applied sciences

Article

A Survey of Detection and Mitigation for Fake Images on Social Media Platforms

Dilip Kumar Sharma [1], Bhuvanesh Singh [2], Saurabh Agarwal [3,4,*], Lalit Garg [5], Cheonshik Kim [6] and Ki-Hyun Jung [4,*]

[1] Department of Computer Engineering and Application, GLA University, Mathura 281406, India; dilip.sharma@gla.ac.in

[2] Graduate Software Programs, University of St. Thomas, St. Paul, MN 55105, USA; bhuvaneshsingh80@gmail.com

[3] Department of Computer Science and Engineering, Amity School of Engineering Technology, Amity University Uttar Pradesh, Noida 201313, India

[4] Department of Software Convergence, Andong National University, Andong-si 36729, Republic of Korea

[5] Computer Information Systems, Faculty of Information & Communication Technology, University of Malta, 2080 Msida, Malta; lalit.garg@um.edu.mt

[6] Department of Computer Engineering, Sejong University, Seoul 05006, Republic of Korea; mipsan@sejong.ac.kr

* Correspondence: saurabhnsit2510@gmail.com (S.A.); khanny.jung@gmail.com or kingjung@anu.ac.kr (K.-H.J.); Tel.: +82-54-820-7968 (K.-H.J.); Fax: +82-54-820-6257 (K.-H.J.)

Abstract: Recently, the spread of fake images on social media platforms has become a significant concern for individuals, organizations, and governments. These images are often created using sophisticated techniques to spread misinformation, influence public opinion, and threaten national security. This paper begins by defining fake images and their potential impact on society, including the spread of misinformation and the erosion of trust in digital media. This paper also examines the different types of fake images and their challenges for detection. We then review the recent approaches proposed for detecting fake images, including digital forensics, machine learning, and deep learning. These approaches are evaluated in terms of their strengths and limitations, highlighting the need for further research. This paper also highlights the need for multimodal approaches that combine multiple sources of information, such as text, images, and videos. Furthermore, we present an overview of existing datasets, evaluation metrics, and benchmarking tools for fake image detection. This paper concludes by discussing future directions for fake image detection research, such as developing more robust and explainable methods, cross-modal fake detection, and the integration of social context. It also emphasizes the need for interdisciplinary research that combines computer science, digital forensics, and cognitive psychology experts to tackle the complex problem of fake images. This survey paper will be a valuable resource for researchers and practitioners working on fake image detection on social media platforms.

Keywords: deep learning; digital image forensic; fake images; generated adversarial networks; multi-modal; image forgery detection

check for updates

Citation: Sharma, D.K.; Singh, B.; Agarwal, S.; Garg, L.; Kim, C.; Jung, K.-H. A Survey of Detection and Mitigation for Fake Images on Social Media Platforms. *Appl. Sci.* **2023**, *13*, 10980. https://doi.org/10.3390/app131910980

Academic Editor: Gianluca Lax

Received: 25 July 2023
Revised: 11 September 2023
Accepted: 12 September 2023
Published: 5 October 2023

1. Introduction

Social networks, which include microblogging platforms like Facebook, Twitter, Instagram, or Weibo, concerning around 3.8 billion people worldwide, have hugely elevated information exchange and subsequently led to the rapid dispersion of public sentiment. Fake content over these platforms has been used to spread malicious intent and sway public opinion to their benefit. Figure 1 illustrates such an example. Fake images have become a social menace now as, at times, their impact is grave. The PRCC US survey [1] shows that around 64% of people need clarification due to false information. Facebook and Twitter are the two sites that distribute false news the fastest, according to a similar study by CIGI-IPSOS and the Internet Society [2]. Global IT companies such as Facebook and Google are creating AI solutions to combat the threat posed by the proliferation of fraudulent images and videos online. According to Buzzfeed Analysis [3], Facebook had more user engagement over fake news than mainstream news on August election day in the US in 2016. After fake news caused mob lynching in India, WhatsApp had to consider the automatic identification of fraudulent photographs and videos on their platform [4]. The hoax image of President Donald Trump endorsing Prime Minster Modi went viral in India (Figure 1).

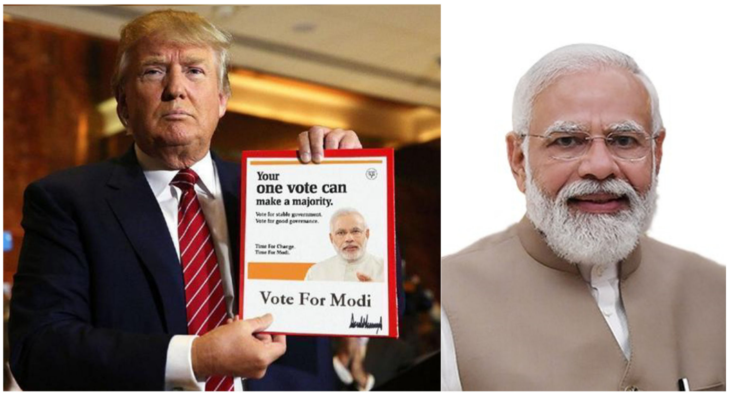

Figure 1. Trump endorsed Modi during the election in India [5].

Similarly, another morphed image, Figure 2, displays Prime Minster Modi bowing down to China's President Xi Jinping. These examples show that fake images have become a powerful medium in the political arena. Coronavirus sufferers started out refusing medication in the United Kingdom due to the spread of false information on social media [6]. According to a poll in Norway 2020 [7], social media sites were the primary source of false information about the coronavirus there. The influence of social media on different media is compared in Figure 3. Therefore, early fraudulent picture identification on social networking sites is essential for effectively avoiding risks and harm.

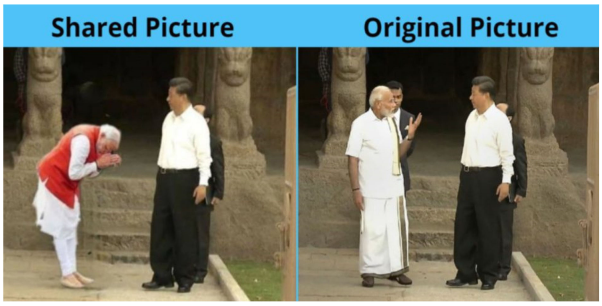

Figure 2. Modi greeting China's President Xi Jinping [8].

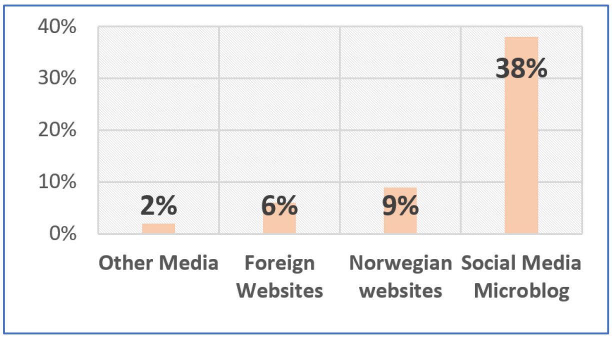

Figure 3. Social media is the highest contributor to fake news [9].

1.1. Motivation

False news was mainly used to propagate rumors, satire, or fun. With time, politicians started using it to sway public sentiments. As stated by the Wall Street Journal, of all color photos published in the United States, 10% of them had been, without a doubt, altered or retouched [9]. In 2002, the photograph of then-president George Bush reading a children's book upturned was shared like wildfire Figure 4. With the arrival of GAN technology, forged images can be cloned/altered very closely to authentic images.

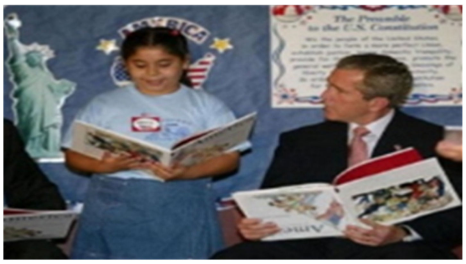

Figure 4. President George Bush reading a book, upturned [10].

The harmful impacts became apparent when they led to grave consequences, including mob lynching, religious disputes, and providing patients with the incorrect treatment counsel. Since the beginning, deepfakes have caused security concerns. Deepfakes have a significant adverse effect known as character assassination. It is noticed that images and videos are spread more to gather additional attention than text. Fake images over fake news must be detected in time, and their dissemination must be mitigated. This paper does a comprehensive survey of digital image tampering detection techniques. The survey considers classical image forgery detection techniques based on forensic features to modern deep learning multi-modal techniques. It also shares this field's current challenges and limitations for further research. Deep learning methods in detecting forged images can be the most efficient solution to this problem.

1.2. Related Study

Fake news detection has been an essential research topic, and many techniques have been employed. However, most of the methods are based on text content. The classification is based on text, sentiment, or user profile analysis. When fake news started creating a nuisance through fake images and videos, a constant effort was made to detect counterfeit images on social media platforms. Multiple types of research proposing diverse solutions were discussed. There has been a continuous effort to review those various image detection techniques from time to time, and a study has been conducted to compare and further guide more research toward fake image detection. Mishra and Adhikary [11] studied various passive techniques. Still, those were more specific to forensic techniques, while later, GAN and deep learning picked up. Later, Mandankandy [12] performed a comparative study of different techniques based on image tampering methods. It also discussed various new techniques and the usage of classifiers. In a media-rich fake news detection, Parikh and Atrey [13] discuss both techniques through text and visual and share information over specific datasets. However, it needed to compare which method is better for text or visual. Tolosana et al. [14] have targeted only deepfakes. It discussed different deepfake tampering types and their detection methods in detail. It also compared various deepfake detection techniques. This survey paper exhausts studies of multiple conventional to modern neural network-based techniques and provides a comparison. It also discusses the issues within each technique.

1.3. Contribution and Organization

This survey research paper contributes significantly to the field of fake image detection, offering valuable insights and uniqueness compared to other surveys. This paper focuses on detecting fake images shared over social media platforms, a crucial aspect of identifying fake news on digital platforms. Other surveys have focused on general image forgery, not on the most impacted area of image forgery—social media platforms. It comprehensively reviews various techniques, from traditional forensics to cutting-edge deep learning approaches, making it distinct and relevant for further research.

The unique contributions of this paper include:

- Comprehensive Coverage: This paper thoroughly examines the fake image detection process, leaving no stone unturned. It explores image tampering techniques, including Generative Adversarial Networks (GANs). It covers various detection methods, encompassing handcrafted forensic features, semantic features, statistical features, web retrievals, neural networks, and multi-modal approaches;
- Performance Comparison: It goes beyond describing these methods by summarizing and comparing their results within each detection category. This performance evaluation aids researchers in selecting the most suitable approach for their specific needs;
- Deep Learning Emphasis: This paper underscores the superiority of deep learning methods for detecting fake images over social media platforms, backed by evidence and comparative analysis. This emphasis provides clear guidance to researchers and practitioners;
- Challenges and Future Scope: It does not shy away from highlighting the current challenges and limitations in the field, shedding light on areas where further research is needed. This forward-looking perspective enhances its value for the research community;
- Dataset and Evaluation Parameters: This paper also provides valuable information on fake image datasets and evaluation parameters, facilitating the replication of experiments and benchmarking new detection methods.

This survey paper is a comprehensive and up-to-date resource for researchers and practitioners interested in fake image detection. Its emphasis on deep learning, comprehensive coverage, performance comparison, and forward-looking perspective make it a unique and valuable contribution to the field, guiding future research efforts and advancements in this critical area. The rest of this paper is organized as follows: Section 2 describes fake image detection processes and their methods; Section 3 provides a brief description of various image tampering techniques; Section 4 details the research work performed to detect fake images using a handcrafted feature set. Their comparisons and issues are also discussed; Section 5 presents detection methods using neural networks. A comparison among multiple methods is also presented; Section 6 targets evaluation parameters and datasets, respectively; Section 7 briefs the challenges and limitations of current work and guides toward future work. The conclusion is provided in Section 8.

2. Fake Image Detection Process

Detection of fake news over digital media has long been challenging. Multiple research works using different techniques are used to detect fake news. Figure 5 illustrates the various methods employed in fake news detection. It can be detected using social context-based, content-based, and user profile-based strategies. Text-based linguistic, image-based, and text style-based approaches are used in content-based. It can be observed from the taxonomy of Figure 5 that even in image-based detection, there are multiple methods. Therefore, this survey's scope is limited to image-based detection methods. Digital signatures and digital watermarking fall under active methods. However, these active methods are not feasible, with many images added over the internet. Passive methods can be categorized into two broad categories. One requires a handcrafted feature set of images, while others are based on neural networks that learn the feature set.

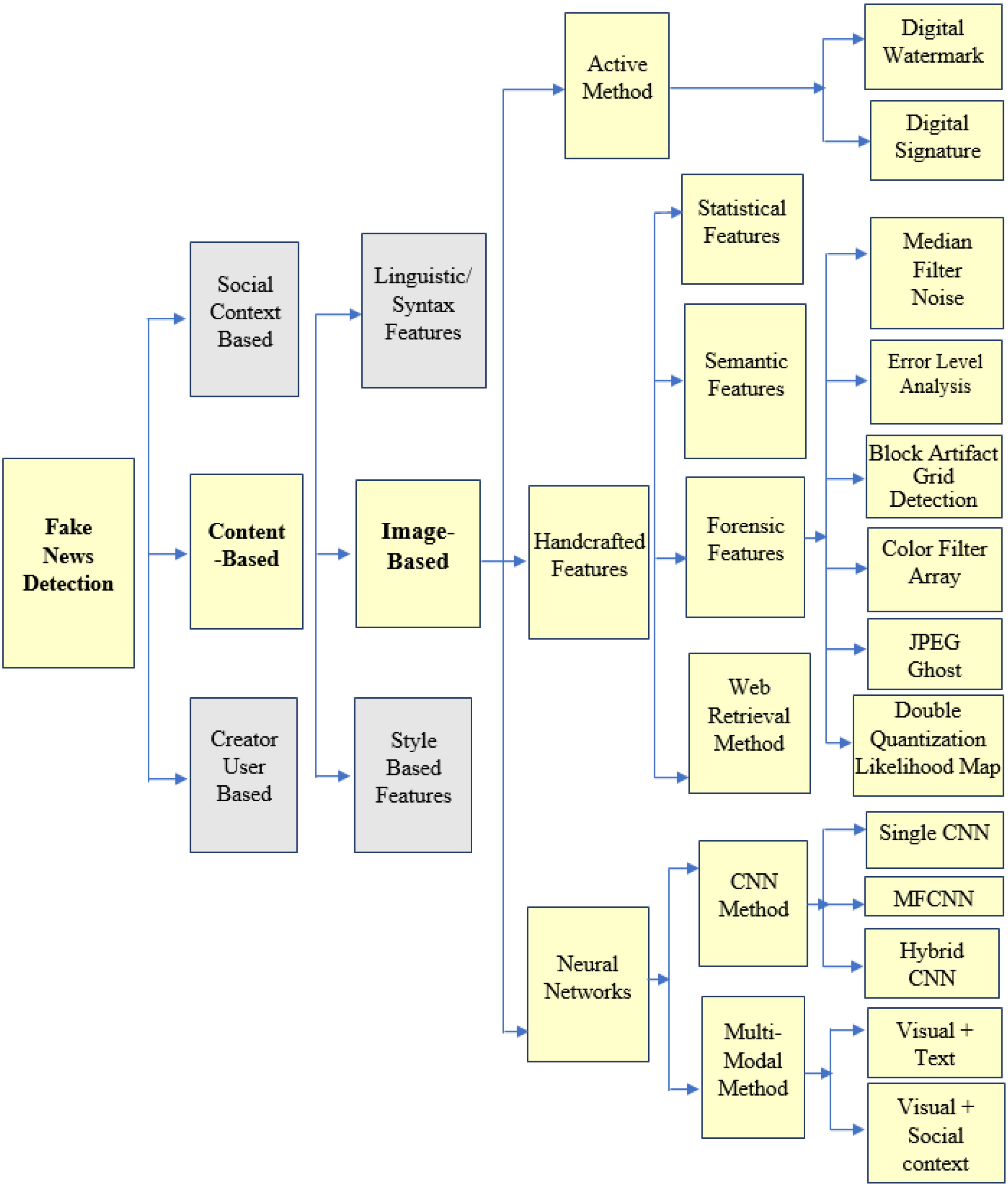

Figure 5. Fake News Detection Taxonomy—Image-Based approaches.

Conventional domain-specific image forensic techniques are used in a handcrafted feature set approach. These forensic techniques are now combined with machine learning for better optimization. Other methods can be used for the semantic and statistical features of the image. Web-Retrieval is another popular method for searching and identifying tampered images. On the other hand, in the neural network-based approach, convolutional neural networks (CNN) are noticed to be very useful in learning the intrinsic features of the manipulated image. Much research was conducted through CNN. The multi-modal approach is currently being applied, i.e., combining images with text, images with social context, etc. These multi-modal approaches also use the same neural networks as CNN. The paper by Wang et al. [15] discusses various techniques.

Fake image detection is a classification problem. The final output is identifying whether the image is fake or not. The process starts with gathering various types of tampered images manipulated using single or multiple alterations and then, after processing, classifying them as real or false. The fake image detection process at a high level comprises handcrafted feature sets and self-learning neural networks. Figure 6a exemplifies a fake image detection workflow using handcrafted features. Initially, a set of tampered images is collected. Then, each image may undergo pre-processing activity, like gray scaling and cropping. In the feature extraction phase, various image features are extracted relating to the image. These features can be device-specific, image-intrinsic, or semantic/statistics characteristics of an image. Forensic methods use handcrafted intrinsic features of images, while other methods use other characteristics. Feature preprocessing may or may not be applied to reduce features to achieve computational efficiency. Figure 6b illustrates the process used by neural networks, which learn the fake image's hidden features. Ultimately, both processes have a classifier applied to mark them as real or fake based on the learnings. Sometimes, based on the detection method capability, image post-processing is also performed. In post-processing, the tampered regions in an image are identified. Forensic feature techniques are proficient in localizing the manipulated areas in a fake image.

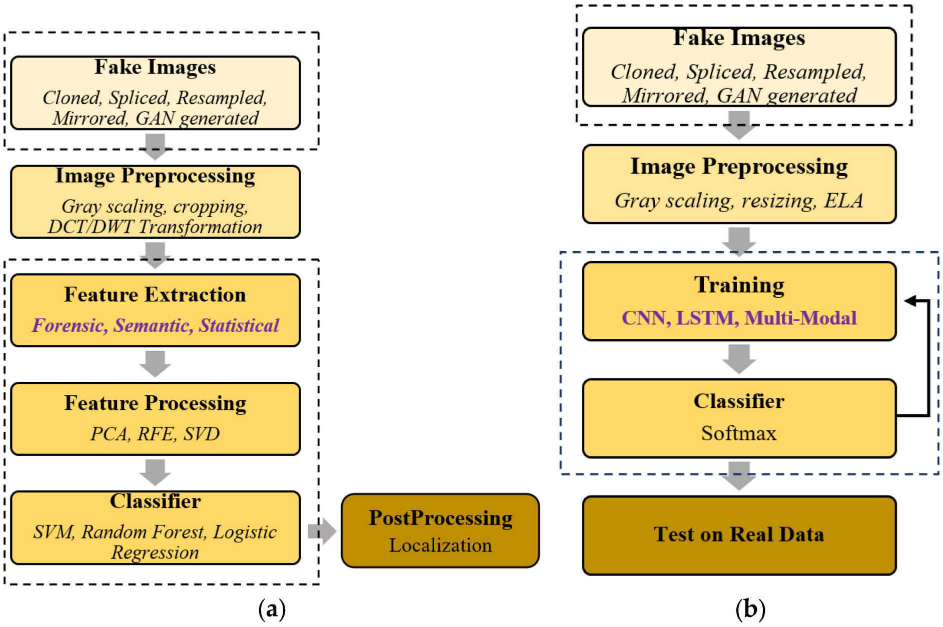
Fake Images
Cloned, Spliced, Resampled, Mirrored, GAN generated
Image Preprocessing
Gray scaling, cropping, DCT/DWT Transformation
Feature Extraction
Forensic, Semantic, Statistical
Feature Processing
PCA, RFE, SVD
Classifier
SVM, Random Forest, Logistic Regression
PostProcessing
Localization
(a)
Fake Images
Cloned, Spliced, Resampled, Mirrored, GAN generated
Image Preprocessing
Gray scaling, resizing, ELA
Training
CNN, LSTM, Multi-Modal
Classifier
Softmax
Test on Real Data
(b)

Figure 6. Fake image detection workflow. (**a**) Fake image detection workflow using handcrafted image features. (**b**) Fake image detection workflow using neural networks.

3. Image Tampering Techniques

Fake images are not new. The first incident goes way back to 1840 by Hippolyte Bayard, who created the first fake photograph. There have been a lot of fake images created since then. In the digital era, tampering with digital photos became very popular as it was effortless to manipulate digital images with photo-editing tools. Some commonly used photo-editing tools are Adobe Photoshop, GIMP, Paint.net, Pixlr, Photoscape X, Fotor, and InPixio. Images can be manipulated in various ways. The primary image tampering methods used are (1) Mirroring, (2) Resampling, (3) Copy-and-Move, (4) Image Splicing, and (5) Generative Adversarial Networks (GAN) generated fake images. These tampering methods have been used well since image altering started, except for GAN, which came into the picture in mid-2014.

3.1. Mirroring

Mirroring is a basic tampering technique. In this technique, the mirror image of the original image is used. These mirrored images are depicted or edited to give different meanings to them. Mirroring is sometimes performed so that fake images are not searched in reverse image searches. Figure 7 shows an example of a deer photo mirrored to create a new image.

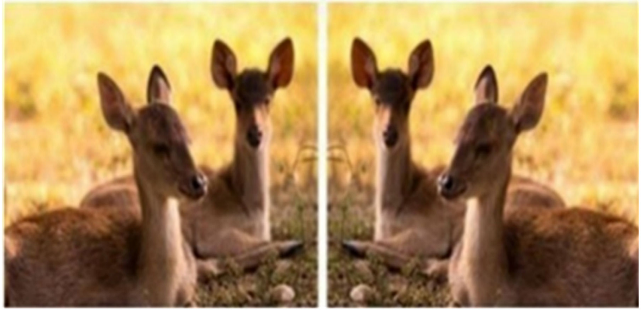

Figure 7. Mirroring of deer [16].

3.2. Resampling

Digital images can be visualized as a grid of evenly spaced pixels. Each pixel can be taken as a subject's sample or amount of light. Resampling is how a tampered picture version can be created with a different height and width in pixels. In upsampling, the size of the image is increased; on the other hand, the reverse is performed in downsampling, where the size is reduced. Image rotation is also achieved through resampling. Figure 8 exemplifies how rotating an image at arbitrary angles transforms a picture.

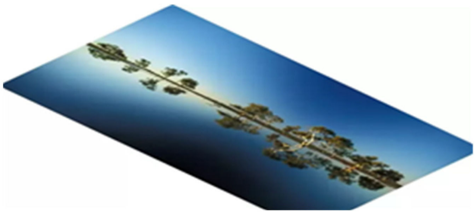

Figure 8. Resampling: rotating a lake image.

3.3. Move/Cloning

In copy–move tampering, a copy of a segment of an image is copied and manipulated and then pasted over the same image but at a different place. This technique is tricky. Figure 9 shows an example of how a forged white car is placed on the right side of the road. The forged car is taken from the same picture on the left side of the road.

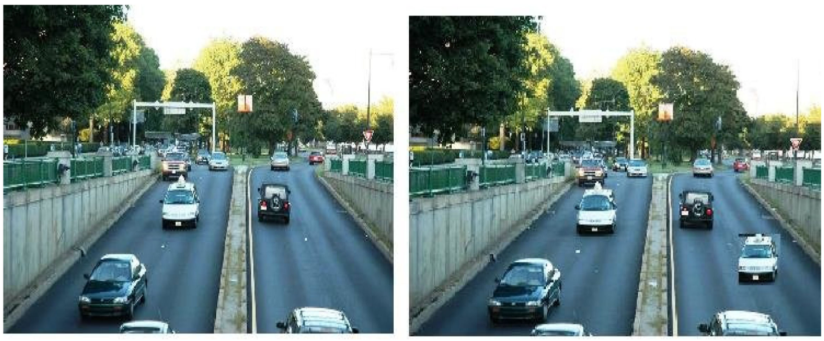

Figure 9. Copy-and-move segment from the same image, a white moving car [17].

3.4. Image Splicing

In this manipulation method, a compound picture takes a few objects from the different images and pastes them over other pictures. This tampering technique is complex and requires excellent skills to create an excellent fake image. Figure 10 displays how a vintage car is placed in front of the Backhoe ground digging machine to depict the car stuck in some deep pit. Here, a vintage car photo is taken from a different picture and pasted precisely in front of a backhoe.

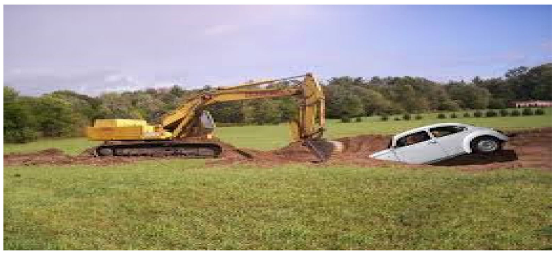

Figure 10. Image Splicing [18].

3.5. Generative Adversarial Networks

GAN, a machine learning framework, can produce new virtual images that appear at least superficially genuine to human observers when trained on different images. The photos have many realistic characteristics. GAN is well known for generating images/videos of fake faces. The altering of faces (deepfakes) can primarily be created in the following ways: full face synthesis; swapping of expression; swapping of identity; and attribute manipulation over the faces. Figure 11 shows an example of a face swap where the celebrity's face is swapped to a different person's image/video.

GAN is a neural network and comes under reinforcement learning. The system here is that the network learns dynamically by tuning actions based on continuous feedback.

GAN components are the generator and discriminator. The generator creates fake images, and the discriminator detects them as fake. The process is repeated until the Nash equilibrium is achieved or nearly achieved [19]. The discriminator's stochastic gradient is updated by ascending to maximize the loss function. On the contrary, the generator's stochastic gradient is updated by descending as there is a need to minimize the loss function.

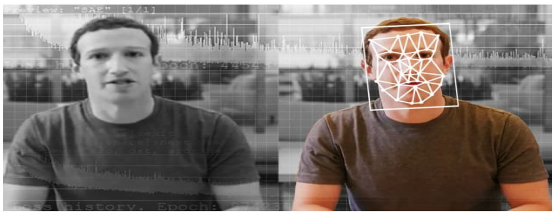

Figure 11. How face swap is made in deepfakes [20].

4. Fake Image Detection Methods—Using Handcrafted Feature Set

As described in Figure 5, various detection methods are used to identify a fake image. The two most prominent methods are based on forensic features and deep learning. This section will discuss methods using feature sets extracted from images. Under each method, a brief comparison is also shown among different research works. It is to be noticed that performance comparison is only made among various techniques that use the same evaluation parameter and dataset.

4.1. Forensic Features Based

Forensic feature-based techniques require the detection of the image using its natural features. Thus, forensic features have different techniques based on image manipulation type. Each manipulation can be detected in its unique way. Below, the research is discussed in fake image detection using forensic features.

4.1.1. Copy-and-Move/Cloning

For cloning detection, two main approaches exist: feature-based and hash-based. Many of the detection algorithms are developed based on features. In contrast, the hash-based method is used only in the case of plain cloning detection, i.e., when the copied image fragment is not transformed/processed. The key gain of hash-based algorithms is to use a low computational complexity. Warif et al. [21] shared a review of various copy-and-move studies but covered only some of the techniques as it was an evaluation comparison paper.

Initially, Fridrich et al. [22] implemented a copy–move tampering detection algorithm using quantized Discrete Cosine Transformation coefficients (DCTs) on small overlying blocks. An image is scanned by BxB block size, and its feature vectors are calculated using DCT. Then, a block comparison is examined after feature vectors are lexicographically sorted. Observing irregular patterns is formed by blocks that match copy–move tampering. Popescu and Farid [23] improvised the DCT-based overlapping block algorithm using principal component analysis (PCA). The small block sizes were passed through PCA to reduce the features. Using PCA, the authors could reduce the features to almost half the feature numbers by Fridrich et al. [22].

The technique is effective and can handle a little noise, but it fails to detect a copy–move image having rotations efficiently. Another method proposed by Li et al. [24] used a Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD) based on a sorted neighborhood approach. The picture is reduced in dimension using the DWT method, and then the SVD is applied over components having low frequency for obtaining the feature vectors. The technique works well even if JPEG is compressed to 70-quality levels but fails if more compression is applied. Bayram et al. [25] applied Fourier Mellin Transform (FMT) properties to detect cloning invariant to scaling, rotation, and translation. The author proposed counting bloom filters rather than lexicographic sorting to improve computational efficiency. It was observed that there is a linear correlation between pixels in the real image to handle rotations, which becomes distorted upon applying any tampering technique [26]. They employed SVD for feature detection extracted from image sub-blocks. This method could detect tampering, such as geometric transformation and brightness alteration. The technique was found robust against rotation. An enhanced DCT-based technique was employed by Huang et al. [27] by including a truncating process that would ignore the blocks having higher frequency coefficients. This process was used to decrease the dimension of the feature vector for fake detection. The results were quite robust to AWGN distortions besides JPEG compression. In another method, instead of square blocks, circular blocks were used [28]. The picture was divided into overlying circular blocks. The feature extraction of the blocks is attained through consistent Local Binary Patterns (LBP), which are invariant to rotation. The method was robust to different transformations like compression, rotation, blurring, flipping, and AWGN. However, the technique fails to spot tampered regions rotated with random angles. Another novel technique proposed by Lee et al. [29] suggested using the Histogram of Oriented Gradients (HOG) for feature detection over overlapping blocks.

The HOG features would detect the forged regions. However, the technique required improvements where tampering was performed over large areas in an image. At the same time, Hussain et al. [30] suggested using a multiscale Weber's law descriptor (WLD) histogram for feature detection. The multi-WLD abstracts feature from chrominance components of the picture. The method used SVM as a classifier and was evaluated over datasets like CASIA 1.0, CASIA 2.0, and Columbia. Another technique used the DCT technique with Gaussian RBF kernel PCA [31] to reduce feature vectors. Overall, this method significantly reduced the feature-length without compromising the results. The results observed were as good as the multi-WLD approach.

Jwaid et al. [17] performed a comparative study of various methods like DWT, LBP, and Scale Invariant Feature Transformation (SIFT), and they found that the SIFT-based approach was better than others. At the same time, Alamro and Nooraini [16] used a fusion of DWT and Speeded Up Robust Features (SURF). DWT is applied to reduce the photo's dimension, and SURF is used to extract the main points from the image. The technique has been verified with JPEG and BMP format images comprising the genuine and fabricated image set. In contrast, Chen et al. [32] proposed a novel method where fractional Zernike moments (FrZMs) are summed to fractional quaternion Zernike moments (FrQZMs) using quaternion algebra. The algorithm considers the FrQZMs as features and uses an enhanced PathMatch algorithm to match the elements. The algorithm worked well on color images and was evaluated over publicly available datasets FAU and GRIP. Dixit and Bag [33] utilized a "Center Surround Extrema" (CenSurE) detector for detecting keypoints within the forged images. The "Local Image Permutation Interval Descriptor" (LIPID) was used to perform the keypoint feature computation. Keypoint feature coupling uses the "k-nearest neighbor" (k-NN) method.

Conversely, Rani, Jain, and Kumar [34] used sophisticated template matching and speeded-up robust features (SURF) approach. The hashing algorithm was used by Tanaka, Shiota, and Kiya [35] to identify picture modifications. This technique may also identify photos that have been compressed after being altered. Tanaka, Shiota, and Kiya [35] refined the hashing method for better outcomes. Yang et al. [36] offered yet another unique approach. Yang and others used a method with two stages. The Grid-Based Filter and the Clustering-Based Filter were the two filters. Tahaoglu et al. [37] employed a textual form of the input image extracted using the suggested approach. Since textual pictures are the source of the SIFT keypoints and descriptors, more robust keypoints and descriptors were used. Keypoint matching identifies suspicious areas and assesses whether the picture is fake. The Ciratefi-based technique is used to localize the fabricated pixel. Uma and Sathya [38] proposed a new CMF detection method that takes into account a few of the strongest KPs, selected from both FAST-corner KPs and "Difference of Gaussian" (DoG)-based KPs, evaluates SIFT descriptors, applied DWT for dimensionality reduction, and uses optimization based on football games (FGBO). The FGBO is a member of the meta-heuristic optimization algorithm family. Gan, Zhang, and Vong [39] have employed SIFT methods for copy–move detection. They used FLM and HSF algorithms to reduce computation and filter out outliers.

Table 1 provides a computational efficiency comparison between the stated above methods. It illustrates that using a dimensionality reduction technique like PCA gives the same or better results with fewer features. Table 2, on the other hand, describes the pros and cons of various methods. It is clear that when images are subjected to any other alteration like compression or rotation, detecting copy–move manipulation becomes harder, and the technique fails to spot them as fake. Multiple manipulations are widespread among images shared over social media.

Table 1. Computational efficiency comparison—copy-and-move.

Methods	Feature Length	Feature
Fridrich et al. [22]	64	DCT
Popescu and Farid [23]	32	PCA
Li et al. [24]	8	DWT + SVD
Bayram, et al. [25]	45	FMT
Huang et al. [27]	16	Improved DCT
Mahmood et al. [31]	10	DCT and KPCA

Table 2. Comparison of techniques used in copy–move detection using forensic features.

S. No	Author	Detection Techniques	Dataset	Results	Pros	Cons
1	Li et al. [24]	DWT + SVD	Self-created Image set	Not provided	Can handle JPEG compression	Fails after 70 factors
2	Gul et al. [26]	SVD	200 Images	86%	Can handle scaling, rotation, and blurring	JPEG compression and con-trast/brightness have low results
3	Li et al. [28]	Local Binary Pattern	200 images from the Internet	Correct detection Ration ~0.9	Can handle rotation, blurring, compression	Images rotated at general angles
4	Lee et al. [29]	Histogram of Oriented Gradients	CoMoFod	FC factor > 90%	Can handle small rotation, blurring, and contrast	High scaling and high rotation
5	Hussain et al. [30]	Multiscale Weber's Law Descriptor MWLD	CASIA 1.0 CASIA 2.0 Columbia	Accuracy 92.08 95.70 94.17	Can handle rotation, compression, noise	Cannot tell localization of tampered image
6	Alamro and Nooraini [16]	DWT + SURF	50 Images from MICC-F2000	95% Accuracy	Good with geometric transformation	Not verified with compression and AWGN noise
7	Chen et al. [32]	FrQZMs	FAU and GRIP	F-Measure of 0.9533 over GRIP and 0.9392 over FAU	Can handle scaling and noise processing	Low results with rotation angles and JPEG compression
8	Tanaka et al. [35]	Robust Hashing	UADFV, CycleGan, StarGan	0.83, 0.97, and 0.99 F-score	Can handle noise, compression, and resizing	Not verified on social media images
9	Gan, Zhang, Vong, [39]	SIFT with HSF algorithm	CMH and GRIP	F-Score 91.50 on CMH. 92.11 on GRIP	Work well on geometrical attacks and post-processing disturbances	Not verified on social media images

4.1.2. Image Splicing

Detecting image splicing is relatively more challenging than copy-and-move tampering. There is comparable lineation of the object of the same image as copy-and-move tampering can have equal transitions, texture, length, and many others, while in image splicing, different former segments are introduced with different textures and image characteristics complexity.

Ng and Chang [40] proposed a method using bicoherence values. The feature values computed from the bicoherence of a spliced image's horizontal and vertical 1-D slices are a detection technique. They observed that image splicing increases the value of the bicoherence magnitude and phase features. The detection accuracy of this model is about 70%. The method does not work well when other non-splicing factors manipulate the image. The alternative method proposed by Popescu and Farid [41] used various CFA interpolations in digital cameras. The correlation of the CFA interpolation is disturbed when tampering is performed. The variance is calculated between the blocks. This method is limited to pictures from digital cameras, which use CFA.

After a detailed evaluation study of image splicing, Chen et al. suggested methods founded on Hilbert–Huang transform in 2006, and the next year [42] improved it with wavelet characteristic functions along with 2-D phase congruency statistics. They observed that splicing leaves traces of image tampering, specifically at locations with sharp image transitions. Wang et al. [43] suggested a method based on a gray-level co-occurrence matrix (GLCM). Here, the feature is extracted based on image edges in the chroma channels. Strong signal denoting the image content was ignored, while the low signals, i.e., the edges of spliced images, were preserved. LIBSVM is used as a classifier. In an improvisation to Wang et al., Zhao et al. [44] suggested using chroma image spaces. The technique uses four directional (0°, 45°, 90°, and 135°) run-length run (RLRN) produced from gray-level run-length pixel number matrices of the de-correlated channel as distinctive features to isolate altered images. A novel framework suggested by Liu et al. [45] uses a photometric uniformity of light in the shadows. Here, the color features of shadows are evaluated by the shadow-matte value. A picture extracts different matte values from shadow boundaries and the penumbra region. Then, consistency is compared. However, this technique works well only on images that have shadows. Improvising over CFA, Farrera et al. [46] proposed a method by calculating the existence of de-mosaicking artifacts left after tampering and then using a statistical model that calculates the geometric mean of the variance at the local level to arrive at the tampering probability of each of two × two image block. However, this method works only over images taken by a digital camera where de-mosaicking algorithms are used. He et al. [47] proposed combining DCT and DWT features. A Markov-based approach was taken where the Markov features were extracted from transition probability matrices in DCT, and additional elements were added from the DWT domain. For feature reduction, the SVM-RFE method is employed. Lastly, the SVM classifier is used. Mazumdar and Bora [48] proposed using an illumination signature to detect images splicing over human faces. The signature is extracted from the face region existing in an image using the "dichromatic reflection model" (DRM). Illumination signature is the "dichromatic plane histogram" (DPH), calculated from the facial region present in an image by applying a 2D Hough Transform. However, this technique is specifically for images that have human faces. Moghaddasi et al. [49] employ PCA over SVD in their implemented model, which uses the SVD-based feature extraction method to extract DCT features from an image. To reduce the feature dimensionality, the author applied Kernel PCA. The study was performed over Columbia datasets with different feature vectors. They found that the best accuracy was observed with 50 dimensions. Sheng et al. proposed a unique method using discrete octonion cosine transform (DOCT) and Markov [50]. The algorithm would first convert the image into the DOCT domain, and then the inter-block and intra-block Markov features are extracted in the DOCT area. LIBSVM is used as a classifier using Markov features. This method gave excellent accuracy results over CASIA ver1 and ver2 datasets but failed when the image size was too small. Jaiswal and Srivastava [51] recently used machine learning logistic regression to identify the image splicing images. The proposed method first converts all images to Grayscale. In the feature extraction stage, it learns four different feature sets: LBP, Laws Texture Energy (LTE), HoG, and DWT (Wavelet Features). As stated above, these feature sets have been used individually in various research works. A logistic regression model is trained and used as a classifier, combining all 142 feature vectors extracted from these feature sets. The model proves its efficiency by giving more than 98% accuracy of CASIA 1.0, CASIA 2.0, and Columbia data sets. But when applied to photographs that have been severely downscaled, texture and clarity are destroyed. Itier et al. [52] proposed a further novel concept by investigating the correlation of image noise over the RGB color channels over a spliced picture. Monika et al. [53] employed a different conventional DCT method to find both modifications. Niyishaka and Bhagvati [54] proposed a framework depending on illumination–reflectance and LBP. The image is transformed into Y and CrCb color space using this technique. The illumination element is then derived using the illumination–reflectance approach. The LBP histogram is produced by illumination in the last stage, and CbCr is used as a function vector for classification. Several machine-learning classification techniques were employed. In their work, Jalab et al. [55] provide a unique Pixel's fractional mean (PFM) approach to improve pictures before classification to improve recognition of image splicing fraud based on texture attributes. Depending on the intensity of each pixel's occurrence, the suggested PFM enhances each pixel independently. The most important elements from allegedly spliced photos are extracted using two texturing algorithms. The SVM classifier then employs these attributes to classify genuine and spliced pictures. In their proposed system, Agarwal et al. [56] used a self-supervised method for training splicing detection/localization models using an image's frequency transform "real-valued fast Fourier transform" (RFFT) algorithm. The deep network developed a representation to capture an image-specific signature by enforcing (image) self-consistency to detect the spliced areas. To solve this issue, the authors suggested an Edge-enhanced Transformer (ET) for tampering area localization. A novel method by Sun et al. [57] proposed a two-branch edge-aware transformer created specifically to include the splicing edge hints into the forgery localization network, creating forgery features and edge features to collect rich tampering traces. Additionally, the authors provided a feature improvement module to draw attention to edge area artifacts in forged features and apply weight values to the resultant tensor in the spatial domain for essential signal amplification and noise reduction.

Table 3 compares the computational efficiency among various methods stated above for detecting splicing. Again, as expected in forensic methods, feature processing techniques like RFE and SVD help drastically reduce the computational power and make the model fast and efficient. Table 4 briefly compares various techniques after 2010 since earlier techniques did not have good accuracy results. Here, it can be observed that the forensic feature set fails to identify the fake image if splicing is fused with other manipulation techniques like compression, retouching, and resizing.

Table 3. Computational efficiency comparison—image splicing.

Methods	Feature Length	Feature
Wang et al. [43]	100	GLCM + BFS
Zhao et al. [44]	60	RLRN
He et al. [47]	100	SVM + RFE
Moghaddasi et al. [49]	50	SVD + DCT
Sheng et al. [50]	972	DOCT + Markov
Niyishaka and Bhagvati [54]	768	Illumination–Reflectance and LBP

4.1.3. Resampling

To detect resampling detection, Popescu and Farid [58] used expectation–maximization (EM) algorithm to evaluate probability maps and spot the image's explicit correlations. Each sample's image is interpreted with its probability of being connected to its neighbors. This technique was verified on basic resampling methods only and cannot be used on compressed images. Fillion and Sharma [59] proposed an approach to detecting content-aware scaling of images using seam carving algorithms. The study was made from seam behaviors—like the distance between seams and energy along the path. The features of a seam are likely to be affected by the seam-carving approach. These seam features were used in the SVM classifier, which delivered an accuracy of 91%. Mahalakshmi et al. [60] used an interpolation-related spectral signature method that spots simple image alterations like resampling. It also detects histogram equalization and contrast enhancements. The fingerprint detection method is used for histogram equalization and contrast enhancement. However, this resampling detection algorithm fails when JPEG compression is performed. Niu et al. [61] recommended using complex-valued invariant features to enhance earlier keypoint-based methods. Multiple clone concerns and geometric transformation problems with earlier keypoint-based approaches were overcome by Niu et al.

Table 4. Comparison of techniques used in splicing detection using forensic features.

S. No	Author	Detection Techniques	Dataset	Results	Pros	Cons
1	Zhao et al. [44]	Run-length run number (RLRN)	CASIA 1.0 Columbia	Accuracy 94.7% 85%	Works well in color and grayscale images	Not verified for any pre/post-processing over images
2	He et al. [47]	Markov features in the DWT and DCT domain	CASIA 1.0 Columbia	Accuracy 89.76% 93.55%	Works well for color and grayscale.	Gives lower accuracy over real-world images, more realistic images
3	Mazumdar and Bora [48]	Illumination-signature using DRM	DSO-1 DSI-1	AUC 91.2%	Good performance for images with faces	Fails on images having sharp contrast skin-tones
4	Moghaddasi et al. [49]	SVD + DCT + PCA	Columbia	Accuracy 80.79% (No PCA) 98.78% (PCA)	Have excellent performance over grayscale images	Not verified for color images. Not verified with Pre/Post-processing
5	Sheng et al. [50]	Markov features of DOCT domain	CASIA 1.0 CASIA 2.0	Accuracy 98.77% 97.59%	Can handle Gaussian blur and white Gaussian noise	Fails over small-size images
6	Jaiswal and Srivastava [51]	Machine learning—logistic regression	CASIA 1.0 CASIA 2.0 Columbia	Accuracy 98.3% 99.5% 98.8%	Can handle pre/post-processing alterations	Fails when images are highly down-sampled
7	Niyishaka and Bhagvati [54]	Illumination reflectance and LBP	CASIA 2.0	Accuracy 94.59%	Can handle down-sampling and resizing	Fails over small size images and images with blurred background
8	Agarwal et al. [56]	RFFT image frequency transform	Columbia	Average Precision 0.918	Can handle down-sampling and resizing	Fails over small-size images

4.1.4. JPEG Compression

JPEG compression is considered a non-malicious manipulation. It is performed to compress the image to meet social platform storage compliance. The three basic processes of JPEG compression are discrete cosine transform, quantization, and entropy coding. On the decoding end, the procedure is reversed. A method for detecting JPEG compression was created by Fan and Queiroz [62]. It would initially determine whether or not a picture has been JPEG compressed. Once the compression signature has been evaluated, compression parameters are estimated. A function to calculate the maximum likelihood for the quantizer step was developed. MLE estimation was devised, which could be used to assess the usage of the quantization table. Krawetz proposed Error Level Analysis (ELA) [63], which uses the fact that the JPEG resaving error is not linear. The method was to resave the JPEG images with a known rate and then compute the difference. In an uncompressed image, all pixels in the picture are not at their local minima, but when compressed, they achieve their local minima. Zhang et al. [64] exemplified a method based on double JPEG2000 compression to spot and locate the manipulated areas in tampered images. The method utilizes the fact that there is a statistical difference in single and double JPEG2000 compression. The difference sums to double quantization of the sub-band DWT coefficients, which brings in specific artifacts visible in the histograms of the Fourier transforms of the DWT coefficient. However, this technique could detect single and double compression only. The method would have a different accuracy if multiple compressions were made. Lin et al. [65] created a fully automatic model for spotting manipulated images by inspecting the Double

Quantization (DQ) effect, which is latent in the DCT coefficients. The technique uses SVM as a classifier. However, the method fails when the original image is not JPEG and if some other tampering has been made, like resampling or splicing. Kwon et al. [66] used a neural network with DCT for JPEG detection.

4.1.5. GAN-Generated Images

It is tough to identify GAN images using forensic methods. Though some reasonable attempts have been made using forensic methods, deep learning methods show better results. The current work in detecting GAN-generated forged images primarily emphasizes using signal-level features for spotting the faux. McCloskey et al. [67] examined the GAN generators, and they observed that the frequency of saturated pixels is limited and that RGB channels are collapsed using weights that are unlike the spectral sensitivities of a digital camera. Based on the frequency of over-/under-exposed pixels, it uses a basic forensic to spot the distinction between GAN-generated and camera imagery. The work introduced intensity noise histograms for classifying authentic and GAN-generated images. As an alternative, Nataraj et al. [68] suggested taking the color co-occurrence matrix as input. The matrix was extracted from the pixel domain's RGB channels for taking spatial correlation features. This feature set is then fed into the CNN framework. The framework was verified against CycleGan and StarGan datasets, with an accuracy of 99%.

For deepfake detection, the "Deepfake Detection Challenge" (DFDC) was organized by the National Institute of Standards and Technology (NIST). Later, Facebook also launched a similar competition. Matern et al. suggested that visual artifacts would be enough to detect deepfakes [69]. The proposed model uses differences in eye color to detect generated faces. Iris pixels of the eye region are used to calculate the color saturation variance. It also checks that the distance between the center of the iris and the center of the eye should be similar for both eyes. However, this method is limited to images that have a human face with bright, open eyes. For deepfake videos, Li et al. [70] proposed that unrealistic eye blinking can be used to detect face-swapping. The model used CNN-based VGG16 to learn this physiological signal of eye blinking using the Eye Aspect Ratio (EAR). The model becomes confused and gives inaccurate results when the eye region is small in the frames. It also needs to be improved for the dynamic pattern of blinking.

The model used residual signals of chrominance components from multi-color spaces. These signals, including HSV, YCbCr, and lab, were passed through a shallow CNN model to learn the representation. In the end, a Random Forest was used as a classifier. The model was verified against images having compression, rotation, noise, and resizing. Zhang et al. proposed a deep learning method using the ELA for face swap detection [71]. The ELA technique uses the principle of having different ratios of image compression. The model suggests using images going through the ELA process before passing it to a CNN model. The CNN learns counterfeit feature vectors from ELA-processed images and identifies them as fake or real. The technique works well for face swap, with compression, but images without compressions. A "pixel-region network" (PRRNet) method to detect face forgery was proposed by Shang et al. [72].

4.1.6. Problems—Forensic Method

The survey presented various techniques used in forensic methods. Forensic methods are specialized methods to be used for some specific manipulation methods. When the images are shared over social networks, the shared image typically undergoes multiple manipulation and transformation. Thus, it becomes very challenging to exploit any one method for detection. High accuracy is observed in the single-manipulation process, but not much efficiency is achieved for multi-manipulation. For example, Figure 12 shows how tiny image manipulations are misclassified as fake using forensic techniques. When images are highly compressed, resized, and cropped, undergo arbitrary rotation, mirroring, and added noise for social media usage, it reduces the overall quality of images, making it hard to discriminate between real and fake. Nowadays, the virtual images generated via adversarial examples have significantly the same image features as the original, and thus, many algorithms fail to spot the tampering. GAN-made pictures are best detected when the deep learning CNN approach is applied along with a forensic approach.

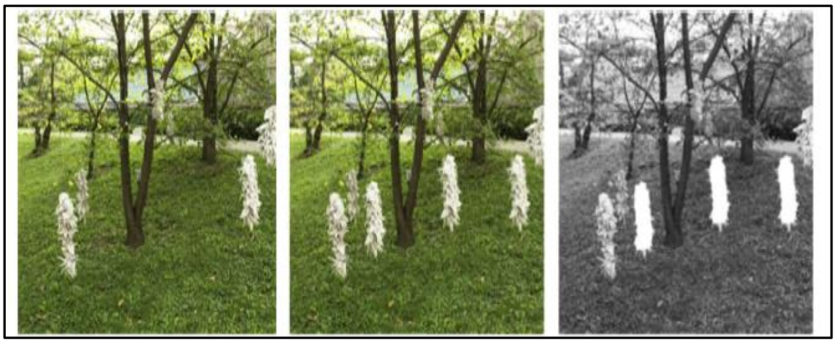

Figure 12. Tampered region identified by forensic technique [29].

Table 5 illustrates performance results between various forensic feature-based techniques. Here, it is to be noted that only those techniques compared, which used "accuracy" as an evaluation parameter, and experiments were performed on similar image-specific datasets. It is obvious that with increasing knowledge and technology, the accuracy of the methods is improving, and the best accuracy is achieved by using machine learning with multiple features. However, as explained above, the forensic method's efficiency fails when the image has undergone multiple manipulations.

Table 5. Accuracy comparison of forensic methods.

Models	Tampering Method	Detection Method	Columbia	CASIA 1.0	CASIA 2.0
Hussain et al. [30]	Copy-and-Move	MultiWLD	94.17%	94.19%	96.61%
Mahmood et al. [31]	Copy-and-Move	DCT+ KPCA	-	92.62%	96.52%
Wang et al. [43]	Image Splicing	GLCM + BFS	-	90.50%	-
Zhao et al. [44]	Image Splicing	RLRN	85.00%	94.70%	-
He et al. [47]	Image Splicing	SVM + RFE	93.55%	-	89.76%
Moghaddasi et al. [49]	Image Splicing	SVD + DCT	98.78%	-	-
Sheng et al. [50]	Image Splicing	DOCT + Markov	-	98.77%	97.59%
Jaiswal et al. [51]	Both	DWT + HOG+ LBP + ML	98.80%	98.30%	99.50%

The issues regarding forensic methods over social media images can be collated as below:

- Specialization: Non-specialized examples as they undergo multiple manipulations;
- Proper resolution: Social Images are deficient in quality due to size constraints over the platforms;
- Compression: They are highly compressed images and have multiple compressions at times;
- Visual Features: Noise addition through the blur and edge removal techniques; thus, features are lost;
- Cropping: Much cropping is carried out to hide the details and highlight emotional content;
- Regions: Images can have large manipulated areas or tiny tampered patches. Figure 12 shows one such example, where a real tree shoot is marked as tampered (last right shoot);
- Source: Sources can be different, like digital cameras, computer-generated, and GAN;
- Formats: Platforms support multiple formats like JPEG, TIFF, GIF, BMP, PNG, and PSB.

4.2. Semantic Features

Fake news is intentionally created to manipulate the individual weaknesses of human beings. Thus, faux images are dramatically exaggerated to incite anger or hate reactions in public, which leads to further dissemination of fake news. These deliberative manipulations have some distinct cues at the semantic level in images that contrast with real news.

Sunstein shares fake news spreaders' emotional and behavioral studies [73]. The research shared real-time examples of herding behavior where the fake news spread is amplified by people sharing the same views or interests. It is also known as the echo chamber effect. It is noticed that people have a perception that complete falsification is not factual, but sensationalist or partisan news does contain some aspects of truth. Faux news generators utilize these behavior patterns to spread their intent fast into society. Based on the above observation, Jin et al. [74] proposed a model built on psychologically triggered visual patterns in fake images. They modeled a domain-transferred deep convolutional neural network with weighted instances and trained over 40k images. Some interesting visual semantic patterns were observed from the results: fake images tend to be more eye-catching, event-centric, disturbing, and low-quality than real ones. These cues confirm observations by Sunstein. Shu et al. [75] used psychological and social theories in combination with data mining. The study shows that fake news detection techniques are primarily based on text or social context content. The social context cues over images can play a significant role in detecting false news. Ghanem et al. [76] suggested utilizing the semantic and stylistic elements of the suspicious account to identify the bogus credibility of the news created from these accounts.

In contrast to the above methods, Huh et al. [38] proposed a self-supervised method that uses an algorithm based on the picture's EXIF metadata as a supervisory signal. These signals are trained in a ResNet50 framework to decide whether a photo is self-consistent. If the photo is self-consistent/untampered, then its constituents should be generated by a single imaging pipeline. This approach has a few limitations, as it depends on the EXIF metadata information of the device. The model is not well-suited for detecting minor splicing over an image. It also becomes confused with underexposed and overexposed regions of the picture. It does not work well with the copy-and-move tampering method, as the manipulation is from the same image. On the other hand, Zhang et al. [77] proposed a method using photo-response non-uniformity (PRNU). Modern PRNU-based forensics techniques often depend on Markov random field modeling with multi-scale trace analysis and result fusion.

Issues with semantic feature detection techniques are their limitation to the semantic features in an image based on behavior or psychological patterns. Each of these patterns, despite knowing, is subjective, and they are updated with evolving technology and public behavior. They will often require domain expertise as the model interpretability is complex. Alone with semantic features, they will need other elements to derive more successful results.

4.3. Image Retrieval/Web Search

Image retrieval or reverse image search is often the most common activity a user performs when it senses any tampering and looks to verify the integrity of the image. Image search engines are now advanced and mostly retrieve other sources with similar images. Commonly used search engines specially designed for these reverse searches are Tineye, Picsearch, Google image reverse search, Yander, and Yahoo image search.

Xiaohui et al. [78] analyzed the survey of over 200 people and tried to predict the user intent in searching images over the web. The study shows that user behavior in searching the web correlates with the intent of his search. Patterns like dwell time, mouse hover, mouse click, and query reformulation can predict the user's intentions. Later, based on user intention, the search engine can provide the exact images the user is searching. Taking learning from the above survey study, next year, Xiaohui et al. [79] again proposed a grid-based evaluation matrix implemented in alternative to Discounted Cumulative Gain

(DCG) or Rank-Biased Precision (RBP), which are traditional list-based metrics. This time, the suggestion was proposed after studying user patterns like middle bias, slower delay, and row skipping. However, this study did not include appearance bias and was conducted on a few people.

Gaikwad and Hoeber proposed an interactive information retrieval process by taking text with visual images over social media platforms. A user study was conducted, and the ImgSEE image database [80] was created, which was designed based on Vakkari's three-stage model of information seeking. The technique is collaborative with exploratory search and sense-making processes. These are useful for an image-search activity with less information, and the user may want to verify what they seek. The study also validated the efficacy of this technique by comparing it with a grid-based search method using candidates' views on usefulness, ease, and satisfaction. At the same time, considering text over an image, a novel technique was employed by Vishwakarma et al., which authenticates the accuracy of text existing over an image by searching for it on the internet and introduces the Reality parameter [81]. The Reality parameter (Rp) is calculated by checking the text's reliability from the top Google search results. The event is marked as real or fake based on the Rp value. However, the technique has limitations when correct text is not extracted from the image using an OCR. If the news depends on geography, it does not gather enough credibility to appear in top searches and will be wrongly termed fake. Issues with web retrieval methods are as follows:

- Not all images can be searched over the reverse web search;
- Images/news not gathering enough highlights will not be ranked in the initial few pages;
- Reverse image searches will also bring images from fake websites wherein such fake images are spread;
- It requires time for the fake image to spread; searching before it becomes viral will not fetch any relevant information;
- Searching fake videos over the web is a tight task and requires effort and time.

4.4. Statistical Features

It is observed that fake images have different statistical distribution cues compared to real news on social media. Gupta et al. [82] studied and found that people naturally share information with photos clicked with them from the incident site. Thus, ideally, the image's authenticity can be checked because various observers would also share other photos. At the same time, if it is fake, there are chances that multiple photos shared will have almost the same content. Thus, visual statistical features can determine the distributional difference between real and fake news and classify it as genuine or false. Huang et al. [83] presented the spatial–temporal structural neural network architecture to model message diffusion from temporal and geographic perspectives for rumor identification. It was effective in spreading rumors, but it did not consider the spread of fraudulent photographs. Chen, Retraint, and Qiao [84] used the GLRT-based statistical method. The detector's architecture is based on a JPEG image's reduced noise model, which considers pixel variance a quadratic function of pixel expectation. Two features of the proposed simplified noise model can be used as camera fingerprints to identify fake images. A training-free "Generalized Likelihood Ratio Test" (GLRT) is created using the framework of hypothesis testing theory, ensuring good detection performance for a predetermined false alarm rate. Jin et al. [85] proposed various statistical features of an image by which this differentiation can be made. This paper suggested the following features:

- Count: The presence of images in fake news. For example, how many images are present?
- Popularity: How popular is the event image over social media, such as comments and re-tweets?
- Dimension: What image size is gaining popularity compared to other images?
- The study suggested specific patterns in these statistical ratios, which are then used to classify the event as real or fake;

- Issues with statistical methods:
    a. Statistical methods need to be researched further. Similar statistical observations can be observed with real news, too;
    b. Also, it does not accurately identify fake images. It only gives a diligent prediction pointing toward fake image probability.

5. Fake Image Detection Method—Using Neural Networks

This section will discuss the techniques for spotting fraudulent images using neural networks. By giving neural networks data to train on, they can discover the hidden properties of a modified image. They can then predict and spot fake images based on their learning.

5.1. Convolutional Neural Network—Image Specific

A convolutional neural network, or CNN, is a deep learning neural network component designed for processing ordered arrays of input, such as photographs. The patterns in the input image, such as lines, gradients, circles, or even eyes and faces, are very well recognized by convolutional neural networks. They can automatically learn the mapping relationship between high-dimensional data and exhibit traits like translation invariance. Because of this characteristic, convolutional neural networks are particularly effective for computer vision issues like image classification, labeling, semantic segmentation, and picture synthesis.

In contrast to earlier computer vision techniques, CNNs may operate directly on a raw image and do not require prior preparation. In convolutional neural networks, many convolutional layers are stacked on top of one another, and each layer can recognize progressively complicated shapes. Three or four convolutional layers are sufficient to detect handwritten numerals, but 25 layers are required to recognize human faces. CNN uses convolutional layers to analyze input images and recognize ever-more-complex qualities like how the human visual cortex is set up.

Initially, a CNN-based model was used for lexical or text-based detection for spotting fake news. CNN models were used to identify counterfeit images based on user profiles and network propagation. Xu et al. [86] proposed deep learning about CNN architecture and long short-term memory (LSTM). The LSTM layer was used before the CNN layer to extract features locally and densely and produce a temporal structure from the input sequence. The training utilized videos rather than single images. The temporal characteristics were collated frame by frame, and the later relationship was established. The technique is limited to face spoofing in videos only. For images, Bayar et al. experimented with a novel CNN model mainly designed to restrain image content and adjusted to learning features to detect tampering. Prediction-level filters are used before passing the image to the convolutional layer [87]. These filters support suppressing the main content and allowing for the manipulated features. The model enforced weight constraint during each iteration after the filter weights had undergone stochastic gradient descent by back-propagating the errors.

Rao and Ni [88] explicitly designed a CNN architecture for cloning and image-splicing detection applications. Unlike a regular procedure, the essential 30 spatial rich models (SRM) filter sets are used to instate the weights at the first layer. This efficiently represses the image contents' characteristics and highlights the low-level artifacts produced by the manipulating attacks. The model was tested against Columbia and CASIA image datasets. Rao et al. proposed an attention-based multi-semantic CRF model for detecting picture counterfeiting [89]. To locate the tampered region, it also applied the CRF approach. The model proved impervious to noise and erosion, although it performed less accurately with JPEG-compressed pictures. The outcomes of the repeated JEPG compression were more decremented.

Salloum et al. used a Multi-task Fully Convolutional Network (MFCN) with a Single Fully CNN. The results were not significant [90]. Two streams of FCNN were used, one for producing a surface probability map and another for an edge probability map. The result of MFCN was that it outperformed existing splicing localization algorithms and could achieve finer localization than the SFCN. The degradation in performance was observed when the images were compressed, or Gaussian noise was added. Improving on it, Bappy et al. [91] employed a hybrid CNN-LSTM deep learning model to differentiate features in manipulated regions of an image. It is observed that discriminative features are present at the boundary of manipulated and non-manipulated regions. The images are passed through a basic convolutional layer at the first level to produce sixteen feature maps. One of the feature maps is passed to the LSTM layer in blocks. LSTM learns the boundary variations between different blocks and generates unique features. This helps in separating the tampered region from the non-tampered region. Now, further layers of CNN learn features from the manipulated regions. Under the Adobe Research program, Zhou et al. [92] employed a novel model using a two-stream Faster R-CNN network, one for RGB and the other for noise. SRM filters the extracted noise features between manipulated and authentic regions. RGB stream is designed to produce tampering feature artifacts like fabricated boundaries and acute color differences. The noise branch captures the features specific to noise using SRM filters, commonly used in steganalysis. A bilinear pooling layer joins features from both streams to further integrate these two modalities' spatial co-occurrence. The model showed slight degradation over copy-and-move tampering as the manipulated region was from the same image. However, compressed images were not taken in the experiment. Working on optimizing training time, Rehman et al. proposed an optimized model, LiveNet [93], based on the data randomization technique, which is like enhanced bootstrapping. In opposition to conventional CNN models, where the training set is randomly arranged once, that paper suggested continuously picking random mini-batches from the full training set at each training iteration. This led to a significant improvement in training time over the datasets. The problem of overfitting is also mitigated with this technique. This model was verified against an inter-database and cross-database containing human faces anti-spoofing data. Xiao et al. proposed another multi-branch framework, a coarse-to-refined convolutional neural net (C2RNet) [94]. In the first stage, there are two cascading CNN models. The first one is Coarse-CNN (C-CNN), a VGG-16-based framework to identify the different manipulated regions in an image. The output of C-CNN is on the coarse level. Therefore, some inaccurately identified areas may be present, especially around the edge of the picture. The resultant C-CNN is then cascaded to the next Refined-CNN (R-CNN) model to train over the image features' differences. R-CNN is based on the VGG-19 framework. The technique proposes an image-level CNN against the commonly used patch-level CNN to decrease computational time. Finally, an adaptive clustering technique is suggested to produce the final detected tampered regions. Adaptive clustering has two stages, an adaptive outlier filtering and a convex full-filling stage. The model achieves good results against the CASIA and Columbia datasets, though degradation is observed when multiple attacks like compression and noise are added. Improving over BusterNet [95], a two-branch Deep Neural Network (DNN), was proposed by Wu et al. for a copy–move fake detection (CFMD). It has two branches, Mini-det and Simi-Det. While the Mini-det was designed to spot tampered regions so that its feature is useful for the property; on the other hand, the Simi-Det was designed to find cloned regions and learn their features. Later, the two branches were merged to estimate pixel-level copy–move masks, distinguishing them from the authentic original image. On similar grounds, a two-stage cascading CNN model was proposed by Bi et al., proposing a CNN-based architecture named Ringed Residual U-Net (RRU-Net) [96], which provides a complete image segregation system. RRU-Net aims to optimize CNN learning through the recall and reorganization process of the human cerebral cortex. To resolve the gradient degradation issue of DNN, the residual block is skipped by one layer and utilized to recall the input vector data. The residual feedback collates the input feature information to discriminate between the true and tampered areas. The RRU-Net is executed on COLUMBIA and CASIA datasets. Liu and Pun [97] proposed a fusion network wherein the multiple layers of denseNet are used. To make the DNN learn fast, the network uses two major assumptions instead of learning from the entire image. The hypothesis has been that noise is observed over the edges where Splicing is applied and compression ratio variations. The FusionNet works well on pre and post-processing tampering as well. Abhishek and Jindal [98] used CNN and semantic segmentation to detect image manipulations. Another suitably lightweight CNN model was proposed by Hosny et al. [99]. It had specific convolutional and max-pool layers after experimentation. 100. Elaskily et al. [100] employed a hybrid model of ConvLSTM for copy-and-move detection using deep neural networks, but they did not include image splicing techniques. Koul et al. [101] proposed a novel method using a slant convolutional neural network (CNN) for automatic copy–move forgery detection.

In parallel, much research has also been conducted for detecting deepfakes. Hsu et al. [102] proposed a model common fake feature network (CFFN) designed explicitly for fake face detection generated by GAN. CFFN is built on reduced DenseNet, having a two-streamed network structure like the Siamese network to allow for pairwise information as the input. Contrastive loss is used to learn the CFFs. As the model created the fake GAN images, it tends to fail on generators using another generation method. Also, videos are not covered by this method. In contrast to previous research on fake face detection, another work proposed by Jeon et al. proposed a novel attuning neural network architecture named “Fake Detection Fine-tuning Network” (FDFtNet) [103]. The model uses a “Fine-Tune Transformer” (FTT), which comprises many self-attention components, which supports reducing CNN’s limitation in attaining long-term dependencies. The architecture uses MobileNet block V3 to determine the picture’s feature vectors through inverted residual structure and linear bottleneck. The model works well on specially trained datasets and needs more generalization. Wang et al. [104] proposed a universal detector technique for finding CNN-based fake faces. Their paper stated that CNN created fake images with common systematic flaws that will never be equivalent to a realistic image. The paper discusses that images produced will always retain detectable fingerprints despite using multiple CNN generators. A suitable image classifier can learn these CNN fingerprints. The study used ResNet-50 as a classifier and the ProGAN dataset to train it. Various data augmentation variants are also used to detect post-processing tampering. Contrary to the above research, Neves et al. propose removing this fingerprinting and implementing GANPrintR [105]. This CNN-based deep learning model removes the fabrication of fake faces and makes them look more natural. The architecture in the study uses an autoencoder, which first learns from the real images, and then the same learning is applied to fabricated fake faces to add extra naturalness. This is achieved by removing the GAN fingerprinting over the synthetic image. The study also verified different artificial face detection techniques like XceptionNet and steganalysis to show a significant drop in the ERR over the dataset created by this model. The model can develop robust artificial faces to improve fake face detection algorithms. In their review, Arora and Soni [106] specifically accounted for false pictures produced by Generative Adversarial Networks. They spoke about several deep learning techniques. Another innovative hybrid approach to identify GAN-generated deepfakes was suggested by Yang et al. [107]. The CNN-LSTM-based model has shown good performance in detecting faked images. LSTM was primarily used with RNN models in detecting fake news using text or sentiments. However, replacing RNN with CNN for images has also provided good results. LSTM sort of stores memory and is used in the prediction. An architectural image of the LSTM cell is depicted below in Figure 13. It describes how previous state memory is stored in each cell of LSTM and can be used for the current iteration for better learning and prediction. Table 6 compares deep learning models used in image tampering detection. A comparative result is displayed for models using similar evaluation parameters and publicly available datasets. RRU-Net achieves the best F-1 score in similar dataset comparisons. RRU is based on Residual propagation, which helps mitigate deep neural network problems of vanishing/exploding gradient. Table 6 shows Zhou et al.’s Noise Net and Late Fusion results [92]. The original literature on these networks is not discussed here. A two-stream CNN model was presented by

Kwon et al. [108]. In one stream, the RGB feature sets of pictures were learned; in another, the DCT feature set was learned. The embeddings were afterward combined for precise categorization. ResNet50 was used by Meena and Tyagi [109] to extract features from the altered photos. The NoisePrint model inflates the manipulation characteristics in the photos before sending them on to ResNet50. Jaiswal and Srivastava [110] proposed a deep learning CNN model using multi-scale input and several convolutional layer stages. These layers are separated into the encoder and decoder blocks. Extracted feature maps from convolutional layers with numerous levels of down sampling are concatenated in the encoder block. Similarly, upsampling and combining extracted feature maps occur in decoder blocks. Using a sigmoid activation function, the final feature map categorizes pixels into forged and non-forged. Zhou et al. [111] suggested a process of self-attention to locate forged areas in forged pictures. A "Channel-Wise High Pass Filter" block was the foundation for the self-attention module (CW-HPF). CW-HPF extracts noise features using high pass filters by correlating features across channels. A self-attention technique dubbed forgery attention is developed based on the CW-HPF to obtain rich contextual dependencies of inherent inconsistency derived from tampered areas. Wu et al. [112] used a noise-based approach. After carefully analyzing the noise caused by online social networks, the authors split it into predictable and unseen noises, which are then modeled independently. Mini-Net was proposed by Tyagi and Yadav [113], which employed the CNN network. Ali, Ganapati, Vu, and Werghi [114] proposed a deep learning CNN model image patch. A pixel in a picture is classified using a patch surrounding it, and then the CNN is used to determine if the pixel is part of the tampered area. The suggested approach accurately predicts the border pixels of the tampered region and the background picture. Singh and Sharma [115] proposed Siteforge, a customized CNN-based deep neural network with high-class filters. Wu et al. [116] used multiple layered CNN networks, ManTraNet, which could detect fake images, and their local anomaly detection network could even identify the tampered regions. Similarly, Hu et al. [117] created a spatial pyramid attention network, a CNN-based network with an attention mechanism for detecting and identifying the tampered regions. But with noisy pictures, there were some false-negative cases. Zhuang et al. [118] created an encoder and decoder-based CNN for detecting image forgery. A similar approach was also used by Biach et al. [119].

A Vision Transformer, often called ViT, is a type of neural network architecture that has gained significant attention and success in computer vision. It was introduced to address image classification tasks, similar to how Convolutional Neural Networks (CNNs) have traditionally been used. ViT is unique because it relies on self-attention mechanisms, previously popular in natural language processing (NLP) tasks like machine translation. Khan et al. [120] have written a survey paper on using transformers in vision classification. They also compared various ViT techniques with recent CNN-based architectures. They found that the ViT-based approach better detects deepfakes and fake videos. Ganguly et al. [121] employed a vision transformer with an exception network (ViXnet) for detecting deepfakes and image forgery. ViXNet has two parts. One part looks at different parts of a face closely to find things that do not match using a special kind of attention and a vision transformer. The other part looks at the whole picture to understand where things are in space using a deep convolutional neural network. Another transformer-based technique was employed by Hao et al. [122]. Dense self-attention coders and dense correction components are the two main parts of their approach. While the latter increases the hidden layer's transparency and corrects the results from various branches, the former is used to model the global context and all pairwise interactions among local patches at various scales. Arshed et al. [123] applied vision transformers over deepfakes and got excellent results on deepfake images shared over Kaggle. Similar results were observed by Heo et al. [124] with deepfake videos. They combine patch-based positioning and vector-concatenated CNN features to interact with all positions to determine the artifact region. The sigmoid function trains the logit for the distillation token using binary cross entropy. The suggested framework is generalized to increase performance by including this distillation.

The advantages of using deep learning for fake image detection are as follows:

- Feature Vectors/Intrinsic characteristics of fake images are learned by themselves. It does not need a feature set;
- It can detect images having multiple manipulations;
- Can detect images having Pre/Post-processing after tampering is applied over the images;
- It can use pre-trained state-of-the-art DNN models, which saves time;
- Provide higher results and better accuracy;
- It can work well on unstructured images/data from various sources and formats;
- ViT's based models work well on deepfakes (GAN) images and videos.

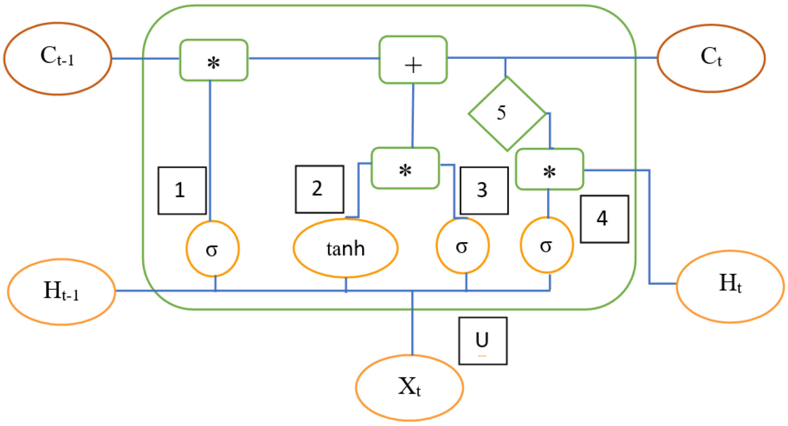

Figure 13. LSTM Cell at a time interval "T" [125].

$$Ct - 1 = \text{Previous Cell Memory; } Ht - 1 = \text{Previous Cell output; } Xt = \text{Input Vector}$$

$$Ct = \text{Current Cell Memory; } Ht = \text{Current Cell output}$$

$$U, W = \text{weights vectors for Candidate (C), Forget\_gate (F), i/p gate (I), o/p gate(O)}$$

$$1 = Ft = \sigma\ (Xt \times Uf + Ht - 1 \times Wf)$$

$$2 = Ct = \tanh\ (Xt \times Uc + Ht - 1 \times Wc)$$

$$3 = It = \sigma\ (Xt \times Ui + Ht - 1 \times Wi)$$

$$4 = Ot = \sigma\ (Xt \times Uo + Ht - 1 \times Wo)$$

So, with the above parameters, Ct and Ht are derived as

$$Ct = Ft \times Ct - 1 + It \times Ct$$

$$Ht = Ot \times \tanh(Ct)$$

Table 6. F1-score comparison of deep learning approach.

Method	Framework	Columbia	CASIA 1.0	CASIA 2.0	NIST 16
Bappy et al. [91]	CNN + LSTM	-	-	-	0.764
Salloum et al. [90]	MFCN	0.611	0.541	-	0.571
Zhou et al. [92]	FRCNN (RBG N)	0.697	0.408	-	0.722
Xiao et al. [94]	C-CNN + R-CNN (C2RNet)	0.695	-	0.675	-
Noise Net [92]	FRNN + SRM Filter	0.705	0.283	-	-
Late Fusion [92]	Fusion FRNN	0.681	0.397	-	-
Wu et al. [95]	Buster Net	-	-	0.759	-
Bi et al. [96]	RRU-Net	0.915	-	0.841	-
Biach et al. [119]	False-Unet	-	0.736	0.695	0.638
Hao et al. [122]	Vision Transformer	-	-	0.620	-

5.2. Multi-Modal Approach

Social media fake news consists of multiple entities like texts, images, videos, audio, links, etc. Sometimes, various objects are combined to propagate fake news, like text over images or text in comments with irrelevant images. Thus, detecting fake images based on the image has some gaps. The efficiency of fake news detection solely based on image analysis only sometimes yields very high accuracy. There are fair chances that the images used in fake news are real and untampered, but the text or audio content is either irrelevant or contains false information.

To overcome this problem, researchers have started applying a multi-modal approach. Besides the image, other content-based features are also considered for detection in the multi-modal approach. Multiple information is received from various streams at the end to classify so that the multi-modal architecture will require a fusion classifier. Some multi-modal approaches are shared below. A multi-modal framework using text and image has performed better than other multi-modal features [126,127].

To combine text and images, Yang et al. created the Text and Image information-based Convolutional Neural Network (TI-CNN) model [128]. By projecting the latent and explicit vectors into integrated vector space, learning the TI-CNN model is simultaneously based on image and text data. In addition to the natural features, the model uniquely employs two parallel CNNs to extract hidden features from visual and textual information. Latent and explicit vectors are projected into an integrated vector space to produce a new presentation of visuals and texts. Finally, the model recommends fusing visual and textual representations concurrently to detect faux news. Event Adversarial Neural Networks (EANN) were suggested by Wang et al. [129] to identify false news, gather features independent of the event, and support fake news detection on newly emerging events. The architecture consists of the multi-modal feature extractor, the false news detector, and the event discriminator. Generating visual and textual features from postings is the primary task of the multi-modal feature extractor. Predicting whether a message is true or false is the aim of the fake news detector. An event discriminator's task is to remove event-specific features while maintaining event-invariant features.

Sentiment-aware multi-modal Embedding (SAME) [130] incorporates users' hidden opinions from their comments into a single deep multi-modal embedding framework as a novel method to detect fake news. The many elements of fake news, such as the name of the publisher, user profiles, and text and image content, are managed by several networks. The adversarial method then educates semantically meaningful spaces for each data modality in the following phase. The model defines a special regularization loss in the final stage to reduce the distance between relevant pair embedding. The SpotFake framework was introduced by Singhal et al. [131] to eliminate sub-task dependencies like event discrimination. The authors' proposed solution detects fake news without depending on other subtasks or finding similarities between modalities. It utilizes both the visual and textual vectors of an article. Bidirectional Encoder Representations from Transformers (BERT) captures contextual text features. VGG-19, a model that has been pre-trained using the ImageNet dataset, was used to learn image vectors. However, many text articles must be used to verify the model. The end-to-end network known as Multimodal Variational Autoencoder (MVAE) was proposed by Khattar et al. [132]. The construction of an autoencoder model was a critical task. The three main modules of the suggested model are the encoder, decoder, and classifier. The model of the encoder component employs two data streams, text and visual, to train its many characteristics. It employs VGG19 to create picture features and Bi-directional LSTM to provide text features. The encoder output is sent to a decoder, which reconstructs and decodes the original post using analogous techniques. This multi-modal variational autoencoder is fused with a classifier for marking the post as true or false. The differential autoencoder using KL divergence loss learns probabilistic hidden variables by minimizing a bound on the marginal similarity of the observed data. Finally, the fake news classifier uses this multi-modal representation generated from the bimodal variational autoencoder to mark the article as real or false.

Zhou et al. [133] proposed the Similarity Aware Fake News (SAFE) framework. SAFE examines visual and textual information in news articles. The algorithm calculates the likelihood of erroneous reports using independent textual and visual learnings. To determine whether or not it is false in the end, it later takes into account both of these probabilities and the estimated similarity index between the text and visual content. In the first stage, an extended version of Text-CNN is used separately to generate and learn textual and visual vectors for news representation. The model also explores the generated vectors' correlation among the different processes. Then, visual and text data representations and their relevance factor are mutually learned and used to identify faux news. This model is advantageous in detecting fake news where the visual and text normally mismatch or the image is irrelevant to the text content. Chen, Cheng, and Shi [134] suggested a hybrid features and semantic reinforcement network (HFSRNet), an encoding and decoding-based network, for picture forgery detection. Long-short-term memory (LSTM) with resampling characteristics has been employed to record traces from the picture patches for discovering manipulation artifacts. The difference between unaltered and altered areas is further amplified by consolidating characteristics taken from spinning residual units. Then, to further include the spatial co-occurrence of these two modalities, the authors hybridize characteristics from them through a concatenation. A similar encoder and decoder-based approach was employed by Biach et al. [119]. Singh and Sharma [135] used efficientNet-B0 and RoBERTa as a multi-modal approach for detecting fake images using the image and text features.

Table 7 provides the "accuracy" comparison between different multi-modal methods. It is observed that multi-modal methods will have different accuracy based on the dataset. The results vary because the dataset's information may have more text than images or vice versa. SAFE has better results in MediaEval, whereas SpotFake works better on Weibo. The results of att-RNN in Table 7 are taken from Khattar et al., and the att-RNN literature is not discussed here. Also, to our best knowledge, we could not find a multi-modal approach combining image and network propagation data.

Table 7. Accuracy comparison of multi-modal approach.

Method	Framework	MediaEval	Weibo
Khattar et al. [129]	att-RNN	66.40	77.90
Wang et al. [129]	EANN	71.50	82.70
Cui et al. [130]	SAME	77.24	81.58
Singhal et al. [131]	SpotFake	77.77	89.23
Khattar et al. [132]	MVAE	74.50	82.40
Zhou et al. [133]	SAFE	87.40	83.80
Singh and Sharma [135]	EfficientNet + BERT	85.30	81.20

6. Evaluation Parameters and Datasets

Fake image detection is primarily a binary classification problem, where the assertion is made about whether the image is tampered with.

6.1. Evaluation Parameters

Besides standard evaluation parameters used in the classification problems, there are multiple other diverse evaluation parameters that researchers are using. This may be due to some of the measurement criteria concerning image features. The researcher usually needs to provide reasoning for choosing specific parameters. It becomes hard to compare the results of different models/proposals if standard parameters are not used. The performance of classification problems is usually evaluated through a confusion matrix (Table 8). The confusion matrix provides a visual representation of the results. Various evaluation parameters are derived based on the confusion matrix's data values (Table 9). Some of the prominent evaluation parameters are described below in Table 9. Some are derived through a confusion matrix, while others are graph-oriented.

Table 8. Confusion matrix.

Confusion Matrix		Actual Values	
		Positive	Negative
Predicted Value	Positive	True-Positive	False-Positive
	Negative	False-Negative	True-Negative

6.2. Datasets

Only some real-world benchmark datasets are available for fake images on social media platforms. However, some well-published text and propagation-based false news databases have been released and made available to the public. Good data sets are available for face detection, but fake social media images are more than artificial faces. We here provide the well-known multimedia datasets used in fake image detection. Some are not from social media platforms, but they provide suitable datasets of tampered images. Primary datasets used for fake images are presented in Table 10. The table also provides information about where these datasets can be accessed. Some are free to use, and some are shared on a paid basis.

Table 9. Evaluation parameters.

S. No.	Parameter Name	Formula	Description
1	Precision (P)	P = TP/(TP + FP)	Measure shown to correct model was in classifying positives.
2	Recall (R)	R = TP/(TP + FN)	Measures how many positives are missed by model. Also called sensitivity or Total Positive Rate (TPR).
3	Accuracy (A)	A = (TP + TN)/ (TP + TN + FP + FN)	Measures how accurately model classifies correctly.
4	F1 (F1)	F1 = 2(PXR)/(P + R)	Measures the harmonic mean of Precision and Recall.
5	False-Positive Rate	FPR = FP/FP + TN	Measure how many negatives are classified as positive. Probability of false alarms. Also called Fallout.
6	False-Negative Rate	FNR = FN/TP + FN	Measures miss rate.
7	Half-Total Error Rate (HTER)	HTER = FPR + FNR/2	Average of FPR and FNR.
8	ROC	Receiver Operating Characteristic plot is used to visualize the performance of a classifier. It is a two-dimensional curve for depicting the system's characteristics.	
9	AUROC	The area under ROC measures the entire area under the ROC curve. It is a total measure of performance across all possible classification thresholds.	
10	mPA	Mean Average Precision is the average of AP. AP is the area under the Precision-Recall curve.	

- BuzzFeed: This source contains data and analysis supporting the BuzzFeed News article, *"These Are 50 of the Biggest Fake News Hits on Facebook in 2017"*, published on 28 December 2017;
- CASIA: Natural color image repository with realistic tampering operations, available for the public for research;
- CelebA: It contains ten thousand celebrity identities, each with twenty images. There are two hundred thousand images in total;
- COCO: The Irnia Holidays—Copydays dataset contains a set of photos that are collected explicitly from personal holidays. Each photo has suffered three kinds of artificial attacks, JPEG, cropping, and "strong";
- Columbia: The original images in this dataset consist of 312 images from the CalPhotos collection and 10 captured using a digital camera. The data set consists of 1845 images;
- CoMoFod: This database contains a total of 13,520 forged images. These images are a set of 260 manipulated images. Of 260 images, 200 images are in a small category, and the rest are in a large category;
- FakeNewsNet: This repository contains fake news articles while traversing the fact-check websites PolitiFact and GossipCop. These articles are then explored over the web pages. The PolitiFact section has 447 real and 336 fake news articles with images, while the GossipCop section contains 16,767 real and 1650 fake articles;
- Fakkedit: The Fakeddit is the latest and largest multi-modal dataset from the real-world social networking website Reddit. It contains over 1 million fake textual news data and over 4 lakh multi-modal samples. The multi-modal samples have text and images. It has both two-way and six-way labeling. Two-way labeling is fake and real. As Reddit collects data from micro-sites like Twitter, Facebook, Instagram, and WhatsApp, this dataset has the largest diversified dataset. In the experiment, we have selected a two-way labeling of fake vs. real. As the images are from multiple platforms, it tests the framework's robustness;
- FNC: Kaggle fake news detection challenge dataset. It has content from 244 websites and includes 12,999 news stories collected from these websites;
- MediaEval: Comprises a total of 413 images, of which 193 cases are of real images, 218 cases are of fake images, and two cases are of altered videos. These images

are associated with 9404 fake and 6225 real tweets posted by 9025 and 5895 unique users, respectively;

- NIST Nimble: The Nimble 16 dataset has approximately 10,000 images with various types of tampering, including the images where anti-forensic algorithms were used to hide minor alterations;
- PGGAN: Consisting of 100K GAN-generated fake celebrity images at 1024 × 1024 resolution;
- PHEME: The rumors and hard facts made on Twitter amid breaking news are collected in this dataset. It contains rumors related to nine events; each is annotated with its veracity value, true, false, or unverified.
- Weibo: Comprises Sina Weibo data collected between 2012 and 2016 from the web and mobile platforms. The collection has domestic and international news.

Table 10. Datasets.

S. No	Data Set	Year	Type	Source	Real Images	Fake Images	Location	Accessed on
1	Buzzfeed	2018	Images, Text	Buzzfeed News	90	80	https://github.com/BuzzFeedNews/2017-12-fake-news-top-50	19 February 2019
2	CASIA1.0	2013	Images	Self made Database	800	921	http://forensics.idealtest.org/	16 March 2019
3	CASIA2.0	2013	Images	Self made Database	7200	5123	http://forensics.idealtest.org/	16 March 2019
4	CelebA	2015	Images	Self made Database	2,00,000	–	https://github.com/tkarras/progressive_growing_of_gans	6 August 2018
5	COCO	2008	Images	Flickr	500	229	http://lear.inrialpes.fr/people/jegou/data.php	14 February 2016
6	COLUMBIA	2004	Images	CalPhotos	933	912	http://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/dlform.html	28 April 2019
7	CoMoFoD	2004	Images	Self-made Dataset (Tralic and Grgic)	260	13,520	https://www.vcl.fer.hr/comofod/comofod.html	4 June 2020
8	FAKENEWSNET	2018	Image, Text	Twitter	447	336	https://github.com/KaiDMML/FakeNewsNet/tree/master/dataset	7 September 2020
9	FNC	2018	Images, Text	Kaggle	–	–	https://www.kaggle.com/c/fake-news/data	7 September 2020
10	MediaEval 2015	2015	Images, Text	Twitter	193	218	https://github.com/MKLab-ITI/image-verification-corpus/tree/master/mediaeval2015	16 March 2019
11	NIST Nimble 16	2017	Images	Self-made database (NIST)	–	10,000	https://www.nist.gov/itl/iad/mig/media-forensics-challenge	22 August 2019
12	PGGAN	2016	Images	GAN generated	–	1,00,000	https://github.com/tkarras/progressive_growing_of_gans	6 July 2019
13	PHEME	2016	Text	Twitter	–	–	https://figshare.com/articles/PHEME_dataset_for_Rumour_DetectionandVeracityClassification/6392078	5 September 2019
14	WEIBO	2016	Images, Text	Sina Weibo	3774	1363	https://drive.google.com/file/d/14LXJ5FCEcN2QrVWHYkKEYDpzluT2XNhw/view	26 December 2019

7. Limitations and Challenges

In the above portion, this survey presented numerous image characteristics and current visual-based techniques for successfully detecting fake images. Although there is a lot of research and models for detecting fake images on social media platforms, some specific challenges still need to be considered. We present them below.

7.1. Labeled Dataset

While existing datasets are available for fake image detection, there is a significant limitation of labeled datasets in deep learning methods. The datasets we have today are not always up-to-date with the latest real-world events, and they often focus on only a limited number of situations [51,129,131]. For instance, datasets that contain fake news data from platforms like Twitter, Weibo, or News websites may not cover the full scope of rapidly evolving multimedia information. This limitation hampers the progress of research in this field. Authors believe that creating new and continuously updated datasets is essential to address this issue. Some recent datasets, like Fakeddit and GSR-Net [127,136], from Reddit, show promise, as they come from a web aggregator that captures a wide range of content. However, the pace of advancement in image tampering techniques requires constant adaptation. Furthermore, deep learning models for fake image detection need regular retraining to stay effective. If they are not updated with the latest data, they can become too specialized and make incorrect predictions, a phenomenon known as overfitting. This presents a significant challenge in the research field.

7.2. Cross-Platform Training

Our current challenge is that existing datasets for fake image detection are tailored to specific social media platforms. Each platform, such as Facebook, Twitter, Instagram, WhatsApp, TikTok, and Weibo, has a unique style, content, and the way information spreads. Consequently, a deep learning model trained solely on one dataset may not perform accurately when dealing with content from other platforms. This limitation poses a significant hurdle in the field of fake image detection. Adopting a more versatile and adaptable approach to address this limitation is essential. We should consider training deep neural networks using a combination of data from various datasets representing different platforms. This multi-source training can help create a more generic model that recognizes fake images across various sources. Additionally, researchers should explore multi-modal approaches [132,133] to account for variations in content and style among different platforms, ultimately improving the model's accuracy across diverse social media sources. By embracing a more holistic and comprehensive training approach, we can significantly overcome the limitations associated with cross-platform fake image detection.

7.3. Satire vs. Fake

Besides being the highest propagator of false news, social media platforms are the biggest platform for sharing satirical or sarcastic views about any topic. Everyone is eager to share their perspective regarding any issue, resulting in images of satire, sarcasm, and jokes. Thus, any image manipulated or tampered with this non-malicious intention will also be marked as fake. The models and proposed methods do not discriminate between fake images posted for misinformation and tampered images posted as jokes. Both are treated as fake. Sharma et al. [137] have created models to weed sarcastic tweets out. Research work can be extended to learn the intrinsic feature differences between satire and fake. More work is needed in detecting sarcastic images from fake images.

7.4. Interpretability

Interpretability is a significant challenge in detecting fake images using deep learning methods. In artificial intelligence, achieving explainability or interpretability is a common and critical issue. This need for interpretability becomes even more pressing as deep learning models, with their inherent complexity, excel in classifying images as fake with impressive accuracy but fall short in providing clear explanations for their decisions. While tools like Grad-Cam and LIME offer insights into image differences, their interpretation still requires domain expertise. This challenge remains a significant obstacle if we intend to release reliable solutions to the public. From the author's perspective, addressing the issue of interpretability in fake image detection is vital for building trust in AI systems and ensuring that users can make informed judgments about the authenticity of visual content.

7.5. Audio Splicing

There is much research on detecting fake news via text analysis or fake images. However, research should also be extended concerning fake audio shared with fake images/videos posted over social media. Recently, many fake images shared did not have text; instead, manipulated audio is shared to convey long messages and gather more attention. Fake audio detection will yield good research and help mitigate fake news during elections or pandemics wherein such fake audio is majorly shared.

7.6. Multi-Modal Detection

Detecting fake news solely through fake image detection has limitations. Sometimes, images are unrelated to the text, leading to misinformation, and manipulations may remain undetected [126]. Additionally, mismatched images from different timelines or locations can blend seamlessly with text, making them difficult to spot. Fake news creators adapt to varying text styles and social contexts, making visual and text combinations unreliable. To address these challenges, a comprehensive approach is needed. Alongside fake image analysis, investigating diffusion network patterns is crucial. Analyzing the propagation of fake content across networks can help uncover subtle cases. Multi-modal research, encompassing image and network analysis, can bridge gaps in detection. This holistic strategy, utilizing cues from image and network patterns, enhances our ability to identify complex instances of fake news dissemination.

7.7. Deepfakes

Deepfake images and videos are increasing day by day. Better generators have created natural, realistic images and videos that are very tough to identify. Though much research has been performed to detect them, no generic model exists. Some are specific to videos, and some are specific to images. They are still based on facial expression methods compared to natural expressions. More research to develop a generic model without intrinsic features can be worked on and researched further.

7.8. Active Methods

Although active methods are of little use considering the humongous amount of data from multiple sources across different platforms, they can surely think of some active method applications now with growing technology and social media companies working on fake news detection. Research can be performed on creating a scalable architecture using AI applications or Blockchain architecture to produce digital signatures over images. Salim et al. [138] have suggested some cryptography-based methods that cannot be scaled with proper cloud architecture. These architectures can be implemented by social media firms and limit the spread of fake news.

8. Conclusions

This research paper on recognizing and reducing fake Images on social media platforms delves into critical issues in this domain. Fake images on social media have become a pressing concern, prompting social media platforms to combat this issue. This survey assesses various techniques for identifying fake pictures on these platforms. This paper discusses different tampering methods for creating fake images, from conventional to recent Generative Adversarial Network (GAN)-generated manipulations. It highlights the fake image detection process, reviewing multiple detection methods, including those using handcrafted features and neural network-based approaches. Performance comparisons for these methods and their respective advantages and limitations are presented. Forensic techniques are noted for accuracy but are less efficient in detecting fake images subjected to multiple manipulations. While they can localize tampered areas, they may not address distinct alterations common in fake photos shared on social media. Semantic and statistical features also have their limitations. This paper emphasizes recent neural network-based detection methods based on Convolutional Neural Network (CNN) architectures. CNN models prove highly resistant to multiple manipulations, effectively identifying fake photos without pre- or post-processing limitations. CNNs are also efficient in detecting deepfake images and videos. However, the vision transformer has shown better results on GAN-generated deepfakes. Multi-modal approaches are discussed, combining visual and text content results, providing enhanced authenticity prediction. Evaluation parameters and dataset information are shared. The research explores various methods, from traditional forensics to deep learning, concluding that deep learning methods outperform others in identifying fake images on social media. However, they depend on extensive labeled datasets, which can be challenging to obtain for fake images. Cross-platform and interpretability issues are highlighted, with multi-modal approaches offering improved accuracy.

This paper calls for further research in multi-modal combinations and the creation of substantial, real-time labeled datasets to support the development of more efficient, generic models. In summary, this study underscores the significance of deep learning in fake image detection while acknowledging data availability and interpretability challenges, advocating for a comprehensive multi-modal approach and more extensive datasets to advance research in this critical area.

Author Contributions: Conceptualization, D.K.S. and B.S.; validation, D.K.S., B.S., S.A., L.G., C.K. and K.-H.J.; formal analysis, D.K.S., B.S., S.A., L.G., C.K. and K.-H.J.; writing—original draft preparation, D.K.S., B.S. and S.A.; writing—review and editing D.K.S., B.S., S.A., L.G. and K.-H.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Brain Pool program funded by the Ministry of Science and ICT through the National Research Foundation of Korea (2019H1D3A1A01101687) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1I1A3049788).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this paper are publicly available, and their links are provided in the reference section.

Acknowledgments: We thank the anonymous reviewers for their valuable suggestions that improved the quality of this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Barthel, M.; Mitchell, A.; Holcomb, J. Many Americans Believe Fake News Is Sowing Confusion. Available online: https://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/ (accessed on 2 May 2020).
2. CIGI-Ipsos Global Survey on Internet Security and Trust. 2019. Available online: https://www.cigionline.org/internet-survey-2019 (accessed on 15 January 2021).
3. Silverman, C. This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook. Available online: https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook#.emA15rzd0 (accessed on 2 May 2020).
4. Gowen, A. As Mob Lynchings Fueled by Whatsapp Messages Sweep India, Authorities Struggle to Combat Fake News. Available online: https://www.washingtonpost.com/world/asia_pacific/as-mob-lynchings-fueled-by-whatsapp-sweep-india-authorities-struggle-to-combat-fake-news/2018/07/02/683a1578-7bba-11e8-ac4e-421ef7165923_story.html (accessed on 2 May 2020).
5. Kudrati, M. This Picture of Donald Trump Endorsing PM Modi Is a Hoax. Available online: https://www.boomlive.in/this-picture-of-donald-trump-endorsing-pm-modi-is-a-hoax/ (accessed on 10 July 2020).

6. Baynes, C. Coronavirus: Patients Refusing Treatment Because of Fake News on Social Media, NHS Staff Warn. Available online: https://www.independent.co.uk/news/uk/home-news/coronavirus-fake-news-conspiracy-theories-antivax-5g-facebook-twitter-a9549831.html (accessed on 15 July 2020).
7. Stoll, J. Reading Fake News about the Coronavirus in Norway 2020, by Source. Available online: https://www.statista.com/statistics/1108710/reading-fake-news-about-the-coronavirus-in-norway-by-source/ (accessed on 2 May 2020).
8. Unnikrishnan, D. Photo of PM Narendra Modi Bowing to Xi Jinping Is Morphed. Available online: https://www.boomlive.in/fake-news/photo-of-pm-narendra-modi-bowing-to-xi-jinping-is-morphed-8579 (accessed on 10 July 2020).
9. Amsberry, C. Alteration of Photos Raise Host of Legal, Ethical Issues. *Wall Str. J.* **1989**, *1*, 26–89.
10. Jaffe, J. Dubya, Willya Turn the Book Over. Available online: https://www.wired.com/2002/11/dubya-willya-turn-the-book-over/ (accessed on 10 July 2020).
11. Mishra, M.; Adhikary, M.C. Digital Image Tamper Detection Techniques—A Comprehensive Study. *arXiv* **2013**, arXiv:1306.6737.
12. Mandankandy, A.A. Image forgery and its detection: A survey. In Proceedings of the International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 17–18 March 2017. [CrossRef]
13. Parikh, S.B.; Atrey, P.K. Media-Rich Fake News Detection: A Survey. In Proceedings of the IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, USA, 10–12 April 2018; pp. 436–441. [CrossRef]
14. Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Morales, A.; Ortega-Garcia, J. Deepfakes and beyond: A Survey of face manipulation and fake detection. *Inf. Fusion* **2020**, *64*, 131–148. [CrossRef]
15. Wang, X.-Y.; Wang, C.; Wang, L.; Yang, H.-Y.; Niu, P.-P. Robust and effective multiple copy-move forgeries detection and localization. *Pattern Anal. Appl.* **2021**, *24*, 1025–1046. [CrossRef]
16. Alamro, L.; Nooraini, Y. Copy-move forgery detection using integrated DWT and SURF. *J. Telecommun. Electron. Comput. Eng. (JTEC)* **2017**, *9*, 67–71.
17. Jwaid, M.F.; Baraskar, T.N. Study and analysis of copy-move & splicing image forgery detection techniques. In Proceedings of the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 10–11 February 2017; pp. 697–702. [CrossRef]
18. Huh, M.; Liu, A.; Owens, A.; Efros, A.A. Fighting Fake News: Image Splice Detection via Learned Self-Consistency. In Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 106–124. [CrossRef]
19. Vadsola, M. The Math behind GANs (Generative Adversarial Networks). Available online: https://towardsdatascience.com/the-math-behind-gans-generative-adversarial-networks-3828f3469d9c (accessed on 10 May 2020).
20. Vincent, J. Facebook's Problems Moderating Deepfakes Will Only Get Worse in 2020. Available online: https://www.theverge.com/2020/1/15/21067220/deepfake-moderation-apps-tools-2020-facebook-reddit-social-media (accessed on 10 July 2020).
21. Warif, N.B.A.; Idris, M.Y.I.; Wahab, A.W.A.; Ismail, N.-S.N.; Salleh, R. A comprehensive evaluation procedure for copy-move forgery detection methods: Results from a systematic review. *Multimed. Tools Appl.* **2022**, *81*, 15171–15203. [CrossRef]
22. Fridrich, J.; Soukal, D.; Lukas, J. Detection of copy-move forgery in digital images. In Proceedings of the Digital Forensic Research Workshop, Cleveland, OH, USA, 6–8 August 2003; pp. 55–61.
23. Popescu, A.C.; Farid, H. *Exposing Digital Forgeries by Detecting Duplicated Image Regions*; Technical Report TR2004-515; Department of Computer Science, Dartmouth College: Hanover, NH, USA, 2004.
24. Li, G.; Wu, Q.; Tu, D.; Sun, S. A Sorted Neighborhood Approach for Detecting Duplicated Regions in Image Forgeries Based on DWT and SVD. In Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, Beijing, China, 2–5 July 2007; pp. 1750–1753. [CrossRef]
25. Bayram, S.; Sencar, H.T.; Memon, N. An efficient and robust method for detecting copy-move forgery. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 1053–1056. [CrossRef]
26. Gul, G.; Avcibas, I.; Kurugollu, F. SVD based image manipulation detection. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 1765–1768. [CrossRef]
27. Huang, Y.; Lu, W.; Sun, W.; Long, D. Improved DCT-based detection of copy-move forgery in images. *Forensic Sci. Int.* **2011**, *206*, 178–184. [CrossRef]
28. Li, L.; Li, S.; Zhu, H.; Chu, S.-C.; Roddick, J.F.; Pan, J.-S. An efficient scheme for detecting copy-move forged images by local binary patterns. *J. Inf. Hiding Multimed. Signal Process.* **2013**, *4*, 46–56.
29. Lee, J.-C.; Chang, C.-P.; Chen, W.-K. Detection of copy–move image forgery using histogram of orientated gradients. *Inf. Sci.* **2015**, *321*, 250–262. [CrossRef]
30. Hussain, M.; Qasem, S.; Bebis, G.; Muhammad, G.; Aboalsamh, H.; Mathkour, H. Evaluation of Image Forgery Detection Using Multi-Scale Weber Local Descriptors. *Int. J. Artif. Intell. Tools* **2015**, *24*, 1540016. [CrossRef]
31. Mahmood, T.; Nawaz, T.; Irtaza, A.; Ashraf, R.; Shah, M.; Mahmood, M.T. Copy-Move Forgery Detection Technique for Forensic Analysis in Digital Images. *Math. Probl. Eng.* **2016**, *2016*, 8713202. [CrossRef]
32. Chen, B.; Yu, M.; Su, Q.; Shim, H.J.; Shi, Y.-Q. Fractional Quaternion Zernike Moments for Robust Color Image Copy-Move Forgery Detection. *IEEE Access* **2018**, *6*, 56637–56646. [CrossRef]
33. Dixit, A.; Bag, S. A fast technique to detect copy-move image forgery with reflection and non-affine transformation attacks. *Expert Syst. Appl.* **2021**, *182*, 115282. [CrossRef]

34. Rani, A.; Jain, A.; Kumar, M. Identification of copy-move and splicing based forgeries using advanced SURF and revised template matching. *Multimed. Tools Appl.* **2021**, *80*, 23877–23898. [CrossRef]
35. Tanaka, M.; Shiota, S.; Kiya, H. A Detection Method of Operated Fake-Images Using Robust Hashing. *J. Imaging* **2021**, *7*, 134. [CrossRef] [PubMed]
36. Yang, J.; Liang, Z.; Gan, Y.; Zhong, J. A novel copy-move forgery detection algorithm via two-stage filtering. *Digit. Signal Process.* **2021**, *113*, 103032. [CrossRef]
37. Tahaoglu, G.; Ulutas, G.; Ustubioglu, B.; Ulutas, M.; Nabiyev, V.V. Ciratefi based copy move forgery detection on digital images. *Multimed. Tools Appl.* **2022**, *81*, 22867–22902. [CrossRef]
38. Uma, S.; Sathya, P.D. Copy-move forgery detection of digital images using football game optimization. *Aust. J. Forensic Sci.* **2020**, *54*, 258–279. [CrossRef]
39. Gan, Y.; Zhong, J.; Vong, C. A Novel Copy-Move Forgery Detection Algorithm via Feature Label Matching and Hierarchical Segmentation Filtering. *Inf. Process. Manag.* **2021**, *59*, 102783. [CrossRef]
40. Ng, T.; Chang, S. A model for image splicing. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Singapore, 24–27 October 2004; pp. 1169–1172.
41. Popescu, A.; Farid, H. Exposing digital forgeries in color filter array interpolated images. *IEEE Trans. Signal Process.* **2005**, *53*, 3948–3959. [CrossRef]
42. Chen, W.; Shi, Y.Q.; Su, W. Image splicing detection using 2-D phase congruency and statistical moments of characteristic function. In *Security, Steganography, and Watermarking of Multimedia Contents IX*; SPIE: Bellingham, WA, USA, 2007; Volume 6505, pp. 281–288. [CrossRef]
43. Wang, W.; Dong, J.; Tan, T. Effective image splicing detection based on image chroma. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 1257–1260. [CrossRef]
44. Zhao, X.; Li, J.; Li, S.; Wang, S. Detecting Digital Image Splicing in Chroma Spaces. In Proceedings of the Digital Watermarking: 9th International Workshop, IWDW 2010, Seoul, Republic of Korea, 1–3 October 2010; pp. 12–22. [CrossRef]
45. Liu, Q.; Cao, X.; Deng, C.; Guo, X. Identifying Image Composites Through Shadow Matte Consistency. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 1111–1122. [CrossRef]
46. Ferrara, P.; Bianchi, T.; De Rosa, A.; Piva, A. Image Forgery Localization via Fine-Grained Analysis of CFA Artifacts. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1566–1577. [CrossRef]
47. He, Z.; Lu, W.; Sun, W.; Huang, J. Digital image splicing detection based on Markov features in DCT and DWT domain. *Pattern Recognit.* **2012**, *45*, 4292–4299. [CrossRef]
48. Mazumdar, A.; Bora, P.K. Exposing splicing forgeries in digital images through dichromatic plane histogram discrepancies. In Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing, Guwahati, India, 18–22 December 2016; Volume 62, pp. 1–8. [CrossRef]
49. Moghaddasi, Z.; Jalab, H.A.; Noor, R.M. Image splicing detection using singular value decomposition. In Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing, Cambridge, UK, 22–23 March 2017; Volume 140, pp. 1–5. [CrossRef]
50. Sheng, H.; Shen, X.; Lyu, Y.; Shi, Z.; Ma, S. Image splicing detection based on Markov features in discrete octonion cosine transform domain. *IET Image Process.* **2018**, *12*, 1815–1823. [CrossRef]
51. Jaiswal, A.K.; Srivastava, R. A technique for image splicing detection using hybrid feature set. *Multimed. Tools Appl.* **2020**, *79*, 11837–11860. [CrossRef]
52. Itier, V.; Strauss, O.; Morel, L.; Puech, W. Color noise correlation-based splicing detection for image forensics. *Multimed. Tools Appl.* **2021**, *80*, 13215–13233. [CrossRef]
53. Monika; Bansal, D.; Passi, A. Image Forensic Investigation Using Discrete Cosine Transform-Based Approach. *Wirel. Pers. Commun.* **2021**, *119*, 3241–3253. [CrossRef]
54. Niyishaka, P.; Bhagvati, C. Image splicing detection technique based on Illumination-Reflectance model and LBP. *Multimed. Tools Appl.* **2020**, *80*, 2161–2175. [CrossRef]
55. Jalab, H.A.; Alqarni, M.A.; Ibrahim, R.W.; Almazroi, A.A. A novel pixel's fractional mean-based image enhancement algorithm for better image splicing detection. *J. King Saud Univ.-Sci.* **2022**, *34*, 101805. [CrossRef]
56. Agrawal, S.; Kumar, P.; Seth, S.; Parag, T.; Singh, M.; Babu, V. SISL: Self-Supervised Image Signature Learning for Splicing Detection & Localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Work-Shops, New Orleans, LA, USA, 19–20 June 2022; pp. 22–32.
57. Sun, Y.; Ni, R.; Zhao, Y. ET: Edge-Enhanced Transformer for Image Splicing Detection. *IEEE Signal Process. Lett.* **2022**, *29*, 1232–1236. [CrossRef]
58. Popescu, A.; Farid, H. Exposing digital forgeries by detecting traces of resampling. *IEEE Trans. Signal Process.* **2005**, *53*, 758–767. [CrossRef]
59. Fillion, C.; Sharma, G. Detecting content adaptive scaling of images for forensic applications. In *Media Forensics and Security II*; SPIE: Bellingham, WA, USA, 2010; Volume 75410. [CrossRef]
60. Mahalakshmi, S.D.; Vijayalakshmi, K.; Priyadharsini, S. Digital image forgery detection and estimation by exploring basic image manipulations. *Digit. Investig.* **2012**, *8*, 215–225. [CrossRef]

61. Niu, P.; Wang, C.; Chen, W.; Yang, H.; Wang, X. Fast and effective Keypoint-based image copy-move forgery detection using complex-valued moment invariants. *J. Vis. Commun. Image Represent.* **2021**, *77*, 103068. [CrossRef]
62. Fan, Z.; de Queiroz, R. Identification of bitmap compression history: JPEG detection and quantizer estimation. *IEEE Trans. Image Process.* **2003**, *12*, 230–235. [CrossRef] [PubMed]
63. Krawetz, N. A Picture's Worth... Hacker Factor Solutions. 2007. Available online: https://www.hackerfactor.com/papers (accessed on 10 May 2020).
64. Zhang, J.; Wang, H.; Su, Y. Detection of Double-Compression in JPEG2000 Images. In Proceedings of the 2008 Second International Symposium on Intelligent Information Technology Application, Shanghai, China, 21–22 December 2008; Volume 1, pp. 418–421. [CrossRef]
65. Lin, Z.; He, J.; Tang, X.; Tang, C.-K. Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. *Pattern Recognit.* **2009**, *42*, 2492–2501. [CrossRef]
66. Kwon, M.-J.; Nam, S.-H.; Yu, I.-J.; Lee, H.-K.; Kim, C. Learning JPEG Compression Artifacts for Image Manipulation Detection and Localization. *Int. J. Comput. Vis.* **2022**, *130*, 1875–1895. [CrossRef]
67. McCloskey, S.; Albright, M. Detecting Gan-Generated Imagery Using Color Cues. *arXiv* **2018**, arXiv:1812.08247.
68. Nataraj, L.; Mohammed, T.M.; Manjunath, B.S.; Chandrasekaran, S.; Flenner, A.; Bappy, J.H.; Roy-Chowdhury, A. Detecting GAN generated Fake Images using Co-occurrence Matrices. *Electron. Imaging* **2019**, *2019*, 532–541. [CrossRef]
69. Matern, F.; Riess, C.; Stamminger, M. Exploiting visual artifacts to expose deepfakes and face manipulations. In Proceedings of the 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), Waikoloa Village, HI, USA, 7–11 June 2019; pp. 83–92.
70. Li, Y.; Chang, M.-C.; Lyu, S. In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. In Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, China, 11–13 December 2018; pp. 1–7. [CrossRef]
71. Zhang, W.; Zhao, C.; Li, Y. A Novel Counterfeit Feature Extraction Technique for Exposing Face-Swap Images Based on Deep Learning and Error Level Analysis. *Entropy* **2020**, *22*, 249. [CrossRef]
72. Shang, Z.; Xie, H.; Zha, Z.; Yu, L.; Li, Y.; Zhang, Y. PRRNet: Pixel-Region relation network for face forgery detection. *Pattern Recognit.* **2021**, *116*, 107950. [CrossRef]
73. Sunstein, C.R. *On Rumors: How Falsehoods Spread, Why We Believe Them, and What Can Be Done*; Princeton University Press: Princeton, NJ, USA, 2014.
74. Jin, Z.; Cao, J.; Luo, J.; Zhang, Y. Image credibility analysis with effective domain transferred deep networks. *arXiv* **2016**, arXiv:1611.05328.
75. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake News Detection on Social Media. *ACM SIGKDD Explor. Newsl.* **2017**, *19*, 22–36. [CrossRef]
76. Ghanem, B.; Ponzetto, S.P.; Rosso, P. FacTweet: Profiling Fake News Twitter Accounts. In Proceedings of the International Conference on Statistical Language and Speech Processing, Cardiff, UK, 14–16 October 2020; pp. 35–45.
77. Zhang, Y.; Tan, Q.; Qi, S.; Xue, M. PRNU-based Image Forgery Localization with Deep Multi-scale Fusion. *ACM Trans. Multimed. Comput. Commun. Appl.* **2023**, *19*, 67. [CrossRef]
78. Xie, X.; Liu, Y.; de Rijke, M.; He, J.; Zhang, M.; Ma, S. Why People Search for Images using Web Search Engines. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Los Angeles, CA, USA, 5–9 February 2018; pp. 655–663. [CrossRef]
79. Xie, X.; Mao, J.; Liu, Y.; de Rijke, M.; Shao, Y.; Ye, Z.; Zhang, M.; Ma, S. Grid-based Evaluation Metrics for Web Image Search. In Proceedings of the The World Wide Web Conference (WWW 2019), San Francisco, CA, USA, 13–17 May 2019; pp. 2103–2114. [CrossRef]
80. Gaikwad, M.; Hoeber, O. An Interactive Image Retrieval Approach to Searching for Images on Social Media. In Proceedings of the Conference on Human Information Interaction and Retrieval (CHIIR, 2019), Glasgow, UK, 10–14 March 2019. [CrossRef]
81. Vishwakarma, D.K.; Varshney, D.; Yadav, A. Detection and veracity analysis of fake news via scrapping and authenticating the web search. *Cogn. Syst. Res.* **2019**, *58*, 217–229. [CrossRef]
82. Gupta, A.; Lamba, H.; Kumaraguru, P.; Joshi, A. Faking Sandy: Characterizing and identifying fake images on Twitter during Hurricane Sandy. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 729–736. [CrossRef]
83. Huang, Q.; Zhou, C.; Wu, J.; Liu, L.; Wang, B. Deep spatial–temporal structure learning for rumor detection on Twitter. *Neural Comput. Appl.* **2020**, *35*, 12995–13005. [CrossRef]
84. Chen, Y.; Retraint, F.; Qiao, T. Image splicing forgery detection using simplified generalized noise model. *Signal Process. Image Commun.* **2022**, *107*, 116785. [CrossRef]
85. Jin, Z.; Cao, J.; Zhang, Y.; Zhou, J.; Tian, Q. Novel Visual and Statistical Image Features for Microblogs News Verification. *IEEE Trans. Multimed.* **2016**, *19*, 598–608. [CrossRef]
86. Xu, Z.; Li, S.; Deng, W. Learning temporal features using LSTM-CNN architecture for face anti-spoofing. In Proceedings of the 3rd Asian Conference on Pattern Recognition, Kuala Lumpur, Malaysia, 3–6 November 2015; pp. 141–145.
87. Bayar, B.; Stamm, M.C. A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, Vigo, Spain, 20–22 June 2016; pp. 5–10.

88. Rao, Y.; Ni, J. A deep learning approach to detection of splicing and copy-move forgeries in images. In Proceedings of the 2016 IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, United Arab Emirates, 4–7 December 2016; pp. 1–6. [CrossRef]
89. Rao, Y.; Ni, J.; Xie, H. Multi-semantic CRF-based attention model for image forgery detection and localization. *Signal Process.* **2021**, *183*, 108051. [CrossRef]
90. Salloum, R.; Ren, Y.; Kuo, C.-C.J. Image Splicing Localization using a Multi-task Fully Convolutional Network (MFCN). *J. Vis. Commun. Image Represent.* **2018**, *51*, 201–209. [CrossRef]
91. Bappy, J.H.; Roy-Chowdhury, A.K.; Bunk, J.; Nataraj, L.; Manjunath, B. Exploiting Spatial Structure for Localizing Manipulated Image Regions. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017. [CrossRef]
92. Zhou, P.; Han, X.; Morariu, V.I.; Davis, L.S. Learning Rich Features for Image Manipulation Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1053–1061. [CrossRef]
93. Rehman, Y.A.U.; Po, L.M.; Liu, M. LiveNet: Improving features generalization for face liveness detection using convolution neural networks. *Expert Syst. Appl.* **2018**, *108*, 159–169. [CrossRef]
94. Xiao, B.; Wei, Y.; Bi, X.; Li, W.; Ma, J. Image splicing forgery detection combining coarse to refined convolutional neural network and adaptive clustering. *Inf. Sci.* **2019**, *511*, 172–191. [CrossRef]
95. Wu, Y.; Abd-Almageed, W.; Natarajan, P. BusterNet: Detecting Copy-Move Image Forgery with Source/Target Localization. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 170–186. [CrossRef]
96. Bi, X.; Wei, Y.; Xiao, B.; Li, W. RRU-Net: The Ringed Residual U-Net for Image Splicing Forgery Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 30–39. [CrossRef]
97. Liu, B.; Pun, C.-M. Exposing splicing forgery in realistic scenes using deep fusion network. *Inf. Sci.* **2020**, *526*, 133–150. [CrossRef]
98. Abhishek; Jindal, N. Copy move and splicing forgery detection using deep convolution neural network, and semantic segmentation. *Multimed. Tools Appl.* **2021**, *80*, 3571–3599. [CrossRef]
99. Hosny, K.M.; Mortda, A.M.; Fouda, M.M.; Lashin, N.A. An Efficient CNN Model to Detect Copy-Move Image Forgery. *IEEE Access* **2022**, *10*, 48622–48632. [CrossRef]
100. Elaskily, M.A.; Alkinani, M.H.; Sedik, A.; Dessouky, M.M. Deep learning based algorithm (ConvLSTM) for Copy Move Forgery Detection. *J. Intell. Fuzzy Syst.* **2021**, *40*, 4385–4405. [CrossRef]
101. Koul, S.; Kumar, M.; Khurana, S.S.; Mushtaq, F.; Kumar, K. An efficient approach for copy-move image forgery detection using convolution neural network. *Multimed. Tools Appl.* **2022**, *81*, 11259–11277. [CrossRef]
102. Hsu, C.-C.; Zhuang, Y.-X.; Lee, C.-Y. Deep Fake Image Detection Based on Pairwise Learning. *Appl. Sci.* **2020**, *10*, 370. [CrossRef]
103. Jeon, H.; Bang, Y.; Woo, S.S. FDFtNet: Facing Off Fake Images Using Fake Detection Fine-Tuning Network. *IFIP Adv. Inf. Commun. Technol.* **2020**, *580*, 416–430. [CrossRef]
104. Wang, S.-Y.; Wang, O.; Zhang, R.; Owens, A.; Efros, A.A. CNN-Generated Images are Surprisingly Easy to Spot. . . for Now. *arXiv* **2020**, arXiv:1912.11035.
105. Neves, J.C.; Tolosana, R.; Vera-Rodriguez, R.; Lopes, V.; Proença, H.P.; Fierrez, J. GANprintR: Improved Fakes and Evaluation of the State of the Art in Face Manipulation Detection. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 1038–1048. [CrossRef]
106. Arora, T.; Soni, R. Arora, T.; Soni, R. A review of techniques to detect the GAN-generated fake images. In *Generative Adversarial Networks for Image-to-Image Translation*; Academic Press: Cambridge, MA, USA, 2021; pp. 125–159. [CrossRef]
107. Yang, J.; Xiao, S.; Li, A.; Lan, G.; Wang, H. Detecting fake images by identifying potential texture difference. *Futur. Gener. Comput. Syst.* **2021**, *125*, 127–135. [CrossRef]
108. Kwon, M.-J.; Yu, I.-J.; Nam, S.-H.; Lee, H.-K. CAT-Net: Compression Artifact Tracing Network for Detection and Localization of Image Splicing. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 375–384. [CrossRef]
109. Meena, K.B.; Tyagi, V. A Deep Learning based Method for Image Splicing Detection. *J. Phys. Conf. Ser.* **2021**, *1714*, 012038. [CrossRef]
110. Jaiswal, A.K.; Srivastava, R. Detection of Copy-Move Forgery in Digital Image Using Multi-scale, Multi-stage Deep Learning Model. *Neural Process. Lett.* **2022**, *54*, 75–100. [CrossRef]
111. Zhuo, L.; Tan, S.; Li, B.; Huang, J. Self-Adversarial Training Incorporating Forgery Attention for Image Forgery Localization. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 819–834. [CrossRef]
112. Wu, H.; Zhou, J.; Tian, J.; Liu, J.; Qiao, Y. Robust Image Forgery Detection Against Transmission Over Online Social Networks. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 443–456. [CrossRef]
113. Tyagi, S.; Yadav, D. MiniNet: A concise CNN for image forgery detection. *Evol. Syst.* **2022**, *14*, 545–556. [CrossRef]
114. Ali, S.S.; Ganapathi, I.I.; Vu, N.-S.; Werghi, N. Image Forgery Localization using Image Patches and Deep Learning. In Proceedings of the 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, 23–24 April 2022; pp. 583–588. [CrossRef]
115. Singh, B.; Sharma, D.K. SiteForge: Detecting and localizing forged images on microblogging platforms using deep convolutional neural network. *Comput. Ind. Eng.* **2021**, *162*, 107733. [CrossRef]

116. Wu, Y.; AbdAlmageed, W.; Natarajan, P. ManTra-Net: Manipulation Tracing Network for Detection and Localization of Image Forgeries with Anomalous Features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9535–9544. [CrossRef]
117. Hu, X.; Zhang, Z.; Jiang, Z.; Chaudhuri, S.; Yang, Z.; Nevatia, R. SPAN: Spatial Pyramid Attention Network for Image Manipulation Localization. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; pp. 312–328. [CrossRef]
118. Zhuang, P.; Li, H.; Tan, S.; Li, B.; Huang, J. Image Tampering Localization Using a Dense Fully Convolutional Network. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 2986–2999. [CrossRef]
119. El Biach, F.Z.; Iala, I.; Laanaya, H.; Minaoui, K. Encoder-decoder based convolutional neural networks for image forgery detection. *Multimed. Tools Appl.* **2022**, *81*, 22611–22628. [CrossRef]
120. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Comput. Surv.* **2022**, *54*, 1–41. [CrossRef]
121. Ganguly, S.; Ganguly, A.; Mohiuddin, S.; Malakar, S.; Sarkar, R. ViXNet: Vision Transformer with Xception Network for deepfakes based video and image forgery detection. *Expert Syst. Appl.* **2022**, *210*, 118423. [CrossRef]
122. Hao, J.; Zhang, Z.; Yang, S.; Xie, D.; Pu, S. TransForensics: Image Forgery Localization with Dense Self-Attention. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 10–17 October 2021; pp. 15035–15044. [CrossRef]
123. Arshed, M.A.; Alwadain, A.; Ali, R.F.; Mumtaz, S.; Ibrahim, M.; Muneer, A. Unmasking Deception: Empowering Deepfake Detection with Vision Transformer Network. *Mathematics* **2023**, *11*, 3710. [CrossRef]
124. Heo, Y.-J.; Yeo, W.-H.; Kim, B.-G. DeepFake detection algorithm based on improved vision transformer. *Appl. Intell.* **2023**, *53*, 7512–7527. [CrossRef]
125. Sanjeevi, M. Available online: https://medium.com/deep-math-machine-learning-ai/chapter-10-1-deepnlp-lstm-long-short-term-memory-networks-with-math-21477f8e4235 (accessed on 10 May 2020).
126. Singh, V.K.; Ghosh, I.; Sonagara, D. Detecting fake news stories via multimodal analysis. *J. Assoc. Inf. Sci. Technol.* **2020**, *72*, 3–17. [CrossRef]
127. Nakamura, K.; Levy, S.; Wang, W.Y. Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection. In Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020), Marseille, France, 11–16 May 2020; pp. 6149–6157. Available online: https://www.aclweb.org/anthology/2020.lrec-1.755 (accessed on 18 December 2020).
128. Yang, Y.; Zheng, L.; Zhang, J.; Cui, Q.; Li, Z.; Yu, P.S. TI-CNN: Convolutional Neural Networks for Fake News Detection. *arXiv* **2018**, arXiv:1806.00749.
129. Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; Gao, J. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018), London, UK, 19–23 August 2018; pp. 849–857. [CrossRef]
130. Cui, L.; Wang, S.; Lee, D. SAME: Sentiment-Aware Multi-Modal Embedding for Detecting Fake News. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM, 2019), Vancouver, BC, Canada, 27–30 August 2019. [CrossRef]
131. Singhal, S.; Shah, R.R.; Chakraborty, T.; Kumaraguru, P.; Satoh, S. SpotFake: A Multi-modal Framework for Fake News Detection. In Proceedings of the 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), Singapore, 11–13 September 2019; pp. 39–47.
132. Khattar, D.; Goud, J.S.; Gupta, M.; Varma, V. MVAE: Multimodal Variational Autoencoder for Fake News Detection. In Proceedings of the The World Wide Web Conference (2019), San Francisco, CA, USA, 13–17 May 2019; pp. 2915–2921. [CrossRef]
133. Zhou, X.; Wu, J.; Zafarani, R. SAFE: Similarity-Aware Multi-Modal Fake News Detection. *Adv. Knowl. Discov. Data Min.* **2020**, *12085*, 354–367. [CrossRef]
134. Chen, H.; Chang, C.; Shi, Z.; Lyu, Y. Hybrid features and semantic reinforcement network for image forgery detection. *Multimed. Syst.* **2022**, *28*, 363–374. [CrossRef]
135. Singh, B.; Sharma, D.K. Predicting image credibility in fake news over social media using multi-modal approach. *Neural Comput. Appl.* **2021**, *34*, 21503–21517. [CrossRef] [PubMed]
136. Zhou, P.; Chen, B.-C.; Han, X.; Najibi, M.; Shrivastava, A.; Lim, S.-N.; Davis, L. Generate, Segment, and Refine: Towards Generic Manipulation Segmentation. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 13058–13065. [CrossRef]
137. Sharma, D.K.; Singh, B.; Agarwal, S.; Kim, H.; Sharma, R. Sarcasm Detection over Social Media Platforms Using Hybrid Auto-Encoder-Based Model. *Electronics* **2022**, *11*, 2844. [CrossRef]
138. Salim, M.Z.; Abboud, A.J.; Yildirim, R. A Visual Cryptography-Based Watermarking Approach for the Detection and Localization of Image Forgery. *Electronics* **2022**, *11*, 136. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.