Copyright 2021 Society of Photo-Optical Instrumentation Engineers (SPIE). One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.
Frederik Temmermans, Deepayan Bhowmik, Fernando Pereira, and Touradj Ebrahimi "JPEG Fake Media: a provenance-based sustainable approach to secure and trustworthy media annotation", *Proceedings of SPIE 11842*, Applications of Digital Image Processing XLIV, 118420L (1 August 2021); https://doi.org/10.1117/12.2597556

JPEG Fake Media: a Provenance-based Sustainable Approach to Secure and Trustworthy Media Annotation

Frederik Temmermans[*a,b], Deepayan Bhowmik[c], Fernando Pereira[d], Touradj Ebrahimi[e]
[a]Vrije Universiteit Brussel, Belgium; [b]imec, Belgium; [c]University of Stirling, United Kingdom;
[d]Instituto Superior Técnico - Universidade de Lisboa and Instituto de Telecomunicações, Portugal;
[e]Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ABSTRACT

Media assets shared via social media can rapidly spread, even if they do not present a true representation of reality. The assets might have been manipulated with photo editing software, artificially created using deep learning techniques or used out of context. At the same time, editing software and deep learning techniques can be used for creative or educational media production. Clear annotation of media modifications is considered to be a crucial element to allow users to assess trustworthiness of media. However, these annotations should be attached in a secure way to prevent them from being compromised. Various organizations have already developed mechanisms that can detect and annotate modified media assets. However, to achieve a wide adoption of such an annotation approach, interoperability is essential. Therefore, the JPEG Committee has initiated the so-called JPEG Fake Media exploration. The objective of this initiative is to produce a standard that can facilitate a secure and reliable annotation of media asset creation and modifications. The standard shall support usage scenarios that are in good faith as well as those with malicious intent. This paper gives an overview of the history in media manipulation, discusses state-of-the-art in media forensics and in the creative industry as well as challenges related to AI-based manipulated media detection methods. In addressing these challenges, the paper introduces the JPEG Fake Media initiative as a provenance-based sustainable approach to secure and trustworthy media annotation.

Keywords: fake media, annotation, provenance, security, trust, interoperability, JPEG, standardization

1. INTRODUCTION

Visual information plays a key role in digital communication, regardless of whether it is for personal (*e.g.*, social network), legal (*e.g.*, trial), or security (*e.g.*, surveillance or police investigation) purposes. Therefore, provenance and authenticity of visual information are important aspects to consider. Since editing of visual information is nowadays easily accessible to the general public, forged contents are becoming more common and increasingly difficult for humans to distinguish from genuine counterparts. Digital forensics is the main field addressing such issues. As in many other fields, in recent years, new approaches based on deep learning have emerged with objectives such as identifying the model of a camera used to capture a picture or detecting falsifications in a video footage. Media modifications can be broadly divided into two categories: benign modifications, where the goal is to improve visual content quality and appearance, such as in denoising, colorization, super resolution, Low Dynamic Range (LDR) to High Dynamic Range (HDR) conversion and so on; and malicious modifications intended for manipulation of visual information, such as image and video inpainting and forgery. The driving force behind such progress has been the recent success of neural network models for several unsupervised image processing tasks, such as Artificial Intelligence (AI)-based denoising and super-resolution. Figure 1 illustrates a few popular examples of the use of AI for the purpose of visual information modifications and creation.

In addition, during the last few years, several learning-based image coding solutions have been proposed and shown promising compression efficiency when compared with state-of-the-art, mostly transform-based, coding solutions. These promise to replace the way visual information is represented, stored and delivered, thanks to efforts in standardization committees such as JPEG aiming at specifications for a learning-based image coding standard [1]. Such learning-based coding methods will be lossy by nature. As a consequence, they may introduce specific artifacts that won't be visible to the human eye but can affect performance of forgery detection systems, since the same types of deep neural networks are also used for media manipulations.

(a) (b) (c)

Figure 1: Examples of AI-based visual information modifications and creation: (a) re-colorization using DeepAI.org (original image courtesy: Unsplash); (b) synthetic generation using generated.photos; (c) deepfake composition.

While researchers made significant contributions towards digital forensics, due to several challenges, the solutions may not necessarily sustainable in the longer term. Therefore, associating provenance to media is increasingly gaining attention to assure trustworthiness in the media consumption chain. This paper proposes an approach that relies on media annotation that is secure and interoperable. Given the complexity of the problem and to achieve wide adoption, the JPEG committee initiated the JPEG Fake Media activity, exploring various scenarios leading to a comprehensive set of requirements. The main objective is to produce a standard that can facilitate secure and reliable annotation of media asset creation and modifications to support usage scenarios that are in good faith as well as those with malicious intent. This paper provides the necessary background in media manipulation and digital forensics and discusses the challenges associated with AI-based detection methods leading to details of activities within the JPEG Fake Media imitative.

The paper is organized as follows. Section 2 discusses the origins and evolution of media manipulations with an emphasis on visual information and their political, social and economic impacts. Section 3 covers the state of the art in visual information modifications in both good and bad faith scenarios. Section 4 focuses on challenges in AI-based visual information manipulations and their detection, particularly in the context of malicious intentions. Section 5 introduces the concept of provenance as a sustainable and effective approach to deal with both good and bad faith usage of media modifications. Section 6 discusses standardization efforts underway by the JPEG standardization committee in this area and presents use cases on which the latter focuses along with their respective requirements and its roadmap. The paper is concluded in Section 7.

2. BACKGROUND

The phenomenon media manipulation dates back over 150 years, to 1860 when a famous portrait of Abraham Lincoln was actually the print of a composite image where Lincoln's head was added on top of Southern politician John Calhoun's portrait [8]. In history there are multiple instances of image manipulations for various purposes, even before the era of sophisticated photo editing software. A brief summary of historical manipulated media is shown in Figure 2 [9]. There might be plethora of reasons for media manipulations including, but not limited to, making models look more attractive, removing female leaders in a newspaper photo in an orthodox society, deceiving others in showing political power, removing former associates/enemies from images to erase an historical trail, forging documents, impersonating to back up an alleged story, etc. The implications have damaging effects at political, social and economic levels, notably:

- **Political:** Often the intention is to sway voters, for example release of any embarrassing or conflicting doctored photo just before an election day may prompt the swing voters to switch sides before it is detected. As shown in Figure 2 (Stalin 1930), often it is also the case when people are being removed from photographs just because the leader no longer wants to be associated with them, or if the original photo projects a weaker image of a politician.
- **Social:** Personal body images of celebrities impact the general public at large and often magazines exploit that. Airbrushing is a common phenomenon that allows to present models/celebrities more attractively. However, such actions create unrealistic expectations of beauty which may results in mental health disorders like *Anorexia*[1].
- **Economic:** There have been instances when misleading images, doctored or used out of context, *e.g.*, taken from a movie scene, have caused riots, and therefore resulted in destruction of properties[2] leading to economic damages.

[1] https://www.nhs.uk/mental-health/conditions/anorexia/overview/

[2] https://www.altnews.in/vicious-cycle-fake-images-basirhat-riots/

Figure 2: History of manipulated media which dates back over 150 years to 1860.

Media manipulation is often discussed in a rather negative context as described earlier. It is noteworthy to observe that the creative industry has hugely benefited from photorealistic media manipulations. While the image editing tools have been available for a while, recent advances in AI and particularly in deep learning-based methods, have made media manipulations distinctly sophisticated. The impact is significant in both the creative industry as well as in spreading misleading information which includes *misinformation*, information that is false but not created with the intention of causing harm; *disinformation*, information that is false and deliberately created to harm a person, social group, organization or country; and *malinformation*, information that is based on reality, used to inflict harm on a person, social group, organization or country [13]. The risks associated with *deepfakes* and their spread include three types of impacts [2], 1) reputational damage, 2) financial and 3) influencing decision making, in three categories of stakeholders, namely, individual, organizational and societal. The risks span from intimidation, defamation, insurance fraud, fabricated court evidence to destabilization of societal cohesion and trust, to name a few.

3. STATE OF THE ART

Multimedia forensics has become even more significant due to the recent rise in both ill intended use cases, such as spreading fake news through doctored media, as well as use cases in the creative industries, where adoption of AI has shown enormous promise. Traditionally, in the era before AI-based media manipulations, digital media forensics were broadly classified in the following categories:

- **Pixel based techniques** where image splicing or copy-move forgeries are detected.
- **Geometric based techniques** which identify geometric distortions due to image editing such as changing text in signs and billboards.
- **Physical based techniques** that investigate physical inconsistencies in the edited images such as shadows and reflections.
- **Camera based techniques** also known as source identification relying on sensor noise patterns to identify specific camera models.
- **Multi-instance compression detection techniques** aiming at identification of multiple compressions (*e.g.*, double JPEG) through the quantization traits.

However, current advances in the deep learning-based techniques have influenced both the manipulated image generation and detection due to its end-to-end computational pipeline and higher accuracy when compared to traditional methods.

Wide availability of software, *e.g.*, Face2Face [25], NeuralTextures [24] and FaceSwap[3] enables creation of deepfakes with near realistic contents, almost indistinguishable from captured content to the human eye. These advances assist in the spread of fake news and therefore pose a new set of challenges within the digital forensics domain. Multiple approaches have been proposed in the recent past to detect deepfakes, including approaches that use common deep learning training [20] on images, identifying the patterns for facial expressions and movements [4], analyzing convolutional traces [10] or finding discrepancies between faces and their context [18].

However, these techniques are also increasingly used in the creative industry. Applications include re-colorization of archived media [15] or black and white photographs [6], virtual reality (VR) content creation [26], style transfer [5] and motion synthesis [11]. Other examples include TV show content creation such as the TV show "For All Mankind" which extensively uses deepfakes to bring characters including Johnny Carson, John Lennon, and Ronald Reagan back to life in an alternate history story line [14] or an alternative production of scenes from the movie "The Irishman" using deep learning methods [27].

While digital forensics has come across a long road, the current approaches may not be perfect and demand an alternative solution. We identify some of these challenges in the next section along with potential solutions in the following sections.

4. CHALLENGES IN AI-BASED MANIPULATED MEDIA DETECTION

Due to the increased spread of doctored or synthetic content relying on AI and their impact on the dissemination of fake news over social networks, detecting AI-generated content has become a major challenge in academia and industry. Recent deepfakes have proven to fool not only the human eye but also detection algorithms developed for this purpose [28] [2][16][17]. This is because deepfakes leverage powerful techniques from machine learning to manipulate or generate synthetic audio and visual content in a manner that looks natural to humans. Major companies have joined forces to organize challenges with the goal of helping in the process of creating widely accessible tools and solutions to detect malicious modifications of media content. One of the most important and recent actions is the Deepfake Detection Challenge organized by Facebook and Microsoft with the involvement of many universities [5]. However, these challenges only aimed at classifying deepfakes (face swap and voice conversion) from genuine data without considering other AI-based media processing that are harmless. When dealing with the current challenges of detecting image and video manipulations using deep learning, one will soon confront a new challenge, namely, which content has been processed by deep learning for malicious purposes and which content for non-malicious ends. If on one hand AI-based processing of multimedia content has brought a major improvement in terms of performance, on the other hand, the presence of AI-originated artifacts in benign visual information makes it more difficult to distinguish it from malicious modifications, such as deepfakes, thus creating false positives that could discourage the usage of deepfake detectors as a mean of information checking.

Another major challenge in AI-based manipulated media detection is in the need for adequate databases of typical manipulations in specific use cases with reliable labelling and containing a large enough corpus, to train AI modules behind such approaches. Unfortunately, access to typical manipulated contents in large enough numbers and in real-life situations is not always possible due to different reasons ranging from confidentiality and privacy concerns to the scarcity of some types of manipulations. Beside efforts in collecting adequate databases, in many situations, augmentation of the dataset or reliance on machine learning approaches that require less, or no training are among potential solutions to explore.

In a general manner, one of the critical weaknesses of most AI-based approaches is that of the black box challenge which makes it difficult to know why and under which circumstances the approach fails and more importantly, how to improve it. One way to address this challenge is to work on explainable AI-based manipulated media detection [1].

An inherent challenge in any detection paradigm is what is often referred to as a cat and mouse game, where on one hand the performance and quality of manipulation improves to the point that it can fool not only human but also machines in their detection capability, until they can learn how to cope with a new category of attack and improve their detection performance. It is a fact that a cat and mouse paradigm is not an effective approach intrinsically, because the detector plays catch up while the attacker always has the initiative. In this context it is also worth noting that the adoption of technology plays a crucial role for such purposes. For example, the media distribution platforms including social media may not be prompt to integrate the most advanced detection solution for various reasons and therefore the purpose of filtering doctored

[3] https://github.com/MarekKowalski/FaceSwap/

media content can be squarely defeated. This puts detection in the weaker position in the game. In several use cases, such as in manipulated media for the purpose of disinformation and misinformation, this could result in significant harm, when considering that in the majority of the cases, fake news, disinformation and misinformation spreads within hours and days and by the time detectors are available to cope with them, it will be too late. This challenge calls for solutions that operate outside of a cat and mouse paradigm such as those addressed in the next section.

5. PROVENANCE AS A SUSTAINABLE APPROACH

Since detection approaches always run behind continuously improving generation methods, they cannot provide a sustainable long-term solution. Therefore, a longer-term vision should rather rely on a sustainable approach, such as signaling associated provenance information of media assets [21]. Also in creative use cases, the ability to signal provenance and authenticity information in an interoperable way enables full transparency about the nature of the content towards consumers. The provenance is defined as a set of information about a media asset including the trail of modifications starting from an actor. An actor is a human or non-human (software or hardware) that participates in the media ecosystem. Common actors include the camera used to capture the asset, the photographer, the editor and the editing software.

The provenance definition implies that the associated provenance information does not necessarily span the entire lifecycle of an asset. Figure 3 shows an illustrative example using the scenario of a professional photographer sharing an image with a client. The media asset origin is the digital camera used to capture the image, the image produced by this camera is the media asset source, which is in this example stored as a camera RAW. The photographer made some enhancements to the source image such as straightening, cropping, increasing the contrast and white balance correction. Thereafter, he/she shares with the client a final high-quality JPEG coded image containing several modifications when compared to the source image. This version is called the digital master. For the photographer, the provenance of the media asset covers the trail of modifications from the origin to the digital master. When the client shares the image on social media, a newly modified version is created that may include additional modifications such as rescaling to a lower resolution and/or transcoding to a lower image quality. The accessible provenance information for the client may not include the provenance information prior to the creation of the digital master. Therefore, the media asset authenticity for the client relies on trust on the photographer who supplied the digital master.

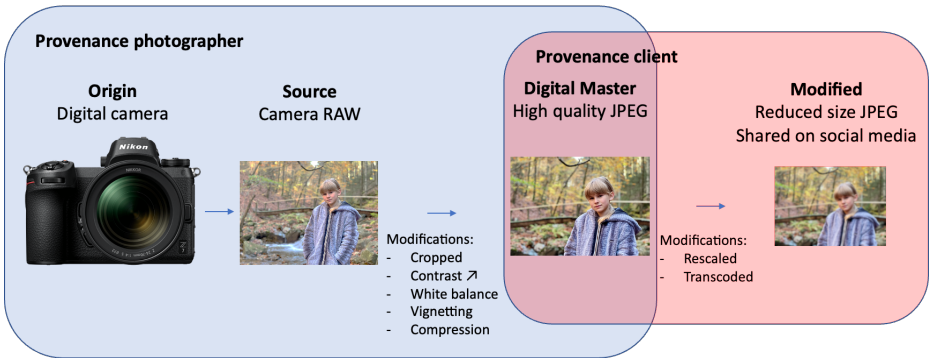

Figure 3: Provenance example in a professional photographer usage scenario.

In a scenario that relies on signaling provenance information, it is crucial that the provenance metadata is associated with the media asset content in an interoperable and secure way. In addition, such a provenance model should allow describing information related to the origin and creation of the asset as well as information related to modifications that have been made during the life cycle of the media asset. In some use cases, additional information related to the actors that interacted with the asset might be required. For example, potentially relevant actors could include the used camera as well as the photographer or an editor. This vision, relying on securely attaching provenance information as metadata directly to media assets, is pursued by several recent initiatives, including the Content Authenticity Initiative (CAI) [19]and Project Origin that recently merged into the Coalition for Content Provenance and Authenticity (C2PA) [19] and the JPEG Fake Media exploration [13].

Nowadays, it is common practice that metadata is stripped from media assets, for example when they are shared on social media. This causes a threat to the provenance model since in such a case the provenance information is detached from the

asset. Therefore, a long-term vision should consider every asset that has no provenance information as untrustworthy. However, at the same time it is important to acknowledge why certain applications strip metadata. The motivation is mostly two-fold. First, some embedded metadata might contain sensitive private information, such as the geolocation. Second, in some cases metadata can be considerably large and as such impact bandwidth usage or storage space. Therefore, future frameworks should support privacy and security mechanisms to protect private information and methods to reference externally hosted metadata to prevent superfluous bandwidth consumption during distribution. Specifically for JPEG media assets, such functionalities could be supported by defining extensions to the JPEG Universal Metadata Box Format (JUMBF). JUMBF is a JPEG Systems specification that provides a framework to facilitate metadata-based extensions within the JPEG ecosystem [23]. JUMBF has already been adopted by the CAI, not only for JPEG assets but also for other media modalities, including PDF documents [19].

6. JPEG FAKE MEDIA: WHY, WHAT AND WHEN

Previous sections have already presented the background and state of the art for media asset creation, modification, manipulation and detection as well as the vision to offer an interoperable, secure and trustworthy annotation ecosystem based on the signaling of the provenance information of media assets.

To address these needs, the JPEG Committee has launched the so-called JPEG Fake Media exploration activity, whose scope is the creation of a standard that can facilitate the secure and reliable annotation of media asset creation and modifications. The standard shall support usage scenarios that are in good faith as well as those with malicious intent [13].

6.1 JPEG Background

The JPEG Fake Media activity has a long, related background within JPEG. In fact, JPEG has a long-lasting set of contributions to make available media related standards to support imaging interoperable ecosystems. The following standards and explorations deserve to be mentioned:

- Multiple JPEG image coding formats, starting with the omnipresent JPEG 1 format, after JPEG 2000, and more recently the JPEG XS and JPEG XL formats.
- JPEG Pleno addresses the efficient coding of emerging plenoptic imaging modalities such as light fields, point clouds and holographic images.
- JPSEC provides a framework, concepts and methodology for securing JPEG 2000 codestreams. JPSEC provides three types of security services for JPEG 2000 images, notably confidentiality, integrity, and authentication.
- JPEG Systems provides a metadata and extensions framework that can be used with any of the JPEG coding formats. Here the JPEG Universal Metadata Box Format (JUMBF) deserves a special mention as the foundation on which metadata-based extensions are built [23].
- JPEG Privacy and Security is a JUMBF based extension focusing on features related to protection and authenticity. For protection, the standard supports tools to protect parts of images and/or associated metadata while retaining backward and forward compatibility. For authenticity, the standard focuses on the use of signatures or hashes to check integrity [22].
- Recent media blockchain and NFT exploration activities investigate standardization needs related to media blockchains and NFT, respectively.

This background was the fuel for the JPEG Fake Media exploration activity which has already identified a set of relevant use cases and requirements, as briefly reviewed in the next subsections.

6.2 Use cases

To better understand the needs related to JPEG Fake Media, it is critical to identify, analyze and learn from the relevant use cases, and there are many. It is important to keep in mind that JPEG Fake Media addresses media asset creation and modification, independently of the good or bad intent.

Based on multiple contributions, the use case clusters shown in Table 1 have been identified; for details on each use case, please see [13]. While a first classification may indicate that the use cases of the left columns are associated to ‘bad intent’ and those on the right columns to ‘good intent’, this type of classification is many times not that sharp as the intent depends on the context, associated text, timing, etc.

Table 1 – JPEG Fake Media use cases [13].

Misinformation and disinformation	**Forgery/Media forensics**	**Media creation**	**Media modification**
● Media usage in breaking news ● Deepfake detection ● Content authenticity checking ● Content usage tracing ● Fraud in academic research ● Photographic framing	● Insurance fraud ● Mileage reporting photo ● Photo for cost charge ● Evidence of trial ● Media sharing on social media ● Credibility of AI training image data sets	● Movie special effects ● Media transcoding ● Chroma keying or silhouette extraction	● Image colorization and restoration ● Photo editing

Naturally, the list of use cases in Table 1 is not exhaustive and will be extended and completed over time. However, it is already representative enough to extract the set of requirements presented in the next subsection.

6.3 Requirements

The use cases above allowed to derive a good set of requirements for JPEG Fake Media which have been clustered in three categories, notably [13]:

- *Media creation and modification descriptions*
- *Metadata embedding and referencing*
- *Authenticity verification*

The key requirements identified related to *media creation and modification descriptions* are:

- The standard shall provide means to describe:
  - How, by whom, where and/or when the media asset was created and/or modified.
  - The type (*e.g.*, transcoding, contrast, brightness, color temperature, adding annotations, *etc.*) and category (*e.g.*, global, local, restoration, enhancement, composition, ...) of modifications.
  - The region of interest (ROI) where the media asset was modified.
  - The purpose of a modification.
  - (Algorithmically or by humans) the probability of the existence of a modification and which method was used to determine that probability.
- The standard shall also provide means to:
  - Reference the asset(s) on which the modifications were applied and/or that were used for its creation.
  - Keep track of the provenance of media assets and/or of specific modifications.
  - Signal the extent of modifications compared to a reference version, for example by providing an objective similarity metric. The standard shall also provide means to signal which method was used.
  - Signal IPR information related to media assets and/or to specific modifications.

The key requirements identified related to *metadata embedding and referencing* are:

- The standard shall comply with the JPEG Systems framework and should retain backwards compatibility.
- The standard shall allow for accommodating non-JPEG formats.
- The standard shall be intelligible as a self-contained structure.
- The standard shall provide means to:
  - Embed provenance, authenticity and IPR information into media assets.
  - Verify the integrity of the media asset by supporting, *e.g.*, various hashing methods; various signing methods; various digital fingerprinting methods; the ability to embed multiple signatures, hashes, or fingerprints with different scope.
  - Protect media asset metadata, including provenance information.

- Provide conditional access to media asset metadata.
- Compress embedded descriptions.
- Embed references to externally hosted descriptions, methods and services.
- Keep track of modifications made to the media asset content and provide means to compare with or rollback to a previous version.
- Keep track of modifications made to the media asset metadata and provide means to compare with or rollback to a previous version.
- Signal what should happen with embedded JUMBF boxes in case modifications are applied: carry over, remove, update, warn the user about potential inconsistencies.

The key requirements identified related to *authenticity verification* are:

- The standard shall support:
  - Registration of media assets, media asset metadata and media asset content.
  - Registration of the actors involved in the media asset creation, modifications and distribution.
  - Decentralized and centralized registration solutions.
  - Explicit denotation of anonymous, obscured, or redacted information; if the information is not provided, then it is considered anonymized.
- The standard shall provide means to:
  - Describe the identify the origin, source or digital master of the media asset while also supporting anonymization or obfuscation of that information if demanded by the use case.
  - Verify the integrity of the media asset.
  - Verify the authenticity of the media asset.
  - A pathway to enable registration of media assets sources, digital masters and modified versions.

As for the use cases, these requirements are under development and may be improved and completed over time.

6.4 Roadmap

In standardization, the most common consequence of a process where use cases and associated requirements are identified is to issue a Call for Proposals asking for relevant technology. Naturally, this will require first to inform and engage with relevant stakeholders. This has been happening, notably through a sequence of dedicated workshops. In addition, appropriate assessment methodologies and metrics to fairly evaluate the proposals need to be developed. After evaluation, the selected technical solutions will be collaboratively improved and completed to address the set of requirements and the outcome might be included in a new standard or might lead to extensions of already available, relevant JPEG standards. For the moment, the detailed timeline is not yet defined but this will likely happen at the October 2021 JPEG meeting.

7. CONCLUSION

The ways to easily create and modify media assets have multiplied in recent years, notably allowing to produce very realistic media assets to the human eye, which may be essentially indistinguishable from genuine media assets, *e.g.*, deepfakes. The bright side of these advanced tools opens opportunities for the creative production of new media in the entertainment, education, and art to mention a few among many. However, its dark side, the intentional or unintentional spread of manipulated media, *e.g.*, modified media with the intention to induce misinterpretation, brings big risks such as social unrest, spread of rumors for political gain or encouragement of hate crimes. In this context, the transparent annotation of media creation and modifications is considered a crucial step to allow users to assess trustworthiness of media. Moreover, these annotations should be attached to the media assets in a secure way to prevent them from being compromised. Finally, to create a powerful, secure and trustworthy ecosystem, interoperability is a must, thus asking for appropriate standards.

This paper has reviewed the background and implications of media manipulation as well as the state of the art of forensic technologies, with emphasis on the challenges associated with AI-based manipulated media detection. Following this context, the paper proposes a sustainable approach to allow creating secure and trustworthy media annotations based on the ability to signal provenance and authenticity information in an interoperable way to enable full transparency about the nature of the content to the users. Finally, a concrete initiative adopting this provenance-based approach is presented, *i.e.*, JPEG Fake Media, which targets the creation of a standard that can facilitate the secure and reliable annotation of media asset creation and modifications, supporting both good faith and malicious intent use cases and associated requirements.

ACKNOWLEDGMENT

The authors acknowledge valuable exchanges with members of the JPEG Fake Media Ad Hoc Group. The last two authors acknowledge support from CHIST-ERA project XAIface (CHIST-ERA-19-XAI-011) with funding from Fundação para a Ciência e Tecnologia (FCT) and Swiss National Science Foundation (SNSF) under grant number 20CH21_195532.

REFERENCES

[1] Adadi, A., and Berrada, M., "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," in IEEE Access, vol. 6, pp. 52138-52160 (2018).
[2] Aengus C. & Touradj E., "Risk governance and the rise of deepfakes", Swiss Federal Institute of Technology Lausanne (EPFL), May 2021, [Online: posted 12-May-2020].
[3] Afchar, D., Nozic, V., Yamagishi, J., and Echizen, I., "MesoNet: a Compact Facial Video Forgery Detection Network," in 2018 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–7, iSSN: 2157-4774 (2018).
[4] Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K. and Li, H., "Protecting world leaders against deep fakes.", in CVPR Workshops, pp. 38–45 (2019).
[5] Aneja, D., Colburn, A., Faigin, G., Shapiro. L., and Mones, B., "Modeling stylized character expressions via deep learning," in Asian conference on computer vision. Springer, pp. 136–153 (2016).
[6] Boutarfass, S. and Besserer, B., "Improving cnn-based colorization of b&w photographs," in IEEE 4th International Conference onImage Processing, Applications and Systems (IPAS), pp. 96–101 (2020).
[7] Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., and Ferrer, C., "The DeepFake Detection Challenge (DFDC) Dataset," arXiv:2006.07397 [cs], Oct. 2020, arXiv: 2006.07397. [Online]. Available: http://arxiv.org/abs/2006.07397
[8] Farid, H., "Digital doctoring: how to tell the real from the fake", Significance, 3(4), 162-166 (2006).
[9] Gray, C., "Using Machine Learning to Identify Fake Images", Bachelor's thesis, University of Stirling (2019).
[10] Guarnera, L., Giudice, O., and Battiato, S., "Deepfake detection by analyzing convolutional traces." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2020).
[11] Holden, D., Jun Saito, J., and Taku Komura, T., "A deep learning frame-work for character motion synthesis and editing," ACM Transactions on Graphics (TOG), vol. 35, no. 4, pp. 1–11 (2016).
[12] ISO/IEC JTC 1/SC29/WG1 N90049, "White Paper on JPEG AI Scope and Framework v1.0", 90th JPEG Meeting, Online, January 2021.
[13] ISO/IEC JTC 1/SC29/WG1 N92018, "JPEG Fake Media: Context, Use Cases and Requirements", 92nd JPEG Meeting, Online, July 2021.
[14] Lindbergh, B., "How they made it: The deeply real deepfakes of 'For all mankind'," March 2021, [Online; posted 5-March-2021].
[15] Mahdi, M. and Behroozi, H., "Context-aware col-orization of gray-scale images utilizing a cycle-consistent generativeadversarial network architecture," Neurocomputing, vol. 407, pp. 94–104 (2020).
[16] Nguyen, H., Fang, F., Yamagishi, J., and Echizen, I., "Multi-task Learning for Detecting and Segmenting Manipulated Facial Images and Videos," in 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–8 (2019).
[17] Nguyen, H., Yamagishi, J., and Echizen, I., "Use of Capsule Network to Detect Fake Images and Videos," arXiv:1910.12467 [cs], Oct. 2019, arXiv: 1910.12467. [Online]. Available: http://arxiv.org/abs/1910.12467
[18] Nirkin, Yuval, et al. "DeepFake Detection Based on Discrepancies Between Faces and their Context." IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).
[19] Rosenthol, L., Parsons, A., Scouten, E., Aythora, J., MacCormack, B., England, P., Levallee, M., Dotan, J., Hanna, S., Farid, H. and Gregory, S., "The Content Authenticity Initiative, Setting the Standard for Digital Content Attribution", (2020).
[20] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M., "Faceforensics++: Learning todetect manipulated facial images," in Proceedings of the IEEE/CVFInternational Conference on Computer Vision, pp. 1–11 (2019).
[21] Temmermans, F., Truyen, F., Dooms, A., Schelkens, P., Vandermeulen, B., "Integrity of Still Images", Image and Research (2018).

[22] Temmermans, F., Ebrahimi, T., Foessel, S., Delgado, J., Ishikawa, T., Natu, A., Schelkens, P., “JPEG Privacy and Security framework for social networking and GLAM services”, EURASIP Journal on Image and Video Processing, 2017, 68 (2017).
[23] Temmermans, F., Kuzma, A., Choi, S., Schelkens, P., “Adopting the JPEG systems layer to create interoperable imaging ecosystems”, Proc. SPIE 11353, Optics, Photonics and Digital Technologies for Imaging Applications VI (2020).
[24] Thies, J., Zollhofer, M., and Nießner, M., “Deferred neuralrendering: Image synthesis using neural textures,”ACM Transactions onGraphics (TOG), vol. 38, no. 4, pp. 1–12 (2019).
[25] Thies, J., Zollhofer, M., Stamminger, M., Christian Theobalt, and Nießner, M., “Face2face: Real-time face capture and reen-actment of RGB videos,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2387–2395 (2016).
[26] Wang, M., W., Lyu, X., Li, Y., and Zhang, F., "VR content creation and exploration with deep learning: A survey", Computational Visual Media, vol. 6, no. 1, pp. 3–28 (2020).
[27] Yarımbaş, N., "Netflix vs Deepfake: The Irishman", September 2020, [Online: posted 9-Sep-2020 on https://medium.com/].
[28] Zhou, P., Han, X., Morariu, V. I., and Davis L. S., “Two-Stream Neural Networks for Tampered Face Detection,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1831–1839 (2017).