EURASIP Journal on Information Security

REVIEW Open Access

Media forensics on social media platforms: a survey

Cecilia Pasquini[1*], Irene Amerini[2] and Giulia Boato[1]

Abstract

The dependability of visual information on the web and the authenticity of digital media appearing virally in social media platforms has been raising unprecedented concerns. As a result, in the last years the multimedia forensics research community pursued the ambition to scale the forensic analysis to real-world web-based open systems. This survey aims at describing the work done so far on the analysis of shared data, covering three main aspects: forensics techniques performing source identification and integrity verification on media uploaded on social networks, platform provenance analysis allowing to identify sharing platforms, and multimedia verification algorithms assessing the credibility of media objects in relation to its associated textual information. The achieved results are highlighted together with current open issues and research challenges to be addressed in order to advance the field in the next future.

Keywords: Media forensics, Social media, Platform provenance analysis, Media verification

1 Intro and motivation

The diffusion of easy-to-use editing tools accessible to a wider public induced in the last decade growing concerns about the dependability of digital media. This has been recently amplified by to the development of new classes of artificial intelligence techniques capable of producing high quality fake images and videos (e.g., Deepfakes) without requiring any specific technical know-how from the users. Moreover, multimedia contents strongly contribute to the viral diffusion of information through social media and web channels, and play a fundamental role in the digital life of individuals and societies. Thus, the necessity of developing tools to preserve the trustworthiness of images and videos shared on social media and web platforms is a need that our society can no longer ignore.

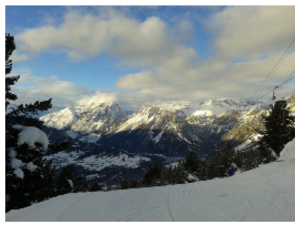
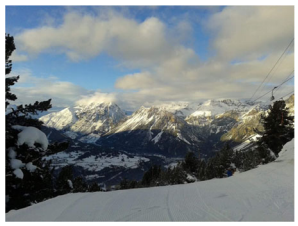
Many works in multimedia forensics have studied the detection of various manipulations and the identification of the media source, providing interesting results in laboratory conditions and well-defined scenarios under different levels of knowledge available to the forensic analyst. Recently, the research community also pursued the ambition to scale multimedia forensics analysis to real-world web-based open systems. Then, potential tampering actions through specialized editing tools or the generation of deceptive fake visual information are mixed and interleaved with routine sharing operations through web channels.

The extension of media forensics to such novel and more realistic scenarios implies the ability to face significant technological challenges related to the (possibly multiple) uploading/sharing process, thus requiring methods that can reliably work under these more general conditions. In fact, during those steps the data get further manipulated by the platforms in order to reduce the memory and bandwidth requirements. This hinders conventional forensics approaches but also introduces detectable patterns. We report in Table 1 two examples of images shared through popular social media platforms and downloaded from there, where it can be seen how the signal gets altered in terms of size and compression quality.

In this framework, being able of, even partially, retrieving information about the digital life of a media object in

*Correspondence: cecilia.pasquini@unitn.it
[1]University of Trento, Via Sommarive 9, 38123, Trento, Italy
Full list of author information is available at the end of the article

Table 1 Examples images before and after being shared and downloaded from social media platforms. The image size and the luminance quantization tables used for the last JPEG compression are reported in each case, showing notable differences between different versions

Native	**Shared with Facebook**	**Shared with Flickr**	**Shared with Twitter**
			
1920 × 2560	1536 × 2048	1536 × 2048	900 × 1200
$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 7 & 5 & 5 & 7 & 11 & 18 & 23 & 28 \\ 6 & 6 & 6 & 9 & 12 & 27 & 28 & 25 \\ 6 & 6 & 7 & 11 & 18 & 26 & 32 & 26 \\ 6 & 8 & 10 & 13 & 23 & 40 & 37 & 29 \\ 8 & 10 & 17 & 26 & 31 & 50 & 47 & 35 \\ 11 & 16 & 25 & 29 & 37 & 48 & 52 & 42 \\ 23 & 29 & 36 & 40 & 47 & 56 & 55 & 46 \\ 33 & 42 & 44 & 45 & 52 & 46 & 47 & 46 \end{bmatrix}$	$\begin{bmatrix} 4 & 3 & 3 & 4 & 7 & 11 & 14 & 17 \\ 3 & 3 & 4 & 6 & 7 & 17 & 17 & 15 \\ 4 & 4 & 4 & 7 & 11 & 16 & 20 & 16 \\ 4 & 5 & 6 & 8 & 14 & 25 & 22 & 17 \\ 5 & 6 & 10 & 16 & 19 & 31 & 29 & 22 \\ 7 & 10 & 15 & 18 & 23 & 29 & 32 & 26 \\ 14 & 18 & 22 & 25 & 29 & 34 & 34 & 28 \\ 20 & 26 & 27 & 28 & 32 & 28 & 29 & 28 \end{bmatrix}$	$\begin{bmatrix} 5 & 3 & 3 & 5 & 7 & 12 & 15 & 18 \\ 4 & 4 & 4 & 6 & 8 & 17 & 18 & 17 \\ 4 & 4 & 5 & 7 & 12 & 17 & 21 & 17 \\ 4 & 5 & 7 & 9 & 15 & 26 & 24 & 19 \\ 5 & 7 & 11 & 17 & 20 & 33 & 31 & 23 \\ 7 & 11 & 17 & 19 & 24 & 31 & 34 & 28 \\ 15 & 19 & 23 & 26 & 31 & 36 & 36 & 30 \\ 22 & 28 & 29 & 29 & 34 & 30 & 31 & 30 \end{bmatrix}$
Native	**Shared with Whatsapp**	**Shared with Telegram**	**Shared with Messenger**
			
1920 × 2560	1200 × 1600	1200 × 1600	1000 × 1334
$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 6 & 4 & 4 & 6 & 10 & 16 & 20 & 24 \\ 5 & 5 & 6 & 8 & 10 & 23 & 24 & 22 \\ 6 & 5 & 6 & 10 & 16 & 23 & 28 & 22 \\ 6 & 7 & 9 & 12 & 20 & 35 & 32 & 25 \\ 7 & 9 & 15 & 22 & 27 & 44 & 41 & 31 \\ 10 & 14 & 22 & 26 & 32 & 42 & 45 & 37 \\ 20 & 26 & 31 & 35 & 41 & 48 & 48 & 40 \\ 29 & 37 & 38 & 39 & 45 & 40 & 41 & 40 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 & 2 & 2 & 3 & 5 & 6 & 8 & 10 \\ 2 & 2 & 2 & 3 & 5 & 6 & 8 & 10 \\ 2 & 2 & 3 & 5 & 6 & 8 & 10 & 12 \\ 3 & 3 & 5 & 6 & 8 & 10 & 12 & 14 \\ 5 & 5 & 6 & 8 & 10 & 12 & 14 & 15 \\ 6 & 6 & 8 & 10 & 12 & 14 & 15 & 15 \\ 8 & 8 & 10 & 12 & 14 & 15 & 15 & 15 \\ 10 & 10 & 12 & 14 & 15 & 15 & 15 & 15 \end{bmatrix}$

terms of provenance, manipulations and sharing operations, can represent a valuable asset, as it could support law enforcement agencies and intelligence services in tracing perpetrators of deceptive visual contents. More generally, it can help in preserving the trustworthiness of digital media and countering misinformation effects by enforcing trustable sources.

This survey aims at describing the work done so far by the research community on multimedia forensic analysis of digital images and videos shared through social media or web platforms, highlighting achieved results but also current open issues and research challenges which still have to be addressed to provide general solutions.

2 Overview and structure

We exemplify in Fig. 1 the main possible steps in the digital life of a media object shared online. While simplified, this representation is sufficiently expressive and allows us to annotate the focus of the different sections of this survey, thus clarifying the paper's structure.

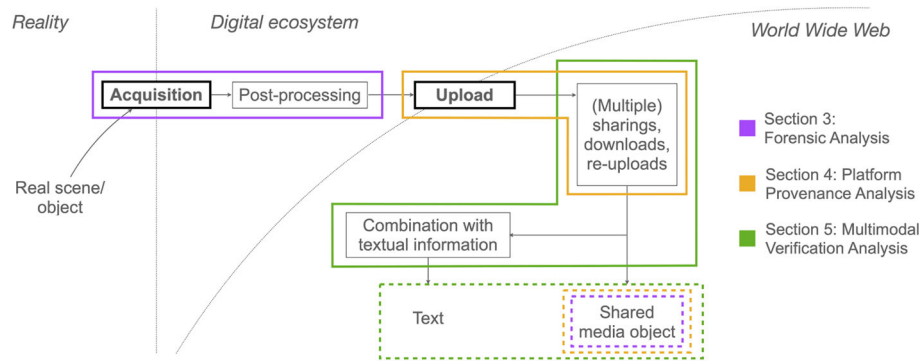


Fig. 1 Visual representation of the research topics reviewed in this survey. Each section corresponds to a color. For each color, solid rectangles highlight the phases studied by the methods reviewed in the corresponding section (i.e., where different hypotheses are defined for the forensic analysis), and dashed rectangles indicate the objects on which the analysis is based

We identify two main milestones, namely the *acquisition* and the *upload* steps, which are reported in bold in Fig. 1. First, at acquisition a real scene/object is captured through an acquisition device and thus enters what we denote as digital ecosystem (i.e., the set of digital media objects). Afterwards, a number of operations can be applied, that we gather under the block "Post-processing", including resizing, filtering, compressions, cropping, semantic manipulations. This is the phase on which most of the research approaches in multimedia forensics operate.

Then, through the upload phase, the object is shared through web services, and thus gets included in the world wide web. Technically, it is very common that acquired media objects are uploaded to web platforms either instantly (through automatic backup services like GooglePhoto or iCloud), or already during the post-processing phase (e.g., through Adobe Creative Cloud), thus squeezing the first and second phase. However, in our work we do not analyze those services (which are primarily conceived for storage purposes), but rather focus on platforms for social networking, dissemination, messaging, that typically process media objects in order to meet bandwidth and storage requirements. This includes popular Social Networks (SN) such as Facebook, Twitter, Instagram, and Google+, as well as messaging services such as WhatsApp, Telegram, and Messenger.

Afterwards, multiple different steps can follow where the object can be either downloaded, re-uploaded, or re-shared through other platforms. In addition, the multimedia content is generally tied to textual information (e.g., in news, social media posts, articles).

The present survey reviews methods performing some kind of multimedia forensic analysis of data that went through the upload phase, i.e., that has been shared (possibly multiple times) through social media or web platforms. We do not target the wider fields of computer forensics [1], but rather aim at reviewing the literature on forensic analysis of media objects shared online, leveraging both signal processing and/or machine learning approaches. Moreover, we focus on real (possibly manipulated) multimedia objects, while discarding specific scenarios such as the diffusion of synthetically generated media.

In this context, a number of forensic approaches have been proposed targeting different phases of the media digital life. A group of techniques addresses typical multimedia forensics tasks that concern the acquisition and post-processing steps (such as source identification or integrity verification) but perform the analysis on shared data. Those are reviewed in Section 3 *Forensic analysis*, which corresponds in Fig. 1 to the purple color. The dashed purple rectangle indicates the object on which the analysis is performed (i.e., the media object after the upload phase), while the solid purple rectangle highlights the phases on which different forensic hypotheses are formulated.

Another problem recently addressed in the literature is the analysis of a shared media object with the goal of reconstructing the steps involved from the upload phase on (i.e., the *sharing history*). This body of work is described in Section 4 *Platform provenance analysis*, and corresponds in Fig. 1 to the yellow color.

Finally, a number of approaches analyze the shared media object in relation to its associated textual information, in order to identify inconsistencies that might indicate a fake source of visual/textual information. This stream of research is indicated in Fig. 1 in green and is treated in Section 5 *Multimodal Verification analysis*.

In addition, we present in Section 6 a complete overview of the datasets created for the above mentioned tasks which include data that have been shared through web channels.

3 Forensic analysis

Major issues in traditional multimedia forensics are the identification of the source of multimedia data and the verification of its integrity. This section reviews the major lines of such research applied on shared media objects (purple boxes in Fig. 1). The prevalence of existing works are mostly dedicated to the analysis of the acquisition source, either targeting the identification of the specific device or the camera model. On the other side, very few contributions are given on forgery detection, most of them reviewing well known methods on new datasets and demonstrating the difficulty of this task when investigated in the wild. The third focus area is the forensic analysis in adversarial conditions, as few approaches deal with counter forensics activities and can be traced back to the source identification issue [2, 3]. In the following subsections, source identification and forgery detection methods will be reviewed separately, thereby the following major topics will be covered:

- Source camera identification - device and model identification
- Integrity verification

Afterwards, a summary and a discussion will be reported at the end of the section.

3.1 Source camera identification

The explosion in the usage of social network services enlarges the variability of image and video data and presents new scenarios and challenges especially in the source identification task, such as: knowing the kind of device used for the acquisition after the upload; knowing the brand and model of a device after the upload as well as the specific device associate to a shared media; dealing with the ability to cluster a bunch of data according to the device of origin; dealing with the ability to associate various profiles belonging to different SNs. Not all of such open questions are equally covered; i.e. very few works exist on brand identification [4]. On the contrary most of the works are mainly dedicated to source camera identification, such as, tracing back the origin of an image or a video by identifying the device or the model that acquired a particular media object. Similarly to what happened in forensic scenarios with no sharing processes, the idea behind these kind of approaches is that each phase of the acquisition process leaves an unique fingerprint on the digital content itself, which should be estimated and extracted. The fingerprint should be robust enough to the modification introduced by the sharing process, so that it is not drastically affected by the uploading/downloading operations and can be still detectable. Several papers use the PRNU (Photo Response Non-Uniformity) noise [5] as fingerprint to perform source identification, as it has proven widely viable for traditional approaches. Some others methods adopt some variants of the PRNU extraction method, and propose to use hybrid techniques or consider different footprints such as video file containers. We decide to split the source camera identification group of techniques in two categories: *perfect knowledge methods* and *limited and zero knowledge methods*, according to the level of information available or assumed on the forensic scenario. The first case, described in Section 3.1.1, is related to the methods employing known reference databases of cameras to perform their task. In the second case (Section 3.1.2) the reference dataset can be partially known or completely unknown and no assumption on the numbers of camera composing the dataset is given. A summary of the papers, that will be described in the following, is reported in Table 2 with details regarding the techniques employed, the SNs involved and the dataset used.

3.1.1 Device and model identification: perfect knowledge methods

The main characteristic of the following set of works is the creation of a reference dataset of camera fingerprint. The source identification is in this case performed in a closed set and we refer to such approaches with the term *perfect knowledge methods*. Most of the works presented hereafter are mainly based on PRNU [5] and they are equally distributed among papers that address the problem of video camera source identification and those interested in the device identification of images. One of the first papers exploring video camera source identification of YouTube videos is [6] that already demonstrates the difficulties to reach a correct identification on shared object, since many parameters that affect PRNU come into play (e.g., compression, codec, video resolution and changes in the aspect ratio). As well as above, in the paper [7], another evaluation of the camera identification techniques proposed by [5] is given, this time considering images coming from social networks and online photo sharing websites. The results show once again that modifications introduced by the upload process make the PRNU detection almost ineffective, thus demonstrating the difficulties in working on shared data. For this reason different papers recently have been tried to improve PRNU estimation in order to achieve a stronger fingerprint in the case of heavy loss [3] and to speed up the computation [8]. The authors of [8], in particular, perform an analysis on stabilized and non-stabilized videos proposing to use the spatial domain averaged frames for fingerprint extraction. A novel method for PRNU fingerprint estimation is presented in [9] taking into account the effects of video compression on the PRNU noise through the selection of blocks of frames with at least one non-null DCT coefficient. In [10], a VGG network is employed as classifier to detect the images according to the Insta-

Table 2 Summary of the *perfect knowledge* source identification approaches

Method	Cues	Methodology	Class	Media	SNs	Dataset
[6]	PRNU	Correlation + threshold	Device	Video	YouTube	–
[8]	PRNU-based	PCE + threshold	Device	Image video	Facebook, YouTube	VISION
[9]	PRNU	PCE + threshold	Device	Video	YouTube	VISION
[11]	Hybrid PRNU	PCE + threshold	Device	Image video	YouTube, Facebook	VISION
[3]	PRNU	PCE + threshold	Device	Video	Wireless Camera	–
[7]	PRNU	PCE + Threshold	Device	Image	Facebook, MySpace, Photobucket	–
[12]	PRNU	PCE + threshold	Device	Image	Facebook, Google+, WhatsApp	–
[13]	Rich features CFA CNN-based	SVM-based approaches, Open-Set Nearest Neighbors classifier	Model	Image	Flickr	FLICKR UNICAMP
[4]	CNN-based	Dense Net (patch based)	Model	Image	Flickr	–
[14]	RGB	CNN	Model	Image	Flickr, Yandex.Fotki, Wikipedia Commons	–
[10]	PRNU	VGG	Device	Image	Instagram	VISION

gram filter applied, aiming at excluding certain images in the estimation of the PRNU and thus improving the reliability of the device identification method for Instagram photos. The VISION dataset [15] is employed by many methods reviewed in this Section. For an overview of public datasets used by each paper, please refer to the Table 2.

Several works proposed the use of PRNU to address a slightly different problem, i.e., to link social media profiles containing images and videos captured by the same sensor [11, 12]. In particular, in [11] a hybrid approach investigates the possibility to identify the source of a digital video by exploiting a reference sensor pattern noise generated from still images taken by the same device. Recently, a new dataset for source camera identification is proposed (the Forchheim Image Database - FODB) [16] considering five different social networks. Two CNNs methods have been evaluated [17, 18] with and without degradation generated on the images by the sharing operation. An overview of the obtained results is shown in Fig. 2 when the two nets are trained on original images and data augmentation is performed with artificial degradations (rescaling, compression, flipping and rotation). The drop in the accuracy is quite mitigated from the employment of a general purpose net like the one proposed in [18].

So far, we have discussed approaches for device identification on shared media objects; however, some interest has been also demonstrated on camera model identification. In particular, [4] and [14] propose the use of DenseNet, a Convolutional Neural Network (CNN) with RGB patches as input, tested on the Forensic Camera-Model Identification Dataset provided by the IEEE Signal Processing (SP) Cup 2018.[1]

3.1.2 *Device and model identification: limited and zero knowledge methods*

In this section, we review the problem of clustering a set of images, according to their source, in case of limited side information about possible reference datasets or about the number of cameras. The first work in this sense involving images computes image similarity based on noise residuals [5] through a consensus clustering [19]. The work in [20] presents an algorithm to cluster images shared through SNs without prior knowledge about the types and number of the acquisition smartphones, as well as in [19] (*zero knowledge* approaches), with the difference that more than one SN have been considered in this case. This method exploits batch partitioning, image resizing, hierarchical and graph-based clustering to group the images which results in more precise clusters for images taken with the same smartphone model.

In [13] and [21], the camera model identification issue in an open set scenario with *limited knowledge* is addressed: the aim in this case is to detect whether an image comes from one of the known camera models of the dataset or from an unknown one.

The paper in [22] faces the problem of profile linking, also addressed by the *perfect knowledge* method in [12];

[1] https://www.kaggle.com/c/sp-society-camera-model-identification

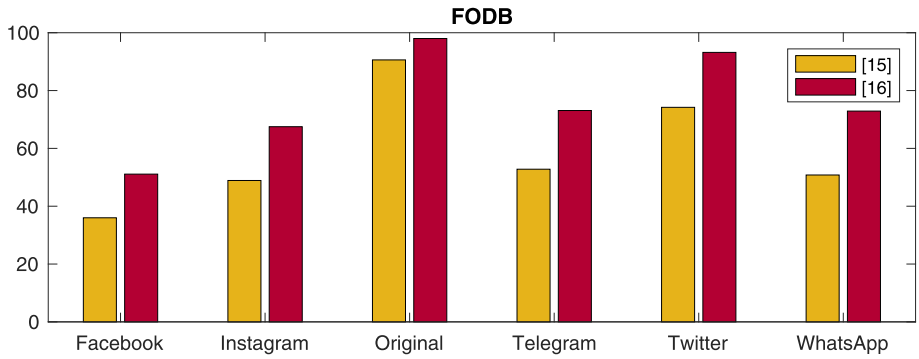


Fig. 2 Camera identification accuracies for the Forchheim Image Database as reported in [16]. Camera-native (original) images and five sharing platforms (Facebook, Instagram, Telegram, Twitter, Whatsapp) have been evaluated

this time an unsupervised approach is used applying a k-medoids clustering.

Differently from the other methods, that mainly use residual noise or PRNU to perform source identification, in [2, 23, 24] video file containers have been considered as hint for the identification and the classification of the device, the brand and the model of a device without a prior training phase. In particular in [24] a hierarchical clustering is employed whereas a likelihood-ratio framework is proposed in [2].

3.2 Integrity verification

An overview of the works dealing with forgery detection on shared data is outlined in this subsection and a summary is given in Table 3, with details about the methodology, the SNs involved and the datasets used.

Some of the works discussed in the previous section addressed also the problem of integrity verification as demonstrated by [2], where the dissimilarity between a query video and a reference file container is searched in order to detect video forgery. Instead, the work in [25], derived from [21], proposes a graph-based representation of an image, named Forensic Similarity Graph, in order to detect manipulated digital images. In detail, a forgery introduces a unique structure into this graph creating communities of patches that are subject to the same editing operation.

In those works the kind of manipulations (splicing, copy-move, retouching, and so on) taken into account are not explicitly given since in both contributions the attention paid to shared media object is very limited.

The alterations that social media platforms apply on images are further investigated in [26, 27] where their impact on tampering detection is evaluated. A number of well-established, state-of-the-art algorithms for forgery detection are compared on different datasets including images downloaded from social media platforms. The results confirm that such operations are so disruptive that sometimes could completely nullify the possibility of a successful forgery identification throughout a detector.

Table 3 Summary of the *limited and zero knowledge* source identification approaches

Method	Cues	Methodology	Class	Media	SNs	Dataset
[22]	PRNU	K-medoids clustering	Device	Image	Facebok, Google+, Telegram, WhatsApp	SDRG
[28]	PRNU	Hierarchical clustering	Device	Image	Flickr	–
[20]	PRNU	Hybrid clustering	Device	Image	Facebook, WhatsApp	VISION
[24]	Video file container	Hierarchical clustering	Device, Model, Brand	Video	WhatsApp, YouTube	VISION
[19]	PRNU	Consensus clustering	Device	Image	Facebook	–
[2]	Video file container	Unsupervised approach (likelihood-ratio framework)	Brand	Video	Facebook, WhatsApp	VISION
[21]	CNN-based	Similarity graph	Model	Image	Reddit	–

3.3 Summary and discussion

To summarize, as previously evidenced, the prevalence of the works discussed in this Section are mostly dedicated to the source camera identification problem and only few contributions are given on the identification of manipulations, demonstrating the difficulties of this particular issue when investigated on shared multimedia objects. Most of the approaches related to source identification address the problem of device camera identification and, to a lesser extent, to the model or brand identification. It has been demonstrated that the existing forensics analysis methods experience significant performance degradation due to the applied post-processing operations. For this reason, it is fundamental that future works will cover this gap in order to achieve successful forgery detection and reliable source identification of digital images and videos shared through social media or web platforms.

An important aspect to close this gap is the creation and diffusion of publicly available datasets to foster real-world oriented research in image forensics, which constitutes a non-trivial task. In fact, major effort should be dedicated to the design of data collections that are comprehensive and unbiased, so that the resulting benchmarks are realistic and challenging enough.

Furthermore, many points still need to be addressed in order to reliably analyze images and videos in the wild, such as the investigation of new kinds of fingerprints and distinctive characteristics: i.e., the PRNU, although very robust in no-sharing scenarios, it has not proven so reliable on shared data. In this context, data-driven approaches based on deep learning might empower more effective strategies for fingerprint extraction, as it has been recently explored in [29].

Another important point that need to be addressed to complete the analysis on the authenticity of images and video in the wild, is related to the Deepfake phenomenon. While there has been a recent burst of new methods for identifying synthetically generated fakes from a pristine media [30], the analysis of such kind of data after a sharing operation is still a rather unexplored problem. In [31] a preliminary analysis on Deepfake detection on Youtube videos is reported. Another point that is underrepresented in the literature so far, is a detailed analysis on adversarial forensics in regards to shared contents, a topic that is necessary to investigate more deeply in the future.

4 Platform provenance analysis

The process of uploading on sharing platforms can represent an important phase in the digital life of media data, as it allows to instantly spread the visual information and bring it to many users. While this sharing process typically hinders the ability of performing conventional media forensics tasks (as evidenced in the previous Section), it also introduces traces that allow to infer additional kind of information. In fact, data can be uploaded in many different ways, once or multiple times on diverse platforms and from different systems.

In this context, the possibility of reconstructing information on the sharing history of a certain object is highly valuable in media forensics. In fact, it could help in monitoring the visual information flow by tracing back the initial uploads, thus aiding source identification by narrowing down the search.

Several studies on this have been conducted in recent years that explore these possibilities. In this section, we collect and review such approaches, that we gather under the name of *platform provenance analysis*. Differently from what discussed in the previous section, platform provenance analysis studies the traces left by the upload phase itself and provides useful insights on the sharing operations applied to the object under investigation. We can broadly summarize the goals of platform provenance analysis as follows:

- Identification of the platforms that have processed the object
- Reconstruction of the full sharing history of the object
- Extraction of information on the systems used in the upload phase.

As a first observation, we report that most of the works addressing platform provenance tasks focus on digital images. To the best of our knowledge, the provenance analysis of videos is currently limited to the approaches proposed in [2], where the structure of video containers is used as cue to identify previous operations. Moreover, a common trait of existing methodologies is the formalization of the addressed provenance-related task as a classification problem, and the use of supervised Machine Learning (ML) as a mean to extract information from the object. In fact, the typical pipeline adopted is reported in Fig. 3: after the creation of a wide dataset containing representative data for the considered scenario, a feature representation carrying certain cues is extracted from each data sample and fed to a machine learning model, which is then trained to perform the desired task at inference phase.

The methods proposed so far in the literature present substantial differences in the way cues are selected and extracted, as well as in the choice of suitable machine learning models that can provide reliable predictions. Given the recurrence of the steps in Fig. 3, in the following we will review existing methods for digital images by examining different aspects of their detection strategies within the depicted pipeline.

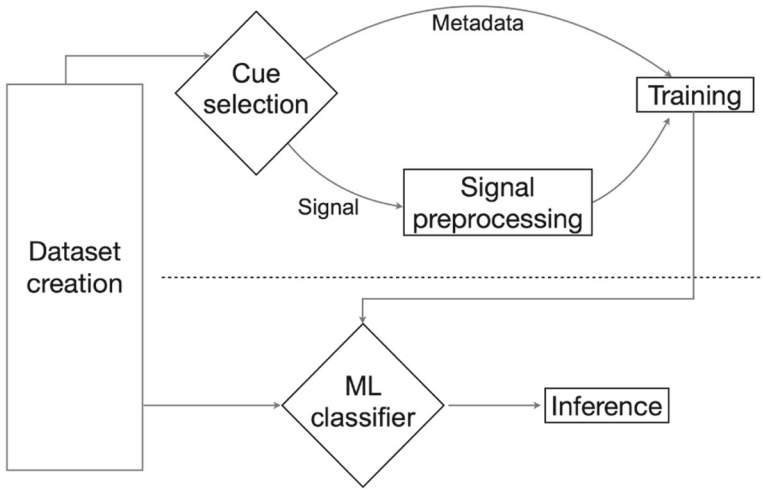


Fig. 3 Pipeline of the approaches proposed for platform provenance analysis

4.1 Dataset creation

In order to analyze the traces left by the sharing operations, suitable datasets must be created by reproducing the conditions of the studied scenario. For platform provenance analysis, images need to be uploaded to and downloaded from the web platforms and SNs under analysis. This can be performed automatically or manually, depending on the accessibility and regulations of the different platforms. For several platforms (such as Facebook, Twitter, Flickr [32]), APIs are available that allow to perform automatically sharing operations with different uploading options, thus significantly speeding up the collection process. Moreover, the platforms often allow to process multiple files in batches, although sharing with different parameters has to be performed manually.

Few works also freely release the datasets used for their analysis, which usually include the versions of each image before and after sharing. We refer to Section 6 for an overview. While some platforms also support other formats (such as PNG or TIFF), such datasets are almost exclusively composed of images in JPEG format, whose specificities are used for provenance analysis.

4.2 Cue selection

The sharing process by means of web platforms and SNs can include several operations leaving distinct traces in the digital image, which can be exposed by means of different cues.

For instance, as firstly observed in [33] for Facebook, compression and resizing are usually applied in order to reduce the size of uploaded images and this is performed differently on different platforms, also depending on the resolution and size of the data before uploading. As it is widely known in multimedia forensics, such operations can be detected and characterized by analyzing the image *signal* (i.e., the values in the pixel domain or in transformed domains), where distinctive patterns can be exposed. This approach is followed in [32, 34–36] for platform provenance analysis, where the image signal is pre-processed to extract a feature representation (see Section 4.3).

Moreover, useful information can be leveraged from the image *metadata*, which provide additional side information on the image. While it can be argued that a signal-based forensic analysis would be preferable (as data structures can be falsified more easily than signals), such cues can play a particularly relevant role in platform provenance analysis. In fact, they typically are related to the software stack used by the platform, rather than to the hardware that acquired the data [37]. In [38], the authors consider several popular platforms (namely Facebook, Google+, Flickr, Tumblr, Imgur, Twitter, WhatsApp, Tinypic, Instagram, Telegram) and show that uploaded files are renamed with distinctive patterns, which occasionally even allow to reconstruct the URL of the web location of the file. Also, they notice platform-specific rules in the way images are resized and/or compressed with the JPEG standard; therefore, they propose a feature representation including the image resolution and the coefficients of the quantization table used for JPEG compression, which can be extracted from the image file without decoding.

Useful evidence for provenance analysis can then be contained in the EXIF information of JPEG files. In fact, sharing platforms usually strip out optional metadata fields (like acquisition time, GPS coordinates, acquisition device), but in JPEG files downloaded from diverse plat-

forms different EXIF fields are retained. This aspect is also explored in [37], where the authors aim at linking JPEG headers of images acquired with Apple smartphones and shared on different apps to their acquisition device; their analysis show that JPEG headers can be used to identify the operating system version and the sharing app used to a certain extent.

Finally, the works in [39, 40] propose a hybrid approach where both signal- and metadata-based features are extracted and used for classification.

4.3 Signal preprocessing

When the signal is used as source of information for the provenance analysis, different choices can be done to preprocess the signal and extract an effective feature representation. The goal is to capture traces left by the sharing operation which, as previously mentioned, usually involves a recompression phase.

To this purpose, a widely investigated solution is to rely on the Discrete Cosine Transform (DCT) domain, as proposed in [32, 34, 39, 40]. In fact, the values of DCT coefficients provide evidence on the parameters used in previous JPEG compression processes and can effectively link a shared image to the (typically last) platform it comes from. In order to reduce the dimensionality of the feature representation, a common strategy is to extract the histogram of a subset of the 64 AC subbands, and further select a range of bins that are discriminative enough.

Alternatively, the approach in [36] explores the use of the PRNU noise as a carrier of traces left by different platforms. To this purpose, a wavelet-based denoising filter is applied to each image patch to obtain a noise residual, that is then fed to the ML classifier. When fused with DCT features, noise residuals can help in raising the accuracy of the provenance analysis, as shown in [35].

Finally, while proposing a methodology to detect different kind of processing operations based simply on image patches in the pixel domain, the authors in [41] show that, as a by-product, their approach can be effective also for provenance analysis.

4.4 Machine learning model

After the feature representation is extracted, different kinds of ML classifiers can then be trained to perform the desired task. Some works employ decision trees [38] or ensemble learning techniques like random forests [32, 37, 39], as well as Support Vector Machines [39, 42] and Logistic Regression [39].

Most recently, researchers focused on deep learning techniques based on CNNs. In [34], one-dimensional CNNs are used to process DCT-based feature vectors, confirming the good performance obtained in [32] on extended datasets. The work in [40] fuses DCT-based and metadata-based features into a single CNN.

A two-dimensional CNN is instead used in [36] in order to learn discriminative representations of the extracted PRNU noise residuals, while in [35] such residual and DCT-based features are combined and fused through a novel deep learning framework (*FusionNet*).

4.5 Addressed tasks

The methods proposed for platform provenance analysis focus on different tasks related to the sharing history of the media object under investigation, concerning diverse kinds of information that can be of interest to aid the forensic analysis. The objectives of provenance analysis can be grouped as follows:

- *Identification of at least one sharing operation.* Depending on the addressed application scenario, there might be no prior information at all on the object under investigation, including whether it was previously shared by means of any platform or not. Thus, a first useful information is to determine whether a sharing operation occurred through a (typically predefined) set of platforms, or whether the data comes directly from the acquisition device or offline editing tools.
- *Identification of the last sharing platform.* Once a sharing operation is detected, thus it is determined that the content does not come natively from a device, it is of interest to identify which platform it was uploaded to. This task is addressed by most of the existing approaches. Although it cannot be excluded that more than one sharing operation was performed, provenance detectors generally identify the *last* platform that processed the data [2, 34].
- *Reconstruction of multiple sharing operations.* In this task, provenance detectors attempt to go beyond the last sharing and identify whether the data underwent more than one sharing operation. It is in fact a common scenario that an image is shared through a certain platform and then subsequently shared by the recipient through another platform [39, 40].
- *Identification of the operating system of the sharing device.* Gathering information on the hardware and software used in the sharing operation can be of interest in the forensics analysis, as it could aid the identification of the person who performed the sharing operation. In [39], it is shown that different operating systems leave distinct traces in the metadata of images shared through popular messaging apps, and can then be identified. Similarly, the authors in [37] observe that JPEG headers of shared images can provide information on the software stack that processed them, while it is harder to get information on the hardware.

4.6 Summary and discussion

In order to provide a clearer overview, Table 4 summarizes the aspects discussed in previous sections, by reporting condensed information for each proposed method. Moreover, Table 5 reports the specific web platforms and SNs included in the analysis of each different method, thus highlighting the diversity of data involved in this kind of studies.

This body of work has exposed for the first time important findings on the effect of sharing operations, and the possibility of effectively identifying their traces and infer useful information. In order to provide a quantitative overview of the methods' capabilities, we report in Fig. 4 selected comparative results from the recent approaches [34–36] on available datasets for the task of identifying the last sharing platform. It emerges that state-of-the-art approaches yield satisfying accuracy, although in rather controlled experimental scenarios.

In fact, while these studies revealed many opportunities for platform provenance analysis, substantial open issues exist and represent challenges for future investigations. First, we can observe that all the proposed approaches are purely inductive, i.e., no theoretical tools are used to characterize specific operations, apart from possible preprocessing steps before feeding a supervised ML model. Therefore, the reliability of the developed detectors strongly rely on the quality of the produced training data, which need to be representative enough for the model to correctly analyze data at inference phase.

Related to that, it is hard to predict the generalization ability of the current detectors when unseen data are analyzed at inference phase. In fact, many factors exist that can induce data variability within the same class, and are currently mostly overlooked. For instance, the traces left by a certain platform or operating system are not constant over time (as observed in [37]), but they might change from version to version. Also, the process of uploading and downloading data to/from platforms is not standardized but can be performed through different tools and pipelines: from mobile/portable devices or computers, using platform-specific APIs or browser functionalities. As a result, a potential class "Shared with Facebook" in a provenance-related task should include many processing variants, for which data need to be collected.

A possible way to alleviate these issues would be to further investigate which component of the software stack (e.g., the use of a specific library) actually leaves the most distinctive traces in the object, and at which point in the overall processing pipeline this happens. For instance, previous studies on JPEG forensics [43, 44] have shown that different libraries for JPEG compression leave specific traces, especially detectable in high quality images. This would help in establishing principled ways to predict whether a further processing variant would impact on the

Table 4 Summary of platform provenance analysis approaches. The column "Analysis" indicates whether the detectors operate on the full image ("Global") or on image patches ("Local")

Method	Cues	Preprocessing	ML classifier	Analysis	Number of sharing	Dataset
[32]	Signal-based	DCT-based feature extraction	Random Forest	Global	Single	MICC social UCID, MICC public UCID
[39]	Both	DCT-based feature extraction + metadata extraction	LR, SVM, RF	Global	Multiple	ISIMA
[34]	Signal-based	DCT-based feature extraction	1D CNN	Local	Single	UNICT-SNIM, MICC social UCID, MICC public UCID
[35]	Signal-based	DCT-based feature extraction + noise residuals extraction	1D CNN + 2D CNN	Local	Single	UNICT-SNIM, MICC social UCID, MICC public UCID, VISION
[36]	Signal-based	Noise residuals extraction	2D CNN	Local	Single	UNICT-SNIM, MICC social UCID, MICC public UCID, VISION
[33]	Metadata-based	Metadata extraction	–	Global	Single	UNICT-SNIM
[38]	Metadata-based	Metadata extraction	Distance-based K-NN + Decision trees	Global	Single	UNICT-SNIM
[40]	Both	DCT-based feature extraction + metadata extraction	CNN + feature fusion	Local	Multiple	R-SMUD, V-SMUD
[37]	Metadata-based	JPEG header analysis	Random forest	Global	Single	–
[42]	Signal-based	Pixel + DCT domain	SVM	Global	Single	–
[41]	Signal-based	Pixel domain	Siamese 2D CNN	Local	Single	–

Table 5 Summary of social media and web platforms analyzed in different studies

Method	Facebook	Twitter	Google+	Flickr	WhatsApp	Telegram	Messenger	Instagram	Others
[32]	✓	✓		✓					
[39]	✓				✓	✓	✓		
[34]	✓	✓	✓	✓	✓	✓		✓	
[35]	✓	✓	✓	✓	✓	✓		✓	
[36]	✓	✓	✓	✓	✓	✓		✓	
[33]	✓								
[38]	✓	✓	✓	✓	✓	✓		✓	Tumblr, Imgur, TinyPic
[40]	✓	✓		✓					
[37]				✓				✓	Camera+, Snapseed
[42]	✓	✓		✓					WeChat
[41]	✓			✓				✓	

traces used in the provenance analysis, but would likely require to reverse engineer proprietary software.

Moreover, as we previously pointed out, it is worth recalling that the platform provenance analysis of videos based on signal properties is essentially unexplored, thus representing a relevant open problem for future investigations. On the other hand, a container-based analysis has been applied in [2, 23].

More generally, studies such as [2, 23, 39, 40] substantially reinforced the role of metadata and format-based cues, which were only marginally considered in multimedia forensics in favor of signal-based approaches but represent a valuable asset for platform provenance identification tasks.

Lastly, we observe that the "platform provenance analysis" as defined here is distinct from the problem of "provenance analysis" as formulated in [45–47], which is rather related to the issues described in the following Section 5. In provenance analysis, a whole set of media object is analyzed, with the goal of understanding the relationships (in terms of types and parameters of transformations) between a set of semantically similar samples and reconstructing a phylogeny graph, thus requiring a substantially different approach. On the other hand, in platform provenance analysis a single object is associated to one or more sharing operations based on a set of objects (not necessarily content-wise similar) that underwent those sharing operations.

5 Multimodal verification analysis

In addition to entertainment purposes (e.g., video streaming services), images and videos typically appear in web pages and platform in conjuction with some form of textual information, which increases their communication - and potentially misinformation - strengths. In fact, the problem of false information originating and circulating on the Web is well recognized [48], and different non-exclusive categories of false information can be identified, including hoaxes, rumors, biased, or completely fake (i.e., fabricated) information.

Visual media can have a key role in supporting the dissemination of these forms of misinformation when coupled with a textual descriptive component, as it happens in popular web channels like online newspapers, social networks, blogs, forums. Therefore, there is a strong interest in developing techniques that can provide indications on the credibility of these information sources in a fully or semi-automatic manner, ideally detecting real-time whether unreliable information is about to be disseminated [49].

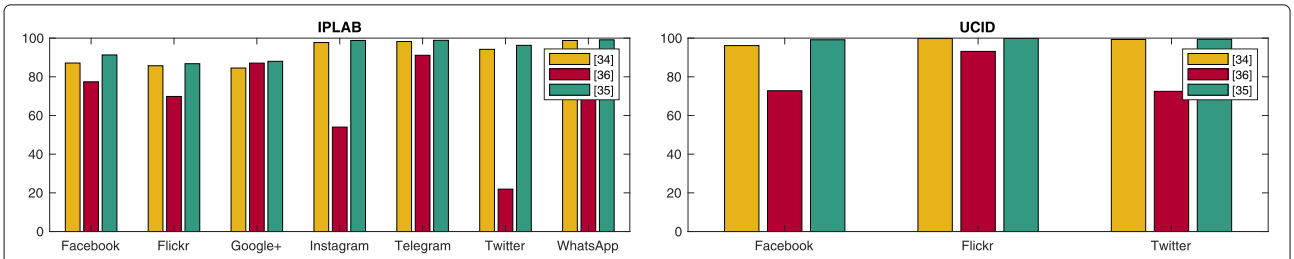


Fig. 4 Accuracy of different state-of-the-art approaches on the task of identifying the last sharing platform, as reported in the respective papers. The left barplot refers to the IPLAB dataset while the right one to the UCID Social dataset, and the analysis is performed on individual 64 × 64 patches

The analysis of images and videos can be functional to assess the credibility of *composite objects*, i.e., pieces of information that contain a textual component, one or more associated media objects, and optional metadata. Examples are given by online news, social media posts, blog articles. In this case, the problem is referred to as *multimedia verification* [50], which is a wide and challenging field that spans several disciplines, from multimedia analysis and forensics to data mining and natural language processing. In a multimedia verification analysis a high variety of factors is involved, for which no rigorous taxonomy is found in the literature. However, different approaches have been recently investigated to characterize patterns of manipulated visual and textual information.

In this context, an inherent difficulty is that the composite objects can be misleading in various ways. In fact, not only images and videos can be semantically manipulated or depict synthetic content (e.g., GAN-based imagery): they can also be authentic but used in the wrong context, i.e., associated to the wrong event, perhaps with incorrect geo-temporal information (Fig. 5).

For this reason, most of the approaches resort to a *multimodal* representation of the analyzed composite object, where different kinds of information are processed together and typically fed to some kind of machine learning classifier or decision fusion system. This includes:

- *Visual cues:* the visual component of the composite object (e.g., the images attached to a Tweet or to an online news), intended as signal and attached metadata;
- *Textual cues:* the textual component of the composite object (e.g., the body of a Tweet or an online news, including hashtags, tags);
- *Propagation cues:* metadata related to the format and dissemination of the composite object through the platform it belongs to (e.g., number and ratio of images per post, number of retweets/reposting, number of comments);
- *User cues:* metadata related to the profile of the posting user (e.g., number of friends/followers, account age, posting frequency, presence of personal information and profile picture).

A number of approaches have addressed the problem of verifying composite objects by relying only on text or categorical data (i.e., textual information, propagation information, user information) and discarding from their analysis the visual component [51–55]. However, in this survey we focus on techniques that explicitly incorporate visual cues in their approach and process the corresponding signal.

Fig. 5 Examples of composite objects containing misleading content. Top: forged image shared online in relation to the Hurricane Sandy in 2012 [51]. Bottom: pictures of vietnamese siblings, erroneously posted in relation to the earthquake in Nepal in 2015 [56]

We first differentiate the methods according to the available information they utilize in their analysis. In fact, in order to automatically verify composite objects, some kind of prior knowledge needs to be built on a set of examples and then be tested on unseen data. Thus, dedicated datasets have been developed for this purpose and represent the starting point for many of the reviewed studies. Relevant examples are given by the datasets developed for the "Verifying Multimedia Use task" (VMU)[2] of the MediaEval Benchmark[3] in 2015 and 2016 containing a collection of tweets, and the dataset collected in [57] through the official rumor busting system of the popular chinese microblog Sina Weibo.

A group of methods perform verification by solely relying on the cues extracted from the composite object under investigation and from one or more of these reference data corpora (typically used for training machine learning models), and those are reviewed in Section 5.1.

Other approaches complement the information provided by the analyzed object and datasets by dynamically collecting additional cues from the web. For instance, they retrieve textually related webpages or similar images through the use of search engines for both text and visual components (e.g., Google search, Google Image search,[4] Tineye[5]). Those methods are reported in Section 5.2.

Finally, in Section 5.3 we focus on the line of research which addresses specifically the detection of media objects that are not manipulated but wrongly associated to the topic or event treated in the textual and metadata component of the composite object they belong (i.e., *media repurposing*).

5.1 Methods based on a reference dataset

The work in [56] proposes an ensemble verification approach that merges together propagation cues, user cues, and visual cues based on image forensics methods. Starting from the data provided in the VMU2016, the authors process the maps provided by the algorithm in [58] for the detection of double JPEG artifacts by extracting statistical features. Two separate classifiers treat forensic-based features and textual/user cues, and an agreement-based retraining procedure is used to correctly fuse their outcome and express a decision (fake, real, or unknown) about each tweet.

In [57], the structure of the data corpus (which is based on different events) is considered to construct a number of features computed on each image, that are intended to describe characteristics of the image distribution and reveal distinctive pattern in social media posts. Inspired by previous work in image retrieval, the authors introduce a visual clarity score, a visual coherence score, a visual similarity distribution histogram, a visual diversity score, which together express how images are distributed among the same or different events. Such values are then combined with propagation-based features on the posts through different classifiers (SVM, Logistic Regression, KStar, Random Forests).

Recently, deep learning approaches have been employed for this problem. In [59], the authors aim at extracting event-invariant features that can be used to discriminate between reliable and unreliable composite objects, and to handle newly emerged events in addition to the ones used in training. To this purpose, they attempt to remove the dissimilarities of the feature representations among different events by letting a feature-extraction network compete with an event discriminator network.

The work in [60] employs Recurrent Neural Network (RNN) strategies to process visual-based information extracted through a pre-trained VGG-19. An attention mechanism is used for training, and textual-based and propagation-based are also incorporated in the model.

In order to capture correlations between different modes, in [61] it is proposed to train a variational autoencoder that separately encodes and decodes textual and visual information, and use multimodal encoded representation for classifying composite objects.

Lastly, the work in [62] rely only on visual information but trains in parallel different CNNs operating both in the pixel domain and in the frequency domain. The authors in fact conjecture that frequency-based features can capture different image qualities and compressions potentially due to repeated upload and download from multiple platforms, while pixel-based features can express semantic characteristics of images belonging to fake composite objects.

Since most of them address the VMU2016 dataset, which comes with predefined experimental settings and metrics, we can report a comparative overview of the results obtained by the different methods on the same data in Table 6.

5.2 Methods based on web-searched information

Due to the abundance of constantly updated data, the web can constitute a valuable source of information to aid the verification analysis.

The work in [63] targets the detection of online news containing one or more images that have been edited. To

Table 6 Comparative results of different approaches in terms of F1-score on the VMU2016 dataset

	[64]	[56]	[60]	[59]	[61]	[62]
F1	0.728	0.911	0.676	0.719	0.758	0.832

[2] http://www.multimediaeval.org/mediaeval2016/verifyingmultimediause/
[3] http://www.multimediaeval.org/
[4] https://images.google.com/
[5] https://tineye.com/

this purpose, starting from a single analyzed online news, a system is proposed that performs textual and visual web searches and provides a number of other online news that are related to the same topic and contain visually similar images. The latter can be successively compared with the original ones in order to discover possible visual inconsistencies. A further step is taken in [65], where a methodology to automatically evaluate the authenticity and possible alterations of the retrieved images is proposed.

In [66], a number of textual features are extracted from the outcome of a web-based search performed on the keywords of the event represented in the analyzed post, and on its associated media objects. Visual features from multimedia forensics are also extracted (namely double JPEG features [58], grid artifacts features [67], and Error Level Analysis[6]) and jointly processed through logistic regressors and random forest classifiers. This approach has been extended in [68] by incorporating textual features used in sentiment analysis, and in [69] by exploiting additional advanced forensic visual features provided by the Splicebuster tool [70]. Moreover, these works are tested on datasets containing different kinds of composite objects, such as Tweets, news articles collected on Buzzfeed and Google News.

5.3 Methods for detecting media repurposing

While the methods previously discussed target the detection of generic manipulations in the visual component of composite objects, a number of approaches focus on the detection of re-purposed media content. Therefore, they do not search for tampering operations in the visual content, but rather for situations where authentic media are used in the wrong context, i.e., incorrectly referred to certain events and discussion topics.

In [71], this problem is tackled by resorting to a deep multimodal representation on composite objects, which allows for the computation of a consistency score based on a reference training dataset. To this purpose, the authors create their own dataset of images, captions and other metadata downloaded from Flickr, and also test their approach on existing datasets like Flickr30K and MS COCO. A larger and more realistic dataset called MEIR (Multimodal Entity Image Repurposing)[7] is then collected in [72], where an improved multimodal representation is proposed and a novel architecture is designed to compare the analyzed composite object with a set of retrieved similar objects.

A peculiar approach is proposed in [73] and improved in [74], where the authors verify the claimed geo-location of outdoor images by estimating the position of the sun in the scene through illumination and shadow effect models. By doing so, they can compare this estimation with the one computed through astronomical procedures starting from the claimed time and location of the picture.

Recently, there has been increasing interest in event-based verification (i.e., the problem of determining whether a certain media object correctly refers to the claimed event), for which specific challenges have also been organized by NIST as part of the DARPA MediFor project.[8] In this context, the work in [75] explores different strategies to apply CNNs for the analysis of possibly re-purposed images. Several pre-trained and fine-tuned networks are compared by extracting features at different layers of the networks, showing that deeper representations are generally more effective for the desired task.

Lastly, the work in [76] addresses the typical lack of training data for repurposing detection by proposing an Adversarial Image Repurposing Detection (AIRD) method which does not need repurposing examples for being trained but only real-world authentic examples. AIRD aims at simulating the interplay between a counterfeiter and a forensic analyst through training adversarially two competing neural networks, one generating deceptive repurposing examples and the other discriminating them from real ones.

5.4 Summary and discussion

To summarize, the body of work presented in this Section faces the problem of multimedia verification, tackled only in recent years by the research community. Here, credibility of composite objects (pieces of information that contain a textual component associated to the media objects) is assessed, allowing to expand the forensic analysis to new challenging scenarios like online news and social media posts.

We described all types of approaches including visual cues in the analysis and processing the relevant signal. We clustered techniques depending on the information exploited: solely relying on the cues extracted from the composite object and from one or more reference dataset, or including additional cues collected exploiting retrieval techniques on the web. Finally, we reviewed algorithms specifically addressing the detection of media repurposing, where media objects are wrongly associated to the described topic or event.

One major challenge in this context is the scarcity of representative and populated datasets, due to the difficulty of collecting realistic data. A reason for this is that recovering realistic examples of rumors or news articles providing de-contextualized media objects is highly challenging and time-consuming, also due to the fact that such composite objects have very short life online. As a result, the risk of overfitting should be carefully accounted for.

[6] http://fotoforensics.com/tutorial-ela.php
[7] https://github.com/Ekraam/MEIR
[8] https://www.nist.gov/itl/iad/mig/media-forensics-challenge-2019-0

Another open issue is the interpretability of the detection tools. Indeed, in this scenario it is often hard to understand which kind of information learning-based system are actually using for providing their outcomes. This is also due to the intrinsic difficulty of the problem, which requires to characterize many different aspects appearing in misleading content. A comprehensive tool providing a reliable analysis on a given media object under investigation is in fact not yet available. An example of the tools currently at disposal is the Reveal Image Verificaton assistant,[9] which only provides a set of maps corresponding to different methodologies applied to the test image.

Again, it is also evident that multimodal video verification is strongly underdeveloped, and no data corpora is nowadays available for this task. Very few approaches were presented for synchronization [77] and human-based verification [78, 79], but signal-based detection is still a challenging open issue.

6 Datasets

In this section, we report an annotated list of the publicly available datasets for media forensics on shared data, with reference to the specific area for which they are created (i.e., forensics analysis, platform provenance or verification analysis). Those are summarized in Table 7. In the first column, the name of each dataset is reported, together with the link for the download (if available). The considered SNs are explicitly stated together with the number of sharing to which images or video are subjected to. An indication of the numerosity of the dataset is also provided with a specification of the devices used.

Datasets built for forensic analysis and/or platform provenance analysis share similar characteristics. VISION [15] is the most widely employed dataset for the source camera identification problem in a whole, and is also used for platform provenance tests. The FLICKR UNICAMP [13] and SDRG [22] datasets have been also proposed with regards to perfect knowledge methods and to limited and zero knowledge methods respectively. Comprehensive datasets to support various forensics evaluation tasks are the Media Forensics Challenge (MFC) [80] dataset with 35 million internet images and 300,000 video clips and the Fake Video Corpus (FVC) [81] that exploits three different social networks. Recently a new dataset has been proposed, the Forchheim Image Database (FODB) [16]. It consists of more than 23,000 images of 143 scenes by 27 smartphone cameras. Each image is provided in the original camera-native version, and five copies from social networks.

In relation to the platform provenance analysis the types of datasets used are more various. The VISION dataset is still used together with MICC UCID social, MICC PUBLIC social [32], and UNICT-SNIM [38]. All of the datasets listed above consider only one sharing throughout various social networks and instant messaging applications.

More recent datasets, like ISIMA [39] and MICC multiple UCID [35], contain images shared two times and R-SMUD, V-SMUD [40] include pictures up to 3 sharings. Images used for these datasets are either acquired personally or taken in their original version from existing datasets. Moreover, data are collected with little attention to the visual content of the shared pictures, as the analysis focuses on properties that are largely content-independent.

As opposed to that, datasets for multimodal verification analysis (such as VMU [51], Weibo [57], MEIR [72]) are built by carefully selecting the visual and textual content, typically requiring manual selection. A common approach in these corpora is to collect data related to selected events (e.g., in VMU 2016 we find "Boston Marathon bombing," "Sochi olympics," "Nepal earthquake"), so that cluster of composite objects related to the same topic are created. Also, images and text descriptions are generally crawled from web platforms and, for the case of Weibo [57], fact checking platforms are used to gather composites objects related to hoaxes. While images are typically shared multiple times and possibly through different platforms, their sharing history is not thoroughly documented as in platform provenance analysis.

7 Conclusions and outlook

In this survey we have described the literature about digital forensic analysis of multimedia data shared through social media or web platforms. Works have been organized into three main classes, corresponding to the different processing considered in the digital life of a media object shared online and evidenced with three different colors in Fig. 1: forensics techniques performing source identification and integrity verification on media uploaded on social networks; platform provenance analysis methodologies allowing to identify sharing platforms both in case of single or multiple sharing; and multimedia verification algorithms assessing the credibility of composite (text + media) objects. Challenges related to the single sub-problems were already revised at the end of relevant sections, while here we highlight the common open issues still requiring effort from the research community, thus providing possible directions for future works.

Just like it happened for the vast majority of computer vision and information processing problems, approaches based on Deep Neural Networks (DNNs) now dominate the field of multimedia forensics. In fact, they proved to deliver significantly superior performance, provided that a number of requirements for their training and deploy-

[9] http://reveal-mklab.iti.gr/reveal/

Table 7 Overview of the datasets on shared media

Name	Sharing	SNs	No. device	Numbers
VISION [82]	Single	Youtube, WhatsApp, Facebook	35 smartphones 11 brands	24,427 images, 1914 videos
Flickr UNICAMP [13]	Single	Flickr (downloaded only)	250 camera models	11,000 images
SDRG [83]	Single	Facebook, Flickr, Google+, GPhoto, Instagram, LinkedIn, Pinterest, QQ, Telegram, Tumblr, Twitter, Viber, VK, WeChat, WhatsApp and WordPress	19 smartphones	100 images from each phone uploaded/downloaded on each SN
MFC (World Data) [84]	Single	–	500 cameras	35 million images, 300,000 video clips
FVC [85]	Single	YouTube, Facebook, Twitter	–	3957 videos annotated as fake, 2458 annotated as real
FODB [86]	Single	Facebook, Instagram, Telegram, Twitter, Whatsapp	27 smartphones	23,000 images
ISIMA [87]	Double	Facebook Messenger, Whatsapp, Telegram	–	2100 images
UNICT-SNIM [88]	Single	Facebook, Google+, Twitter, Tumblr, Flickr, Instagram, Imgur, Tinypic, WhatsApp, Telegram	4 cameras	2720 images
MICC social UCID [89]	Single	Facebook, Flickr, Twitter	From UCID dataset [90]	40,140 images
MICC public UCID [89]	Single	Facebook, Flickr, Twitter	–	3000 images
MICC multiple UCID [89]	Double	Facebook, Flickr, Twitter	From UCID dataset [90]	120,420 images
R-SMUD [91]	Multiple	Facebook, Flickr, Twitter	From RAISE dataset [92]	900 images (shared 3 times on 3 SNs)
V-SMUD [91]	Multiple	Facebook, Flickr, Twitter	From VISION dataset	510 images (shared 3 times on 3 SNs)
VMU2016 [93]	Multiple	Twitter	–	193 real images, 218 misused images, 2 misused videos, 6225 real and 9404 fake tweets posted by 5895 and 9025 users
Weibo [57]	Multiple	Twitter	–	10,231 real images, 15287 misused images, 23,456 real and 26,257 fake tweets posted by 21,136 and 22,584 users
MEIR [94]	Multiple	Flickr	–	82,156 real images and captions, 57,940 manipulated images and captions

ment are met. In this context, the use of DNNs represents the most promising research direction for the forensic analysis of digital images and they have been also recently applied to spatio-temporal data (e.g., videos). Nevertheless, current approaches and solutions suffer from several shortcomings, that compromise their reliability and feasibility in real-world applications like the ones tackled in this survey. This includes the need of large amounts of good quality training data, which typically requires a time-consuming data collection phase, in particular in the context of shared media objects.

Indeed, a major challenge is the need to create even more comprehensive and un-biased realistic data corpora, able to capture the diversity of the shared media (and composite) objects. This, coupled with more theoretical works able to characterize specific operations happening in the sharing process, could support the generalization ability of the detectors when unseen data are analyzed.

Moreover, although research effort in this area has been increasingly devoted in the recent years, another important aspect is that the forensic analysis of digital videos currently lies at a much less advanced stage than for still images, thus representing a relevant open problem for future investigations. This is a crucial issue, since digital videos strongly contribute to the viral diffusion of information through social media and web channels, and

nowadays play a fundamental role in the digital life of individuals and societies. Advances in artificial intelligence and computer graphics made media manipulation technologies widely accessible and easy-to-use, thus opening unprecedented opportunities for visual misinformation effects and urgently motivating a boost of forensic techniques for digital videos.

At a higher level, a consideration clearly emerging from our literature survey is that the forensic analysis of multimedia data circulating through web channels poses a number of new issues, and will represent an increasingly complex task. First, one questions whether a hard binary classification as "real" or "manipulated" is still representative enough when dealing with such a variety of possible different manipulations and digital histories. The definition of authenticity becomes in fact variegated, depending on the targeted applications. Arguably, systems with the ambition of treating web multimedia data will either narrow down the scenarios to strict definitions, or envision more sophisticated *authenticity indicators* expressing in some form different information on the diverse aspects of the object under investigation. This will likely encompass the application of many different tools possibly spanning multiple disciplines, whose synergy could be the key for advancing the field in the next future.

Concerning multimedia analysis, the design of signal-processing-oriented methods on top of data-driven AI techniques could mitigate part of the current shortcomings affecting deep learning-based approaches, such as the need of high data amount needed for training and the low interpretability of the outcomes. More generally, it is clear that a powerful analysis of other forms of information (e.g., text, metadata) can strongly aid the multimedia analysis and provide more complete indications of semantic authenticity, thus calling in the future for stronger connections between research in multimedia forensics, data mining, and natural language processing.

Acknowledgements
No acknowledgements to be reported.

Authors' contributions
CP designed the focus and structure of the manuscript, in consultation with IA and GB. CP, IA, and GB individually carried out a literature search on the selected topics and identified relevant contributions to be reviewed. They together formalized overall achievements, limitations, and open challenges of the state of the art in the field. All co-authors contributed in writing the manuscript and have approved the submitted version.

Funding
This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR00112090136 and by the PREMIER project, funded by the Italian Ministry of Education, University, and Research (MIUR).

Availability of data and materials
Not applicable.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]University of Trento, Via Sommarive 9, 38123, Trento, Italy. [2]Sapienza University of Rome, Via Ariosto 25, 00185, Roma, Italy.

Received: 5 October 2020 Accepted: 23 March 2021
Published online: 01 May 2021

References

1. R. Böhme, F. C. Freiling, T. Gloe, M. Kirchner, in *Computational Forensics*, ed. by Z. J. M. H. Geradts, K. Y. Franke, and C. J. Veenman. Multimedia forensics is not computer forensics (Springer, Berlin, Heidelberg, 2009), pp. 90–103
2. M. Iuliani, D. Shullani, M. Fontani, S. Meucci, A. Piva, A video forensic framework for the unsupervised analysis of mp4-like file container. IEEE Trans. Inf. Forensic Secur. **14**(3), 635–645 (2019)
3. S. Chen, A. Pande, K. Zeng, P. Mohapatra, Live video forensics: Source identification in lossy wireless networks. IEEE Trans. Inf. Forensic Secur. **10**(1), 28–39 (2015)
4. A. M. Rafi, U. Kamal, R. Hoque, A. Abrar, S. Das, R. Laganière, M. K. Hasan, Application of Densenet in camera model identification and post-processing detection. arXiv preprint 1809.00576 (2018)
5. J. Lukas, J. Fridrich, M. Goljan, Digital camera identification from sensor pattern noise. IEEE Trans. Inf. Forensic Secur. **1**(2), 205–214 (2006)
6. W. van Houten, Z. Geradts, Source video camera identification for multiply compressed videos originating from youtube. Digit. Investig. **6**(1), 48–60 (2009)
7. A. Castiglione, G. Cattaneo, M. Cembalo, U. Ferraro Petrillo, Experimentations with source camera identification and online social networks. J. Ambient Intell. Humanized Comput. **4**(2), 265–274 (2013)
8. S. Taspinar, M. Mohanty, N. Memon, Camera fingerprint extraction via spatial domain averaged frames. arXiv preprint arXiv:1909.04573 (2019)
9. E. K. Kouokam, A. E. Dirik, Prnu-based source device attribution for youtube videos. Digit. Investig. **29**, 91–100 (2019)
10. Y. Quan, X. Lin, C.-T. Li, in *Data Mining*, ed. by R. Islam, Y. S. Koh, Y. Zhao, G. Warwick, D. Stirling, C.-T. Li, and Z. Islam. Provenance analysis for instagram photos (Springer, Singapore, 2018), pp. 372–383
11. M. Iuliani, M. Fontani, D. Shullani, A. Piva, Hybrid reference-based video source identification. Sensors. **19**(3), 649 (2019)
12. F. Bertini, R. Sharma, A. Iannì, D. Montesi, in *International Database Engineering & Applications Symposium*. Profile resolution across multilayer networks through smartphone camera fingerprint, (2015), pp. 23–32
13. P. R. Mendes Júnior, L. Bondi, P. Bestagini, S. Tubaro, A. Rocha, An in-depth study on open-set camera model identification. IEEE Access. **7**, 180713–180726 (2019)
14. A. Kuzin, A. Fattakhov, I. Kibardin, V. I. Iglovikov, R. Dautov, in *2018 IEEE International Conference on Big Data (Big Data)*. Camera model identification using convolutional neural networks (IEEE, New York, 2018), pp. 3107–3110
15. D. Shullani, M. Fontani, M. Iuliani, O. A. Shaya, A. Piva, Vision: a video and image dataset for source identification. EURASIP J. Inf. Secur. **2017**(1), 1–16 (2017)
16. B. Hadwiger, C. Riess, The Forchheim image database for camera identification in the wild. arXiv preprint 2011.02241 (2020)
17. A. M. Rafi, T. I. Tonmoy, U. Kamal, Q. M. J. Wu, M. K. Hasan, Remnet: Remnant convolutional neural network for camera model identification. arXiv preprint 1902.00694 (2020)
18. M. Tan, Q. Le, in *Proceedings of the 36th International Conference on Machine Learning*. EfficientNet: Rethinking model scaling for convolutional neural networks, vol. 97, (2019), pp. 6105–6114
19. F. Marra, G. Poggi, C. Sansone, L. Verdoliva, Blind PRNU-based image clustering for source identification. IEEE Trans. Inf. Forensic Secur. **12**(9), 2197–2211 (2017)
20. R. Rouhi, F. Bertini, D. Montesi, X. Lin, Y. Quan, C. Li, Hybrid clustering of shared images on social networks for digital forensics. IEEE Access. **7**, 87288–87302 (2019)

21. O. Mayer, M. C. Stamm, Forensic similarity for digital images. arXiv preprint 1902.04684 (2019)
22. R. Rouhi, F. Bertini, D. Montesi, C. Li, in *IEEE International Workshop on Biometrics and Forensics (IWBF)*. Social network forensics through smartphones and shared images (IEEE, New York, 2019), pp. 1–6
23. P. Yang, D. Baracchi, M. Iuliani, D. Shullani, R. Ni, Y. Zhao, A. Piva, Efficient video integrity analysis through container characterization. IEEE J. Sel. Top. Sig. Process. **14**(5), 947–54 (2020)
24. R. Ramos López, E. Almaraz Luengo, A. L. Sandoval Orozco, L. J. G. Villalba, Digital video source identification based on container's structure analysis. IEEE Access. **8**, 36363–36375 (2020)
25. O. Mayer, M. C. Stamm, Exposing Fake Images With Forensic Similarity Graphs. IEEE J. Sel. Top. Sig. Process. **14**(5), 1049–1064 (2020)
26. M. Zampoglou, S. Papadopoulos, Y. Kompatsiaris, in *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. Detecting image splicing in the wild (web), (2015), pp. 1–6
27. M. Zampoglou, S. Papadopoulos, Y. Kompatsiaris, Large-scale evaluation of splicing localization algorithms for web images. Multimed. Tools Appl. **76**(4), 4801–4834 (2017)
28. X. Jiang, S. Wei, R. Zhao, R. Liu, Y. Zhao, Y. Zhao, in *Image and Graphics*. A visual perspective for user identification based on camera fingerprint (Springer, Cham, 2019), pp. 52–63
29. M. Kirchner, C. Johnson, in *IEEE International Workshop on Information Forensics and Security (WIFS)*. SPN-CNN: boosting sensor-based source camera attribution with deep learning (IEEE, New York, 2019), pp. 1–6
30. L. Verdoliva, Media forensics and deepfakes: an overview. IEEE J. Sel. Top. Sig. Process. **14**(5), 910–932 (2020). in press
31. S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, H. Li, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. Protecting world leaders against deep fakes (IEEE, New York, 2019)
32. R. Caldelli, R. Becarelli, I. Amerini, Image origin classification based on social network provenance. IEEE Trans. Inf. Forensic Secur. **12**(6), 1299–1308 (2017)
33. M. Moltisanti, A. Paratore, S. Battiato, L. Saravo, in *International Conference on Image Analysis and Processing*. Image manipulation on facebook for forensics evidence (Springer, Cham, 2015), pp. 506–517
34. I. Amerini, T. Uricchio, R. Caldelli, in *IEEE Workshop on Information Forensics and Security (WIFS)*. Tracing images back to their social network of origin: A CNN-based approach (IEEE, New York, 2017), pp. 1–6
35. I. Amerini, C.-T. Li, R. Caldelli, Social network identification through image classification with CNN. IEEE Access. **7**, 35264–35273 (2019)
36. R. Caldelli, I. Amerini, C. T. Li, in *European Signal Processing Conference (EUSIPCO)*. PRNU-based image classification of origin social network with CNN (IEEE, New York, 2018), pp. 1357–1361
37. P. Mullan, C. Riess, F. Freiling, Forensic source identification using jpeg image headers: The case of smartphones. Digit. Investig. **28**, 68–76 (2019)
38. O. Giudice, A. Paratore, M. Moltisanti, S. Battiato, in *Image Analysis and Processing - ICIAP 2017*. A classification engine for image ballistics of social data (Springer, Cham, 2017), pp. 625–636
39. Q. Phan, C. Pasquini, G. Boato, F. G. B. De Natale, in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*. Identifying image provenance: An analysis of mobile instant messaging apps (IEEE, New York, 2018), pp. 1–6
40. Q. Phan, G. Boato, R. Caldelli, I. Amerini, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Tracking multiple image sharing on social networks (IEEE, New York, 2019), pp. 8266–8270
41. A. Mazumdar, J. Singh, Y. S. Tomar, P. K. Bora, in *Pattern Recognition and Machine Intelligence*. Detection of image manipulations using siamese convolutional neural networks (Springer, Cham, 2019), pp. 226–233
42. W. Sun, J. Zhou, in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. Image origin identification for online social networks (osns) (IEEE, New York, 2017), pp. 1512–1515
43. B. Lorch, C. Riess, in *ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'19)*. Image forensics from chroma subsampling of high-quality JPEG images (ACM, New York, 2019)
44. C. Pasquini, R. Bohme, in *IEEE International Conference on Image Processing (ICIP)*. Towards a theory of jpeg block convergence (IEEE, New York, 2018), pp. 550–554
45. D. Moreira, A. Bharati, J. Brogan, A. Pinto, M. Parowski, K. W. Bowyer, P. J. Flynn, A. Rocha, W. J. Scheirer, Image provenance analysis at scale. IEEE Trans. Image Process. **27**(12), 6109–6123 (2018)
46. A. Bharati, D. Moreira, P. Flynn, A. Rocha, K. Bowyer, W. Scheirer, Learning transformation-aware embeddings for image forensics. arXiv preprint 2001.04547 (2020)
47. A. Bharati, D. Moreira, J. Brogan, P. Hale, K. Bowyer, P. Flynn, A. Rocha, W. Scheirer, in *IEEE Winter Conference on Applications of Computer Vision (WACV)*. Beyond pixels: Image provenance analysis leveraging metadata (IEEE, New York, 2019), pp. 1692–1702
48. S. Zannettou, M. Sirivianos, J. Blackburn, N. Kourtellis, The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. J. Data Inf. Qual. **11**(3), 1–37 (2019)
49. J. Cao, P. Qi, Q. Sheng, T. Yang, J. Guo, J. Li, in *Disinformation, Misinformation, and Fake News in Social Media*. Exploring the role of visual content in fake news detection (Springer, New York, 2020)
50. C. Boididou, S. Papadopoulos, D. T. Dang Nguyen, G. Boato, M. Riegler, A. Petlund, I. Kompatsiaris, in *Proceedings of CEUR Workshop*. Verifying multimedia use at mediaeval 2016, (2016)
51. C. Boididou, S. Papadopoulos, M. Zampoglou, L. Apostolidis, O. Papadopoulou, Y. Kompatsiaris, Detection and visualization of misleading content on twitter. Int. J. Multimed. Inf. Retr. **7**(1), 71–86 (2018)
52. A. Gupta, H. Lamba, P. Kumaraguru, A. Joshi, in *International Conference on World Wide Web. WWW '13 Companion*. Faking Sandy: Characterizing and identifying fake images on twitter during hurricane Sandy, (2013), pp. 729–736
53. C. Boididou, S. Papadopoulos, Y. Kompatsiaris, S. Schifferes, N. Newman, in *International Conference on World Wide Web*. Challenges of computational verification in social multimedia, (2014), pp. 743–748
54. C. Maigrot, V. Claveau, E. Kijak, in *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. Fusion-based multimodal detection of hoaxes in social networks (ACM, New York, 2018), pp. 222–229
55. F. Yang, Y. Liu, X. Yu, M. Yang, in *ACM SIGKDD Workshop on Mining Data Semantics*. Automatic detection of rumor on sina weibo (ACM, New York, 2012)
56. C. Boididou, S. E. Middleton, Z. Jin, S. Papadopoulos, D. Dang-Nguyen, G. Boato, Y. Kompatsiaris, Verifying information with multimedia content on Twitter - A comparative study of automated approaches. Multimed. Tools Appl. **77**(12), 15545–15571 (2018)
57. Z. Jin, J. Cao, Y. Zhang, J. Zhou, Q. Tian, Novel visual and statistical image features for microblogs news verification. IEEE Trans. Multimed. **19**(3), 598–608 (2017)
58. T. Bianchi, A. Piva, Image forgery localization via block-grained analysis of JPEG artifacts. IEEE Trans. Inf. Forensic Secur. **7**(3), 1003–1017 (2012)
59. Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, J. Gao, in *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Eann: Event adversarial neural networks for multi-modal fake news detection (ACM, New York, 2018), pp. 849–857
60. Z. Jin, J. Cao, H. Guo, Y. Zhang, J. Luo, in *ACM International Conference on Multimedia*. Multimodal fusion with recurrent neural networks for rumor detection on microblogs (ACM, New York, 2017), pp. 795–816
61. D. Khattar, J. S. Goud, M. Gupta, V. Varma, in *The World Wide Web Conference*. Mvae: Multimodal variational autoencoder for fake news detection, (2019), pp. 2915–2921
62. P. Qi, J. Cao, T. Yang, J. Guo, J. Li, in *IEEE International Conference on Data Mining*. Exploiting multi-domain visual information for fake news detection (IEEE, New York, 2019), pp. 518–527
63. C. Pasquini, C. Brunetta, A. F. Vinci, V. Conotter, G. Boato, in *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. Towards the verification of image integrity in online news (IEEE, New York, 2015), pp. 1–6
64. A. Gupta, P. Kumaraguru, C. Castillo, P. Meier, in *International Conference on Social Informatics (SOCINFO)*. Tweetcred: a real-time web-based system for assessing credibility of content on Twitter (Springer, New York, 2019), pp. 1–6
65. S. Elkasrawi, A. Dengel, A. Abdelsamad, S. S. Bukhari, in *IAPR Workshop on Document Analysis Systems (DAS)*. What you see is what you get? Automatic image verification for online news content (ACM, New York, 2016), pp. 114–119

66. Q.-T. Phan, A. Budroni, C. Pasquini, F. G. De Natale, in *CEUR Workshop*. A hybrid approach for multimedia use verification (CEUR, Aachen, 2016)
67. W. Li, Y. Yuan, N. Yu, Passive detection of doctored JPEG image via block artifact grid extraction. Sig. Process. **89**(9), 1821–1829 (2009)
68. F. Lago, Q. Phan, G. Boato, in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*. Image forensics in online news (IEEE, New York, 2018), pp. 1–6
69. F. Lago, Q.-T. Phan, G. Boato, Visual and textual analysis for image trustworthiness assessment within online news. Secur. Commun. Netw. (2019)
70. D. Cozzolino, G. Poggi, L. Verdoliva, in *IEEE International Workshop on Information Forensics and Security (WIFS)*. Splicebuster: A new blind image splicing detector (IEEE, New York, 2015), pp. 1–6
71. A. Jaiswal, E. Sabir, W. AbdAlmageed, P. Natarajan, in *ACM International Conference on Multimedia*. Multimedia semantic integrity assessment using joint embedding of images and text (ACM, New York, 2017), pp. 1465–1471
72. E. Sabir, W. AbdAlmageed, Y. Wu, P. Natarajan, in *Proceedings of the 26th ACM International Conference on Multimedia*. Deep multimodal image-repurposing detection (ACM, New York, 2018), pp. 1337–1345
73. P. Kakar, N. Sudha, Verifying temporal data in geotagged images via sun azimuth estimation. IEEE Trans. Inf. Forensic Secur. **7**(3), 1029–1039 (2012)
74. X. Li, X. Qu, W. Xu, S. Wang, Y. Tong, L. Luo, Validating the contextual information of outdoor images for photo misuse detection. arXiv preprint 1811.08951 (2018)
75. M. Goebel, A. Flenner, L. Nataraj, B. S. Manjunath, Deep learning methods for event verification and image repurposing detection. Electron. Imaging Media Watermarking Secur. Forensic. **2019**(5), 530–15307 (2019)
76. A. Jaiswal, Y. Wu, W. AbdAlmageed, I. Masi, P. Natarajan, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Aird: Adversarial learning framework for image repurposing detection (IEEE, New York, 2019)
77. G. Pinheiro, M. Cirne, P. Bestagini, S. Tubaro, A. Rocha, in *IEEE International Conference on Image Processing (ICIP)*. Detection and synchronization of video sequences for event reconstruction (IEEE, New York, 2019), pp. 4060–4064
78. D. Teyssou, J.-M. Leung, E. Apostolidis, K. Apostolidis, S. Papadopoulos, M. Zampoglou, O. Papadopoulou, V. Mezaris, in *First International Workshop on Multimedia Verification (MuVer)*. The inVID plug-in: Web video verification on the browser (ACM, New York, 2017), pp. 23–30
79. A. Axenopoulos, V. Eiselein, A. Penta, E. Koblents, E. La Mattina, P. Daras, A framework for large-scale analysis of video "in the wild" to assist digital forensic examination. IEEE Secur. Priv. **17**(1), 23–33 (2019)
80. H. Guan, M. Kozak, E. Robertson, Y. Lee, A. N. Yates, A. Delgado, D. Zhou, T. Kheyrkhah, J. Smith, J. Fiscus, in *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*. MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation (IEEE, New York, 2019), pp. 63–72
81. O. Papadopoulou, M. Zampoglou, S. Papadopoulos, I. Kompatsiaris, A corpus of debunked and verified user-generated videos. Online Inf. Rev. **43**(1) (2019)
82. VISION Dataset. https://lesc.dinfo.unifi.it/en/datasets. Accessed 16 Apr 2021
83. SDRG Dataset. http://smartdata.cs.unibo.it/datasets#images. Accessed 16 Apr 2021
84. MFC Dataset. https://mfc.nist.gov/. Accessed 16 Apr 2021
85. FVC Dataset. https://mklab.iti.gr/results/fake-video-corpus/. Accessed 16 Apr 2021
86. FODB Dataset. https://faui1-files.cs.fau.de/public/mmsec/datasets/fodb/. Accessed 16 Apr 2021
87. ISIMA Dataset. http://loki.disi.unitn.it/ISIMA/. Accessed 16 Apr 2021
88. IPLAB Dataset. https://iplab.dmi.unict.it/DigitalForensics/social_image_forensics/. Accessed 16 Apr 2021
89. MICC Dataset. http://lci.micc.unifi.it/labd/2015/01/trustworthiness-and-social-forensic/. Accessed 16 Apr 2021
90. G. Schaefer, M. Stich, in *SPIE Storage and Retrieval Methods and Applications for Multimedia, vol. 5307*. Ucid: an uncompressed color image database (SPIE, Bellingham, 2004), pp. 472–480
91. R-SMUD, V-SMUD Dataset. http://loki.disi.unitn.it/~rvsmud/. Accessed 16 Apr 2021
92. RAISE Dataset. http://loki.disi.unitn.it/RAISE/. Accessed 16 Apr 2021
93. VMU2016 Dataset. https://github.com/MKLab-ITI/image-verification-corpus. Accessed 16 Apr 2021
94. E. Sabir, W. AbdAlmageed, Y. Wu, P. Natarajan, in *ACM on Multimedia Conference*. Deep multimodal image-repurposing detection (ACM, New York, 2018), pp. 1337–1345

Publisher's Note