

Visual Mis/disinformation in Journalism and Public Communications: Current Verification Practices, Challenges, and Future Opportunities

T.J. Thomson , Daniel Angus , Paula Dootson , Edward Hurcombe & Adam Smith

To cite this article: T.J. Thomson , Daniel Angus , Paula Dootson , Edward Hurcombe & Adam Smith (2020): Visual Mis/disinformation in Journalism and Public Communications: Current Verification Practices, Challenges, and Future Opportunities, Journalism Practice, DOI: [10.1080/17512786.2020.1832139](https://doi.org/10.1080/17512786.2020.1832139)

To link to this article: <https://doi.org/10.1080/17512786.2020.1832139>



Published online: 19 Oct 2020.



Submit your article to this journal [↗](#)







View related articles [↗](#)



View Crossmark data [↗](#)



Visual Mis/disinformation in Journalism and Public Communications: Current Verification Practices, Challenges, and Future Opportunities

T.J. Thomson , Daniel Angus , Paula Dootson , Edward Hurcombe  and Adam Smith

Digital Media Research Centre, Queensland University of Technology, Brisbane, Australia

ABSTRACT

Social media platforms and news organisations alike are struggling with identifying and combating visual mis/disinformation presented to their audiences. Such processes are complicated due to the enormous number of media items being produced, how quickly media items spread, and the often-subtle or sometimes invisible-to-the-naked-eye nature of deceptive edits. Despite knowing little about the provenance and veracity of the visual content they encounter, journalists have to quickly determine whether to re-publish or amplify this content, with few tools and little time available to assist them in such an evaluation. With the goal of equipping journalists with the mechanisms, skills, and knowledge to be effective gatekeepers and stewards of the public trust, this study reviews current journalistic image verification practices, examines a number of existing and emerging image verification technologies that could be deployed or adapted to aid in this endeavour, and identifies the strengths and limitations of the most promising extant technical approaches. While oriented towards practical and achievable steps in combating visual mis/disinformation, the study also contributes to discussions on fact-checking, source-checking, verification, debunking and journalism training and education.

KEYWORDS

Visual misinformation; visual disinformation; fake images; image verification; vetting images; image accuracy; visual media forensics; journalistic verification

Introduction

“For almost every globally significant news event of 2019—and of these there were many—it is possible to identify misleading video and pictures that were shared on social media in the aftermath”, said Reuters employee Hazel Baker, global head of user-generated content newsgathering for the organisation. Baker (2019) continued:

This year the Reuters newsroom discovered, among many other examples, that a harrowing video shared during Cyclone Idai had been shot in Libya five years earlier. We watched videos going viral following Brazil’s Brumadinho dam collapse that were from different, unrelated incidents. And during conflict between India and Pakistan in February this year, we saw numerous false clips circulating, including one taken from a video game. As social media

has evolved to become a critical part of newsgathering, it has also become a minefield, where misinformation travels rapidly. Newsrooms have had to adapt, growing journalists' verification skills in order to be able to filter powerful, authentic eyewitness media from that which distorts the truth.

Visual mis/disinformation is proliferating and journalists are often complicit in amplifying visual information with unknown provenance and unknown accuracy. As examples, when devastating floods surged Queensland, Australia, in early 2019, people posted and shared photos of alleged crocodiles on the streets of the affected area only for it to be later discovered that the images were from 2014 and depicted American alligators in Florida, USA (see [Figure 1](#)). Other images purporting to be of the 2019 flood in actuality originated from other geographic contexts and time periods but were also shared and re-tweeted, sometimes by journalists.

In the US, a video posted in September 2019, which received 4.4 million views and purported to depict Hurricane Dorian swirling above the Bahamas, was, in reality, a composite image of a storm cell in Kansas and a photo of Miami Beach that had been animated (see [Figure 4](#)). Altered images of politicians, such as ones released through US President Donald Trump's official social media accounts that show him slimmed down and with elongated fingers (see [Figure 5](#)), have broken precedents (Novak 2019) and further complicate the public's ability to discern fact from fiction. Even edits that leave pixel positions alone but alter their values—such as increasing the saturation of Donald Trump's tanned skin (see [Figure 8](#))—have engendered fierce controversy about the nature of truth and the role of the visual in mediating reality.

The examples of misattributed, doctored, and faked imagery shared earlier attest to the importance of accuracy, transparency, and trust in the arena of public discourse. People vote and make decisions based rarely on first-person experience but, rather, through mediated depictions that come to them from friends and family, politicians, organisations, and journalists. The visual information we see and interact with is especially potent considering its emotional pull and its persuasive influence (Joffe 2008). While journalists who create visual media haven't been immune to ethical breaches¹ (Lester 2016), the practice toward incorporating more user-generated and crowd-sourced visual content into news



Figure 1. American alligators in Australia.

Note: Australia-based Nine News was complicit in amplifying a photo of what was allegedly a crocodile in the streets of Townsville, QLD, Australia, during a 2019 flood but, in reality, depicted an American alligator in Florida, USA, from six years prior.

reports is only growing (Matatov et al. 2018) as the number of traditionally more trustworthy and credible professional visual journalists decreases (Thomson 2018).

The number of visuals created daily (in excess of 3.2 billion photos and 720,000 hours of video daily), the speed at which they are produced, published, and shared, the digital and visual literacy of those who see them, and economic realities and ideological differences can all impact whether visual mis/disinformation is detected, how quickly, by whom, and with what effects. While the creation of synthetic media is not inherently problematic (for example, in the cases of art, satire, or parody), issues emerge when those media are presented without transparency and masquerade as reality. The complexity surrounding such issues is compounded when we consider the broad continuum of what Wardle (2018) terms “information disorder”. In delineating such a conception, Wardle outlines seven categories, including (1) satire or parody, where no intent to harm exists but the media can still fool those who see it; (2) false connection, when, for example, stock photos are packaged with a news story about a natural disaster but the severity or attributes of the stock photo don’t match the current situation; (3) misleading content, such as when images are cropped to inaccurately frame an issue or individual; (4) false context, such as when dated photos are used along with text that claims they originate from a current situation; (5) imposter content, when genuine sources are impersonated, such as when Photoshopped tweets or screenshots are made that alter the original content; (6) manipulated content, such as adding or removing elements from a photo or video; and (7) fabricated content, when 100 percent of the content is false and the intent is to deceive or do harm.

Worthy of mention, too, is a brief discussion of why people manipulate visuals in the first place. The reasons will differ based on the type of person manipulating the media (e.g., satirist, artist, professional media worker, etc); however, at a societal level and considering disinformation,² specifically, much of the motivation for creating, editing, or sharing inauthentic visual content is politically or economically motivated (Bakir and McStay 2018).

What disinformation seeks ... is not necessarily to convince the public to believe that its content is true, but to impact on agenda setting (on what people think is important) and to muddy the informational waters in order to weaken rationality factors in people’s voting choices. (Ireton and Posetti 2018, 10)

Such campaigns seek to distract, confuse, manipulate, and sow division, discord, and uncertainty and are more prevalent in highly polarised nations where socioeconomic inequalities, disenfranchisement, and propaganda are common (ibid). They can also, for economic reasons, capitalize on the attention and awareness people pay to sensational but false content over real but less engaging content. Journalists, thus, not only have to verify whether an image has been doctored, but also to consider who doctored the image in which way(s) and for what purpose(s). Yet the tools available to the everyday news consumer or journalist attempting to verify a media item’s provenance and accuracy are few and ill-suited for some types of forgeries, as will be further detailed later on.

The purpose of this study is to map journalists’ current social media verification techniques alongside the verification tools that they have available, to identify emerging computer vision techniques that can detect visual mis/disinformation, and to offer recommendations for those in journalism on which tools are most effective for which

circumstances. In order to do so, we first explore and examine common examples of problematic edits within the visual news realm before discussing techniques to identify and combat them. The review is aimed at a general audience of journalists, considering that any solutions proposed must take into account the diverse backgrounds of stakeholders involved in the production, editing, and sharing of the visual media in question. Likewise, any solutions offered need to accommodate journalists' production and editing workflows.

Image Ethics and Verification Strategies in Visual Journalism and Public Communications

Numerous books have been written on ethics related to visuals presented in news contexts, in particular, and on image ethics in the digital age, more broadly (see Silva and Eldridge 2020). Of special note are Gross, Katz, and Ruby's (2017) and Lester's (2016) books, which provide an overview of relevant issues from a visual media perspective and consider both still and moving images. Within the visual journalism sphere, Lester addresses ethics both in the field, such as privacy concerns and coverage of victims, as well as in the editing suite, such as colour manipulations, airbrushing content out, repositioning elements within the composition, altering their size, or combining two or more images without acknowledgement of such. Gross and colleagues address the politics of visual representation and also situate their enquiry more fully in the internet age, addressing issues created or compounded by dissemination speed, a desire for spectacle, and computer-generated images. Both texts acknowledge that the potential for deception has existed and been exercised throughout each technology's inception and evolution. More recently, however, public awareness and concern for manipulated media has been kindled and the need for tools to identify deceptive visuals has grown in response.

Scholars have given significant attention to mis- and disinformation, especially that which is text-based rather than visual, on social media platforms. Most notable in this regard is "fake news"—a contentious and politically loaded term commonly used to describe fabricated or deliberately misleading news content (Rose 2017). Here, we present a review of current scholarly and grey literature on how journalists have responded to social media-based mis- and disinformation, in particular visual media in news contexts.

While there is significant extant research on the social media verification practices of journalists, there is less research on the verification of visuals, especially during time-sensitive, breaking news contexts (Matatov et al. 2018). This is despite the ever-growing list of high-profile instances of misattributed or manipulated imagery circulating on platforms during such circumstances. In addition, the literature indicates that responses by journalists to platform-based mis- and disinformation (of any kind) have to be grounded within the context of an industry increasingly reliant on platforms for both news content and news distribution (Gottfried and Shearer 2016; Caplan and boyd 2018), and which is attempting to balance maximising monetisable audience metrics with responsible reporting, while also grappling with severe economic disruption (Franklin 2014). Moreover, current verification practices are shaped by longstanding industry demands, such as breaking news and the 24-hour news cycle.

In 2017, a global quantitative study performed by the International Centre for Journalists (ICFJ) found a very low usage of social media verification tools in newsrooms. The ICFJ surveyed over 2,700 journalists and newsroom managers in over 130 countries, and found that only 11% of those surveyed used social media verification tools. While the ICFJ found that many newsrooms considered verification an issue for journalists, with 46% of the newsrooms surveyed providing training in social media research and verification, only 22% of journalists identified such training “helpful” (2017).

Other studies that have focused on journalists’ social media verification practices are largely qualitative, interview-based, and conducted in a European context. Brandtzaeg et al. (2016), for example, interviewed 24 journalists from major European news organisations, and identified tensions between verification processes and the demands for fast-paced publishing. They argued that journalists would need “efficient and easy-to-use” tools to deal with such demands (2016, 323). Tolmie et al., had similar findings to Brandtzaeg et al., in their interviews with a Swiss news organisation (2017). The authors found that the journalists they interviewed stated that they wished for tools that could provide indications of veracity “at a glance”, especially when news desks were busy. Manual verification and non-automated fact-checking processes were very time-consuming, if possible at all. Due to this, the journalists interviewed had a strong preference towards using social media content produced by “trusted” users, such as other news organisations or government organisations, as they felt some degree of veracity could be assumed. Likewise, Pantti and Sirén (2015) studied image verification of “amateur” (that is, user-generated) images in Finnish newsrooms. They found the news organisations had few means for verifying these images in breaking news events, although they found that image verification was a growing problem for these organisations.

Researchers and industry organisations alike have also created a number of tools and training manuals in an attempt to assist journalists and ordinary citizens with social media verification. For instance, on the tools front, the REVEAL project introduced two tools to assist journalists with verification tasks (Zampoglou et al. 2016). One was the Journalist Decision Support System (JDSS) which classifies Twitter-based UGC in real-time, such as by identifying whether tweets contain false claims based on linguistic patterns associated with previous false claims in posts (Middleton 2017). Another was an Image Verification Assistant, a free browser-based tool designed for journalists to test visuals for image-tampering. The tool also provided a metadata analysis (which doesn’t work for images posted on social media platforms as such metadata are stripped upon uploading) and reverse image search functionality (2019); however, websites like 4chan or Reddit aren’t included in Google’s consumer-facing reverse image search results, despite the prevalence of problematic images on such sites (Matatov et al. 2018). Additionally, reverse image search processes only return exact or near-exact matches so, for example, search engines would have trouble identifying as a match a flipped version of an image (ibid). As of the time of writing, only the Image Verification Assistant is still functioning.

Another tool, still (which isn’t available to the public), is the “News Provenance Project”, announced in 2019 by the *New York Times* and IBM with the goal of providing audiences “with a way to determine [through blockchain] the source of a photo, or whether it had been edited after it was published” (Koren 2019). In addition to working only for certain of images (i.e., those that have been previously published in a raw/unedited state), the project’s leaders also acknowledged the following year that “effective solutions would

require large-scale cooperation and commitment, across several industries”, including with camera makers, news publishers, and platforms, such as Google, Facebook, Twitter, and Apple, and apps like WhatsApp and Signal (Tameez 2020, n. pag.), which makes the project’s future uncertain.

On the training manuals front, media forensics expert Hany Farid (Smith 2018) provides four current manual detection strategies, including:

1. using a reverse image search, which relies on original and unedited images being publicly available online,
2. examining image metadata, which requires the images in question not to have been sourced from social media as, for example, Facebook and Twitter strip such information whenever a photo is uploaded to their platforms,
3. examining the light and shadows in a scene to see if they make visual sense, which requires shadows to be visible in the image, and
4. using image editing software to adjust the media’s contrast, brightness, and exposure, which can only potentially reveal certain types of forgeries and requires software that can be expensive or can require technical skill to operate and interpret.

Many of these methods are also included in a training guide written by an International Fact-Checking Network employee and published by the Poynter Institute (Tardáguila 2020) but they rely on inconsistencies that are visible to the naked eye or require copies of the original, unaltered media item to also be available elsewhere online.

In 2020, Reuters partnered with Facebook, the platform where the highest number of false political stories are shared (McCabe and Alba 2020), to launch a new, free, online short course on “Identifying and Tackling Manipulated Media”. The course provides a module with tips on how to identify deep fakes and also a module with a suggested verification workflow, which bears similarities to the already referenced International Fact-Checking Network guide (Tardáguila 2020), published by Poynter. While such initiatives can raise awareness of manipulated media and provide tools to potentially combat it, they only works for certain types of forgeries and only when journalists have adequate time and resources to conduct in-depth forensic analyses. Further, the verification work required is often duplicative as many news organisations and individual journalists must each independently vet and verify images that aren’t “featured” on fact-checking websites like Snopes.com.

Journalists have also produced resources for other news professionals, most notably *The Verification Handbook* (Silverman 2014), which was updated in 2020. The handbook provides information and tips from a range of news professionals on how to verify social media user-generated content, as well as providing a list free verification tools for journalists to use. However, it is not currently clear whether the tools listed in the Handbook have had a significant uptake—as the ICFJ study (2017) found many journalists still seem to either lack the time, access, knowledge, or suitable training to use verification tools. Similarly, with the recent release of the Reuters/Facebook short course, it is too early to tell what kind of adoption it will have. Most importantly, training manuals and guides don’t address visual fakes that are invisible to the naked eye or that exist without a copy of the original, unaltered media item online to serve as a verification check. For example, a reverse image search wouldn’t be able to help journalists detect the

retouching on images posted to US President Donald Trump's official social media accounts (see Figure 5) unless the original, unedited versions of those same photos had also been previously published. For these reasons, journalists must be aware of and seek to use, when appropriate, both passive and/or active digital forensic techniques (detailed below) during their verification workflows.

Recent studies corroborate to an extent claims that expertise, digital literacy and cautious scrutiny are significant factors in identifying suspicious imagery: as Shen, et al, indicated, photo-editing experience, Internet skills, and prior political convictions all play a role in how journalists determine the veracity of different images (2019). Still, although such news articles (Koettl 2018) work to assure readers of the capability of journalists to deal with malicious and misleading social media activity, they lack a response to the findings in the scholarship relayed above: that accurate and verified reporting, although occurring, is significantly impeded by structural issues within the news industry and a lack of digital literacy among journalists.

There is thus a great opportunity, here, to study if and in what ways journalists deal with visuals from social media that appear to be manipulated, misattributed or otherwise malicious. Although, as the literature on journalistic verification practices indicates, such a study would need to pay acute attention to the technological, industrial, professional and occupational contexts in which media workers live. From such contextualised research, researchers can begin to develop useful and successful tools to assist journalists in verifying visuals on social media.

Types of Issues and Corresponding Detection Methods

As photo and video editing technology becomes more widespread and sophisticated, developed society is increasingly being exposed to more imagery with unknown provenance and accuracy. Journalists, media outlets, and law enforcement often cannot establish the veracity of an image by simply probing its source, and now look to use passive "digital forensics" tools in order to know whether an image's contents can be trusted (Songcharoen, Bite, and Clay 2014; Farid 2018). This section presents an overview of types of manipulation and opportunities journalists have to identify and evaluate them.

Considering the above-noted issues concerning the media literacy required to investigate visual fakery, we present below five common types of image manipulations alongside a grid (see Table 1) for an example of each category within a journalistic context alongside the best approach(s) to detect it.

- **Copy-move.** A copy-move or cloning operation takes a region of pixels within an image, copies it, and moves this copy into another area within the image. Such was the case with a 1964 photo of Dr Martin Luther King Jr, which was edited using the copy-move technique to make it appear that MLK was flipping off the camera (see Figure 2). Copy-move operations are among the most common image forgery techniques, as to the human eye they uphold the internal consistency of an image. Copy-move operations are also among the most commonly detected manipulations. They are often identified by looking for JPEG compression inconsistencies, CFA interpolation inconsistencies, contrast inconsistencies, and are detected by most deep-learning based general classifiers. In a journalistic context, peak bodies, such as the

Table 1. Typology of common forgery operations in journalistic and public communications contexts.

Type of forgery	Example in a public communications context	Method best suited for detection
Saturation/desaturation	A photo is posted on Twitter that appears to have been edited to make the difference between US President Donald Trump's tanned and untanned skin more extreme.	Colour / Lighting inconsistency, DNNs. Investigate the correlation between channels.
Artificial blurring	The US National Archives edits a news photo by Mario Tama for Getty of the 2017 Women's March by blurring out certain words, such as a reference to US President Donald Trump in a sign that read, "God Hates Trump", among other edits.	CFA interpolation. Look for inconsistencies in interpolation.
Copy/move and splicing	A photo of MLK Jr is edited to make it appear he is flipping off the camera or multiple photos are combined together to make it look like a girl standing in water is in front of flames, wearing a gas mask, and is holding a koala during the 2019–20 Australian bushfire crisis.	DNNs. Look for double-jpeg compression, CFA interpolation inconsistencies, noise inconsistencies etc. with statistical tools or general classifier.
Retouching	Photos shared on US President Donald Trump's official Instagram and Facebook accounts were edited to show him with elongated fingers, a tightened waistline and higher crotch, a slimmed neck and shoulder, and tightened hair.	CFA interpolation, noise inconsistency. Identify small local inconsistencies in correlations between neighbouring pixels.
Misattribution	A 2014 of an American alligator in Florida, USA, makes its rounds on social media in 2019 with people incorrectly claiming it depicts a crocodile in the streets of Townsville, Queensland, Australia.	Limited information about the source can be derived from JPEG metadata, CFA interpolation method, sensor noise.
Cropping	The White House crops official photos of US President Donald Trump's inauguration to make the crowd appear larger.	JPEG compression inconsistency

Associated Press in the US or the Media Entertainment & Arts Alliance in Australia provide guidance to their members about ethical conduct related to editing and presenting news visuals. Relevant to both copy-move and splicing operations is guidance from both organisations that states: "No element should be digitally added to or subtracted from any photograph" (Associated Press 2020) and "Present pictures and sound

**Figure 2.** MLK comparison.

Note: In the original photo, at right, Dr Martin Luther King Jr reacts in St Augustine, Florida, after learning that the US Senate passed the civil rights bill on June 19, 1964. An altered version of the image, using the copy-move technique, clones part of the background over MLK's second finger to make it appear that he is flipping off the camera. The edited image was shared widely on Twitter, Reddit, and on the white supremacist website *Daily Stormer*.

which are true and accurate. Any manipulation likely to mislead should be disclosed” (MEAA 2020). The US-based National Press Photographers Association encourages its members to “maintain the integrity” of the visuals’ content and context and to “not manipulate images or add or alter sound in any way that can mislead viewers or misrepresent subjects” (National Press Photographers Association 2020).

- **Splicing.** Splicing, also known as image composition, is similar to copy-move. Cloned regions from one source are copied into another to form a composite image (see [Figure 3](#)). Splicing is similarly detected by identifying JPEG compression inconsistencies, CFA interpolation inconsistencies, contrast inconsistencies, and by using general classifiers.
- **Resampling.** Resampling is often used in the process of resizing, rotating, stretching, or skewing in order to create images with different pixel densities. In a journalistic context, the Associated Press’s editing and presentation principles allow backgrounds to be removed and figures to be overlaid over a neutral background for graphics and stipulates “such compositions must not misrepresent the facts and must not result in an image that looks like a photograph—it must clearly be a graphic” (Associated Press 2020). Often during a copy-move or splicing operation, copied regions that don’t have identical dimensions need to be resampled in order to create convincing forgeries. The algorithms used to resample images can leave artefacts and statistical differences within resampled regions. This is often detected by looking for a CFA interpolation inconsistencies, noise inconsistency, and by using general classifiers ([Figure 4](#)).
- **Retouching.** Retouching is the process of making small adjustments to emphasise or obfuscate features of an image, such as blemishes, acne, or scars. Retouching is often

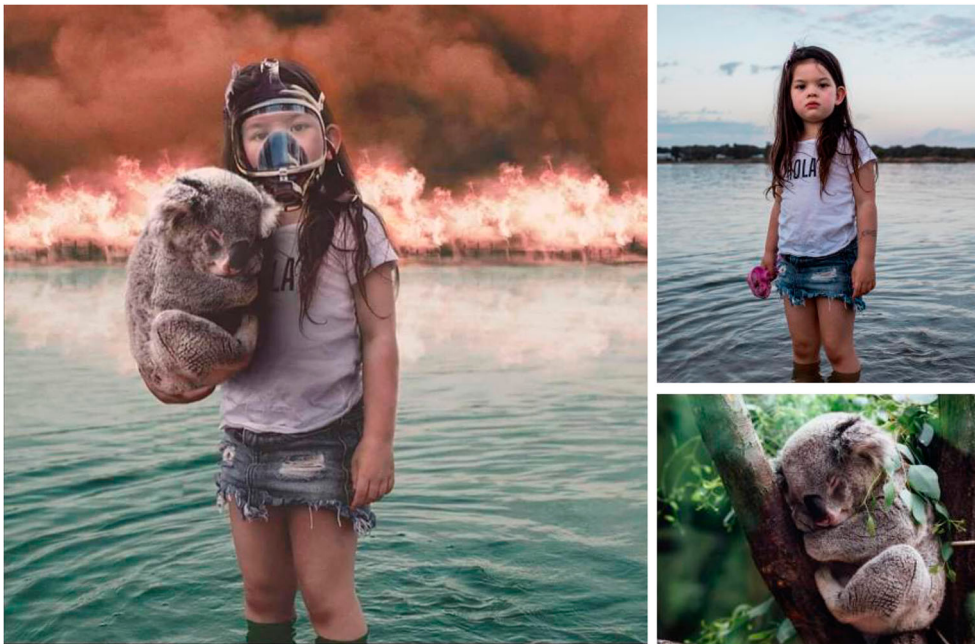


Figure 3. Australian bushfire composite.

Note: The photo on the left, which was shared widely on social media during the 2019–20 Australian bushfire crisis, is a composite of at least three other images that have been digitally combined in Photoshop using the splicing technique.



Figure 4. Hurricane Dorian composite.

Note: The composite image, at left, which uses multiple manipulation techniques, and was shared on social media with a caption that claimed it portrayed Hurricane Dorian swirling over Florida, used resampling techniques to make the horizontal photo of a storm cell in Kansas fit the vertical composition at left.

done as part of the post-production process for publishing photos of people, and recently has become popular among amateur photographers on social media. Such retouching has been done on photos of US President Donald J Trump that have been shared to his official Facebook and Instagram accounts (see [Figure 5](#)) in order to make him appear slimmer and with more flattering hair. Within a journalistic context, the only retouching allowed under Associated Press principles is “to eliminate dust on camera sensors and scratches on scanned negatives or scanned prints” (Associated Press 2020). Retouching is often detected by looking for CFA interpolation inconsistencies, noise inconsistency, and by using general classifiers ([Figure 5](#)).

- **Cropping.** Cropping removes unwanted areas at the edges of an image. Cropping is a standard, widespread, and “acceptable” form of news image editing (Zhang 2018) unless the intent is to deceive. It is often used in image forgery, however, to hide objects or conceal context. Such was the case in 2017 when, under pressure from the White House, the U.S. National Parks Service cropped official photos taken from the Washington Monument that showed the size of US President Donald J Trump’s inauguration crowd size. The cropped photos provided less overall context and made the overall crowd size less apparent ([Figure 6](#)). Such cropped edges are often highlighted when looking for JPEG compression inconsistencies.



Figure 5. Donald J. Trump comparison.

Note: The original photo of Donald J Trump (left) and the edited version of him (right) that he shared on his official Instagram and Facebook accounts and that shows him with elongated fingers, a tightened waistline and higher crotch, a slimmed neck and shoulder, and tightened hair.

Both active and passive techniques can be used to verify the authenticity of an image. Active techniques like digital watermarking are able to identify individual photographers or camera equipment, as well as uphold the veracity of an image. Digital watermarking invisibly embeds authentication data into original images, which can later be derived to verify said image (Shih 2017). The most prominent obstacle to the widespread use of digital watermarking is that photographers and media organisations don't always have incentives to use it, and it places the burden of proof and cost upon the media creators who need to use additional third-party tools during post-production (Zhou and Lv 2011).

In contrast to active techniques, passive (sometimes called blind) techniques look for visual artefacts and anomalies in the underlying composition and derived statistics of images to identify image manipulation or forgery. These tools are powerful because they can be applied to images with unknown provenance, and identify an ever-expanding list of image manipulation techniques. They are limited however, as they cannot comprehensively "prove a negative", and assessing the accuracy of these techniques can be challenging (Marshall and Paige 2018).

There are many existing literature reviews and surveys in the field of image forensics (Farid 2009; Birajdar and Mankar 2013; Qureshi and Deriche 2015), although few that consider newer deep-learning techniques, and none that do so in the context of journalism and public communication. These reviews exist primarily in the computer science and media forensics fields, and while not entirely divorced from the contexts that their

evaluated techniques can be used in, their findings are intended to be broadly applicable and don't necessarily reflect pragmatic usage.

Passive Detection Methods

Passive statistical and algorithmic techniques look for manipulation artefacts and analyse the underlying statistics of an image. Although visual artefacts are not always recognisable or visible, the underlying statistics of images are changed when they are manipulated. [Table 2](#) provides an overview of each approach alongside its strengths and weaknesses.

Table 2. Typology of statistical and algorithmic methods for detecting common forgery types.

Name of method	Type of forgery suited to detect	Strengths of the approach	Weaknesses of the approach
JPEG compression inconsistency	Copy-move, Splicing, Cropping	<ul style="list-style-type: none"> - Clear outputs in positive cases - Can be localised (e.g., can point to a specific area of the image that is problematic) - Can be used to derive information about the source (e.g., type of camera used to make the image) 	<ul style="list-style-type: none"> - Vulnerable to attack by adding precise noise - Compounded JPEG-compression of internet-sourced images with unknown provenance raises false alarms - Format-specific (e.g., doesn't work for PNGs, TIFFs, etc)
CFA interpolation inconsistency	Copy-move, Splicing, Resampling, Retouching	<ul style="list-style-type: none"> - Can be used to derive information about the source - Works in lossy & lossless contexts - Can be localised 	<ul style="list-style-type: none"> - Requires deducing the algorithms used to compose a given image (may raise false alarms in unusual cases) - Vulnerable to attack by resampling onto a CFA then re-interpolating
Contrast and lighting inconsistency	Copy-move, Splicing, Retouching	<ul style="list-style-type: none"> - Unaffected by some standard attacks - Can be localised 	<ul style="list-style-type: none"> - Limited effectiveness in JPEG-compressed images - Vulnerable to attack by global adjustments (e.g., uniformly brightening/darkening the image)
Noise inconsistency	Splicing, Retouching	<ul style="list-style-type: none"> - Can be used to derive information about the source - Can be localised 	<ul style="list-style-type: none"> - Limited effectiveness on copy-moves or splices with identical noise - Limited effectiveness in JPEG-compressed images - Vulnerable to attacks that add global noise
Convolutional neural networks	Copy-move, Splicing, Retouching	<ul style="list-style-type: none"> - Can be trained on specific artefacts - Can be trained to identify previously discussed inconsistencies - Can be localised 	<ul style="list-style-type: none"> - Vulnerable to same attacks that all neural architectures are - Complex architecture - Opaque results - May be influenced by regular features of an image

- **JPEG compression inconsistency.** The JPEG format is used to provide a universally accessible way to compress and display images with varying compression ratios (Marqués, Menezes, and Ruiz-Hidalgo 2009). JPEG sees widespread use online because it significantly reduces an image's file size with only minimal changes to its appearance. JPEGs' discrete cosine transform (DCT) compression process is "lossy", meaning that accurate pixel information is lost during compression. Splicing one JPEG image into another with photo editing software causes "double JPEG compression" (Popescu and Farid 2004), as the image is compressed a second time. This second round of compression introduces changes such as periodic artefacts in the DCT coefficient distribution of blocks within the image. This unusual periodicity can be detected using statistical tools (Fan and De Queiroz 2003). It should be noted, however, that double-jpeg compression is not in itself evidence of manipulation, as it could also be caused by incidental compression of an already compressed image. This is common in scenarios where for example, a photo is taken on a camera and saved as a JPEG, and then uploaded to a social media service where it is compressed again to decrease file size. This makes its universal usage not ideal for the contexts that this study is considering, as it would have a high false-positive rate. It would however be a useful tool in cases where a JPEG file is deceptively presented as having been unaltered from its original state, such as in visual journalism competitions, like World Press Photo, where the amount of post-production can be limited if allowed at all, or in the earlier example (see Figure 6) of the US National Parks Service cropping official photos of US President Donald Trump's 2017 inauguration. Fu, Shi, and Su (2007) demonstrate

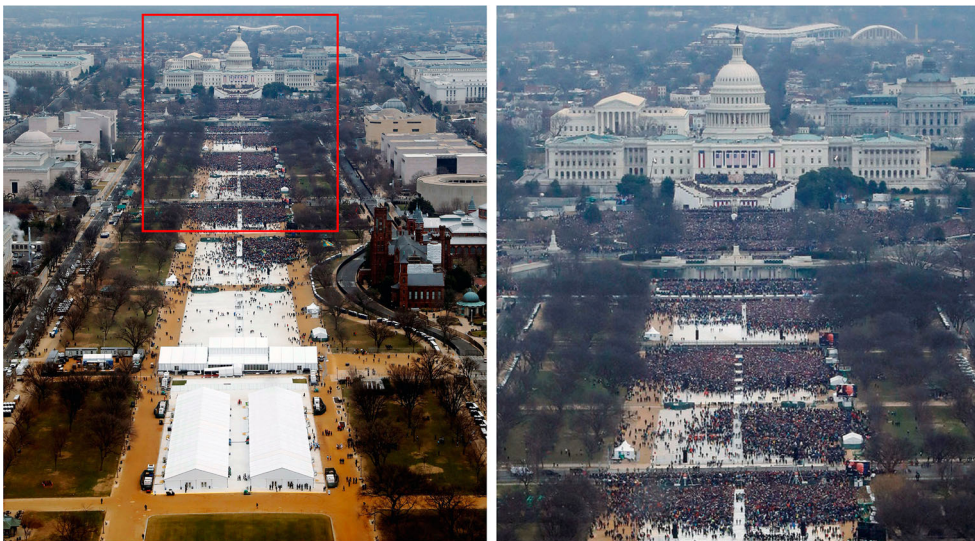


Figure 6. Trump inauguration comparison.

Note: A US National Parks Service photographer edited official pictures of Donald Trump's inauguration to make the crowd appear bigger following a personal intervention from the president, according to documents obtained through a Freedom of Information Act request lodged by The Guardian in 2018. The NPS photographer cropped out empty space "where the crowd ended" for a new set of pictures requested by Trump on the first morning of his presidency, after he was angered by images showing his audience was smaller than Barack Obama's in 2009. Such edits could have been revealed through an analysis of the JPEG file's compression details.

a way to determine whether an image has been JPEG compressed, and to estimate its compression quality (1–100), which can give clues about its provenance. Farid (2006, 2008) also found that the way that a JPEG's quantisation table (part of the compression process and reflected in the file's headers) is indicative of the make and model of camera used to create the image, which can be used for authentication.

- **CFA interpolation inconsistency.** In digital photography, images are created by capturing light in arrays of small cells known as photosites or photoreceptors. When a camera's shutter is open for a given exposure time, each cell captures and quantifies an amount of light which is then digitised. This value in each of the cells is combined and used to create an image. In order to include colours in images, varying filters are applied to each photosite that will block out unwanted light, often such that only one colour is quantified per cell (Malpas 2007). This arrangement of cells is known as the Colour Filter Array. Because only one colour is quantified per cell, neighbouring cells are used to create a pixel with multiple colour channels through estimation algorithms by sampling surrounding cells, a process known as interpolation. As noted by Popescu and Farid (2005), the various algorithms used for interpolation introduce statistical correlations within images that are disrupted or made inconsistent when an image is forged, for example by copy-move or splicing operations. They demonstrate that statistical inconsistencies caused by tampered regions can be detected and thus locally identified. Goljan and Fridrich (2015) similarly note that there is usually a stronger correlation between colour channels than merely among neighbouring pixels, which can be analysed to detect manipulations. Such statistical inconsistencies can also reveal artificial blurring of images, such as when the US National Archives in 2020 blurred images critical of US President Donald Trump (see Figure 7). The archives did not disclose the edit nor did it respond to requests for a list of other edits it made "so as not to engage in current political controversy", which was the rationale offered for the change.
- **Contrast and lighting inconsistency.** An image's contrast is the range of difference in brightness or colour between its elements. This is sometimes edited in extreme ways for political effect. A historic example is the darkening of OJ Simpson's mugshot on the 1994 cover of *TIME* magazine and a more contemporary example is a photo of US Donald Trump (see Figure 8) that was posted on Twitter and appears to have increased the saturation and/or contrast of the president's face, which accentuated the difference between tanned and untanned parts of his skin. The Tweet attracted widespread attention and even drew a mention by Trump himself who claimed it was Photoshopped.

Adjusting contrast is also often changed when attempting to splice two images with varying contrasts together in order to appear genuine (see Figure 4). Forgeries that don't change contrast can often be detected by the human eye and can also be detected using statistical tools. Lin, Li, and Hu (2013) demonstrate that adjusting an image's contrast is detectable by analysing the correlation between colour channels as discussed in CFA interpolation inconsistency. Stamm and Liu (2010) also demonstrate that contrast enhancement can be detected by analysing the "smoothness" of the frequency of change in colour across images (which appears more uniform in genuine photographs), and furthermore, that the degree of contrast enhancement can be derived through an iterative algorithm. Image light sources can be hard to manipulate in image forgery as they can be difficult to derive with the human eye.



Figure 7. Women’s March edits.

Note: The US National Archives sparked controversy in 2020 when it edited a news photo by Mario Tama for Getty (pictured) of the 2017 Women’s March by blurring out certain words, such as a reference to US President Donald Trump in a sign that read, “God Hates Trump”, among other edits.

Johnson and Farid (2005, 2007) show that when correctly derived with computer vision techniques, splicing operations can be detected.

- **Noise inconsistency.** Noise is a common and, depending on lighting conditions and the camera’s settings, reasonably invisible component of digital photographs that results from statistical variations in the camera sensor being translated into visual artefacts. As noise is often (but not always) uniform in genuine photographs, image forgery can be detected by looking for inconsistencies in noise across the image (see Figure 7). Pan, Zhang, and Lyu (2012) demonstrate this by estimating global noise variance across the image and thus determining if it has been manipulated. Chen et al. (2008) additionally show that an image’s origin can be verified by analysing the noise on the image and comparing it to the noise expected from a given sensor.

Deep Learning Methods

Deep learning for the identification of image manipulation is growing in popularity, largely due to the boom in the generalised use of machine learning and artificial intelligence. Deep learning refers to a subclass of neural network algorithms (deep neural network, or DNN) that can be trained to classify input information (in this case pictures) into specific predefined categories, such as photos of Donald Trump versus photos of Justin Trudeau, for example. As opposed to traditional forensic algorithms (ie, the “passive” methods detailed above), DNNs don’t require prior understanding of precisely how images are manipulated. Instead, they require a representative sample of manipulated and unmanipulated images (training data). These training data are fed iteratively



Figure 8. Trump skin comparison.

Note: A photo, at left, which depicted an extreme difference between the tanned and untanned portions of US President Donald Trump's face, attracted widespread attention and drew condemnation from Trump himself. Other photos taken the same day, such as the one at right, by Al Drago for Bloomberg, show a similar but less extreme difference. Passive forensic methods, such as evaluating the image for contrast/lighting inconsistency, could provide certainty to the claims that the image at left was Photoshopped.

through the network and a learning algorithm is used to modify a vast array of internal parameters to gradually maximize classification accuracy on these samples. Training finishes when the classification accuracy remains unchanged, known as convergence, indicating that the DNN is now “trained”. The assumption behind the “train-by-example” approach is that, if the training examples represent a comprehensive enough sample of images of a phenomenon, then an adequately resourced and configured DNN can create an abstract internal representation of what image qualities combine to indicate manipulation (Patterson and Gibson 2017). This abstracted internal model means that the trained DNN can then confidently identify manipulated images in a generalised way (Rao and Ni 2016). A trained model can therefore identify both the common artefacts that are left behind after image manipulation operations and underlying statistical anomalies.

The most common and accessible DNN architecture for image classification is the convolutional neural network (CNN). Outside digital forensics, CNNs are typically used to identify and classify discrete objects in a given image (such as in, for example, Facebook's facial recognition photo tagging functionality). In digital forensics, CNNs can similarly be used to identify the manipulated regions of an image (Zhou et al. 2018) (see Figure 9). However as Chen et al. (2015) found, generic CNN models are heavily affected by object edges and textures, and may not be suited to identify small image manipulation artefacts, such as airbrushing out a zit or enlarged pore in contrast to adding or removing an entire discrete object.

Due to their generalised approach, under certain circumstances deep-learning-based general image manipulation detection tools may demonstrate an advantage compared to specific image manipulation detection tools. For example, rather than a journalist having to perform multiple individual tests to identify whether the image had been cropped, selectively blurred, or made more or less saturated, a single deep learning

method could, in theory, identify all three of those edits as well as other ones that might not be apparent (such as retouching, resampling, splicing, or copy-moving). And in the contexts that this review considers, they can present a clear usability advantage in that there is usually a single (binary, scalar, or visual) output. Additionally, because a single tool can be used to investigate a variety of manipulations, it is simpler to account for false positives and negatives than when using an assortment of single-purpose tools. There are, however, clear disadvantages to this approach. From the user's perspective, the results can be visual and localised so the viewer can helpfully see which region or regions of the image are problematic; however, the underlying logic related to why certain regions are classified the way they are isn't transparent. The way that data are sourced to train deep-learning based classifiers is also problematic given the relatively small size of popular training datasets and the limited availability of datasets demonstrating the diverse range of sophisticated image manipulation techniques, as well as adversarial³ ones.

The adversarial issue is most pertinent though, as following the development of forensic methods that establish the trustworthiness of images, anti-forensic methods have also been developed expressly to thwart traditional forensic methods. These adversarial

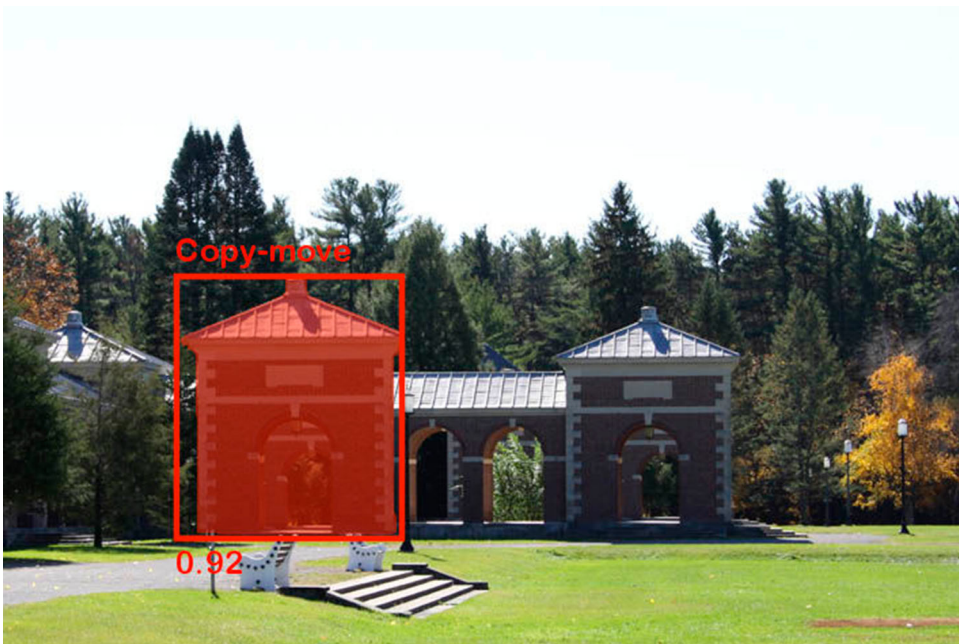


Figure 9. Mask R-CNN example.

Note: This image, taken from a standardised image library designed to train and test machine vision classifiers in the detection of image manipulation (Silva et al. 2015), demonstrates how a deep learning approach might work in practice. "Masking" (the area shaded in red) is a technique used in machine vision that seeks to separate (segment) different objects in an image or a video. Given an input image, a masking algorithm provides object bounding boxes, labels, and a mask. In the case of image manipulation, these masks could be used to highlight regions of suspected pixel manipulation, the form of manipulation believed to have taken place, and a confidence estimate. In this example, the algorithm has identified roughly 20 percent of the image as potentially having been forged, has labelled the type of potential forgery as a copy-move manipulation, and estimates with 92 percent confidence that the forgery has taken place.

methods present critical challenges when using image-manipulation tools practically. Anti-forensic methods rely on knowing how particular forensic methods work, either through reverse engineering or using available knowledge. It is then possible to derive and test adversarial examples which modify images such that they won't be flagged by classifiers.

Stamm et al. (2010) and Stamm and Liu (2011) demonstrate that methods detecting JPEG compression inconsistencies can be fooled by modifying the intrinsic statistical properties of images in ways that allow them to appear both visually and statistically ordinary. This can be done by adding precise dither (i.e., noise) to the underlying DCT coefficients, such that their distribution (as discussed in the JPEG compression inconsistency section) doesn't appear irregular or blocky. This highly effective technique severely undermines the efficacy of JPEG-compression based forgery identifiers and analysers.

Anti-forensic tools are also effective when used against deep-learning forgery detection methods. Szegedy et al. (2013) found that by making imperceptibly small adjustments to the pixel-contents of an image, deep neural network-based classification models can be misled. And as Nguyen, Yosinski, and Clune (2015) found, DNNs can even be fooled into misclassifying images that are visually meaningless, examples including images that appear to a human eye as noise, that are classified by a DNN as objects such as animals, fruit and household objects. While these techniques describe adversarial examples for traditional DNNs, the principals are similarly applied to forgery classifiers.

Even though developing and using anti-forensic techniques is non-trivial for a forger, the methods' considerable efficacy and versatility demonstrate the extreme importance for consideration to be given to them when developing new classification tools. Furthermore, in practical contexts, the fallibility and limitations of image forgery classifiers needs to be communicated to journalists in order to give them realistic expectations and mitigate so-called "automation bias". An ongoing commitment to the development of such tools is also essential for journalists and the public so that the tools remain relevant in the "digital arms race" of using computational approaches both to verify and frustrate verification efforts.

Implications for Journalistic Practice

Considering the wide variance of content that can fall on Wardle's (2018) "information disorder" spectrum (everything from satire and false connection to false context and fabricated content), the firehose of content journalists confront on a day-to-day basis, and the intense time pressures they have, especially on social media where breaking news unfolds, journalists face serious challenges when it comes to verifying visuals online. It's disheartening to recall past research that has found that only 11 percent of journalists globally use social media verification tools and to consider how much content of unknown provenance or authenticity they might be complicit in sharing, publishing, and amplifying.

Some tools, such as TinEye (a reverse image search engine), Foto Forensics, or Iztiru, exist, but they would have to be used in combination and journalists have little time to manually run an image through each tool, compare the results, and decide. Additionally,

reverse image search engines are not foolproof solutions, as major options like Google don't crawl the entirety of the web, such as images on 4chan or Reddit, rely on original and unedited versions of images being available, and can also be fooled by relatively simple edits, such as flipping the image's orientation from left to right. As such, without abdicating their verification duties entirely, journalists need to outsource some of the initial verification labour to machines that are more efficient with automated tasks to quickly run a battery of tests and return a decision—with rationale—to the journalist for final vetting and verification.

As an example: when US President Donald Trump posted on his official Facebook and Instagram accounts in 2018 and 2019 photos of himself that had been digitally edited to make him appear slimmer and to change the length of his fingers (Novak 2019), a statement like, "There is an inconsistency in the way that the smooth hue transition around the hands and waist, respectively, has been interpolated, which indicates a forgery in these regions of the image", which is clear and falsifiable, is more useful in a journalistic context than "Our model says that there's an 88 percent certainty that this region of the image is forged". Machines could process large amounts of data on journalists' behalf and raise red flags for journalists to further investigate and decide on. By using such a funnel approach, journalists are able to dedicate their precious time to the most suspect of cases and machines are able to save journalists the time and labour of winnowing from the field various low-risk images. These efforts don't have to be siloed and duplicative, however. The experimental and prototypical DeJaVu tool (Matatov et al. 2018), which isn't yet publicly available, proposes a database approach to allow journalists to flag problematic images and to be notified via a browser-based extension when such problematic images surface on their screens.

Vetting processes would need to happen at two critical junctures in the news gathering and dissemination process: first, **at the point of discovery**, such as when a journalist is trawling social media and considering amplifying a post by sharing it, and second, **at the point of distribution/re-distribution on the associated news organisation's website**, such as when a journalist embeds a Tweet into a news article or when a journalist uploads media into the organisation's content management system (CMS). This would provide verification opportunities both for user-generated content on social media as well as content that freelancers or staffers produce as part of their official duties.

Regarding the point of discovery, room for additional public-private partnerships exists. Some of the biggest tech companies are the ones best positioned to tackle these issues as they have the largest training datasets available and the technical infrastructure and engineering teams to handle a tool, ideally co-designed among developers, journalists, and members of the public, that could be integrated directly into social media platforms that would allow a user to, for example, click a button next to a post and have a report generated in return that provides an evaluation of the content's provenance and veracity. However, it is important to note that platform companies that integrate news into their offerings, such as Google, Facebook, Yahoo, and Twitter, are disproportionately benefiting from the advertising revenue that the more engaging mis/disinformation brings to their companies (Bakir and McStay 2018) and so would have to balance the potential revenue loss against the fear of tighter government regulation or fines for allowing mis/disinformation to flourish on their platforms.

Discussion and Conclusion

Countries with advanced and developing economies alike are concerned with the effects of mis/disinformation on democratic institutions, political discourse, trust, and social harmony (Levush 2019). Digital platforms are primary news avenues for many in connected nations; however, these same citizens are also increasingly concerned about what is real or fake on the internet (Park et al. 2020). The scope of the problem is underscored in a recent study which tracked 96 separate foreign influence disinformation campaigns, targeting 30 countries between 2013 and 2019, that sought to defame public figures, persuade the public or polarise debates (Martin, Shapiro, and Ilhardt 2020). Indeed, the global media ecosystem is rife with mis/disinformation produced by political actors and ordinary citizens through user-generated content, which is sometimes also shared and amplified by individual journalists and mainstream media organisations alike. Journalists therefore face the jointly Herculean and Sisyphean tasks of using digital platforms to share their original reporting in an ecosystem that sometimes promotes mis/disinformation ahead of accurate and truthful accounts (Bechmann 2018), creating the seemingly insurmountable dual task of providing truthful and impactful content while also fact checking and verifying content in service of the public.

Fact-checking, source-checking, verification, and debunking have long been journalistic practices; however, when faced with the firehose of user-generated content online, these seem to fall by the wayside more than they should, especially when it comes to visuals. Journalists, media outlets, and law enforcement often cannot establish the veracity of an image by simply probing its source, as journalists and news organisations are often the ones re-Tweeting, embedding, or otherwise amplifying user-generated content into their news reporting without a full understanding of how accurate that content is. This needn't be the case, though. As the myriad examples presented earlier attest, journalists have a responsibility for the vision they embed into their news coverage and amplify on social media platforms, especially during crises and times when sharing a visual or amplifying an image could result in the potential for harm. Likewise, they also have a responsibility to increase their media literacy and technical acumen to ensure they can perform their verification and debunking mission with digital tools.

As has been previously argued, manual detection methods aren't enough, as not all fraudulent edits are visible to naked vision; however, journalists can't outsource their critical thinking skills entirely to computers, either. Even if bespoke solutions are integrated into production and presentation workflows, perhaps through embedding as a CMS or browser-based plugin, and every bit of content featured is scanned, journalists still have to rely on their intuition, news judgment, and willingness to question the established narrative to identify potentially problematic content and perform their due diligence to evaluate whether they're sharing or amplifying visual mis- and disinformation, especially when the threshold for harm is high.

Traditional statistical methods present clear and highly accurate techniques for detecting common types of forgery, and are sufficiently expressive in their outputs as to be used authoritatively. Emerging methods that use machine learning are similarly accurate and useful but require an ongoing investment by news organisations and other actors in the public trust to ensure that they remain accurate and useful even as fraudulent media continues to become more and more sophisticated. The need for such

development and ongoing investment in tools for combatting visual mis/disinformation is critical as the ease of use, affordability, and ubiquity of tools for generating visual disinformation have skyrocketed in the past three years. Since surfacing to public consciousness in 2017 (Westerlund 2019), deepfakes, for example, have progressed from requiring a large set of training data (in the range of thousands of images), significant technical skill and time, and processing power, to now being possible—with various levels of sophistication—almost instantly through a humble smartphone and a free app like Zao or ReFace that can create a type of deep fake with only a single photo. These apps enjoy widespread popularity, with ReFace claiming the top spot in the Apple App Store’s “entertainment” category and are increasingly being embedded in ordinary peoples’ cultural practices for entertainment as well as, potentially, for more nefarious purposes. Recalling again that the purpose of disinformation isn’t necessarily to fool the viewer but can also be to simply confuse, distract, or sow distrust. That these consumer apps can be used to create media that can be instantly and anonymously shared across borders and, importantly, across legal systems and potentially beyond the repercussions of them, underscores the pressing task faced by journalists and others with an interest in the veracity of their visual information.

Similarly, the datasets used for machine learning need to become more representative in order to become more useful in journalistic and public communications contexts. For example, much of the data in machine learning training sets are themselves computer-generated or drawn from only a very narrow slice of publicly available imagery. Diversifying these data sets with vision from news organisations, for example, from Reuters, AFP, or AP, would go a long way to ensuring emerging methods are trained with actual news images and are able to detect fakes in them.

The accuracy of these visual disinformation detection tools is high, but their fallibility needs to be considered when designing the systems that they are implemented in. There is limited benefit to using all-in-one style classifiers on their own, as using a combination of tools that can be keenly examined provides journalists with the context and certainty to correctly assess the integrity of images with unknown provenance. Provided that journalists are given enough information about the fundamentals of the way that these tools work and the reasons for their limited fallibility, the acuity of image forgery detection tools presents a compelling argument for their usage.

Notes

1. Manipulations involving visual media, including by journalists or photo editors themselves, have a long history and include activity being staged or stylised in the field as well as pixels being altered in post-production through analog or, more recently, digital means. Notable examples include Roger Fenton’s 1855 “Valley of the Shadow of Death” photograph, *National Geographic’s* 1982 cover which featured a horizontal photo that was altered to fit a vertical magazine cover, and former AP photographer Narciso Contreras’s photo in which he airbrushed out in 2014 a fellow journalist’s video camera. All these examples are relevant to note but, with this specific paper, we choose to focus on manipulations created, edited, or circulated initially by non-journalists on social media.
2. We operationalize in this paper *misinformation* as inaccurate information that is most likely shared without the intent to harm and *disinformation* as inaccurate information shared with the intent to deceive.

3. Adversarial training reverses the logic of classification by attempting to “fool” a system through targeted manipulations of input data to try to exploit weaknesses in the classification logic. Under adversarial conditions the system designer tries to generate (often using a generative neural network) examples of images that are misclassified, and these images themselves can be used to assist in the generation of further images that may perform even worse.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Bushfire and Natural Hazards Cooperative Research Centre.

ORCID

T.J. Thomson  <http://orcid.org/0000-0003-3913-3030>

Daniel Angus  <http://orcid.org/0000-0002-1412-5096>

Paula Dootson  <http://orcid.org/0000-0002-8020-8762>

Edward Hurcombe  <http://orcid.org/0000-0002-5838-2019>

References

- Associated Press. 2020. Standards for Visuals. <https://www.ap.org/about/news-values-and-principles/telling-the-story/visuals>.
- Baker, H. 2019. Introducing the Reuters Guide to Manipulated Media, in Association with the Facebook Journalism Project [Press release]. <https://www.reuters.com/article/rpb-hazeldeepfakesblog/introducing-the-reuters-guide-to-manipulated-media-in-association-with-the-facebook-journalism-project-idUSKBN1YY14C>.
- Bakir, V., and A. McStay. 2018. “Fake News and the Economy of Emotions: Problems, Causes, Solutions.” *Digital Journalism* 6 (2): 154–175.
- Bechmann, A. 2018. The Epistemology of the Facebook News Feed as a News Source. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3222234>.
- Birajdar, G. K., and V. H. Mankar. 2013. “Digital Image Forgery Detection Using Passive Techniques: A Survey.” *Digital Investigation* 10 (3): 226–245.
- Brandtzaeg, P. B., M. Lüders, J. Spangenberg, L. Rath-Wiggins, and A. Følstad. 2016. “Emerging Journalistic Verification Practices Concerning Social Media.” *Journalism Practice* 10 (3): 323–342.
- Caplan, R., and D. boyd. 2018. “Isomorphism Through Algorithms: Institutional Dependencies in the Case of Facebook.” *Big Data & Society* 5 (1): 1–12.
- Chen, M., J. Fridrich, M. Goljan, and J. Lukás. 2008. “Determining Image Origin and Integrity Using Sensor Noise.” *IEEE Transactions on Information Forensics and Security* 3 (1): 74–90.
- Chen, J., X. Kang, Y. Liu, and Z. J. Wang. 2015. “Median Filtering Forensics Based on Convolutional Neural Networks.” *IEEE Signal Processing Letters* 22 (11): 1849–1853.
- Fan, Z., and R. L. De Queiroz. 2003. “Identification of Bitmap Compression History: JPEG Detection and Quantizer Estimation.” *IEEE Transactions on Image Processing* 12 (2): 230–235.
- Farid, H. 2006. Digital Image Ballistics from JPEG Quantization. *Dept. Comput. Sci., Dartmouth College, Tech. Rep. TR2006-583*.
- Farid, H. 2008. Digital Image Ballistics from JPEG Quantization: A Followup Study. *Department of Computer Science, Dartmouth College, Tech. Rep. TR2008-638, 7, 1–28*.
- Farid, H. 2009. “Image Forgery Detection.” *IEEE Signal Processing Magazine* 26 (2): 16–25.
- Farid, H. 2018. “Digital Forensics in a Post-Truth age.” *Forensic Science International* 289: 268–269.
- Franklin, B. 2014. “The Future of Journalism.” *Journalism Studies* 15 (5): 481–499.

- Fu, D., Y. Q. Shi, and W. Su. 2007. "A Generalized Benford's Law for JPEG Coefficients and its Applications in Image Forensics." *Proc. SPIE 6505, Security, Steganography, and Watermarking of Multimedia Contents IX*, 65051L (27 February 2007). doi:10.1117/12.704723.
- Goljan, M., and J. Fridrich. 2015. "CFA-Aware Features for Steganalysis of Color Images." *Proceedings of the SPIE: Media Watermarking, Security, and Forensics*, vol. 9409, pp.94090V. <http://ws2.binghamton.edu/fridrich/Research/color-spie-2015-8.pdf>.
- Gottfried, J., and E. Shearer. 2016. "News Use Across Social Media Platforms 2016." *Pew Research Centre*, 26 May. Accessed 21 June 2019. <https://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/>.
- Gross, L., J. Katz, and J. Ruby. 2017. *Image Ethics in the Digital Age*, 1–370. Minneapolis: University of Minnesota Press.
- International Centre for Journalists. 2017. *The State of Technology in Global Newsrooms*. Report.
- Ireton, C., and J. Posetti, Eds. 2018. *Journalism, 'Fake News' & Disinformation: United Nations Educational, Scientific and Cultural Organization*.
- Joffe, H. 2008. "The Power of Visual Material: Persuasion, Emotion and Identification." *Diogenes* 55 (1): 84–93.
- Johnson, M. K., and H. Farid. 2005. "Exposing Digital Forgeries by Detecting Inconsistencies in Lighting." In *Proceedings of the 7th Workshop on Multimedia and Security*, 1–10. ACM.
- Johnson, M. K., and H. Farid. 2007. "Exposing Digital Forgeries in Complex Lighting Environments." *IEEE Transactions on Information Forensics and Security* 2 (3): 450–461.
- Koettl, C. 2018. "Satellite Images and Shadow Analysis: How The Times Verifies Eyewitness Videos." *The New York Times*, 4 September. Accessed 21 July 2019. <https://www.nytimes.com/2018/09/04/reader-center/social-media-video-how-to-verify.html>.
- Koren, S. 2019. "Introducing the News Provenance Project." *NYT Open*. <https://open.nytimes.com/introducing-the-news-provenance-project-723dbaf07c44>.
- Lester, P. M. 2016. *Photojournalism: An Ethical Approach*. Abingdon, UK: Routledge.
- Levush, R. 2019. *Government Responses to Disinformation on Social Media Platforms: Comparative Summary*. Library of Congress. <https://www.loc.gov/law/help/social-media-disinformation/compsum.php>.
- Lin, X., C. T. Li, and Y. Hu. 2013. "Exposing Image Forgery Through the Detection of Contrast Enhancement." In *2013 IEEE International Conference on Image Processing*, 4467–4471. IEEE.
- Malpas, P. 2007. *Basics Photography 03: Capturing Colour*. London, UK: Bloomsbury Publishing.
- Marqués, F., M. Menezes, and J. Ruiz-Hidalgo. 2009. "How are Digital Images Compressed in the Web?" In *Applied Signal Processing*, edited by Thierry Dutoit and Ferran Marqués, 265–310. Boston, MA: Springer.
- Marshall, A. M., and R. Paige. 2018. "Requirements in Digital Forensics Method Definition: Observations from a UK Study." *Digital Investigation* 27: 23–29.
- Martin, D., J. Shapiro, and J. Ilhardt. 2020. "Trends in Online Influence Efforts." *Empirical Studies of Conflict Project*. https://drive.google.com/file/d/18QIENHZsINloKvOu72iEjG6RgWL1Dww/_view.
- Matatov, H., A. Bechhofer, L. Aroyo, O. Amir, and M. Naaman. 2018. "DejaVu: A System for Journalists to Collaboratively Address Visual Misinformation." In *Computation+ Journalism Symposium*. Miami, FL. https://scholar.harvard.edu/files/oamir/files/dejavu-system-journalists_3.pdf.
- McCabe, D., and D. Alba. 2020. "Facebook Says It Will Ban 'Deepfakes'." *The New York Times*. <https://www.nytimes.com/2020/01/07/technology/facebook-says-it-will-ban-deepfakes.html>.
- MEAA. 2020. *MEAA Journalist Code of Ethics*. <https://www.meaa.org/meaa-media/code-of-ethics/>.
- Middleton, S. 2017. "Journalist Decision Support System (JDSS)." *Reveal: Social Media Verification*. Accessed 19 July 2019. <https://revealproject.eu/journalist-decision-support-system-jdss/>.
- National Press Photographers Association. 2020. "Code of Ethics." *NPPA*. <https://nppa.org/code-ethics>.
- Nguyen, A., J. Yosinski, and J. Clune. 2015. "Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 427–436.

- Novak, M. 2019. *President Trump Posts Altered Photos To Facebook And Instagram That Make Him Look Thinner*. Gizmodo. <https://www.gizmodo.com.au/2019/01/president-trump-posts-altered-photos-to-facebook-and-instagram-that-make-himlook-thinner/>
- Pan, X., X. Zhang, and S. Lyu. 2012. "Exposing Image Splicing with Inconsistent Local Noise Variances." In *2012 IEEE International Conference on Computational Photography (ICCP)*, 1–10. IEEE.
- Pantti, M., and S. Sirén. 2015. "The Fragility of Photo-Truth." *Digital Journalism* 3 (4): 495–512.
- Park, S., C. Fisher, J. Y. Lee, K. McGuinness, Y. Sang, M. O'Neil, M. Jensen, K. McCallum, and G. Fuller. 2020. "Digital News Report: Australia 2020." University of Canberra. <https://apo.org.au/sites/default/files/resource-files/2020-06/apo-nid305057.pdf>.
- Patterson, J., and A. Gibson. 2017. *Deep Learning: A Practitioner's Approach*. Sebastopol, CA: O'Reilly Media, Inc.
- Popescu, A. C., and H. Farid. 2004. "Statistical Tools for Digital Forensics." In *International Workshop on Information Hiding*, 128–147. Berlin: Springer.
- Popescu, A. C., and H. Farid. 2005. "Exposing Digital Forgeries in Color Filter Array Interpolated Images." *IEEE Transactions on Signal Processing* 53 (10): 3948–3959.
- Qureshi, M. A., and M. Deriche. 2015. "A Bibliography of Pixel-Based Blind Image Forgery Detection Techniques." *Signal Processing: Image Communication* 39: 46–74.
- Rao, Y., and J. Ni. 2016. "A Deep Learning Approach to Detection of Splicing and Copy-Move Forgeries in Images." In *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, 1–6. IEEE.
- Rose, J. 2017. "Brexit, Trump, and Post-Truth Politics." *Public Integrity* 19 (6): 555–558.
- Shen, C., M. Kasra, W. Pan, G. A. Bassett, Y. Malloch, and J. F. O'Brien. 2019. "Fake Images: The Effects of Source, Intermediary, and Digital Media Literacy on Contextual Assessment of Image Credibility Online." *New Media & Society* 21 (2): 438–463.
- Shih, F. Y. 2017. *Digital Watermarking and Steganography: Fundamentals and Techniques*. Boca Raton, FL: CRC Press.
- Silva, E., T. Carvalho, A. Ferreira, and A. Rocha. 2015. "Going Deeper into Copy-Move Forgery Detection: Exploring Image Telltales via Multi-Scale Analysis and Voting Processes." *Journal of Visual Communication and Image Representation* 29: 16–32.
- Silva, M. F. S., and S. A. Eldridge II. 2020. *The Ethics of Photojournalism in the Digital Age*. London: Routledge.
- Silverman, C. Ed. 2014. *Verification Handbook: An Ultimate Guideline on Digital Age Sourcing for Emergency Coverage*. European Journalism Centre.
- Smith, B. 2018. *Fake News, Hoax Images: How to Spot a Digitally Altered Photo from the Real Deal*. ABC News. <https://www.abc.net.au/news/science/2018-02-11/fake-news-hoax-images-digitally-altered-photos-photoshop/9405776>
- Songcharoen, S. J., U. Bite, and R. P. Clay. 2014. "Caveat Spectator: Digital Imaging and Data Manipulation." In *Mayo Clinic Proceedings*, Vol. 89, No. 8, pp. 1036–1041. Elsevier.
- Stamm, M. C., and K. R. Liu. 2010. "Forensic Estimation and Reconstruction of a Contrast Enhancement Mapping." In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1698–1701. IEEE.
- Stamm, M. C., and K. R. Liu. 2011. "Anti-forensics of Digital Image Compression." *IEEE Transactions on Information Forensics and Security* 6 (3): 1050–1065.
- Stamm, M. C., S. K. Tjoa, W. S. Lin, and K. R. Liu. 2010. "Anti-forensics of JPEG compression." In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1694–1697. IEEE.
- Szegedy, C., W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. 2013. Intriguing Properties of Neural Networks. *arXiv preprint arXiv:1312.6199*.
- Tameez, H. 2020. "Here's How The New York Times Tested Blockchain to Help You Identify Faked Photos on Your Timeline." *NiemanLab*. <https://www.niemanlab.org/2020/01/heres-how-the-new-york-times-tested-blockchain-to-help-you-identify-faked-photos-on-your-timeline/>.
- Tardáguila, C. 2020. "How to Use Your Phone to Spot Fake Images Surrounding the U.S.-Iran Conflict." *Poynter*. <https://www.poynter.org/fact-checking/2020/how-to-use-your-phone-to-spot-fake-images-surrounding-the-u-s-iran-conflict/?fbclid=IwAR3ADZi2k4Vb13Fy-M-HKLSi40luv4b04z1IKKs8z1EeDrCcDn2SGzkyg>.

- Thomson, T. J. 2018. "Freelance Photojournalists and Photo Editors." *Journalism Studies* 19 (6): 803–823. doi:10.1080/1461670X.2016.1215851.
- Tolmie, P., R. Procter, D. W. Randall, M. Rouncefield, C. Burger, G. Wong Sak Hoi, A. Zubiaga, and M. Liakata. 2017. "Supporting the Use of User-Generated Content in Journalistic Practice." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3632–3644.
- Wardle, C. 2018. "The Need for Smarter Definitions and Practical, Timely Empirical Research on Information Disorder." *Digital Journalism* 6 (8): 951–963.
- Westerlund, M. 2019. "The Emergence of Deepfake Technology: A Review." *Technology Innovation Management Review* 9 (11). <https://timreview.ca/article/1282>.
- Zampoglou, M., S. Papadopoulos, Y. Kompatsiaris, R. Bouwmester, and J. Spangenberg. 2016. "Web and Social Media Image Forensics for News Professionals." *Social Media in the Newsroom: Technical Report WS-16-19*.
- Zhang, M. 2018. "Trump Had Inauguration Crowd Photos Edited, Report Claims." *PetaPixel*. <https://petapixel.com/2018/09/07/trump-had-inauguration-crowd-photos-edited-report-claims/>.
- Zhou, P., X. Han, V. I. Morariu, and L. S. Davis. 2018. "Learning Rich Features for Image Manipulation Detection." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1053–1061.
- Zhou, G., and D. Lv. 2011. "An Overview of Digital Watermarking in Image Forensics." In *2011 Fourth International Joint Conference on Computational Sciences and Optimization*, 332–335. IEEE.