame techniques, and (2) intra-frame techniques. To deal with temporal information, inter-frame video forensics techniques are employed. The intra-frame video forensics techniques are almost similar to the image forensics techniques as they deal with individual frames, and does not analyse the temporal information of the video. We briefly describe the two forgeries below.

Inter-frame Video Forgery Detection

Inter-frame video forgery is carried out in the temporal domain, for example, (1) frame insertion, (2) frame deletion, (3) frame shuffling, and (4) frame duplication. Typically, the inter-frame forgeries are employed to tamper, twist, conceal, or falsify the information present inside a video.

A number of different techniques were proposed by the scientific research community to detect inter-frame video forgeries by utilising diverse set of features as described in [138], for example,

- Compression Artifacts: Compression related artifacts/abnormalities are used to detect the traces of forgery applied to the video.
- Noise Artifacts: Sensor noise fingerprints are analysed to detect traces of forgery.

TABLE 2. Percentage of the newsrooms, journalists use verification and fact-checking tools, at least weekly from journalists and news managers from 149 countries, reported by International Centre For Jounralists (ICFJ) in 2019 [5].

Tool & Use Case	Newsrooms	Journalists
Tools used to identify trustworthy news sources (e.g., Google Fact Check Tools, Facebook Fact Checker)	49%	52%
The use of reverse image search to identify the source of photos (e.g., Google Image Search, TinEye)	48%	40%
Use of fact-checking websites (e.g., Factcheck.org, Politifact)	41%	38%
Use of tools to verify photos and videos (e.g., Google Earth Pro, Tin Eye)	40%	35%
Tools to detect plagiarism (e.g., Grammarly, Copyleaks)	38%	30%
Use of social media analytics platforms (e.g., Storyful, Dataminr)	36%	25%
Recording apps for interviews with sources (e.g., Google Automatic Call Recorder)	36%	33%
Tools to identify fake news websites (e.g., KnowNews)	32%	25%
Consult fact-checking and verification resources (such as Verification Handbook, Verification Junkie, First Draft News)	32%	24%
Tools to track and find contact details of content uploaders (e.g., Pipl)	27%	19%

- Motion Features: Forgery performed on a video may interfere with the motion features of the video, resulting in changing the relation between different adjacent frames. Motion related features (optical flow etc) can thus be used to detect intra-frame video forgeries.
- Statistical Features: Pixel-based or statistical feature-based methods to detect video forgeries analyse statistical properties of objects, pixel-level inconsistencies and correlations between different frames of the video.
- Machine Learning Techniques: Machine learning, deep learning models (i.e., reacquiring huge amount of training data) are employed. New deep learning models are can automatically learn complex patterns from the data to detect image forgery, without requiring any hand-crafted.

Intra-frame Video Forgery Detection

The intra-frame forgeries are carried out in the spatial domain, i.e., single frame present inside a video is manipulated using the image manipulation techniques, for example, copy/move, splicing or cropping etc. Intra-frame forgeries are used to add or remove a portion or an object from within one or multiple frames of any given video to conceal or misrepresent content of the video.

These forgeries are similar to image forgeries, since individual frames within a video are manipulated and thus can be detected using passive image forensics techniques as

described in previous sections. However, some of the techniques take temporal features into account in order to detect intra-frame forgeries. For example, [139] proposed to employ optical flow (helps in tracking the movement of objects) related inconsistencies in order to detect intra-frame copy-move video forgery.

F. ACTIVE FORENSICS

The forensics techniques presented in earlier sections are ''passive'' in nature, i.e., do not require any prior information about the visual content which is being analysed [38]. Active forensics is another family of forensics techniques which analyse visual content by examining specific watermarks, or signatures embedded during acquisition or processing stages.

A limitation of active approaches is that these approaches fail to work in cases where there is no prior information available about the image being verified, for example, if the information about the watermark/signature is not available, or if there is no watermark/signature embedded into the image. Also, when the images shared on social media platforms are uploaded/downloaded several times, the image compression rate gets affected severely, influencing watermark/signature embedded in the image initially [140]. Furthermore, if the watermarks or signatures are added during image acquisition phase, the camera must be equipped with a special watermarking chip or digital signature chip [38].

G. CONTENT AUTHENTICITY INITIATIVE

Content Authenticity Initiative (CAI) is a new project aimed at developing an end-to-end secure system for digital content (image/video) provenance and attribution. Through this initiative several big tech companies like Adobe, Microsoft are working collectively with big media houses including BBC, AFP, The Washington Post etc. to combat visual mis-/disinformation [141].

The initiative's goal is to include a layer of verifiable trust within all types of digital content i.e., photos, videos by employing provenance and attribution solutions. Although this initiative is at its evolutionary stages, it can prove to be extremely useful in fighting visual mis-/disinformation online. The initial version of the CAI will appear in the beta version of Adobe Photoshop, a widely popular Adobe's ubiquitous photo editing software. Eventually, the CAI might help transform the social media feeds or news websites by filtering out content which is ''possibly'' inauthentic.

IV. MAPPING

In this section, we present an overview of the verification tools media practitioners employ, the limitations associated with these tools, and the future prospect of visual UGC verification tools.

The employment of computational verification tools and resources is growing [5]. In 2017, ICFJ states that only 11% of the interviewed journalists and news managers were using some kind of computer tools to verify content shared on social media platforms. The figure in 2019, however, shows a remarkable increase in this number with around 33% of the interviewed journalists and news managers utilising computational tools and resources to verify UGC [5]. The ICFJ's 2019 report reveals that more than half of the surveyed journalists use digital fact-checking tools [5]. This upwards trend of using verification tools is due to the speed and the scale at which visual mis-/disinformation is disseminated. Table 2 presents the percentage of the UGC verification tools used in the media industry according to the ICFJ's 2019 report.

Journalists and fact-checkers typically use basic tools such as reverse image search, Exif data viewers and online maps with known limitations as discussed in Section II. A variety of multimedia forensics tools such as *forensically*, *fotoforensics*, *WeVerify - InVID* verification plugin, *MeVer*, *DeDigi* [142], and other similar tools are available online which can assist journalists in detecting possible tampering operations an image might have underwent. Some classical image forgeries such as copy-move and image splicing forgeries can be detected using these tools. However, such tools are not widely employed by media practitioners in visual UGC verification. The reason might be because these tools require technical knowledge and training to be used properly. Moreover, most of the available multimedia forensics tools do not take any contextual information into account when used to verify a piece of visual UGC. These caveats might be the reason why most of the news media professionals are reluctant in trusting automated verification tools.

For deepfake detection, although there are some tools available online they mostly do not work as expected. The available deepfake detection tools just provide a binary, ''real'' or ''fake'' answer without providing any insights on why and how the decision has been made.

A study to address the issue of contextual information was proposed in [143], describing a system called *Seriously Rapid Source Review (SRSR)*. SRSR is able to provide contextual cues from different sources allowing media practitioners to find and analyse sources relating to breaking news events [143]. A similar tool called Journalistic Decision Support System (JDSS) [26] was also developed under the REVEAL project [144]. JDSS is free to use, and provides diverse set of functionalities to crawl Twitter for useful content in order to carry out verification. In [145], also under the REVEAL project, a web-based image verification system, which featured metadata visualisation, and image tampering detection tools was proposed. In [27], *Context Aggregation and Analysis Tool* to verify user generated videos was proposed. The tool is claimed to be capable of automatically collecting and calculating several different contextual verification cues for a given video. The cues include, (1) contextual information about the video (e.g., comments, thumbnails, Twitter context), (2) if the video has already been debunked in the past. Authors also used machine learning models trained on real and fake video data to automatically analyse a given video.

TABLE 3. A list of useful tools for visual UGC verification, and some of their limitations. The associated visual UGC verification elements described in Section II are also presented in this table, where 1 = Provenance, 2 = Source, 3 = Date, 4 = Location, 5 = Motivation and 6 = Multimedia Forensics.

Tool	Use Case	Element	Limitations
WeVerify - InVID https://tinyurl.com/mtfcj59s	Image/Video Analysis, Metadata Analysis, Frame Extraction	1, 3, 4, 6	Struggles against heavy compression, requires some level of training to be used.
TrulyMedia https://www.truly.media/	Contextual Image/Video Analysis, Identity Verification	1, 2, 3, 4, 5	Restricted access, no forensics tools are made available, no documentation available.
MeVer https://caa.iti.gr/	Contextual Visual Content Analysis, Metadata Analysis	1, 2, 3, 4, 5, 6	Relies heavily on the already available information on the web, not useful when there is no related information available about fairly recently surfaced fake visual content.
FotoForensics http://fotoforensics.com/	Image Analysis, Metadata Analysis, String Extraction	1, 3, 4, 6	No dedicated copy-move detector, struggles against heavy compression, does not allow customized forensics filters.
Forensically https://29a.ch/photo-forensics/	Image Analysis, Metadata Analysis, String Extraction	1, 3, 4, 6	Struggles against heavy compression, requires some level of training to be used.
Ghiro https://www.imageforensic.org/	Image Analysis, Metadata Analysis, GPS Localization	1, 3, 4, 6	No copy-move detector, struggles against heavy compression, does not allow customized forensics filters.
DeDigi http://www.dedigi.tech/	Image Analysis, Metadata Analysis, GPS Localization	1, 3, 4, 6	Struggles against heavy compression, user-interface can be improved.
Deepware https://deepware.ai/	Deepfake Detection	6	Only analyzes videos with duration of less than 10 minutes, the available deepfake detection models can be improved.
Snopes https://www.snopes.com/	Fact Checking	1, 2, 3, 4	Only helps if the image/video being verified has already been fact-checked.
Google Image Search https://www.google.com/imghp	Reverse Image Search	1, 2, 3	Will not help if the visual UGC being verified has been shared for the first time, or fairly recently.
TinEye https://tineye.com	Reverse Image Search	1, 2, 3	Will not help if the visual UGC being verified has been shared for the first time, or fairly recently.
RevEye https://tinyurl.com/3hvx3ne5	Reverse Image Search	1, 2, 3	Will not help if the visual UGC being verified has been shared for the first time, or fairly recently.
TweetDeck https://tweetdeck.twitter.com	Twitter Anaytics	2, 5	Only useful if the source being verified has a Twitter profile.
Twitonomy https://www.twitonomy.com/	Twitter Anaytics	2, 5	Only useful if the source being verified has a Twitter profile, paid subscription required to use premium features.
TweetBeaver https://tweetbeaver.com/	Twitter Anaytics	2, 5	Only useful if the source person being verified has a Twitter profile.
BotSentinel https://botsentinel.com/	Bot Detection	2, 5	Only works for Twitter, not always 100% accurate.
CrowdTangle Search https://www.crowdtangle.com/	Facebook/Instagram/Reddit Analytics	2, 5	Only keeps track of verified accounts with a certain amount of followers i.e., celebrities, politicians, journalists etc.
SPOKEO https://www.spokeo.com/	Identity Verification	2, 5	Only provides details of people residing in USA, requires paid subscription to utilize its full functionality.
Webmii https://webmii.com/	Identity Verification	2, 5	No content filtering capability, only available in English.
Pipl https://pipl.com/	Identity Verification	2, 5	Paid subscription, restricted entry.
Online Exif Viewer http://exif-viewer.com/	Metadata Analysis	1, 3	Not useful for images having no metadata information.
Exifdata https://exifdata.com/	Metadata Analysis	1, 3	Not useful for images having no metadata information.
YouTube Data-Viewer https://tinyurl.com/yckp89jc	Metadata Analysis, Thumbnail Extraction, Reverse Image Search	1, 3	Only works for YouTube Videos, does not provides too many details about the video i.e., only provides the assoicated tags.
WolframAlpha https://www.wolframalpha.com/	Weather Information	3, 4	Requires paid version to access some functionalities, not designed specifically for weather information analysis.
SunCalc https://www.suncalc.org/	Weather Information	3, 4	No mobile version available, the web service is not properly maintained.
Google Earth https://earth.google.com/web/	Location Information	4	Satellite imagery is not real-time i.e., takes month/s or even years to update maps, streetview data mostly available for developed countries only.
Wikimapia http://Wikimapia.com	Location Information	4	No street view available, less amount of functionality, user-interface can be improved.
Viewdns.info https://viewdns.info/	DNS Analytics	2	Requires expert technical knowledge to be used properly.

Other similar popular projects succeeding REVEAL are InVID, and WeVerify which are focused on building a platforms to detect and verify visual content. These projects aim at developing tools for image/video metadata analysis, key-frame extraction, reverse image search, magnifier, forensic analysis and contextual data analysis [18], [146], [147]. Under the InVID project, researchers have also developed a social media analytics dashboard to find and track trending stories across several social media websites [148].

Considering the importance of visual UGC verification, and the lack of available trustworthy tools and resources, media industry is joining hands with the research community to address these issues [149]. It is true that no automated verification tool can verify a piece of visual content with 100% accuracy [150] but tools can make the verification process

more efficient by reducing the burden on the fact-checkers. Tools which provide contextual information about a given visual UGC item, while analysing its veracity by using both content and context based features will be extremely helpful for the media practitioners. By using such tools, journalists and fact-checkers will have all of the required information from different sources in one place, which will enable them to carry out verification effectively. Also, such tools will help reduce the need to look at different sources online manually to gather more information, resulting in efficient verification.

It should be stressed that the final decision is to be made by the person (journalist/fact-checker/editor) who is using the tool, and not by the automated tool itself. New tools should be tailored to provide all of the required information at one place, and let the user to decide if the content is genuine, fake, tampered or re-contextualised. Table 3 presents a variety of available verification tools journalists and fact-checkers typically use to verify UGC along with some of their limitations.

V. CONCLUSION

In this paper we presented an overview of visual UGC verification in journalism, i.e., we described in detail 5 elements of UGC verification, along with the computational tools journalists and fact-checkers employ in order to verify visual UGC shared online. In addition to the 5 basic pillars of UGC verification, in this study we propose a 6th pillar which we call ''Multimedia Forensics'', which could potentially benefit the news media professionals in verifying manipulated visual content. Besides this, from a technical perspective we also analysed a variety of visual content forgeries, and the forensics techniques proposed by the computer science community to detect these forgeries. In the end, we presented a mapping of the available computational tools media professionals frequently employ in order to verify visual UGC, the available multimedia forensics tools which are not commonly used by the journalists and fact-checkers, and the limitations of the available tools.

Based on our analysis of the journalistic UGC verification practices, we conclude that (semi-)automated verification tools are required in order to aid media professionals and newsrooms in their fight against an increasing amount of visual mis-/disinformation online. We also suggest that multimedia forensics tools should be incorporated into the basic journalistic verification workflows. In addition to that, to properly make use of forensics tools, journalists and fact-checkers should be trained.

From a computer science perspective, we believe that more user-friendly, explainable forensics tools are required in order to gain the confidence of media professionals in using multimedia forensics tools in their day-to-day routine. Additionally, most of the available multimedia forensics tools carry out content based analysis only, and does not take into account the contextual information while verifying a piece of visual UGC. We suggest that new forensics tools should be designed in a way so that they can take advantage of the contextual information acquired from different sources relating to the UGC item being verified. We believe this will further enhance the verification process, and will gain confidence from the media industry to use such tools since they will then be able to see on what basis the tool has made a certain decision.

Generating visual mis-/disinformation and detecting it is an ongoing arms race. The researchers propose new solutions to detect manipulated visual content, and the adversaries propose new techniques to evade the detection algorithms while generating more and more realistic looking fakes. We expect that this will result in extremely realistic fake visual content that it will not be possible to detect such fake content using passive techniques anymore. We therefore think that active forensics techniques will be more useful in the future to detect fake content. The active forensics techniques require special signatures, watermarks to be inserted into the visual content at the time of creation. Such signature, watermarks can be used to check whether the content has been manipulated or not. Content Authenticity Initiative briefly discussed in section III is a step towards contemporary form of active forensics, and we foresee it as a vital apparatus in the fight against visual mis-/disinformation in the future.

ACKNOWLEDGMENT

The authors would like to thank Verdens Gang (VG), Bergens Tidende, Faktisk, and Bellingcat for their valuable insights on visual content verification from journalistic perspective.

REFERENCES

[1] P. Tolmie, R. Procter, D. W. Randall, M. Rouncefield, C. Burger, G. W. S. Hoi, A. Zubiaga, and M. Liakata, ''Supporting the use of user generated content in journalistic practice,'' *Proc. Conf. Hum. Factors Comput. Syst.*, 2017, pp. 3632–3644.

[2] S. Schifferes, N. Newman, N. Thurman, D. Corney, A. Göker, and C. Martin, ''Identifying and verifying news through social media,'' *Digit. Journalism*, vol. 2, no. 3, pp. 406–418, Jul. 2014.

[3] D. Turner. (2018). *Truth in the Age of Social Media*. [Online]. Available: https://niemanreports.org/wp-content/uploads/2014/03/Summer-2012.pdf

[4] C. Silverman, Ed., *Verification Handbook: An Ultimate Guideline on Digital Age Sourcing for Emergency Coverage*. European Journalism Centre, 2014.

[5] (2019). *State of Technology in Global Newsrooms: International Centre For Journalists (ICFJ)*. [Online]. Available: https://www.icfj.org/sites/default/files/2019-10/2019%20Final%20Report.pdf

[6] M. Walker and K. E. Matsa. (2021). *News Consumption Across Social Media in 2021*. [Online]. Available: https://tinyurl.com/yeammb92

[7] C. Ireton and J. Posetti, *Journalism, Fake News & Disinformation: Handbook for Journalism Education and Training*. Paris, France: UNESCO Publishing, Sep. 2018.

[8] T. Thomson, D. Angus, P. Dootson, E. Hurcombe, and A. Smith, ''Visual mis/disinformation in journalism and public communications: Current verification practices, challenges, and future opportunities,'' *J. Pract.*, vol. 16, no. 5, pp. 1–25, 2020.

[9] S. Urbani. (2019). *Verifying Online Information*. [Online]. Available: https://firstdraftnews.org/long-form-article/verifying-online-information/

[10] M. Looney. (2018). *The BBC's Best Practices for Verifying User-Generated Content*. [Online]. Available: https://ijnet.org/en/story/bbcs-best-practices-verifying-user-generated-content

[11] D. Teyssou. (2018). *InVID: A Horizon 2020 European Project*. [Online]. Available: https://www.invid-project.eu/invid-at-the-ieee-workshop-on-information-forensics-and-security-2017/

[12] H. T. Sencar, L. Verdoliva, and N. Memon, *Multimedia Forensics* (Advances in Computer Vision and Pattern Recognition). Singapore: Springer, doi: 10.1007/978-981-16-7621-5_1.
[13] C. François. (2019). *Actors, Behaviors, Content: A Disinformation ABC*. [Online]. Available: https://tinyurl.com/2csdjxde
[14] A. Alaphilippe. (2020). *Adding a 'D' to the ABC Disinformation Framework*. [Online]. Available: https://www.brookings.edu/techstream/adding-a-d-to-the-abc-disinformation-framework/
[15] J. Pamment. (2020). *The ABCDE Framework*. [Online]. Available: https://www.jstor.org/stable/pdf/resrep26180.6.pdf
[16] J. Gray and S. Terp. (2019). *Misinformation: We're Four Steps Behind Its Creators*. [Online]. Available: https://tinyurl.com/2p834esy
[17] S. Blazek. (2021). *SCOTCH: A Framework for Rapidly Assessing Influence Operations*. [Online]. Available: https://tinyurl.com/3jd4d6bc
[18] D. Teyssou, J.-M. Leung, E. Apostolidis, K. Apostolidis, S. Papadopoulos, M. Zampoglou, O. Papadopoulou, and V. Mezaris, "The InVID plug-in: Web video verification on the browser," in *Proc. Int. Workshop Multimedia Verification (MuVer)*, 2017, pp. 23–30.
[19] (2006). *Altered Images Prompt Photographer's Firing*. [Online]. Available: https://www.nbcnews.com/id/wbna13165165
[20] (2008). N. Shachtman. *Iran Missile Photo Faked*. [Online]. Available: https://www.wired.com/2008/07/iran-missile-ph/
[21] (2017). D. Trotta. *Crowd Controversy: The Making of an Inauguration Day Photo*. [Online]. Available: https://www.reuters.com/article/us-usa-trump-inauguration-image-idUSKBN1572VU
[22] C. Wardle and H. Derakhshan. (2018). *Thinking About Information Disorder: The Seven Formats of Mis- and Dis-Information*. [Online]. Available: https://tinyurl.com/bdddjuv3
[23] (2019). BBC. *The Fake Video Where Johnson and Corbyn Endorse Each Other*. [Online]. Available: https://www.bbc.com/news/av/technology-50381728
[24] C. Pasquini, I. Amerini, and G. Boato, "Media forensics on social media platforms: A survey," *EURASIP J. Inf. Secur.*, vol. 2021, no. 1, pp. 1–19, Dec. 2021, doi: 10.1186/s13635-021-00117-2.
[25] V. Conotter, D.-T. Dang-Nguyen, G. Boato, M. Menéndez, and M. Larson, "Assessing the impact of image manipulation on users' perceptions of deception," *Proc. SPIE*, vol. 9014, pp. 239–247, Feb. 2014.
[26] S. Middleton. (2017). *Journalist Decision Support System (JDSS)*. [Online]. Available: https://revealproject.eu/journalist-decision-support-system-jdss/
[27] O. Papadopoulou, D. Giomelakis, L. Apostolidis, S. Papadopoulos, and Y. Kompatsiaris, "Context aggregation and analysis: A tool for user-generated video verification," in *Proc. SIGIR Workshop Reducing Online Misinf. Exposure*, Rome, Italy, vol. 25, Jul. 2019.
[28] M. Kharrazi, H. T. Sencar, and N. Memon, "Blind source camera identification," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, Oct. 2004, pp. 709–712.
[29] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, "A Bayesian-MRF approach for PRNU-based image forgery detection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 4, pp. 554–567, Apr. 2014.
[30] B. Paris and J. Donovan. (2019). *DEEPFAKES AND CHEAP FAKES: The Manipulation of Audio and Visual Evidence*. [Online]. Available: https://datasociety.net/library/deepfakes-and-cheap-fakes/
[31] R. F. Check. (2022). *Fact Check-Animation Miscaptioned as if to Show Video of Ukrainian Fighter Jet Shooting Down Russian Plane*. [Online]. Available: https://tinyurl.com/6n57muhc
[32] D. Evon. (2021). *Is Viral Heart-Shaped Sunset Photo Real?*. [Online]. Available: https://www.snopes.com/fact-check/heart-shaped-sunset/
[33] S. Sheth. (2020). *A Photo of Trump and Other Leaders Staring at Putin is Going Viral—But it's Fake*. [Online]. Available: https://www.businessinsider.com/fake-viral-photo-trump-putin-g20-2017-7
[34] B. Posters and D. Howe. (2019). *Verifying Online Information*. [Online]. Available: http://billposters.ch/the-zuckerberg-deepfake-heard-around-the-world/
[35] N. Pham. (2015). *Haunting 'Nepal Quake Victims' Photo From Vietnam*. [Online]. Available: https://www.bbc.com/news/world-asia-32579598
[36] H. Denham. (2020). *Another Fake Video of Pelosi Goes Viral on Facebook*. [Online]. Available: https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/
[37] W. Luo, Z. Qu, F. Pan, and J. Huang, "A survey of passive technology for digital image forensics," *Frontiers Comput. Sci. China*, vol. 1, no. 2, pp. 166–179, May 2007, doi: 10.1007/s11704-007-0017-0.
[38] A. Piva, "An overview on image forensics," *Int. Scholarly Res. Notices*, vol. 2013, pp. 1–22, 2013.
[39] K. S. Choi, E. Y. Lam, and K. K. Y. Wong, "Source camera identification by JPEG compression statistics for image forensics," in *Proc. IEEE Region 10th Conf. (TENCON)*, Nov. 2006, pp. 1–4.
[40] S. Bayram, H. Sencar, N. Memon, and I. Avcibas, "Source camera identification based on CFA interpolation," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2005, p. 69.
[41] S. Bayram, H. T. Sencar, N. Memon, and I. Avcibas, "Improvements on source camera-model identification based on CFA interpolation," in *Proc. WG*, 2006, vol. 11, no. 9.
[42] J. Lukäš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006, doi: 10.1109/TIFS.2006.873602.
[43] T. Filler, J. J. Fridrich, and M. Goljan, "Using sensor pattern noise for camera model identification," *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 1296–1299.
[44] C.-T. Li, "Source camera identification using enhanced sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 280–287, Jun. 2010.
[45] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, "Blind PRNU-based image clustering for source identification," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 9, pp. 2197–2211, Sep. 2017.
[46] H. Farid, "Digital image forensics," *Sci. Amer.*, vol. 298, no. 6, pp. 66–71, 2008.
[47] D. Freire-Obregón, F. Narducci, S. Barra, and M. Castrillón-Santana, "Deep learning for source camera identification on mobile devices," 2017, *arXiv:1710.01257*.
[48] P. Yang, R. Ni, Y. Zhao, and W. Zhao, "Source camera identification based on content-adaptive fusion residual networks," *Pattern Recognit. Lett.*, vol. 119, pp. 195–204, Mar. 2019.
[49] Y. Liu, Z. Zou, Y. Yang, N.-F.-B. Law, and A. A. Bharath, "Efficient source camera identification with diversity-enhanced patch selection and deep residual prediction," *Sensors*, vol. 21, no. 14, p. 4701, Jul. 2021.
[50] R. Caldelli, R. Becarelli, and I. Amerini, "Image origin classification based on social network provenance," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 6, pp. 1299–1308, Jun. 2017.
[51] I. Amerini, T. Uricchio, and R. Caldelli, "Tracing images back to their social network of origin: A CNN-based approach," in *Proc. IEEE Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2017, pp. 1–6.
[52] I. Amerini, C. Li, and R. Caldelli, "Social network identification through image classification with CNN," *IEEE Access*, vol. 7, pp. 35264–35273, 2019.
[53] A. Bharati, D. Moreira, J. Brogan, P. Hale, K. Bowyer, P. J. Flynn, A. Rocha, and W. J. and Scheirer, "Beyond pixels: Image provenance analysis leveraging metadata," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1692–1702.
[54] P. Mullan, C. Riess, and F. Freiling, "Forensic source identification using JPEG image headers: The case of smartphones," *Digit. Invest.*, vol. 28, pp. S68–S76, Apr. 2019.
[55] A. Black, T. Bui, S. Jenni, V. Swaminathan, and J. P. Collomosse, "VPN: Video provenance network for robust content attribution," in *Proc. Eur. Conf. Vis. Media Prod.*, 2021, pp. 1–10.
[56] I. Amerini, A. Anagnostopoulos, L. Maiano, and L. R. Celsi, "Learning double-compression video fingerprints left from social-media platforms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 2530–2534.
[57] M. C. Stamm and K. J. R. Liu, "Blind forensics of contrast enhancement in digital images," in *IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2008, pp. 3112–3115.
[58] C. Rathgeb, A. Botaljov, F. Stockhardt, S. Isadskiy, L. Debiasi, A. Uhl, and C. Busch, "PRNU-based detection of facial retouching," *IET Biometrics*, vol. 9, no. 4, pp. 154–164, Jul. 2020.
[59] C. Rathgeb, C.-I. Satnoianu, N. E. Haryanto, K. Bernardo, and C. Busch, "Differential detection of facial retouching: A multi-biometric approach," *IEEE Access*, vol. 8, pp. 106373–106385, 2020.
[60] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. Efros, "Detecting photoshopped faces by scripting photoshop," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10071–10080.
[61] A. Jain, R. Singh, and M. Vatsa, "On detecting GANs and retouching based synthetic alterations," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Oct. 2018, pp. 1–7.

[62] R. D. Singh and N. Aggarwal, "Video content authentication techniques: A comprehensive survey," *Multimedia Syst.*, vol. 24, no. 2, pp. 211–240, Mar. 2018.

[63] J. Swaine. (2018). *Trump Inauguration Crowd Photos Were Edited After he Intervened*. [Online]. Available: https://www.theguardian.com/world/2018/sep/06/donald-trump-inauguration-crowd-size-photos-edited

[64] Y. Mirsky and W. Lee, "The creation and detection of deepfakes," *ACM Comput. Surv.*, vol. 54, no. 1, pp. 1–41, Jan. 2022, doi: 10.1145/3425780.

[65] X. Feng and G. J. Doërr, "JPEG recompression detection," *Proc. SPIE*, vol. 7541, pp. 188–199, Jan. 2010.

[66] W. Luo, Z. Qu, J. Huang, and G. Qiu, "A novel method for detecting cropped and recompressed image block," *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 2, Apr. 2007, pp. 217–220.

[67] H. C. Nguyen and S. Katzenbeisser, "Detecting resized double jpeg compressed images–using support vector machine," in *Communications and Multimedia Security*. Berlin, Germany: Springer, 2013.

[68] M. Fanfani, M. Iuliani, F. Bellavia, C. Colombo, and A. Piva, "A vision-based fully automated approach to robust image cropping detection," *Signal Process., Image Commun.*, vol. 80, Feb. 2020, Art. no. 115629.

[69] W. Li, Y. Yuan, and N. Yu, "Passive detection of doctored JPEG image via block artifact grid extraction," *Signal Process.*, vol. 89, no. 9, pp. 1821–1829, 2009.

[70] D.-K. Hyun, S.-J. Ryu, H.-Y. Lee, and H.-K. Lee, "Detection of upscale-crop and partial manipulation in surveillance video based on sensor pattern noise," *Sensors*, vol. 13, no. 9, pp. 12605–12631, Sep. 2013.

[71] H. Farid, "Image forgery detection: A survey," *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 16–25, Mar. 2009. [Online]. Available: https://farid.berkeley.edu/downloads/publications/spm09.pdf

[72] A. Parveen, Z. H. Khan, and S. N. Ahmad, "Block-based copy–move image forgery detection using DCT," *Iran J. Comput. Sci.*, vol. 2, no. 2, pp. 89–99, Jun. 2019.

[73] K. Cabeen and P. Gent. *Image Compression and the Discrete Cosine Transform*. Accessed: Sep. 23, 2022. [Online]. Available: https://tinyurl.com/bdp6kybk

[74] A. C. Popescu and H. Farid. (2004). *Exposing Digital Forgeries by Detecting Duplicated Image Regions*. [Online]. Available: https://farid.berkeley.edu/downloads/publications/tr04.pdf

[75] G. Li, Q. Wu, D. Tu, and S. Sun, "A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD," in *Proc. IEEE Multimedia Expo Int. Conf.*, Jul. 2007, pp. 1750–1753, doi: 10.1109/ICME.2007.4285009.

[76] J. Fridrich, D. Soukal, and J. Lukás, "Detection of copy-move forgery in digital images," in *Proc. Digit. Forensic Res. Workshop*, 2003. [Online]. Available: http://www.ws.binghamton.edu/fridrich/research/copymove.pdf

[77] M. Bashar, K. Noda, N. Ohnishi, and K. Mori, "Exploring duplicated regions in natural images," *IEEE Trans. Image Process.*, early access, Mar. 25, 2019, doi: 10.1109/TIP.2010.2046599.

[78] S. Bayram, H. Taha Sencar, and N. Memon, "An efficient and robust method for detecting copy-move forgery," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 1053–1056.

[79] A. D. Warbhe, R. V. Dharaskar, and V. M. Thakare, "A survey on keypoint based copy-paste forgery detection techniques," *Proc. Comput. Sci.*, vol. 78, pp. 61–67, Jan. 2016.

[80] H. Huang, W. Guo, and Y. Zhang, "Detection of copy-move forgery in digital images using SIFT algorithm," in *Proc. IEEE Pacific–Asia Workshop Comput. Intell. Ind. Appl.*, vol. 2. Dec. 2008, pp. 272–276, doi: 10.1109/PACIIA.2008.240.

[81] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A SIFT-based forensic method for copy-move attack detection and transformation recovery," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 1099–1110, Sep. 2011.

[82] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, L. Del Tongo, and G. Serra, "Copy-move forgery detection and localization by means of robust clustering with J-Linkage," *Signal Process., Image Commun.*, vol. 28, no. 6, pp. 659–669, Jul. 2013.

[83] X. Bo, W. Junwen, L. Guangjie, and D. Yuewei, "Image copy-move forgery detection based on SURF," in *Proc. Int. Conf. Multimedia Inf. Netw. Secur.*, 2010, pp. 889–892.

[84] L. Jing and C. Shao, "Image copy-move forgery detecting based on local invariant feature," *J. Multimedia*, vol. 7, no. 1, pp. 90–97, Feb. 2012.

[85] Y. Rao and J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2016, pp. 1–6.

[86] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting residual-based local descriptors as convolutional neural networks: An application to image forgery detection," in *Proc. 5th ACM Workshop Inf. Hiding Multimedia Secur.*, Philadelphia, PA, USA, Jun. 2017, pp. 159–164, doi: 10.1145/3082031.3083247.

[87] Y. Liu, Y. Zhong, and Q. Qin, "Scene classification based on multiscale convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7109–7121, Dec. 2018, doi: 10.1109/TGRS.2018.2848473.

[88] J.-L. Zhong and C.-M. Pun, "An end-to-end dense-InceptionNet for image copy-move forgery detection," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2134–2146, 2020.

[89] Y. Zhu, Ch. Chen, G. Yan, Y. Guo, and Y. Dong, "AR-Net: Adaptive attention and residual refinement network for copy-move forgery detection," *IEEE Trans. Ind. Informat.*, vol. 16, pp. 6714–6723, 2020.

[90] N. Goel, S. Kaur, and R. Bala, "Dual branch convolutional neural network for copy move forgery detection," *IET Image Process.*, vol. 15, no. 3, pp. 656–665, Feb. 2021.

[91] A. C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3948–3959, Oct. 2005.

[92] H. Farid, "Exposing digital forgeries from JPEG ghosts," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 1, pp. 154–160, Mar. 2009.

[93] Z. He, W. Lu, W. Sun, and J. Huang, "Digital image splicing detection based on Markov features in DCT and DWT domain," *Pattern Recognit.*, vol. 45, no. 12, pp. 4292–4299, 2012.

[94] M. K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *Proc. 7th Workshop Multimedia Secur.*, Aug. 2005, pp. 1–10.

[95] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 758–767, Feb. 2005.

[96] A. Swaminathan, M. Wu, and K. J. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 101–117, Mar. 2008.

[97] Y.-F. Hsu and S.-F. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2006, pp. 549–552.

[98] D. Cozzolino, G. Poggi, and L. Verdoliva, "Splicebuster: A new blind image splicing detector," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Nov. 2015, pp. 1–6.

[99] Q. Zhang, W. Lu, R. Wang, and G. Li, "Digital image splicing detection based on Markov features in block DWT domain," *Multimedia Tools Appl.*, vol. 77, no. 23, pp. 31239–31260, 2018.

[100] N. Kanwal, A. Girdhar, L. Kaur, and J. S. Bhullar, "Digital image splicing detection technique using optimal threshold based local ternary pattern," *Multimedia Tools Appl.*, vol. 79, nos. 19–20, pp. 12829–12846, May 2020.

[101] P. Niyishaka and C. Bhagvati, "Image splicing detection technique based on illumination-reflectance model and LBP," *Multimedia Tools Appl.*, vol. 80, no. 2, pp. 2161–2175, Jan. 2021.

[102] A. K. Jaiswal and R. Srivastava, "A technique for image splicing detection using hybrid feature set," *Multimedia Tools Appl.*, vol. 79, nos. 17–18, pp. 11837–11860, May 2020.

[103] Y. Rao, J. Ni, and H. Zhao, "Deep learning local descriptor for image splicing detection and localization," *IEEE Access*, vol. 8, pp. 25611–25625, 2020.

[104] B. Ahmed, T. A. Gulliver, and S. alZahir, "Image splicing detection using mask-RCNN," *Signal, Image Video Process.*, vol. 14, no. 5, pp. 1035–1042, Jul. 2020.

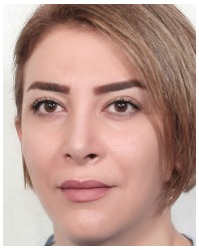
[105] M.-J. Kwon, I.-J. Yu, S.-H. Nam, and H.-K. Lee, "CAT-Net: Compression artifact tracing network for detection and localization of image splicing," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 375–384.

[106] K. B. Meena and V. Tyagi, "A deep learning based method for image splicing detection," *J. Physics: Conf. Ser.*, vol. 1714, no. 1, Jan. 2021, Art. no. 012038.

[107] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proc. ICML Unsupervised Transf. Learn.*, 2012, pp. 37–49.

[108] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 27, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2014, doi: 10.1145/3422622.
[109] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*.
[110] C. Bodnar, "Text to image synthesis using generative adversarial networks," 2018, *arXiv:1805.00676*.
[111] Y. Mirsky, T. Mahler, I. Shelef, and Y. Elovici, "CT-GAN: Malicious tampering of 3D medical imagery using deep learning," in *Proc. USENIX Secur. Symp.*, 2019, pp. 461–478.
[112] J. Damiani. (2019). *A Voice Deepfake was Used to Scam a CEO Out of $243,000*. [Online]. Available: https://bit.ly/3secirD
[113] Y. Jia, Y. Zhang, R. J. Weiss, Q. Wang, J. Shen, F. Ren, Z. Chen, P. Nguyen, R. Pang, I. Lopez-Moreno, and Y. Wu, "Transfer learning from speaker verification to multispeaker text-to-speech synthesis," in *Proc. Conf. Neural Inf. Process. Syst. (NIPS)*, 2018, pp. 4485–4495.
[114] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, "Defending against neural fake news," 2019, *arXiv:1905.12616*.
[115] T. Fagni, F. Falchi, M. Gambini, A. Martella, and M. Tesconi, "TweepFake: About detecting deepfake tweets," *PLoS ONE*, vol. 16, no. 5, May 2021, Art. no. e0251415.
[116] J. Wakefield. (2022). *Deepfake Presidents Used in Russia-Ukraine War*. [Online]. Available: https://www.bbc.com/news/technology-60780142
[117] D. Guera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2018, pp. 1–6.
[118] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A compact facial video forgery detection network," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2018, pp. 1–7.
[119] Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, "Thinking in frequency: Face forgery detection by mining frequency-aware clues," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 86–103.
[120] S. A. Khan and H. Dai, "Video transformer for deepfake detection with incremental learning," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 1821–1828.
[121] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, "Emotions don't lie: An audio-visual deepfake detection method using affective cues," in *Proc. 28th ACM Int. Conf. Multimedia*, Seattle, WA, USA, 2020.
[122] U. A. Ciftci, I. Demir, and L. Yin, "FakeCatcher: Detection of synthetic portrait videos using biological signals," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jul. 15, 2020, doi: 10.1109/TPAMI.2020.3009287.
[123] S. Agarwal, H. Farid, T. El-Gaaly, and S.-N. Lim, "Detecting deep-fake videos from appearance and behavior," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2020, pp. 1–6.
[124] X. Zhu, H. Wang, H. Fei, Z. Lei, and S. Z. Li, "Face forgery detection by 3D decomposition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2928–2938.
[125] I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, "Deepfake video detection through optical flow based CNN," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1205–1207.
[126] X. Zhang, S. Karaman, and S.-F. Chang, "Detecting and simulating artifacts in GAN fake images," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2019, pp. 1–6.
[127] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, "Detection of GAN-generated fake images over social networks," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Apr. 2018, pp. 384–389.
[128] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting residual-based local descriptors as convolutional neural networks: An application to image forgery detection," in *Proc. 5th ACM Workshop Inf. Hiding Multimedia Secur.*, Jun. 2017, pp. 159–164.
[129] D.-T. Dang-Nguyen, G. Boato, and F. G. B. D. Natale, "Discrimination between computer generated and natural human faces based on asymmetry information," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2012, pp. 1234–1238.
[130] N. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen, "Distinguishing computer graphics from natural images using convolution neural networks," in *Proc. IEEE Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2017, pp. 1–6.
[131] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in *Proc. 4th ACM Workshop Inf. Hiding Multimedia Secur.*, Jun. 2016, pp. 5–10.
[132] L. Nataraj, T. M. Mohammed, S. Chandrasekaran, A. Flenner, J. H. Bappy, A. K. Roy-Chowdhury, and B. S. Manjunath, "Detecting GAN generated fake images using co-occurrence matrices," 2019, *arXiv:1903.06836*.
[133] S. Aneja, C. Midoglu, D. T. Dang-Nguyen, M. A. Riegler, P. Halvorsen, M. Nießner, B. Adsumilli, and C. Bregler, "MMSys'21 grand challenge on detecting cheapfakes," in *Proc. 12th ACM Multimedia Syst. Conf. (MMSys)*, Istanbul, Turkey, 2021.
[134] S. Aneja, C. Bregler, and M. Nießner, "COSMOS: Catching Out-of-Context misinformation with self-supervised learning," 2021, *arXiv:2101.06278*.
[135] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "EANN: Event adversarial neural networks for multi-modal fake news detection," *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2018, pp. 849–857.
[136] L. Shang, Y. Zhang, D. Zhang, and D. Wang, "FauxWard: A graph neural network approach to fauxtography detection using social media comments," *Social Netw. Anal. Mining*, vol. 10, no. 1, pp. 1–16, Dec. 2020.
[137] D. Khattar, J. S. Goud, M. Gupta, and V. Varma, "MVAE: Multimodal variational autoencoder for fake news detection," in *Proc. World Wide Web Conf.*, 2019, pp. 2915–2921.
[138] N. A. Shelke and S. S. Kasana, "A comprehensive survey on passive techniques for digital video forgery detection," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 6247–6310, Feb. 2021.
[139] O. I. Al-Sanjary, A. A. Ahmed, A. A. B. Jaharadak, M. A. M. Ali, and H. M. Zangana, "Detection clone an object movement using an optical flow approach," in *Proc. IEEE Symp. Comput. Appl. Ind. Electron. (ISCAIE)*, Apr. 2018, pp. 388–394.
[140] R. Thakur and R. Rohilla, "Recent advances in digital image manipulation detection techniques: A brief review," *Forensic Sci. Int.*, vol. 312, Jul. 2020, Art. no. 110311.
[141] (2019). *Content Authenticity Initiative*. [Online]. Available: https://contentauthenticity.org/
[142] C.-H. Tran, Q.-T. Tran, Q.-C. Long-Vu, H.-S. Nguyen, A.-D. Tran, and D.-T. Dang-Nguyen, "DeDigi: A privacy-by-design platform for image forensics," in *Proc. 3rd ACM Workshop Intell. Cross-Data Anal. Retr.*, Jun. 2022, pp. 58–62.
[143] N. A. Diakopoulos, M. D. Choudhury, and M. Naaman, "Finding and assessing social media information sources in the context of journalism," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2012, pp. 2451–2460.
[144] I. Matzakou. (2013). *REVEAL—REVEALing Hidden Concepts in Social Media*. [Online]. Available: https://revealproject.eu/
[145] M. Zampoglou, S. Papadopoulos, Y. Kompatsiaris, R. Bouwmeester, and J. Spangenberg, "Web and social media image forensics for news professionals," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 10, no. 2, Aug. 2021, pp. 159–166.
[146] V. Mezaris. (2016). *InVID: In Video Veritas—Verification of Social Media Video Content for the News Industry*. [Online]. Available: https://www.invid-project.eu/
[147] (2012). *Weverify: Wider and Enhanced Verification for You*. [Online]. Available: https://weverify.eu/
[148] A. Scharl, A. Hubmann-Haidvogel, M. C. Göbel, T. Schäfer, D. Fischl, and L. J. B. and Nixon, "Multimodal analytics dashboard for story detection and visualization," in *Video Verification Fake News Era*. Cham, Switzerland: Springer, 2019.
[149] C. Trattner, D. Jannach, E. Motta, I. C. Meijer, N. Diakopoulos, M. Elahi, A. L. Opdahl, B. Tessem, N. Borch, M. Fjeld, L. Øvrelid, K. De Smedt, and H. Moe, "Responsible media technology and AI: Challenges and research directions," *AI Ethics*, vol. 2, no. 4, pp. 585–594, Nov. 2022.
[150] D. Giomelakis, O. Papadopoulou, S. Papadopoulos, and A. Veglis, "Verification of news video content: Findings from a study of journalism students," *J. Pract.*, vol. 2021, pp. 1–30, Aug. 2021.

SOHAIL AHMED KHAN received the M.Sc. degree in cybersecurity and artificial intelligence from the University of Sheffield, U.K. He is currently pursuing the joint Ph.D. degree with MediaFutures and the University of Bergen. Prior to joining MediaFutures, he worked as a Research Assistant at the Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates. Before that, he worked as a Remote Research Assistant at the CYENS Centre of Excellence, Nicosia, Cyprus. His research interests include deep learning, computer vision, and multimedia forensics. He is also associated with MediaFutures Work Package 3, Media Content Analysis, and Production.



GHAZAAL SHEIKHI received the master's degree in biomedical engineering from the Amirkabir University of Technology, Teheran, Iran, and the Ph.D. degree in computer engineering (machine learning) from Eastern Mediterranean University, North Cyprus. She is currently a Postdoctoral Fellow at MediaFutures. Her research interests include machine learning, natural language processing, and textual content analysis.

ANDREAS L. OPDAHL received the Ph.D. degree from the Norwegian University of Science and Technology (NTNU), in 1992. He is currently a Professor in information systems development at the University of Bergen, Norway, where he heads the Research Group for Intelligent Information Systems (I2S). His research interests include ontologies and knowledge graphs, enterprise, and IS modeling and their applications to media production. He is the author, the coauthor, or a co-editor of more than a 100 peer-reviewed and widely cited research papers. He is a member of IFIP WG5.8 on Enterprise Interoperability and WG8.1 on Design and Evaluation of Information Systems. He serves as an associate editor or renowned international journals and as an organizer of renowned international conferences and workshops.

FAZLE RABBI received the Doctor of Philosophy (Ph.D.) degree in software engineering from the University of Oslo. He is currently an Associate Professor at the University of Bergen. He has long and varied experience with software development in smaller and larger projects within a large spectrum of domain areas and technological solutions. His research interests include model-based software engineering, data mining, and machine learning, with emphasis on addressing the information science problems in healthcare applications, and software engineering related research: workflow modeling and its verification, metamodeling, building decision support systems, multi-agent systems, and process engineering.

SERGEJ STOPPEL received the Ph.D. degree in computer science in the field of visualization from the University of Bergen. He is currently the Chief Innovation Officer of Wolftech Broadcast Solutions, where he is driving the innovation of a collaborative news and media production tool that is used by more than 18000 of users on a daily basis. He is also working in areas of data science and analytics, deep learning, and natural language processing. He was awarded with the EuroVis Best Ph.D. Award, in 2019.

CHRISTOPH TRATTNER received the B.Sc. degree in computer science and telematics and the M.Sc. and Ph.D. degrees (Hons.) from the Graz University of Technology, Austria. He is currently a Professor at the University of Bergen and the Center Director of the Research Centre for Responsible Media Technology & Innovation—SFI MediaFutures worth around 26 million EUR. His research interest includes two central specializations in the information science research field. The first is "Behavioral Data Analytics" and the second is "Recommender Systems." He is a Senior Member of ACM and a Former Austrian Research Promotion Agency (FFG) Fellow and a Marshall Plan and European Research Consortium for Informatics and Mathematics (ERICM) Fellow.

DUC-TIEN DANG-NGUYEN (Member, IEEE) is currently an Associate Professor in computer science at the Department of Information Science and Media Studies, University of Bergen. His research interests include multimedia forensics, lifelogging, multimedia retrieval, and computer vision. He is a member of MediaFutures WP3—Media Content Analysis and Production in Journalism and The Nordic Observatory for Digital Media and Information Disorder (NORDIS). He is the author or coauthor of more than a 100 peer-reviewed and widely cited research papers. He is a PC member in a number of conferences in the fields of lifelogging, multimedia forensics, and pattern recognition. He is a co-organizer of over 40 special sessions, workshops, and research challenges from ACM MM, ACM ICMR, NTCIR, ImageClef, and MediaEval during the last ten years. He is also the General Chair of MMM 2023.