Hindawi
Security and Communication Networks
Volume 2022, Article ID 9291971, 27 pages
https://doi.org/10.1155/2022/9291971

Review Article

Overview of Image Inpainting and Forensic Technology

Kai Liu[1,2] **Junke Li**[2,3,4] **and Syed Sabahat Hussain Bukhari**[5]

[1]*College of Educational Science, Qiannan Normal University for Nationalities, Duyun 558000, Guizhou, China*
[2]*Key Laboratory of Machine Learning and Unstructured Data Processing of Qiannan, Duyun 558000, Guizhou, China*
[3]*School of Computer and Information, Qiannan Normal University for Nationalities, Duyun 558000, Guizhou, China*
[4]*Key Laboratory of Complex Systems and Intelligent Optimization of Guizhou, Duyun 558000, Guizhou, China*
[5]*College of Computer Science, Neijiang Normal University, Neijiang 641100, China*

Correspondence should be addressed to Junke Li; ljk2006ljk@163.com

Received 23 September 2021; Revised 27 October 2021; Accepted 30 March 2022; Published 20 May 2022

Academic Editor: Farhan Ullah

At present, digital image is the most significant information carrier on the Internet, which provides convenience for people to exchange information. With the progress of image processing theory, digital image inpainting technology has also developed rapidly. It not only enhances the image expression but also provides tools for image forgery. Misusing the tools may trigger a crisis of trust in the image. To eliminate the negative impact of image forgery, many scholars have conducted in-depth research on it and proposed a series of detection methods called image forensics for institutions or communities to identify such forged images. Current forensic technology still has many limitations, such as relying on specific features and data distribution which makes forensic technology lag behind image inpainting technology. So far, academia still lacks a unified understanding of image forgery and detection technology, and the research architecture of inpainting and forensic technology is not clear. This article reviews the development of image inpainting and forensic technology and systematically summarizes and scientifically classifies the current research work. Finally, the forensic technology is summarized and the future research direction is prospected to guide subsequent scholars to further promote the progress of image forensic technology.

1. Introduction

As early as decades after the birth of photos, image processing technology has emerged [1]. In the film age, it is time-consuming and hard to modify images. However, with the development of society and the progress of science and technology, the popularization of digitization and informatization, the rapidity of digital images have brought many conveniences to various industries, including news photography. While using images, people often need to modify the images appropriately to make up for the defects leftover from the camera. For example, photo processing such as removing red eyes, making up for insufficient exposure, color scale adjustment, and sharpening enhancement is used to enhance the expressiveness of images and make digital images clearer and beautiful. Today, with the highly developed Internet, digital images have become the significant carrier of information on the Internet. People can publish the captured digital images at any time by using social network platforms. These published digital images convey important information to people. At the same time, news image is no longer the patent of news photographers but has become people's daily skill.

Limited by the carrier of news image, it cannot be released according to the original size at the time of collection on some occasions, so the image needs to be compressed. With the development of digital image processing technology and the continuous emergence of more and more image processing and editing software (such as Photoshop, AutoCAD), the modification, editing compression, and storage of digital images become very simple and imperceptible. While people enjoy the convenience of the digital images from networks, it also takes opportunities to some individuals and institutions with ulterior motives. For various purposes, they unintentionally or deliberately, or even maliciously, release digital images that are beneficial to

themselves to deceive and confuse the public. Therefore, people also begin to face the great test of image authenticity judgment especially in the field of news photography. Even in the news reports of some well-known media at home and abroad, there are many fake "news images," which often become the fuse of network rumors. For example, the faked image of Indian Prime Minister Narendra Modi inspecting the flood disaster in southern India by plane in the official news released by the Indian information and Information Bureau in 2015 makes this fraud widely questioned by the public [2]. Chinese Sichuan officials publicly apologized for publishing synthetic images [3]. In recent years, face-changing technology represented by Deepfakes [4] and Zao app [5] has begun to rise on the Internet and provide mass face-changing entertainment services, which has posed a significant challenge to the judicial system. The problem of image tampering has widely existed in various fields, such as news media, fashion magazines, social media, online auction websites, and academic research journals. [6]. When the forged image is applied in the process of the monitoring system, publication, crime investigation, and court proof as evidence, its authenticity identification is very significant.

The authenticity of the image plays a more significant role in the quality of the event. For example, whether the authenticity of the moon landing photos in the U.S. Lunar Landing Plan is real or not [7]. For another example, seek the suspect who assassinated U.S. President Kennedy [8]. To make digital images truly reflect objective facts and news reports more authentic, how to distinguish the authenticity of digital images has become a key problem to be solved. The digital image processing technology is a "double-edged sword" which is originally produced to facilitate the use of images. Under normal circumstances, image processing not only does not affect the information of image exchange but also makes the information contained in the image clearer and more accurate. This "modification" is acceptable to people. For counterfeiters who maliciously tamper with images, forge scenes, and distort facts, this "tampering" is unacceptable. The reasons for forging images are complex, including political reasons, interest-driven, expressing opinions, and simple entertainment. In general, modifying the image can convey a certain intention or attract more people's attention, and finally achieve a certain purpose. Some image tampering behavior will bring adverse effects, introduce unstable factors to society, and cause immeasurable serious consequences. Therefore, how to identify the authenticity of image content which is also called digital image forensics is an urgent issue related to people's production and life, national and social stability, and so on. Malicious image inpainting causes digital image forensics. To avoid image forensic recognition, malicious attackers have developed anti-forensic-related technologies. Their relationship between them is listed in Figure 1, and the structure of this context is based on it. At present, some summaries of image inpainting [9, 10] and forensics [11–14] are still lacking in this field. There is only an urgent need to systematically sort out and scientifically summarize and classify the existing works to promote the research in this field.

The rest of this paper is organized as follows: Section 2 introduces digital image inpainting technology. Section 3 reviews digital image forensic technology. Section 4 discusses the game of image forgery and detection technology. Section 5 summarizes the current datasets for forgery research. Finally, Section 6 provides the summary and prospects of the work.

2. Digital Image Inpainting

With the advent of the Internet era, multimedia presents a colorful virtual world for people, in which digital images carry a huge amount of information and become an indispensable role. Therefore, it has attracted many scholars in companies and academia to study digital images. With the increasing demand for the repair of old and damaged photos and the coloring of black-and-white photos, there is also research on image inpainting in the field of image research. At present, according to the evolution of image inpainting technology, it can be divided into image editing inpainting, image synthesis inpainting, and deep learning-based image inpainting, as shown in Figure 2.

2.1. Image Editing Inpainting. There are many kinds of image editing operations, such as splicing, deletion, copy-move, blurring, contrast enhancement, compression, scaling, median filtering, and adding noise. Splicing refers to cutting an object from another image and inserting it into the image to be edited. Deleting operation refers to deleting one or more objects in the image to be edited. Copy-move is a common image inpainting method. It hides important information or forges false scenes by copying an area in the image, scaling, rotating, and pasting it to other locations of the same or other images. Usually, to achieve the purpose of attack, attackers will perform preprocessing operations such as rotation and scaling on the editing area, or smooth the edge of the editing area. These operations will change the image content to varying degrees, among which splicing, deletion, and copy-move are the most common image editing operations. Through image editing operation, the tension of image performance can be enhanced, black-and-white photos can be colored, and the original appearance of damaged photos can be restored.

2.2. Image Synthesis Inpainting. Before the development of deep learning technology, image inpainting was mostly realized by synthetic inpainting algorithm, which was deeply studied by many scholars. Image synthesis inpainting refers to the process of visually filling the damaged area of the image to restore the integrity of the image, and it is difficult for the observer to detect the damaged area afterward. The main idea of image synthesis inpainting is to use the information of multiple reference areas of the image to fill or synthesize the area to be restored and to keep the structure and texture of the image before and after the restoration as consistent as possible. Criminisi et al. [15] proposed an inpainting method based on image block texture synthesis, which used a block-based sampling process to achieve the filling of texture and

Figure 1: The structure of context.

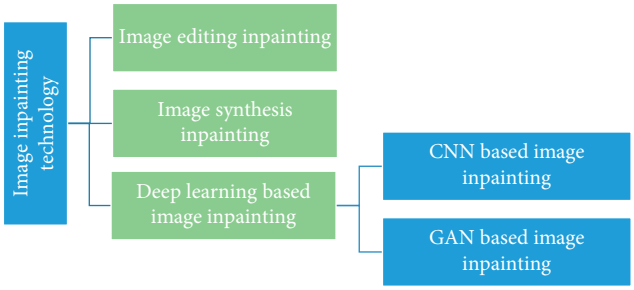

Figure 2: Classification of image inpainting technology.

structure information. Its process to repair the image is shown in Figure 3. Let the repaired area be O and the reference area be F as shown in Figure 3(a). It first calculates the priority of each point on the boundary ($\delta\Omega$) of the repaired area in the image to obtain the highest priority point p and then selects the image block ψp centered on p as the current block to be repaired as shown in Figure 3(b). Second, the image block Ψ_q' that most matches ψp is searched in the reference area of the image as shown in Figure 3(c). Again, the part to be filled in ψ_p is replaced with the pixel at the corresponding position in the best matching block ψ_q' as shown in Figure 3(d). Finally, the edge is updated and the above process is repeated until the O is repaired. At present, most image inpainting algorithms improve the algorithm in the [15]. For example, the algorithm in [16] has improved the fast priority dropping and visual continuity of [15]. The algorithm in [17] has improved the structural discontinuity and texture inconsistency of [15]. Jiao et al. [18] and Su et al. [19] have improved computing efficiency and the visual effect of repair respectively.

2.3. Image Inpainting Based on Deep Learning. At present, image synthesis inpainting methods mainly fill the missing part according to the statistical information of the residual image content, that is, each time a pixel of the missing part is constructed using the similarity principle to maintain its consistency with the surrounding pixels. However, when the missing part is large or complex, the traditional image inpainting methods often fail because they cannot obtain higher level semantic, texture, and other deep features from the original image. In 2006, Hinton [20] first put forward the relevant viewpoints and concepts of deep learning to learn the deep feature expression of things by creating a multilayer network. With the rapid progress of machine learning technology, various neural networks emerge, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs). Among them, CNNs and GANs have been shown their effectiveness for image-related tasks. CNN is a kind of feedforward neural network with convolution operation and deep structure. Typical CNN architectures include AlexNet [21], ResNet [22], VGGNet [23], Xception [24], GoogleNet [25], DenseNet [26], and SENet [27]. CNN is one of the most typical architectures of deep learning and has achieved certain results in image inpainting [28, 29].

The current advanced image inpainting methods mainly include two categories: (1) the methods based on deep CNN proposed by [30, 31] of NVIDIA company; and (2) the method based on GAN proposed by [32].

2.3.1. CNN-Based Image Inpainting. In the early days when deep learning was applied to image inpainting, Pathak et al. [30] combined the encoder and decoder structure with CNN to design a context coder in 2016 to solve the problem that CNN depended on a large number of labeled data and the semantic understanding problem contained in the image to be completed. It used multiple convolution layers in the codec structure, which can complement the semantic-sensitive content of the image scene in a parametric way. Therefore, it can synthesize high-level features in a large spatial range and provide better features for the inpainting method based on the nearest neighbor. This method realizes the application of CNN in image inpainting for the first time.

In order to meet the visual and semantic rationality of the repaired image, Zeng et al. [33] proposed a pyramid-context encoder network called PEN-NET for image inpainting. Based on U-Net structure, they used deep generative models to restore an image by encoding contextual semantics from full-resolution input and by decoding the learned semantic features back into images. To solve the above problem, Liu et al. [31] proposed to apply the partial convolution algorithm to the image inpainting problem and used some convolution layers to replace all convolution layers in the U-Net structure to complete the missing part of

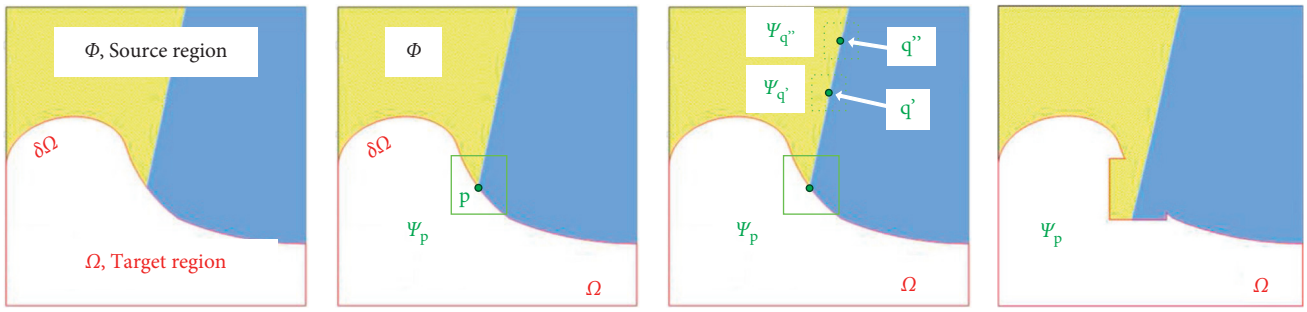

Figure 3: The process of algorithm of [15].

the image. This method only inputted the effective pixels of the unmissed area in the picture and added a mask update process after each layer, which realized the image inpainting that was independent of the size of the initial missing part and did not require any postprocessing. This is the first time CNN has been used to complete the missing irregular shapes in images. Yu et al. [34] optimized partial convolution by using gated convolution. They abandoned the hard mask of partial convolution and optimized it into a flexible mask which was automatic learning from data to better complete the images with irregular shapes of incomplete parts. Gated convolution network accelerated the training speed and improved flexibility through flexible mask and dynamic feature selection mechanism. It can improve the repair effect under the condition of free-form masks and user-guided input.

To repair the wrong texture generated by content-aware filling and apply image inpainting to higher resolution images, Yang et al. [35] proposed a high-resolution image inpainting method of multiscale neural patch synthesis. There are two networks in its structure: one is a content generation network and the other is a texture generation network. The former network directly generates images and infers the possible contents of missing parts. The latter network enhances the output texture of the content network. The obvious disadvantage of this algorithm is that its execution process consumes a lot of computing resources and takes a long time to generate the results. In addition, this method does not extend the application to the image inpainting with irregular missing areas.

For the image inpainting problem with poor processing edge position effect of the missing part of the image, DFNet (deep fusion network) proposed by Hong et al. [36] can harmoniously integrate the completed image area with the existing incomplete image, especially in terms of pixel transition and semantic structure consistency of the boundary area, so as to better realize image inpainting.

Both Yan et al. [37] and Wang et al. [38] are dedicated to repairing the central part, without considering the continuity between pixels. From the semantics perspective, they did not consider the continuity of features, resulting in color fault or line fault. For solving the above problem, Liu et al. [39] proposed a coherent semantic attention (CSA) model, which has a rough repair step and a fine repair step. The network of rough repair used a U-net structure. The repair speed is fast and the effect is good.

To solve the problem that the mean and variance shifts caused by full-spatial feature normalization (FN) limit the image inpainting network training, Yu et al. [40] proposed a spatial region-wise normalization named region normalization (RN). It divides spatial pixels into different regions according to the input mask and computes the mean and variance in each region for normalization. Two kinds of RN for image inpainting network are introduced: (1) Basic RN (RN-B), which normalizes pixels from the corrupted and uncorrupted regions separately based on the original inpainting mask to solve the mean and variance shift problem. (2) Learnable RN (RN-L), which automatically detects potentially corrupted and uncorrupted regions for separate normalization and performs global affine transformation to enhance their fusion.

In summary, Table 1 lists the categories, models, structures, advantages/disadvantages, and the best performance in the article of the above deep CNN-based image inpainting methods.

2.3.2. GAN-Based Image Inpainting. GAN is a deep learning method proposed by Goodfellow et al. [41], which uses the mutual game of generative models and discriminant methods to produce better output results. Its network structure is shown in Figure 4, and it includes two parts: generator *G* and discriminator D. G takes random noise as input and generates some fake samples similar to real ones, whereas the *D* has to learn to determine whether samples are real or fake. If *D* judges correctly, it is necessary to adjust the parameters of *G* to make the generated false data more realistic. If the judgment of *D* is wrong, the parameters of *D* are adjusted to avoid making a wrong judgment next time. GAN directly performs feature extraction and image generation on samples by considering global information. Therefore, the generation time of the target is shorter, and the speed is relatively fast, and the generated images are more realistic. These make GAN an extremely high-performance generation method, which has become the basis of many image inpainting algorithms.

With the continuous development of GAN, a series of GAN models has been developed, which correspond to

Table 1: Comparison of deep CNN-based image inpainting methods.

Category	Model	Structure	Advantage/disadvantage	Perf.
CNN	Traditional CNN	Convolution layer; subsampling layer; full link layer	Efficient processing; automatic feature selection; good classification effect.	
Deep CNN	Context encoder	Encoder and decoder structure combined with convolutional neural network	Unmarked image, irregular shape of missing area; can obtain the semantic information of the image; good repair effect, high processing speed.	17.58 dB (PSNR)
	PEN-NET	U-net structure with deep generative models	Visual and semantic coherence for image inpainting; fast convergence in training	9.94 (L1 loss)
	Partial convolution	Applying the partial convolution to replace all convolutions in the U-Net structure.	Stable performance; suitable for image inpainting with irregular shape.	19.04 dB (PSNR)
	Gated convolution	Using gated convolution to optimize partial convolution	Improved the flexibility; with mask and guiding input, the repair effect can be improved.	1.6% (mean L1 error)
	High-resolution	Content network and texture network	Consume a lot of computing resources and take a long time; only the patch in the picture is used instead of the data in the whole dataset.	18.00 dB (PSNR)
	Shift-net	U-net-based architecture	Image center restoration; continuity between pixels is not considered.	26.51 dB (PSNR)
	CSA	U-net architecture	Fast speed; good repair effect	26.54 dB (PSNR)
	RN	Encoder and decoder structure	Good repair effect	28.16 dB (PSNR)

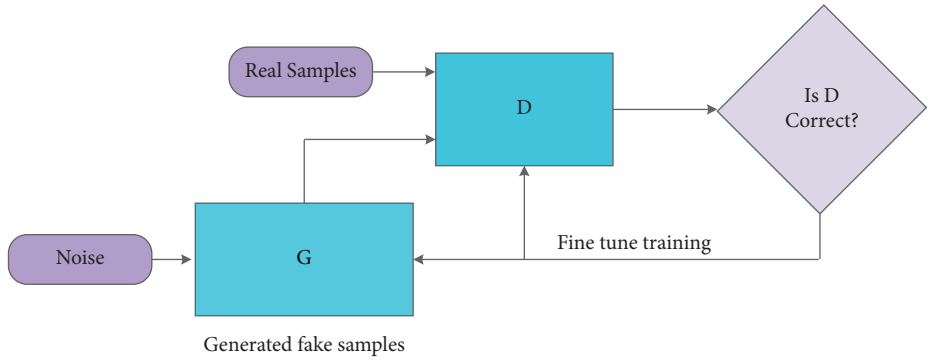

Figure 4: The structure of GAN.

different computing performance and application requirements. Generally, GANs can be divided into two categories: structure change-based GAN and loss function-based GAN [42]. Among them, structure change-based GAN mainly covers several categories: conditional GAN, deep convolution GAN, combined GAN, self-attention GAN, style-based GAN, etc. The loss function-based GAN mainly covers several categories, such as least-squares GAN and Wasserstein GAN.

(1) Conditional GAN. In order to solve the problem that the original GAN has almost no constraints on the generator, Mehdi et al. [43] imposed constraints on it and proposed conditional GAN (CGAN) to make the network generate samples in a given direction. In the training process of CGAN, constraints were added to both generator and discriminator to guide the generation of data. But CGAN did not solve the problem of unstable training. CGAN is mainly used in image enhancement, such as image generation and image inpainting [44].

In the original GAN, the input of the generator G is generally a continuous single random noise Z, and the generator G will perform highly entangled processing on Z. In this case, we cannot specify the samples generated by G. For this, Chen et al. [45] proposed interpretable representation learning by information maximizing GAN (info-GAN). It added a latent code c based on GAN to control the attributes of the generated image by controlling the variables of the corresponding dimensions. Info-GAN can make the network learn interpretable features and generate specified images, but the mutual information between data needs to be calculated in the training process, which increases the training complexity.

(2) Deep Convolution GAN. The image inpainting method based on context encoder predicts the missing area through context, and the effect of the generated image is not ideal. Therefore, the deep convolution GAN model (DCGAN) in Yu et al. [46] is proposed by combining traditional CNN with GAN and applying traditional CNN to generators and

discriminators. It used convolution layer instead of full connection layer and fractional step convolution instead of the upper sampling layer, and its generator and discriminator were both using batch normalization layer. This method has no artifact in the inpainted area, the image edge is clearer and has achieved good results in solving the problem of inpainting large missing areas.

In 2017, the semantic image inpainting method based on DCGAN proposed by Yeh et al. [32] gave a better solution. It used DCGAN as the basic model structure and combined the semantic repair method. Then, it adjusted and improved the loss function of DCGAN in a targeted manner. Similar to the context encoder, this method used unlabeled data during training and testing. But compared to the context encoder, one of the main advantages of this method is that no mask is required for training. Because the training process does not rely on masks, this method is more suitable for missing regions of arbitrary structure than context encoders.

Although DCGAN is used for missing areas of arbitrary structure, it is difficult to cope with image inpainting in many different scenes. To improve the coordination among the inpainting area, the surrounding area, and the global and make the inpainting model meet the image generation tasks of high-resolution, irregular missing areas and multiple scenes at the same time, Iizuka et al. [47] proposed a globally and locally consistent image inpainting (GL) algorithm by combining GAN and CNN. This method took an inpainting network and two discriminator (a global discriminator and a local discriminator) networks as the main architecture, which greatly improved the quality of image inpainting. Although the GL model can complete a variety of scenes, the effect of this method will be poor if the mask of the image contains a large area of structural objects, such as people and animals.

(3) Combined GAN. In order to solve some complex problems, a method of splitting complex problems into small parts and then training them one by one is proposed. This method achieves better results by translating, stacking, or combining multiple GANs.

Synthesizing high-quality images through text description is a difficult problem in computer vision. For this reason, Zhang et al. [48] proposed a stacked GAN (StackGAN) model, which was essentially a combination of two layers of CGAN. It used a two-stage generation method: the first stage drew the basic outline and color of the image according to the given text and generated a low-resolution image; and the second stage used the low-resolution image and text description generated in the first stage as input, corrected the defects in the results of the first stage, and added some details on this basis to generate high-resolution images. The StackGAN model stabilizes the training of CGAN very well and has made significant improvements in generating realistic images based on text descriptions. StackGAN++ in Zhang et al. [49] is also used to generate high-resolution images for text.

Zhu et al. [50] proposed cycle consistent generative network (CycleGAN) in 2017. CycleGAN realized the conversion of images from source domain x to target domain y and did not need pairs of images as training data. The model of CycleGAN is a ring structure, which is composed of two generators and two discriminators. CycleGAN can learn the mapping relationship and transformation method between artworks and real photos. Its disadvantage is that although its cyclic mechanism can ensure that the imaging will not deviate too much, some information will be lost in the cyclic conversion, resulting in the low quality of the generated image.

In order to extract information from areas far from the image, expand the receptive field of computer vision, and improve the stability of network training, Yu et al. [51] proposed a deep generative model-based approach (DGM) based on the GL method. The method has global and local discriminators, and its generator is composed of a feed-forward generation network with two stages. A narrow and deep network is used in the generator. In this method, the coarse network and fine network used for image completion are "connected in series," and the two-generation networks with different functions in the front and back stages are combined into a generator to make the model structure "narrow." The two networks are respectively responsible for the completion tasks of the first and second stages and are composed of deep expansion convolution layers, so the structure is "deep." Similar to the GL method, this method uses a deeper extended convolution layer in both stages of image completion from coarse to fine, but the number of parameters is significantly less than that of the GL method.

In order to maintain higher sample diversity, obtain multiple inpainting results, and improve the completion effect of existing methods in large missing areas, Zheng et al. [52] proposed a pluralistic image completion (PICNet) method with two parallel paths based on probability principle in 2019 and used short and long-term attention layer in the model. PICNet can generate multiple different completion schemes with trusted content for a single masked input. The use of short-term and long-term perception layers improves the reliability of completion. PICNet has high quality in various datasets, especially for large missing areas.

(4) StyleGAN. When the resolution of the generated image is very high, the discriminator can easily recognize that the image generated by the generator is fake but this makes the generator difficult to train. The described above problem makes DCGAN can only generate 64×64 images. When DCGAN generated a larger resolution image, the output seriously lost more details. To solve this problem, Karras et al. [53] proposed progressive growing of GAN (PGGAN) by introducing a progressive training method through gradually generating images from 4×4 to 1024×1024. The generator and discriminator in PGGAN grow incrementally by gradually adding new layers to make the network model more complicated to learn better-detailed features. This method can not only accelerate the training but also make the training more stable. The progressive method of PGGAN allows the training to first discover the distribution of images with large-scale structures and then gradually shifts the focus to better scale details, instead of having to learn all scales at the same time. Moreover, the generator and the

discriminator compete with each other to promote the training of both. With the gradual growth of the GAN network, most of the iterations are completed at a lower resolution, which reduces the training time.

StyleGAN proposed by the well-known technology company NVIDIA uses the idea of style conversion to design a new generator architecture in Karras et al. [54], which has a better effect when generating a single object. The network structure of StyleGAN consists of two parts: the first is the mapping network, which is used to control the style of the generated image; and the second is the synthesis network, which is used to generate the image and is used to control the style of the generated image. The entire network structure of Style-GAN still maintains the structure of PGGAN. However, the images generated by StyleGAN sometimes contain speckle-like artifacts. For this reason, StyleGAN v2 is proposed in Karras et al. [55].

(5) Self-Attention GAN. Deep CNN can improve the details of GANs to generate high-resolution images, but due to the limitations of the local receptive field of convolutional networks, if you want to generate long-range dependency regions, CNN will have problems. For this reason, the researchers proposed to apply the attention mechanism to make the designated and refined area has a larger field of vision.

Zhang et al. [49] built a variant of GAN that is attention driven and process dependent, called self-attention GAN (SAGAN). It combined the self-attention in the attention model with the original GAN, thus realized the application of all feature points in the low-resolution image to generate high-resolution detail features and used spectral normalization to improve the training effect of the generator in the training process. Both generative network and discriminant network of SAGAN adopt attention mechanisms. SAGAN can well model the attention-driven process dependency and coordinate the relationship between each position and each end detail when generating the image. The discriminator of SAGAN can also more accurately realize the complex geometric constraints on the global image structure.

In order to solve the problem of successfully generating high-resolution and diverse samples from complex datasets, BigGAN in Brock [56] was proposed in 2018. Based on SAGAN, BigGAN uses a set of TPUs to improve accuracy by leaps and bounds. BigGAN uses a larger number of channels in the design of the convolutional layer and uses a large batch size, which can maximize the performance of GAN. BigGAN can improve the model performance and make the training more stable by using the truncation trick, but it needs to balance the sample diversity and fidelity. BigGAN can ensure the stability of training through the collection of existing and other novel technologies, but the accuracy will also decline. It is necessary to balance the performance and training stability. BigGAN also has the disadvantages of a large model, many parameters, and high training cost.

In summary, given the shortcomings of the original GAN, different scholars have changed its structure to obtain different types of GANs and achieved good results, but there are still shortcomings. Table 2 shows the comparison of different types of GAN models based on structural change. The performance mentioned in Table 2 is the best in each model.

2.3.3. Loss Function-Based GAN

(1) Least-Squares GAN. In order to further solve the problem of gradient disappearance during GAN training, Mao et al. [57] proposed least-squares GAN (LSGAN), which replaced the cross-entropy loss function in the GAN discriminant network with the least square loss function. The advantage of least-squares loss is that it allows the samples generated by the generator to pull the generated image to the decision boundary on the premise of cheating the discriminator. Compared with the original GAN, LSGAN generates higher quality samples and its training process is more stable. The defect of LSGAN is that it does not solve the problem of gradient dispersion when the discriminator is good enough, because the least square function does not meet the Lipschitz continuity condition, and its derivative has no upper bound.

(2) Wasserstein GAN. In order to solve the problem of the disappearance of the training gradient of the original GAN, Arjovsky et al. [58] proposed Wasserstein GAN (WGAN) in 2017, which analyzed the reasons for the instability of the GAN in the training process, but also theoretically solved the mode collapse. It proposed to use Wasserstein distance to replace the JS divergence used in the loss functions of various GANs before. This method does not need generator and discriminator to maintain balance in the training process. It can generate more kinds of images to ensure the diversity of generated samples. WGAN promotes the development of GANs. However, in the experiment, it is found that WGAN still has the problems of difficult training and slow convergence. In the process of limiting the weight, if the design is inappropriate, the gradient explosion or gradient disappearance of the network will occur.

Given the problems existing in WGAN, Gulrajani et al. [59] introduced a gradient penalty on the basis of WGAN called WGAN-GP. It discards the weight clipping in WGAN and directly adds the gradient of the discriminator as a regular term to the loss function of the discriminator, that is, the gradient penalty replaces the weight clipping operation. WGAN-GP solves the problem of vanishing/exploding gradients while generating higher quality samples. However, the experiment found that the method has a slow convergence speed. Under the same dataset, WGAN-GP requires more training times to converge.

(3) Context-Aware Semantic Inpainting. In 2018, Li et al. [60] proposed a context-aware semantic inpainting method (CASI) based on the GAN framework. Its generator used a complete convolution network to replace the full connection layer to better preserve the spatial structure and combined the improved joint loss function to capture the high-level semantics in the context. The complete convolution network can save the spatial information of the image so that the network can accept the input of any dimension, and the

Table 2: Comparison of different types of structural change-based GAN.

Category	Model	Structural features	Advantage/disadvantage	Perf.
Initial model	GAN	Generator and discriminator	High-definition images; unstable training; mode collapses; disappearing gradient.	225 (mean log likelihood)
Conditional GAN	CGAN	Adding constraints to the input layer of GAN to guide generation	High requirements for the dataset; tagged dataset; low quality image.	—
	Info-GAN	Use generator and discriminator and add hidden code in the generator	Extending the theory and increasing the interpretability of GAN model; large amount of computation; and poor diversity.	—
Deep convolutional GAN	DCGAN	Symmetrical discriminator and the generator; using fractional-strided convolution instead of upsampling.	Stability training; No artifacts; clearer edges; data distribution affect the effect; unsuitable for complex scenarios.	85.95 (accuracy)
	GL	A inpainting network and two discriminators network	The result is locally and globally consistent; can complete any scene; and unsuitable for heavily structured objects.	96.5% (accuracy)
Combined GAN	DGM	Global and local discriminators; two stages feedforward generation network.	Expanding the receptive field and improving the stability of network training; less parameters.	18.91 (PSNR)
	StackGAN	Superimpose two CGANs	High-resolution images; stable training.	8.45 (IS)
	Cycle-GAN	Circular mechanism to achieve two domain conversions and constraints	Low data requirements; no one-to-one paired images; low resolution is not high.	0.58 (accuracy)
	PICnet	Parallel generating paths and reconstructing paths; variation encoder structure of the generator.	The reliability of complementary content; high-quality images ; suitable for the images with arbitrary of incomplete parts.	20.10 (PSNR)
Pattern-based GAN	PGGAN	Gradually grow generators and discriminators.	Reducing the training time; generating better high-resolution images.	0.2838 (MS-SSIM)
	Style-GAN	Including mapping network and synthesis network; style-based structure of generator.	Can control the visual features from coarse features to fine details.	4.40 (FID)
	Style-GAN2	Replace the gradient network structure of style GAN with fixed network training.	Focus on repairing artifacts and further improving the quality of generated images	2.32 (FID)
Self-attention GAN	SAGAN	Both generative network and discriminant network adopt attention mechanism.	Getting good balance between improving the receptive field and reducing the amount of parameters.	18.65 (FID)
	BigGAN	More channels in convolutional layer and use truncation and orthogonal regularization	Model training is stable and can generate ultra-clear images; large amount of parameters; difficult to train	7.4 (FID)

spatial information of the image can be retained without dimensionality reduction. In addition, without the full connection layer, the complete convolution network occupies less memory and can learn and predict more effectively. The CASI model can successfully reduce semantic errors, infer semantically valid content from the image context, and generate images that conform to the context of the image. Compared with the normal model using the fully connected layer, it can better grasp the spatial information of the image and make the structure of the completed image more reasonable.

To sum up, good results have been achieved in changing the original GAN loss function, but there are still deficiencies. Table 3 shows the comparison of loss function-based GAN. The performance in Table 3 is the best in each paper. Table 4 lists various image inpainting methods and their advantages and disadvantages. The use of various image inpainting techniques under normal morals and ethics reproduces the classics of the past, makes up for the defects of the pictures, and brings more and more laughter to people. However, when the repaired image goes beyond the original intention of the image, it becomes image forgery and image malicious attack, which also provides technical conditions for attackers to use the inpainting image for malicious purposes. The multicategory and realistic high-resolution color images generated by forged technologies [41, 48, 49, 53, 56, 61] not only destroy the authenticity of the images, but also convey false information to the public and may even cause serious social harm. As a technology, how to use it, whether it brings happiness or disaster to people depends entirely on people themselves, and not on tools. Therefore, it is necessary to detect the images inpainting by these technologies.

3. Digital Image Forensics

How to identify the forgery and malicious attacks brought by image inpainting has attracted the attention of a large number of people in government, associations, and academia. Then, they put forward many solutions, resulting in digital image forensic technology [62–65].

Table 3: Comparison of GAN model based on variant of loss function.

Category	Model	Structural features	Advantage/disadvantage	Perf.
Initial model	GAN	Generator and discriminator	High-definition images; unstable training; mode collapses; disappearing gradient.	225 (mean log likelihood)
GAN model based on loss function variation	LSGAN	Replacing the cross-entropy loss function with the least square loss function	Limit the unlimited modeling capabilities of GAN; unsolved the problem of gradient dispersion in the generator.	6.47 (IS)
	WGAN	Using Wasserstein distance instead of JS divergence	Stable training; theoretically solves the model collapse; gradients disappear and explode; slow convergence speed.	—
	WGAN-GP	Using gradient penalty mechanism instead of weight interception operation	Stable training; solves the problem of gradient disappearance and explosion; higher quality samples; slow convergence speed.	7.86 (IS)
	CASI	Fully convolutional network; introduced perceptual loss; joint loss function.	Reduces semantic errors; generates images that conform to the contextual content.	20.37 dB (PSNR)

Table 4: Comparison of various image inpainting methods.

Method	Advantage	Disadvantage
Image editing inpainting	Widely popularized; easy to operate; easy to learn	The effect only stays at the visual level and depends on the experience of the operator; the repair speed of multiple pictures is slow.
Image synthesis inpainting	The inpainting speed is faster and the algorithm is relatively stable. The inpainting can be completed under the condition of a small number of images.	The abstraction process of information and data only stays at a shallow level; higher level features cannot be obtained. The resolution is low.
Image inpainting based on deep learning	High-level features of the image can be obtained; the method is independent and anti-interference; the image inpainting rate is higher and the effect is better.	Rely on a large amount of labeled data. The inpainting speed depends on the configuration of the device. The accuracy of the calculation results and the robustness of the method cannot be obtained at the same time. The gradient tends to be unstable.

In the government and associations, the Defense Advanced Research Projects Agency (DARPA) launched a research project called media forensics to develop an end-to-end digital media forensic platform that automatically evaluates the integrity of images and videos [66]. IEEE launched the digital image forensic competition in 2013 to promote technological progress in this field [67]. Recent forensic tools such as FotoForensics [68] and MMC Image Forensic Tool [69] are created to identify the image.

In academia, digital image forensic technology realizes the objectives of identifying the source, detecting the processing operation of the image, and judging the authenticity of content by analyzing the inconsistency and tampering traces of image content. Digital image forensics approximately describes the process of applying image processing operation to a digital image in the form of a mathematical model. Through the analysis and understanding of the mathematical model, it is proposed that the features of the above operation process can be characterized. Finally, the corresponding feature extraction algorithm is designed to detect whether the operation process exists in the image. This research process requires theoretical knowledge such as digital image processing, signal processing, mathematical-statistical analysis, computer vision, pattern recognition, machine learning, etc. In addition, the research on image forensics is helpful to further understand the change law of statistical features of images before and after the operation, thus improving the performance of image processing operation. The research of digital image forensics is based on "digital fingerprint," which can be the information actively added by the forensics, or the unique traces introduced by image processing or tampering. These two types of digital fingerprints correspond to active forensics and passive forensics, respectively. The classification of forensics is shown in Figure 5. The active forensic method is to actively add a digital watermark [70, 71] or digital signature [72] to the original image. When copyright and authentication problems occur, the extraction algorithm is used to extract this information to provide copyright proof or content authenticity and integrity authentication. Passive forensics, also known as blind forensics, means to verify its authenticity and source only by analyzing the image without adding any information in advance. The methods of image verification can use image generation features, image operation features, and deep learning technology. At present, passive forensics has a wide range of applications and is the focus of scholars' attention, so this article mainly focuses on passive forensics.

Understanding the image generation process is helpful to better study passive forensics. Therefore, Figure 6 shows the generation process of a digital image. In a specific scene in the real world, the camera image is generated by the lens, optical filtering, CFA interpolation, and compression coding in the camera, and then the final image is obtained after

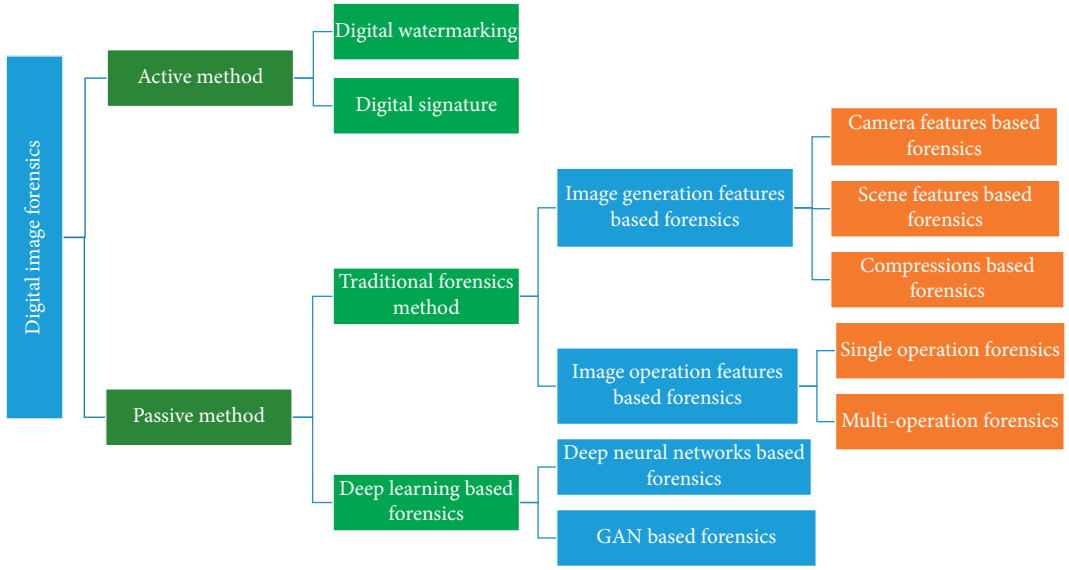

Figure 5: Classification of digital image forensic technology.

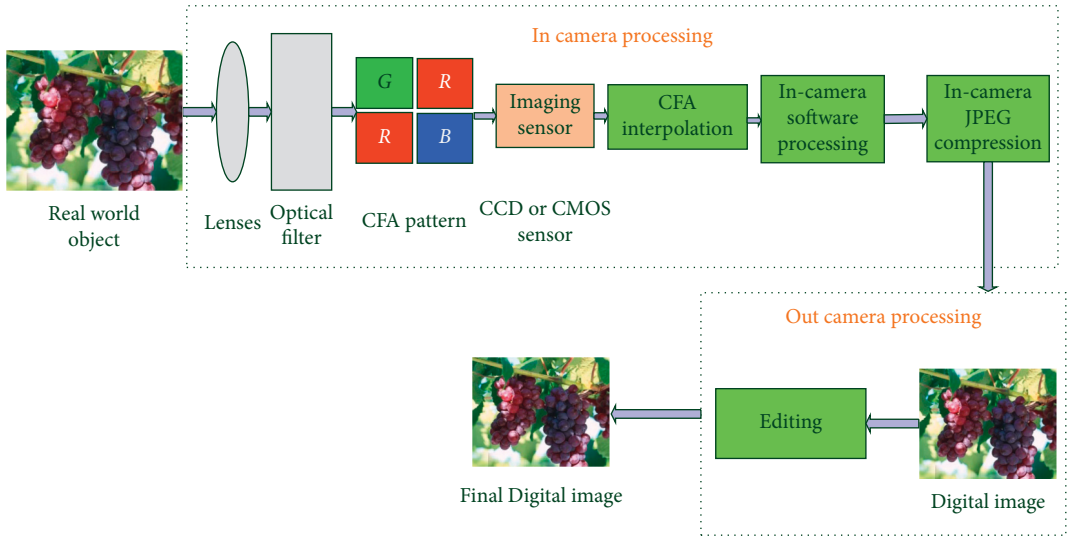

Figure 6: Generation process of a digital image.

editing and processing out of the camera. For each of the above stages, scholars have carried out a large number of forensic research. Forensic efforts are devoted to finding traces left in each step of the above image generation process. These traces may come from light and shadow in the scene, an imaging component of the camera, image editing operations, or anti-forensic operations added to cover up image editing operations. These traces can be regarded as the digital fingerprint of the operation, which uniquely corresponds to the operation performed so that the fingerprints are different from each other. These digital fingerprints are the design basis of the forensic detection algorithm. According to the digital fingerprints generated in different stages, digital image forensic technology can be divided into the following categories: traditional forensics and deep learning-based forensics. Traditional forensic method mainly includes image generation feature-based forensics and image operation feature-based forensics. Deep learning-based forensics mainly includes deep neural network-based forensics and GAN-based forensics. These forensics methods will be introduced as follows.

3.1. Traditional Forensics

3.1.1. Image Generation Feature-Based Forensics. According to the process of image generation, the specific scene is imaged by the camera, and then the captured images are compressed and stored or published. Since the camera, specific scene, and generated compressed image all have their features, obtaining the common features of the generated image can detect whether the image has undergone the operation. Based on detecting features, the image generation feature-based forensic technology has developed into

scene feature-based forensics, camera feature-based forensics, and compression-based forensics.

The scene feature-based forensics is often based on fact that the characteristics of illumination, shadow, reflection, geometric characteristics, and perspective in the same scene are the same, while the lighting and shadow generated in different scenes are different, and even the geometric characteristics of objects from different scenes do not match each other.

Therefore, based on the inconsistency of the above characteristics in the image, many forensic technologies have been developed. Johnson et al. [73] conducted forensic research on illumination features, [74] is on shadow features, [75, 76] are on texture features, [76, 77] are on vector methods in image objects, [78] is on the inconsistency of objects, and height ratio features in images.

The camera feature-based forensics mainly adopts inconsistent characteristics such as the distribution or relationship of camera color [75, 79], camera pattern noise [80, 81], photo response non-uniformity (PRNU) [82–84], and color filter array (CFA) [85, 86] to realize image forensics.

For facilitating image storage and transmission, images are usually undergoing compression encoding. If the image has undergone tampering, it is inevitable to save the image in a compressed format again. Therefore, the detection of single and multiple compression encoding operations has become an important forensic problem. For compression detection, DCT transformation is one of the steps. The blocking execution of JPEG compression will lead to a blocking effect, and the higher the compression intensity, the more significant the blocking effect. Luo et al. [87] counted the integral of the DCT coefficient histogram in the region and calculated the ratio between them. If the ratio is less than a certain threshold, it is judged that the image has undergone JPEG compression. Because the image tampering operation must save the image again, this will lead to two or more JPEG compression. Huang et al. [88] considered that two compression operations adopt the same quantization step size and detect two JPEG compression operations by analyzing the number of changed DCT coefficients. Milani et al. [89] used the distribution of the first number of DCT coefficients to detect double JPEG compression.

3.1.2. Image Operation Feature-Based Forensics. Whether it is to achieve image modification or to cover up the traces of image tampering, people often use various image processing operations. The study of these operations can deeply explore the impact of various operations on the statistical law of the original signal. The detection of these operations can effectively expose the fact of image tampering and provide more detailed evidence. According to the types of operations to modify or tamper with images, it can be divided into single operation forensics and multioperation forensics. Single operation forensics can be divided into image enhancement, median filtering, resampling, and content modification. Multioperation forensics is a combination of the above single forensic technologies.

(1) Single Operation Forensics. The image enhancement operation can improve the image quality or mask the trace of image modification by modifying the brightness distribution, smoothing or sharpening, noise removal, etc. The main image enhancement operations include contrast enhancement, histogram equalization, median filtering, and so on. Contrast enhancement operation will produce "peak-valley" effect in the histogram. Stamm et al. [90] detected this feature by Fourier transform of image gray histogram. Cao et al. [91] further used the zero-height gray-level features in the histogram to detect homology contrast-enhanced image synthesis. Zou et al. [92] exploited GAN to process the contrast-enhanced image and make it indistinguishable from the unaltered one in the pixel domain. Then, a designed histogram-based loss to enhance the attack effectiveness in the histogram domain and the gray-level co-occurrence matrix domain and pixel-wise loss to keep the visual enhancement effect of the processed image are introduced.

Median filtering can smooth the image while maintaining the edges of the image through nonlinear operations. It is widely used for image denoising and smoothing, and even for anti-forensic operations [93]. Kirchner et al. [94] calculated the ratio of the number of zeros to the number of ones in the difference graph and detected the median filter for the probability of occurrence of zeros. To resist the postprocessing of JPEG compression, Liu et al. [90] used the SPAM feature in steganalysis. In order to avoid the influence of image content on the statistical characteristics of image difference, Popescu et al. [95] proposed the concept of median filter residual (MFR) and adopted an autoregressive model to capture the statistical characteristics of MFR.

Resampling is an indispensable process after the image is scaled, rotated, and affine transformed. Therefore, it has attracted the attention of many forensic researchers. Resampling operation will introduce correlation between local samples, and this correlation can be used as an important basis for testing resampling [95]. Mahdian et al. [96] designed detection algorithms based on the mean and variance periodicity of the second-order difference of the image after image resampling to realize the judgment. Vázquez-Padín et al. [97] performed singular value decomposition on the image matrix to extract eigenvalues and distinguish whether the image has undergone resampling operation according to the distribution characteristics of eigenvalues.

Modifying the image content is the ultimate goal of the tamper, for the most commonly used copy-move technique, so the forensics of image copy-move is the focus of image forensics. An algorithm for identifying copy-move tampering in images was proposed by Avidan et al. [98] for the first time. It divided the image into partially overlapping blocks and traversed all the blocks for block matching search. If there are two identical blocks, it is judged that there is a copy-move operation. The idea points out the direction for related research work in this field in the future. Based on this idea, many forensic methods that extract features from image blocks and then conduct block matching search are proposed, mainly including forensic methods based on scale-invariant feature transform [99], DCT coefficient

[99, 100], residual image [101], JPEG compression [102], and Fourier-Mellin transform [103].

The content-aware image resizing (CAIR) technology proposed by Avidan et al. [98] is another content modification operation technology. To detect the CAIR operation, Chen et al. [104] designed an improved method based on calibrated adjacency density (CNJD) to identify seam carving for JPEG images. Liu et al. [105] proposed a detection method based on large-scale feature mining, which used an integrated learning classifier when dealing with high-dimensional problems to avoid overfitting of traditional classifiers in detection. Liu et al. [98] proposed a hybrid detection method based on large-scale feature mining. Taspinar et al. [106] detected the tampered image with seam carving using the PRNU feature of the camera. Liu et al. [107] extracted high-dimensional features in the spatial domain and frequency domain. Then, they used an integrated learning algorithm to detect whether the JPEG image has undergone the seam carving operation. Ye et al. [108] proposed a detection method based on mixed features, including subtracted pixel adjacency model, Markov transition probability, and local derivative mode. Zhang et al. [109] proposed a blind detection method for image seam carving for low zoom ratios. Zhang et al. [110, 111] proposed a method for detecting image seam carving based on the Unified Local Binary Pattern. The CAIR operation will affect the local texture of the image, resulting in changes in the amplitude distribution of the pixel intensity difference in the local neighborhood, and the correlation of the pixels in the local neighborhood will be destroyed [112, 113].

(2) Multioperation Forensics. For forgers, when a single operation is difficult to achieve purpose, they will often consider using multiple operations. The concept of multiple operations was put forward earlier in Comesaña et al. [114]. Thereafter, different methods are proposed. Chen et al. [115] researched multioperation forensics for the image undergoing JPEG compression, then zooming, rotating, or both operations on the image, and finally saved it as a JPEG format. It proposed a deformed block-effect grid feature extraction algorithm and DCTR feature to detect multioperation forgery. Chen et al. [116] proposed a random feature selection strategy for multiple operations and proved the effectiveness of the strategy theoretically. Conotter et al. [117] constructed model for multiple operations of JPEG compression and linear filtering to describe the DCT coefficients in JPEG images after linear filtering of the entire image. Ferrara et al. [118] proposed two methods to detect the multiple operations of double JPEG compression with linear contrast enhancement between them. Its detectors used the "peak-valley" effect of the DCT coefficients and the distribution characteristics of the first digits. Bianchi et al. [119] used the idea of reverse engineering to detect multiple operations of double JPEG compression with scaling between them. Chu et al. [120] proposed a method to define the order detection problem as a multi-hypothesis test problem. Baracchi et al. [121] studied the anti-forensic method and pointed out that the image single operation (e.g., JPEG compression, sensor pattern noise, histogram statistics) could negatively affect other traces. Based on this observation, they could generate a pristine image with native camera metadata, JPEG structure, quantization tables, preview, thumbnails, their corresponding quantization tables, single compression statistics, and any in-camera (even proprietary) processing under a given tampered picture. They called this method camera obscura.

3.2. Deep Learning-Based Forensics. Deep learning has the features of powerful learning ability, strong adaptability, and good expression ability, which has caused evidence collectors to introduce it into the field of forgery detection and forensics. According to the deep learning technologies applied to forensics, it can be divided into deep neural network-based forensics and GAN-based forensics.

3.2.1. Deep Neural Network-Based Forensics. According to the features and structure of the forensic method used, the deep neural network-based forensics can be divided into three categories, namely image generating feature-based method, operation feature-based method, and CNN structure or loss function changed method. Table 5 lists the classification, method, detection type, and the best performance in each article of deep neural network-based forensics. Their specific descriptions are as follows:

(1) Image Generating Feature-Based Method. For the double JPEG compression operation, Wang et al. [122] designed a deep convolution neural network by extracting the histogram of DCT coefficients. On the premise of no manual careful design of features, the deep network is used to automatically learn the most essential changes of histogram features before and after tampering. Finally, tamper detection based on deep learning is demonstrated. Bunk et al. [123] proposed two methods to detect and localize image manipulations based on a combination of resampling features and deep learning. The first method used the resampling features, deep learning classifiers, a Gaussian conditional random field model, and random walker. The second method used resampling features and LSTM-based network. Experimental results show that both are effective in detecting and localizing digital image forgeries.

The intrinsic model features of the camera can also be used for image forensics in combination with deep learning technology. Bondi et al. [124] used CNN to extract the intrinsic model features of the camera from the image patch. Then, these features were analyzed by clustering technology to detect image tampering and locate tampering. In Chen et al. [125], the camera response function (CRF) was used for splicing and copy-move detection and positioning. Local descriptors based on image noise residuals are widely used in image forensics, such as tamper detection and location. Cozzolino et al. [126] showed a residual-based descriptor, which can actually be considered as a simple constrained CNN. Zhou et al. [127] proposed a double stream faster R-CNN network to detect tampered images. One branch of the two streams is the RGB stream, which is used to learn strong contrast differences. Another stream is used for noise

Table 5: Comparison of deep neural network-based forensics.

Category	Source	Method	Detection type	Perf.
Image generating feature-based method	[122]	Histograms of discrete cosine transform (DCT) coefficients	Double JPEG compression	0.98 (AUC)
	[123]	Resampling features, deep learning, and LSTM	Image manipulation	0.9138 (AUC)
	[124]	Camera models	Image authenticity.	0.908 (AUC)
	[125]	Camera response functions	Splicing and copy-move operations	0.97 (accuracy)
	[126]	Residual-based local descriptors	Image manipulation	100.00% (accuracy)
	[127]	Tampered artifacts and noise features	Image manipulation	0.95 (AUC)
	[128]	LSTM and CNN; similar patches	Image object removal	97.89% (accuracy)
Operation feature-based method	[129]	Median filtering residual; Modified CNN;	Compressed image	96.84% (accuracy)
	[130]	Hybrid CNN-LSTM; resampling features	Copy-clone, object splicing, and removal	94.8% (accuracy)
	[131]	Block-like features; self-correlations between different blocks,	Copy-move	0.7572 (F-score)
	[132]	BusterNet; visual artifacts; visual similarities	Copy-move	80.48% (accuracy)
	[133]	Similarity between feature vectors of image overlapped subblocks	Copy-move	0.93 (F-measure)
	[134]	Dense-InceptionNet	Copy-move	0.7058 (precision)
	[135]	Adaptive attention and residual refinement Network ; Atrous spatial pyramid pooling ; correlation maps	Copy-move	0.8488 (AUC)
CNN structure or loss function changed method	[136]	CNN with weighted cross-entropy loss function	Patch-based operation.	98.3% (AP)
	[137]	Decoder network and feature pyramid network	Patch-based operation	98.99% (TPR)
	[138]	Encoder-decoder network structure; using a label matrix and the weighted cross-entropy as the loss function	Recapture image	99.74% (accuracy)
	[139]	CNNs and the segmentation-based multiscale analysis	Splicing and copy-move	0.4063 (F1-score)
	[140]	Segmentation-based key point distribution strategy and adaptive over segmentation method	Copy-move	0.7627 (F1-measure)
	[141]	CNN-based framework; full-resolution information	Splicing, copy-move	0.886 (AUC)
	[142]	Modified CNN by demosaicing algorithms; mosaic inconsistencies	Splicing, inpainting, or copy-move	0.926 (AUC)
	[143]	Constrained R-CNN	Splicing, copy-move, and removal.	0.992 (AUC)
	[144]	Dense fully convolutional encoder-decoder architecture with dense connections and dilated convolutions	Commonly used editing tools and operations in photoshop	0.99 (AUC)
	[145]	ResNet; image residuals	Remove; CNN inpainting	97.97 (precision)
	[146]	Multibranch CNN architecture	Copy-move	0.920 (F1-score)
	[147]	RGB-N, MSCNNs, DCNNs	Splicing, copy-move	0.7328 (precision)
	[148]	Feature pyramid network, stagewise-weighted cross-entropy Loss function	JPEG compression, scaling	0.9967 (F1-score)
	[149]	CNN with CRF-based attention model	Splicing, copy-move	0.804 (F1-score)
	[150]	Dense self-attention encoders	Copy-move	0.883 (AUC)

extraction to find the noise inconsistency between the real and tampered regions. Then, the bilinear fusion of the features from the two streams is conducted to enhance the tamper identification ability. The algorithm achieved better detection performance than each stream and the comparison method. For detecting patch-based image inpainting, Wang

et al. [151] proposed a mask regional convolutional neural network (Mask R-CNN) approach by adjusting the sizes of the anchor scales due to the inpainting images and then by replacing the original non-maximum suppression single threshold with an improved non-maximum suppression (NMS) to reduce the missed detection areas and improve detection accuracy. Lu et al. [128] proposed an image inpainting forgery detection method based on LSTM-CNN. This method used the strong learning ability of CNN to search abnormal similar blocks and used the LSTM network to eliminate the influence of texture consistent region on the detection results.

(2) Operation Feature-Based Method. Chen et al. [129] made the first layer be the filter layer that received the input of images and outputs them as media filtering residual to improve the effect of CNN. The author proposed to use deep learning technology to solve the problem of median filter forensics for the first time, which is the first work of deep learning in the field of forensics, and also played a guiding role in the combination of follow-up deep learning and image forensics.

To efficiently segment various types of manipulations including copy-move, object removal, and splicing, Bappy et al. [130] proposed a high-confidence manipulation localization architecture, which included CNN architecture to provide spatial feature maps of manipulated objects and LSTM to observe the transition between manipulated and nonmanipulated patches by using resampling features of the patches, and a decoder network to learn the mapping from low-resolution feature maps to pixel-wise predictions for image tamper localization. Aiming at the problem of copy-move forgery location, Wu et al. [131] used CNN to extract block features from the image. Then, the feature points were matched according to the autocorrelation between different blocks. Finally, the forgery location results were generated by a deconvolution neural network. Wu et al. [132] added another branch structure based on the similarity detection network [122] to detect the forged target area of the image. Then, they proposed a two-way fusion BusterNet model, which can preliminarily distinguish the original area of the image, the forged source area, and the forged target area, but there is still a problem of low detection accuracy. Muzaffer et al. [133] used AlexNet to extract the feature vectors of images and studied the similarity between them for copy-move forgery detection and location. Zhong et al. [134] proposed using a dense network structure to solve the problem of copy-move forgery detection and location. First, the pyramid feature extractor was used to extract multidimensional and multiscale features. Second, the feature correlation matching module was used to independently learn the correlation of features. Finally, the forgery location results were obtained through the postprocessing module. Zhu et al. [135] introduced the adaptive attention mechanism and the neural network structure of residual optimization into the copy-move forgery location task. Based on obtaining the coarse mask, the result was optimized through the residual subdivision module to obtain the final forgery location result.

(3) CNN Structure or Loss Function Changed Method. Zhu et al. [136] proposed a CNN-based approach which was built following the encoder-decoder network structure to detect patch-based inpainting operation by predicting the inpainting probability for each pixel in an image. By assigning a class label for each pixel of an image can guide the CNN to automatically learn the inpainting features. Using the weighted cross-entropy as the loss function can supervise the CNN to capture the manipulation information. Zhu et al. [137] proposed an image inpainting forensic method based on deep CNN. This method used an encoder network to extract image features and a decoder network to restore the feature map extracted by the encoder network to the original image size. Then, the feature pyramid network is used to supplement the information of the feature map in the upsampling process. Yang et al. [138] proposed a Laplacian CNN algorithm to detect and reacquire images which put signal enhancement layer into CNN structure, and Laplacian filter was used in the signal enhancement layer. Liu et al. [139] proposed a CNN and segmentation-based multiscale localization to analyze the tampered area in a digital image. The author designed the unified CNN architecture to get a set of tampering detectors for different scales to generate a series of complementary tampering possibility maps. Then, the proposed segmentation-based method fused these maps and generated the final decision map. Liu et al. [140] used convolution kernel network for feature extraction in copy-move forgery detection and location tasks and used adaptive segmentation methods to obtain the forgery location results. In order to solve the problem that the current deep learning model adjusts the size of the input image to destroy the valuable high-frequency details and seriously affect the forensic effect, Marra et al. [141] proposed a CNN-based image forgery detection framework based on full-resolution information gathered from the whole image. Using gradient check pointing, the framework was trainable end-to-end with limited memory resources and weak supervision, allowing for the joint optimization of all parameters. Bammey et al. [142] designed a CNN structure based on demosaicing algorithms that can train directly on unlabeled and potentially forged images to point out local mosaic inconsistencies and then classify image blocks by their position in the image. To address the deep learning-based models lacking poor universality of handcrafted features and only focusing on manipulation localization and overlooking manipulation classification, Yang et al. [143] proposed constrained R-CNN for complete and accurate image forensics. It included a learnable manipulation feature extractor to create a unified feature representation of various content manipulation directly from data and a two-stage architecture to simulate the coarse-to-fine forensic process in the real world.

To well capture tampering traces left by Photoshop, Li et al. [144] proposed a fully convolutional encoder-decoder architecture with dense connections and dilated convolutions for achieving better localization performance. For effectively training a model in the case of insufficient tampered images, it designed a training data generation strategy by resorting to Photoshop scripting, which can

imitate human manipulations and generate large-scale training samples. Li et al. [145] found that in the residual region of the image, the transfer probability value of adjacent pixels in the region repaired by deep learning is much lower than that of adjacent pixels in the unrepaired region, indicating that the repaired image contains fewer high-frequency components. According to this law, the author first set up a learnable prefiltering module to extract the image residual, then used four consecutive ResNet V2 modules for feature extraction, and finally carried out upsampling to realize the accurate positioning of the repaired area. Barni et al. [146] designed a multibranch CNN architecture that solved the hypothesis testing problem by learning a set of features capable to reveal the presence of interpolation artifacts and boundary inconsistencies in the copy-moved area. Zhang et al. [147] proposed a hybrid architecture called Pixel-level Image Tampering Localization Architecture (PITLArc) by integrating the advantages of top-down detection-based methods and bottom-up segmentation-based methods. To evaluate the effectiveness, they used one outstanding detection-based method (dual-stream faster region-based convolutional neural network (RGB-N)) and two segmentation-based methods (multiscale convolution neural networks (MSCNNs) and dual-domain convolutional neural networks (DCNNs)) to implement the PITLArc. Zhang et al. [148] improved U-shaped net to migrate feature pyramid network (FPN) for multiscale inpainting feature extraction and designed a stagewise-weighted cross-entropy loss function to take advantage of both the general loss and the weighted loss to improve the prediction rate of inpainted regions of all sizes. Rao et al. [149] found the key observation that the boundary transition artifacts arising from the blending operations are ubiquitous in various image forgery manipulations, which is well characterized with CRF (conditional random field)-based attention model. Therefore, an image forgery detection and localization scheme is proposed based on a deep convolutional neural network (CNN) and CRF-based attention model. Hao et al. [150] proposed a novel image localization method inspired by transformers called TransForensics. It had two major components which were dense self-attention encoders and dense correction modules. The dense self-attention encoders were used to model global context and all pairwise interactions between local patches at different scales. The dense correction modules were used for improving the transparency of the hidden layers and correcting the outputs from different branches. Experiments showed that TransForensics could not only capture discriminative representations and obtain high-quality mask predictions but were also not limited by tampering types and patch sequence orders.

3.2.2. GAN-Based Forensics. According to the different purposes of GAN-based forensics, it can be divided into forensics for classification and forensics for localization. Table 6 lists the classification, source, detection type, methods, and the best performance in each article of GAN-based forensics. Their specific descriptions are as follows.

(3) GAN-Based Forensics for Classification. Li et al. [152] statistically analyzed each component of the original pixel domain and residual domain of the image in three color spaces (RGB, HSV, and YCbCr). Then, they found the difference between the GAN tampered image and the real image in the residual domain of different color spaces. After analyzing, they got the component with the most significant difference (i.e., H, S, Cb, and Cr) from the statistical results. According to the above statistical results, it first selected the required channels from different color spaces. Second, the residual characteristics of color components and the co-occurrence matrix were calculated. Third, connected them into feature vectors and finally trained classifiers to predict whether the image is real or generated by the deep network. Nataraj et al. [153] proposed a method to detect GAN-tampered images using a three-channel co-occurrence matrix. In this method, the author did not calculate the co-occurrence matrix on the residual domain. Instead, they directly used the three channels of the input image to calculate the co-occurrence matrix, then inputted it into the convolution network to extract features, and finally used the fully connected layer for classification. McCloskey et al. [154] extracted features from color and saturation space to detect GAN-generated images. For the color feature, the INH network is used to classify GAN-generated images by using the bivariate histograms of r, g components in the standard rg chromaticity space. For the saturation feature, the SVM is used to classify GAN-generated images by extracting two groups of saturation. According to the report, saturation statistics provided better performance. From the perspective of artificial fingerprints, Marra et al. [155] explored a PRNU-based scheme for Cycle-GAN, ProGAN, and Star-GAN-generated image detection. The results demonstrated that those GAN schemes would leave artificial fingerprints into the generated images. Ning et al. [156] proposed a method to realize multiclassification through fingerprint matching. The author believed that each GAN network model will have its unique fingerprint affected by training data, network structure, loss function, parameter setting and other factors. The fingerprint of the GAN model will affect its generated image and leave a unique image fingerprint. Therefore, they designed a forensic method to extract GAN model fingerprint and image fingerprint and then matched them to realize image classification. No matter what types of images are generated by various GAN and what network structure is used, the synthetic false images have the same defects. Wang et al. [157] explored how to use a single GAN model to identify other GAN-generated images. First, 11 GAN models were used to construct a large-scale synthetic image authentication database. Thereafter, only using a single ProGAN model to train can show good generalization performance on the above database. Experiments show that data enhancement as a postprocessing method and the diversity of training data is the key factor to success. Mo et al. [158] expected that the main difference between the original and GAN-generated images would be reflected on the residual domain. Therefore, they presented a three-layer CNN with a Laplacian filter preprocessing to identify fake face images generated by PGGAN. Marra et al.

Table 6: Comparison of GAN-based forensics.

Category	Source	Detection type	Methods	Performance
Classification	[152]	Deep network-generated images	Residual domain of chrominance components	100% (ACC)
	[153]	Cycle-GAN Star-GAN	Co-occurrence matrices	93.42 (accuracy) 99.49 (accuracy)
	[154]	GAN-generated images	INHNet Saturation feature	0.56 (AUC) 0.7 (AUC)
	[155]	Cycle-GAN, ProGAN, Star-GAN	GAN fingerprints	0.999 (AUC)
	[156]	GAN-generated images	GAN fingerprints	99.93% (accuracy)
	[157]	11 GAN models	Single ProGAN	93.0 (mAP)
	[158]	PGGAN	Lap-CNN	96.3%
	[159]	Cycle-GAN	Eight methods	>83.58%
	[160]	GAN-generated human face images	CNN with self-attention mechanism	99.3 (accuracy)
	[161]	Five GAN models	EM; fingerprint	99.65 (accuracy)
	[162]	PGGAN	Shallow CNN architecture	99.99% (AUC)
	[163]	PGGAN	Xception	99.99% (accuracy)
	[164]	PGGAN, WGAN, Style-GAN, LSGAN, DCGAN	Global and local feature, ArcFace loss, CNN	99.99% (accuracy)
	[165]	BigGAN, Style-GAN2, PGGAN	R, *G*, and B components, DWT, SVM	98.45% (accuracy)
Localization	[166]	Criminisi, SN-PatchGAN	Multitask deep learning network based on mask R-CNN	97.8 (mAP)
	[167]	GAN-based face manipulation	Gray-scale fakeness prediction map; encoder-decoder architecture with attention mechanism	99.95 (ACC)

[159] evaluated the performance of several image forensic detectors and popular computer vision CNN architectures on GAN-generated images detection. More specifically, the authors used four image forensic detectors and four CNN architectures. Experimental results showed that XceptionNet has the highest average detection accuracy. Mi et al. [160] found that the upsampling process conducted by the Transposed Convolution operation was the most vulnerable link in GAN generators. The transposed convolution in the process will cause the lack of global information in the generated images. Then, the authors proposed a detection method with CNN and self-attention to improve detection accuracy. Guarnera et al. [146] found that a sort of fingerprint left in the image generation process is hidden in images. The authors extracted forensic traces through the EM algorithm and then used naive classifiers to get the classification.

In recent years, different approaches have been revealed for detecting GAN-generated images. Lee et al. [162] proposed the face manipulation detection pipeline which included facial image preprocessing and the proposed Shallow-FakeFaceNet (SFFN) detection model. The detection pipeline had two stages. The first stage was to perform facial image preprocessing to (1) crop the face region, (2) filter cropped faces, (3) apply upscaling methods, and (4) augment the data. The second stage is to pass the preprocessed faces to the SFFN models to train and detect facial manipulations. This method only used RGB channel information from the image to distinguish facial manipulated images. Based on the discovery that both the luminance and chrominance components play an important role, and the RGB and YCbCr color spaces achieve better performance than the HSV and Lab color spaces, Chen et al. [163] designed the dual-stream network to detect image by integrating both the luminance component and chrominance components of dual color spaces (RGB and YCbCr). Then, they improved Xception model by introducing the convolutional block attention module and multilayer feature aggregation module to enhance its feature representation power and aggregate multilayer features, respectively. Chen et al. [164] pointed out that the local features had played significant roles in the field of face recognition and face synthesis. For improving the generalization capability, Chen et al. [164] strengthened the learning on important local areas and combined the global and local features. Then, they proposed the method including the feature learning step and classification learning step. In the learning step, important local areas containing many face key points are learned by combining the global and local features. In the classification learning step, the metric based on the ArcFace loss is applied to extract common and discriminative features. Finally, the extracted features are fed into the classification module to detect GAN-generated faces. Tang et al. [165] proposed a detecting GAN-synthesized fake image method named fake image discriminator (FID) which used the strong spectral correlation in the imaging process of natural color images. FID includes the feature extraction stage and classification stage. In the feature extraction stage, the color image is converted into three color components of *R*, G, and B, then discrete wavelet transform (DWT) is applied to RGB components separately. In the classification stage, the correlation coefficient between the subband images was used as an input feature vector to SVM for authenticity classification.

(4) GAN-Based Forensics for Localization. Wang et al. [166] tried to use an improved mask R-CNN to detect the images generated by [15] and GAN-generated networks. Based on

the traditional mask R-CNN, the author added the local binary pattern channels to the network input, and combined feature pyramid networks and back connections to extract more features. Experiments show that the proposed method has good performance.

Using the imperfection of the upsampling procedure in all the GAN-based partial face manipulation methods and full-face synthesis methods, Huang et al. [167] proposed the FakeLocator method. First, ground truth fakeness maps are produced. Second, the real images or fake images are inputted to the encoder-decoder architecture network by introducing the attention mechanism to output gray-scale fakeness prediction maps. Last, the gray-scale fakeness prediction maps and the ground truth fakeness maps are combined to calculate the loss.

In summary, with the continuous emergence of various new technologies, forensic technology has made considerable progress and achieved good results. As the conclusion from Gragnaniello et al. [168] shows that we are still very far from having reliable tools for GAN image detection after analyzing and comparing seven detectors for GAN image detection. Table 7 shows the advantages and disadvantages of various forensic methods.

4. Game between Image Inpainting and Forgery

While forensic technology is making progress, forgers have not waited to die. After knowing the forensics algorithm, they will try their best to cover up the traces of image forgery or image operation and carry out the anti-forensic operation on the forensics so that the forensic algorithm can get wrong judgment results. At present, anti-forensic algorithms can be divided into orientation methods and general methods [169]. The orientation method is to remove a specific trace to invalidate the method of relying on this trace to obtain evidence. The general method modifies the features that the forensic method depends on to make the features consistent with the features of the image without operation, to make the forensics fail.

For the orientation method, Wang et al. [157] found that different GAN-generated fake images all leave specific fingerprint features. Although the generalization ability of detectors that rely on fingerprint feature training is not good, preprocessing the training data, such as adding JPEG compression, blur, and other operations, greatly improves the generalization performance of the model. At the same time, postprocessing the image during detection can increase the robustness of the model. Neves et al. [170] designed an automatic encoder, which can remove the fingerprint and other information from the synthetic forged image, making the existing forgery detection system invalid. Zhang et al. [171] looked for common traces of GAN to improve the robustness of the detector. The existing detectors have a strong dependence on data and lack generalization. Du et al. [172] used the automatic encoder with local sensitivity to realize the detection, making the model focus on the tampered area and have stronger universality. Cozzolino et al. [173] proposed the SpoC method which used a GAN-based approach that injected camera traces into synthetic images to deceive forensic detectors. To conceal or eliminate the traces left by multiple manipulating operations such as JPEG compression and median filtering successively, Wu et al. [174] investigated multioperation anti-forensics with GANs. Its generator network is trained to automatically learn the visual and statistical features of the original images by applying appropriate loss functions in the process of optimization. Then two strategies are given to validate the effectiveness.

In the general method, Marra et al. [159] simulated the detection of tampered pictures in the social network scene. The results showed that the existing detectors perform poorly in the real network confrontation environment. Brockschmidt et al. [175] evaluated the antagonism of deep forgery detectors (Xception [24] and Mesonet [159]). The author used six forgery datasets to detect the reliability of the detector. The results showed that the detector can achieve a very high detection rate on the same distributed datasets. However, in the dataset of unknown tamper type, only the datasets with high feature coincidence had good mobility, otherwise, the detection effect was very poor. Huang et al. [176] drew on the idea of adversarial samples, carried out adversarial attacks on these neural network-based detectors, and designed two methods—a single adversarial attack and a general adversarial attack—to make the tampering classification and positioning of the detector invalid. Most of the deep forgery detection algorithms used neural network technology, and the neural network itself had anti-sample attacks [177, 178]. The anti-sample attack was a technology that disturbed the model input and made the model misjudge, which made the deep forgery technology can hide some of its features to bypass the detection. Although many detectors performed well on some datasets, attackers could still improve the generation method and hid some symbolic features to bypass the detector, which was a long-term attack and defense game process.

When the anti-forensics is aware of the existence of forensics, they have devised different methods to achieve their goals. Chen [116] detected multiple operations with anti-forensic operations. It is assumed that the anti-forensics can obtain the forensic decision plane, modify the target image to reach the decision plane, or even surpass the decision plane, thereby deceiving the detector. It proposed a strategy of random feature selection and theoretically proved the effectiveness of the strategy. Considering the problem of camera source identification based on device fingerprints, Zeng et al. [179] studied the confrontation between analysts and fingerprint copy attackers in the case of incomplete information. The author established a Bayesian game model through the noise-level estimation algorithm. The results showed that the gains obtained by the forensics when they have incomplete information are lower than those obtained when they have complete information. A stronger confrontation is considered in the game theory framework of [180, 181], that is, anti-forensics can destroy training samples to interfere with forensic analysis. Evidence collectors and anti-forensics were both working towards their side, making both parties enter a ring with no end. Ding et al. [182] proposed an anti-forensic method based on the GAN

Table 7: Advantages and disadvantages of various detection methods.

Methods			Advantages	Disadvantages
Traditional forensic method	Image generation feature-based forensics		Mature technology and interpretable features	Dependence on specific Features; weak universality
	Image operation feature-based forensics	Single operation forensics	Pay attention to the local information of the image	Dependent on specific features
		Multioperation forensics	Learning local information of image	Weak universality and insufficient robustness
Deep learning-based forensics		Deep neural network-based forensics	Large amount of data; more learning information and high accuracy	Dependence on distributed datasets and unknown types have a great impact on performance
		GAN-based forensics	Focus on GAN fingerprint information; high accuracy	Strong data dependence; poor interpretability

model with two input channels. One is for the real frames and the other is for the corresponding Deepfake ones. Paired images are inputs to the GAN model for adversarial training. In order to a better result, anti-forensic abilities are defined and a loss function is designed. Xie et al. [183] proposed an anti-forensic framework called dual-domain generative adversarial network (DDGAN). It consisted of a generator and two discriminators working on the operation-specific feature domain which helps to conceal the artifacts from the perspective of forensic analysis for the target task, and the spatial domain which facilitates taking advantage of machine-learned features from the scratch as a supplementary.

5. Dataset for Image Forensics

To study the feasibility, effectiveness, and robustness of forensic algorithms, different researchers, scientific research institutions and companies have released datasets for forensics, covering all categories. According to the purpose of the dataset, it can be divided into splicing, camera-based forensics, double JPEG compression, copy-move, GAN-generated image forensics, and various manipulations. Their detailed list is shown in Table 8 and its introduction is as follows.

5.1. Dataset for Splicing. The Columbia dataset is one of the first ones made available to the forensic community. The Columbia dataset contains the gray version of Columbia [187] released in 2004 and the color version of Columbia [188] released in 2006. The gray version of the dataset contains 1845 image blocks (128×128 pixels) with different contents extracted from the images on the CalPhotos website, including 933 real image blocks and 912 mosaic image blocks as well as a small group of 10 images taken by the authors. The number of real images and splicing images in the dataset is roughly the same. The real image and the splicing image are separate local blocks of fixed size (128×128 pixels). The blocks are kept at a reasonable size to ensure that the empirical data of each block can be used to estimate sufficiently accurate statistical characteristics. Color version of Columbia dataset comprises a total of uncompressed 363 images. 183 of them are authentic images and 180 are spliced ones. For the splicing image, there was not any type of postprocessing put on the spliced region and it can be identified easily by our eyes. Based on the above reason, fine-tuning, training, and testing are not recommended to perform.

For this reason, CASIA [189] and DSO-1 [190] datasets are released. For the CASIA dataset, it has 1.0 and 2.0 versions. The 1.0 version provided splicing with sharp boundaries and is easily detectable. To meet practical needs, the 2.0 version is released. The tampered image in CASIA 2.0 preprocesses the tampered part before tampering: for example, scaling, deformation, rotation, and other operations are performed on the tampered part before pasting to the target image; some blur operations are carried out on the edge of the tampered area or inside the tampered area. Therefore, the dataset can better represent most of the spliced tampered images in daily life. DSO-1 [190] realistic dataset is a subset taken from the IEEE Image Forensics Challenge. It provided the images with uncompressed PNG format, but most of them had been JPEG compressed before. Unfortunately, DSO-1 only provided fixed image resolution and missed information such as how to create a dataset and how many cameras are used.

5.2. Dataset for Camera-Based Forensics. Dataset for camera-based forensics mainly includes the dataset proposed in [184], Dresden [191], and IMD2020 [192]. For identifying the camera from sensor fingerprints, [184] utilized Flickr to form the dataset which contained 1053580 pictures taken by 6896 cameras covering 150 camera models to evaluate the camera identification based on sensor fingerprint. Dresden database in [191] is specially used for the development and benchmarking of camera-based digital forensic technology. A total of 73 digital cameras collected more than 14000 images of various indoor and outdoor scenes. The camera models involve 25 to ensure that device-specific and model-specific characteristics can be separated and studied separately. In addition, auxiliary images for estimating device-specific sensor noise patterns are collected for each camera. To study the model-specific JPEG compression algorithm, another set of images has been compiled for each model. In Novoz´amsk´y et al. [192], the authors identified the majority of camera brands and models on the market, which resulted in 2,322 camera models. Then, they collected a dataset named IMD2020, which included 35,000 real images captured by these

Table 8: The detailed list of datasets.

Forensic type	Dataset name		Started year	No. original/forged	Image size	Format
Splicing	Columbia	gray	2004	933/912	128×128	BMP
		Color	2006	183/180	757 × 568~1152 × 768	Raw, BMP
	DSO-1		2013	100/100	2048 × 1536	PNG
	CASIA	1.0	2013	800/921	384 × 256	JPG
		2.0	2013	7491/5123	240 × 160 ~ 800 × 600	BMP, TIF, JPG
Camera-based forensics	[184]		2009	1053580/—	1600 × 1200 ~ 3072 × 2048	JPEG
	Dresden		2010	Over 14000/—	2592 × 1944 ~ 4032 × 3024	Raw, JPEG
	IMD2020		2020	37000/—	Various	—
Double JPEG compression	[185]		2012	100/110	256 × 256 ~ 1024 × 1024	TIFF
	[186]		2018	18946	256 × 256	RAW
Copy-move	MICC F220		2011	110/110	722 × 48 ~800 × 600	JPG
	MICC F2000		2011	1,300/700	2048 × 1536	JPG
	FAU		2012	48/48	2362 × 1581~ 3888 × 2592	PNG, JPG
	CoMoFoD		2013	260/160	512 × 512 ~ 3 000 × 2 000	PNG, JPG
	GRIP		2015	80/80	768 × 1024	PNG
	COVERAGE		2016	100/100	400 × 486	TIF
GAN-generated forensics	[159]		2018	More than 18,432/18,432	256 × 256	—
	FFHQ		2019	70000/-	1024 × 1024	PNG
	[160]		2020	3,000/3,000	256 × 256	JPG
	IMD2020		2020	37,000/37,000	Various	—
Various manipulations	NC2017		2017	2667/1410	160 × 120~8000 × 5320	Raw, PNG, BMP, JPG
	PS-battles		2018	11,142/102,028	130 × 60~ 10,000 × 8558	PNG, JPG
	MFC2018		2018	14,156/3,265	128 × 104~ 7952 × 5304	Raw, PNG, BMP, JPG, TIF
	MFC2019		2019	10,279/5,750	160 × 120~2624 × 19680	Raw, PNG, BMP, JPG, TIF
	DEFACTO		2019	–/229,000	240 × 320~ 640 × 640	TIF

camera models. The authors also created a dataset of 2,000 real-life manipulated images and corresponding real images.

5.3. Dataset for Double JPEG Compression. For testing double JPEG compression, Bianchi et al. [185] provided a realistic dataset that was captured by three different digital cameras. In 2018, for training deep networks, Yang et al. [138] released a dataset by collecting 1120 different quantization tables from actual requested images which were generated by mixing all quality factors.

5.4. Dataset for Copy-Move. For copy-move, two ground truth databases for CMFD algorithms, called MICC F220 and MICC F2000 consisting of 200 and 2000 images, respectively, were released by Park et al. [186].In each of these datasets, half of the images have been tampered with. The image size is 2048 × 1536 pixels. The type of processing on the copy-move forgeries is limited to rotation and scaling. Additionally, the source files are not available. Thus, adding noise or other artifacts to the copied region is not feasible. To address these issues, many datasets have been released by [99, 193–196]. Some of them simply copy one object from one position to paste it to another position in the image, and some of them copy an object and then use rotation, resizing, and change of illumination operation on it to express the actual operation as truthfully as possible.

5.5. Dataset for GAN-Generated Image Forensics. For the detection of GAN-generated images, the current datasets mainly include FFHQ, CelebA-HQ, IMD2020, and other GAN-generated datasets [159]. The following mainly introduces FFHQ, CelebA-HQ, and IMD2020.

FFHQ (Flickr-faces-high quality), originally created as the benchmark of GAN, is also used in the training dataset of StyleGAN [54, 55] and is open-sourced by NVIDIA in 2019. FFHQ is a high-quality face dataset containing 70,000 high-definition face images in PNG format with a resolution of 1024 × 1024. They are rich and diverse in age, race (many face attributes), and image background and have obvious differences.

The datasets in Mi et al. [160] focused on human faces which consisted of the real face set with 30,000 images from the CelebA-HQ dataset [197], and the fake face set (GAN-generated human face images) with 30,000 images that NVIDIA open-sourced as the exemplary images generated by PGGAN [180]. In order to control variables and rule out possible interference, the 1024 × 1024 images from both sets are resized to 256 × 256 using bilinear interpolation.

Marra et al. [159] built a large dataset of samples of different categories by image-to-image translation using the code available online (https://github.com/junyanz/CycleGAN). The dataset included more than 36K 256 × 256 color images; and half of them are real images, and the other half are GAN generated.

In the study by Novoz´amsk´y et al. [192], the authors also created the same number of digitally manipulated images by using a large variety of main image manipulation methods as well as advanced ones such as GAN or inpainting resulting in a dataset of 70,000 images. The real versions of these images are also provided. The authors also manually created binary masks localizing the exact manipulated areas of these images.

5.6. Dataset for Various Manipulations. To test the robustness of the algorithms, NC2016, NC2017, MFC2018, and MFC2019 [197] are released by the U.S. National Institute of Standards and Technology (NIST). For the NC2016 has some redundancies, for example, each spliced photo is presented four times, JPEG compressed at low and high quality, and with and without postprocessing on the splicing boundaries. Therefore, NIST published NC2017, MFC2018, and MFC2019 datasets in 2017, 2018, and 2019 without the above redundancies. These datasets presented the various types of manipulations, resolutions, formats, compression levels, and acquisition devices. Moreover, multiple manipulations are often carried out on the same image and even on the same objects. Overall, they represent a very challenging and reliable testbed for new proposals. Heller et al. [198] presented the PS-Battles dataset which was gathered from a large community of image manipulation enthusiasts and provided a basis for media derivation and manipulation detection in the visual domain. It consisted of 102,028 images grouped into 11,142 subsets, each containing the original image as well as a varying number of manipulated derivatives.

Mahfoudi et al. [199] presented a dataset named DEFACTO for image and face manipulation detection and localization. The DEFACTO was automatically generated using Microsoft common object in context database (MSCOCO) to produce semantically meaningful forgeries. It generated Splicing forgeries, copy-move forgeries, object removal forgeries, and morphing forgeries. Over 200000 images have been generated, and each image is accompanied by several annotations allowing precise localization of the forgery and information about the tampering process.

6. Summary and Prospect

6.1. Summary. With the development of image inpainting technology, image forgery technology has become more and more sophisticated. This brought great negative impact to the country and society. At the national level, the negative impact of forged pictures will have an irreversible impact on the government or politicians. If the evidence of forged pictures is accepted by the court, it will doubt the credibility of the government. At the social level, forged pictures of individuals seriously damage personal reputation. The negative of forged pictures make citizens lose confidence in the stock market. Forged pictures are certified by equipment, which has a huge security risk to personal credit and assets. These risks also hide deeper social issues such as national security and stability, ethics, economic development, and trust crises, and more effective countermeasures are urgently needed.

6.2. Prospect. At present, the forensic method based on specific features has formed the forensic theory and achieved fruitful research results, which has gotten rid of the trap of forgery to a certain extent. However, we should also be aware that there are still many key problems to be solved in this field, especially the rapid development of image inpainting technology, which makes identification and forensics in a passive and disadvantageous position. To avoid this dilemma and look forward to the future, we can consider exploring the feasible direction of detection and forensics in the future from multiple angles and levels. At the national level, relevant laws and regulations are promulgated to clearly stipulate the responsibility for forging and tampering with images and to clarify the behavioral norms of forgers. At the social level, for researchers, the universal and robust detection algorithms are studied, as many deep forgery types as possible are explored, and the common features are found. By learning the common features, the detection model can be applied to more forgery types. Second, scattered data resources are concentrated and a unified detection and forensic group is established, so that researchers can make better use of existing resources and achievements, regularly hold academic seminars and competitions, and increase researchers' attention to the field of deep forgery detection. Finally, with the development of existing traceability technology and blockchain technology, forensics can gain some inspiration and combining them with current forensic technology can achieve complementary advantages.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant no. 61862051; the Science and Technology Foundation of Guizhou Province under Grant no. [2019]1447; the Youth Philosophy and Social Science Foundation of Guizhou Province under Grant no. 18GZQN36; the Nature Science Foundation of Educational Department under Grant nos. [2022]094 and [2022]100; and the Nature Science Foundation of Qiannan Normal University for Nationalities under Grant no. QNSYRC201714.

References

[1] H. Farid, "Digital image forensics," *Scientific American*, vol. 298, no. 6, pp. 66–71, 2008.
[2] wwwchinadailycom, "www.chinadaily.com," 2015, http://www.chinadaily.com.cn/interface/yidian/1120783/2015-12-04/cd.22630382.html.
[3] politics.people.com, "politics.people.com," 2015, http://politics.people.com.cn/n/2015/0303/c70731-26629561.html.
[4] Deepfakes, "Deepfakes," 2019, https://github.com/deepfakes/faceswap.
[5] ZaoApp, "Zao App," 2019, https://zao-app.com/.

[6] H. Farid, "How to detect faked photos," *American Scientist*, vol. 105, no. 2, pp. 77–81, 2017.

[7] E. Kee and J. F. O, Brien and H. Farid, "Exposing photo manipulation from shading and shadows," *ACM Transactions on Graphics*, vol. 33, no. 5, pp. 1–21, 2014.

[8] S. Pittala, E. Whiting, and H. Farid, "A 3-d stability analysis of lee Harvey Oswald in the backyard photo," *Journal of Digital Forensics, Security and Law*, vol. 10, no. 3, 2015.

[9] H. F. Tang and Y. F. Dong, "Survey of image inpainting algorithms based on deep learning," *Computer Science*, vol. 47, no. S2, pp. 151–164, 2020.

[10] M. H. Yap, N. Batool, and C. Ng, "A Survey on Facial Wrinkles Detection and Inpainting: Datasets, Methods, and Challenges," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, 2021.

[11] M. C. Stamm, M. Min Wu, and K. J. R. Liu, "Information forensics: an overview of the first decade," *IEEE Access*, vol. 1, pp. 167–200, 2013.

[12] E. Nowroozi, A. Dehghantanha, R. M. Parizi, and K.-K. R. Choo, "A survey of machine learning techniques in adversarial image forensics," *Computers & Security*, vol. 100, Article ID 102092, 2021.

[13] Z. P. Chen, *Research on Digital Image Forensics of Processing History*, Beijing Jiaotong University, Beijing, China, 2019.

[14] Y. Liang, *Research on Key Technologies of Digital Image Forensics Based on Deep Neural Network*, Xidian University, Xi'an, China, 2019.

[15] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[16] H. Wang, L. Jiang, R. Liang, and X.-X. Li, "Exemplar-based image inpainting using structure consistent patch matching," *Neurocomputing*, vol. 269, pp. 90–96, 2017.

[17] Z. Chen, C. Dai, L. Jiang et al., "Structure-aware image inpainting using patch scale optimization," *Journal of Visual Communication and Image Representation*, vol. 40, pp. 312–323, 2016.

[18] A. S. M. Jiao, P. W. M. Tsang, and T.-C. Poon, "Restoration of digital off-axis Fresnel hologram by exemplar and search based image inpainting with enhanced computing speed," *Computer Physics Communications*, vol. 193, pp. 30–37, 2015.

[19] C. Su, T. Fu, X. Zhang, J. Ren, and L. Jin, "Adaptively-weighted blind image restoration algorithm based on energy constraint," *Acta Optica Sinica*, vol. 38, no. 2, Article ID 0210001, 2018.

[20] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," *Lecture Notes in Computer Science*, vol. 9, no. 1, pp. 599–619, 2012.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Advances in Neural Information Processing Systems*, pp. 1097–1105, MIT Press, Cambridge, MA, USA, 2012.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[23] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014, https://arxiv.org/abs/1409.1556.

[24] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition (CVPR)*, pp. 1251–1258, Honolulu, HI, USA, July 2017.

[25] C. Szegedy, W. Liu, and Y. Jia, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, June 2015.

[26] G. Huang, Z. Liu, and L. Van Der Maaten, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition (CVPR)*, pp. 4700–4708, Honolulu, HI, USA, July 2017.

[27] H. Jie, S. Li, and S. Albanie, "Squeeze-and-Excitation Networks," 2017, https://arxiv.org/abs/1709.01507.

[28] J. Gu, Z. Wang, J. Kuen et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[29] K. Sasaki and S. Iizuka, "Joint gap detection and inpainting of line drawings," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5725–5733, IEEE, Honolulu, HI, USA, June 2017.

[30] D. Pathak, P. Krahenbuhl, and J. Donahue, "Context encoders: feature learning by inpainting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536–2544, IEEE, Las Vegas, NV, USA, June 2016.

[31] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proceedings of the European Conferenceon Computer Vision (ECCV)*, pp. 89–105, Glasgow, UK, August 2018.

[32] R. A. Yeh, C. Chen, and T. Y. Lim, "Image inpainting with deep generative models," in *Proceedings of the 2017 IEEE-Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5485–5489, IEEE, New York, NY, USA, June 2017.

[33] Y. Zeng, J. Fu, H. Chao, and B. Guo, "Learning pyramid-context encoder network for high- quality image inpainting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1486–1494, IEEE, Long Beach, CA, USA, June 2019.

[34] J. Yu, Z. Lin, and J. Yang, "Free-form image inpainting with gated convolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4471–4480, Long Beach, CA, USA, June 2019.

[35] C. Yang, X. Lu, and Z. Lin, "High-resolution image inpainting using multi-scale neural patch synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6721–6729, IEEE, New York, NY, USA, June 2017.

[36] X. Hong, P. Xiong, and R. Ji, "Deep fusion network for image completion," in *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 2033–2042, France, October 2019.

[37] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan, "Shift-net: image inpainting via deep feature rearrangement," 2018, https://arxiv.org/abs/1801.09392.

[38] N. Wang and J. Li, "MUSICAL: Multi-Scale Image Contextual Attention Learning for Inpainting," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 1–19, Yokohama, August 2019.

[39] H. Liu, "Coherent Semantic Attention for Image Inpainting," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)* IEEE, Seoul, Korea (South), June 2019.

[40] T. Yu, Z. Guo, X. Jin et al., "Region normalization for image inpainting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, Article ID 12733, 2020.

[41] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[42] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks in computer vision," *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–38, 2022.

[43] M. Mehdi and O. Simon, "Conditional generative adversarial nets," 2014, https://arxiv.org/abs/1411.1784.

[44] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proceedings of the International Conference on Machine Learning (PMLR)*, pp. 2642–2651, Paris, France, July 2017.

[45] X. Chen and Y. Duan, "Infogan: interpretable representation learning by information maximizing generative adversarial nets," in *Proceedings of the 30th International Conference on Neural Information Processing Systems (NeuralIPS)*, pp. 2180–2188, Seoul, Korea (South), December 2016.

[46] Y. Yu, Z. Gong, P. Zhong, and J. Shan, "Unsupervised representation learning with deep convolutional neural network for remote sensing images," in *Proceedings of the International Conference on Image and Graphics*, pp. 97–108, Springer, China, September 2017.

[47] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–14, 2017.

[48] H. Zhang, T. Xu, and H. Li, "Stackgan: text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer vision (ICCV)*, pp. 5907–5915, IEEE, Venice, Italy, October 2017.

[49] H. Zhang, T. Xu, H. Li et al., "Stackgan++: realistic image synthesis with stacked generative adversarial networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1947–1962, 2019.

[50] J. Y. Zhu, T. Park, and P. Isola, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer vision (ICCV)*, pp. 2242–2251, IEEE, Venice, Italy, August 2017.

[51] J. Yu, Z. Lin, and J. Yang, "Generative image inpainting with contextual attention," in *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition (CVPR)*, pp. 5505–5514, Salt Lake City, UT, USA, October 2018.

[52] C. Zheng, T. J. Cham, and J. Cai, "Pluralistic image completion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1438–1447, Long Beach, CA, USA, June 2019.

[53] T. Karras, T. Aila, and S. Laine, "Progressive Growing of gans for Improved Quality, Stability, and Variation," 2017, https://arxiv.org/abs/1710.10196.

[54] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4401–4410, Long Beach, CA, USA, June 2019.

[55] T. Karras, S. Laine, and M. Aittala, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8110–8119, Seattle, WA, USA, June 2020.

[56] A. Brock, J. Donahue, and K. Simonyan, "Large Scale GAN Training for High Fidelity Natural Image Synthesis," 2018, https://arxiv.org/abs/1809.11096.

[57] X. Mao, Q. Li, and H. Xie, "Least squares generative adversarial networks," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2813–2821, IEEE, Istanbul, Turkey, January 2018.

[58] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, https://arxiv.org/abs/1701.07875.

[59] I. Gulrajani, F. Ahmed, and M. Arjovsky, "Improved Training of Wasserstein GANs," 2017, https://arxiv.org/abs/1704.00028.

[60] H. Li, G. Li, L. Lin, H. Yu, and Y. Yu, "Context-aware semantic inpainting," *IEEE Transactions on Cybernetics*, vol. 49, no. 12, pp. 4398–4411, 2019.

[61] H. Zhang, I. Goodfellow, and D. Metaxas, "Self-attention generative adversarial networks," in *Proceedings of the International Conference on Machine Learning (PMLR)*, pp. 7354–7363, Paris, France, May 2019.

[62] A. R¨ossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: learning to detect manipulated facial images," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1–11, Seoul, Korea (South), November 2019.

[63] H. Naeem, F. Ullah, M. R. Naeem et al., "Malware detection in industrial internet of things based on hybrid image visualization and deep learning model," *Ad Hoc Networks*, vol. 105, Article ID 102154, 2020.

[64] L. Zhu, J. Wang, X. Luo, and Y. Zhang, "PRUDA: A Novel Measurement Attribute Set towards Robust Steganography in Social Networks," *Security and Communication Networks*, vol. 2021, Article ID 9864833, 13 pages, 2021.

[65] H. Naeem, "Detection of malicious activities in internet of things environment based on binary visualization and machine intelligence," *Wireless Personal Communications*, vol. 108, no. 4, pp. 2609–2629, 2019.

[66] wwwdarpamil, "www.darpa.mil," 2019, https://www.darpa.mil/program/media-forensics.

[67] T. C. Ifs, "The 1st IEEE IFS-tc image forensics challenge," 2013, http://ifc.recod.ic. unicamp.br/fc.website/index.py.

[68] N. Krawetz, "FotoForensics, Neal Krawetz," 2020, http://fotoforensics.com.

[69] Multimedia Computing Laboratory, "MMC Image Forensic Tool image Watermarking Tool," 2015, http://rtlab.kaist.ac.kr.

[70] S. Duan, H. Wang, and Y. Liu, "A Novel Comprehensive Watermarking Scheme for Color Images," *Security and Communication Networks*, vol. 2020, Article ID 8840779, 15 pages, 2020.

[71] W. Qi, Y. Liu, S. Guo, and X. Wang, "An Adaptive Visible Watermark Embedding Method Based on Region Selection," *Security and Communication Networks*, vol. 2021, Article ID 6693343, 10 pages, 2021.

[72] P. Zhang, L. Wang, and W. Wang, "A Blockchain System Based on Quantum-Resistant Digital Signature," *Security and Communication Networks*, vol. 2021, Article ID 6671648, 13 pages, 2021.

[73] M. K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *Proceedings of the 7th Workshop on Multimedia and Security*, pp. 1–10, New York, NY, USA, August 2005.

[74] Q. Qiguang Liu, X. Xiaochun Cao, C. Chao Deng, and Xiaojie Guo, "Identifying image composites through shadow matte consistency," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1111–1122, 2011.

[75] H. Farid and J. Kosecka, "Estimating planar surface orientation using bispectral analysis," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2154–2160, 2007.

[76] Y. Zhang, T. Liu, C. Cattani, Q. Cui, and S. Liu, "Diffusion-based Image Inpainting Forensics via Weighted Least Squares Filtering Enhancement," *Multimedia Tools and Applications*, vol. 80, pp. 1–15, 2021.

[77] U. A. Ciftci, I. Demir, and L. Yin, "Fakecatcher: Detection of Synthetic Portrait Videos Using Biological Signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, 2020.

[78] H. Yao, S. Wang, Y. Zhao, and X. Zhang, "Detecting image forgery using perspective constraints," *IEEE Signal Processing Letters*, vol. 19, no. 3, pp. 123–126, 2012.

[79] V. Itier, O. Strauss, L. Morel, and W. Puech, "Color noise correlation-based splicing detection for image forensics," *Multimedia Tools and Applications*, vol. 80, no. 9, Article ID 13215, 2021.

[80] J. Luka, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.

[81] A. K. Jaiswal and R. Srivastava, "Forensic image analysis using inconsistent noise pattern," *Pattern Analysis & Applications*, vol. 24, no. 2, pp. 655–667, 2021.

[82] F. Bellavia, M. Fanfani, C. Colombo, and A. Piva, "Experiencing with electronic image stabilization and PRNU through scene content image registration," *Pattern Recognition Letters*, vol. 145, pp. 8–15, 2021.

[83] D. Cozzolino, F. Marra, D. Gragnaniello, G. Poggi, and L. Verdoliva, "Combining PRNU and noiseprint for robust and efficient device source identification," *EURASIP Journal on Information Security*, vol. 2020, no. 1, 2020.

[84] X. Lin and C. T. Li, "On constructing A better correlation predictor for PRNU-based image forgery localization," in *Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, Shenzhen, China, July 2021.

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

[86] P. Ferrara, T. Bianchi, A. De Rosa, and A. Piva, "Image forgery localization via fine-grained analysis of CFA artifacts," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1566–1577, 2012.

[87] W. Luo, J. Huang, and G. Qiu, "JPEG error analysis and its applications to digital image forensics," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 480–491, 2010.

[88] F. Huang, J. Huang, and Y. Q. Shi, "Detecting double JPEG compression with the same quantization matrix," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 848–856, 2010.

[89] S. Milani, M. Tagliasacchi, and S. Tubaro, "Discriminating multiple JPEG compressions using first digit features," *APSIPA Transactions on Signal and Information Processing*, vol. 3, no. 1, 2014.

[90] M. C. Stamm and K. J. R. Liu, "Forensic detection of image manipulation using statistical intrinsic fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 492–506, 2010.

[91] G. Cao, Y. Zhao, R. Ni, and X. Li, "Contrast enhancement-based forensics in digital images," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 515–525, 2014.

[92] H. Zou, P. Yang, R. Ni, and Y. Zhao, "Anti-Forensics of Image Contrast Enhancement Based on Generative Adversarial Network," *Security and Communication Networks*, vol. 2021, Article ID 6663486, 8 pages, 2021.

[93] J. Wu, T. Tong, and Y. Chen, "An adversarial learning framework with cross-domain loss for median filtered image restoration and anti-forensics," *Computers & Security*, vol. 112, Article ID 102497, 2021.

[94] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," in *Media forensics and security II* vol. 7541, International Society for Optics and Photonics, Article ID 754110, 2010.

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

[96] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 529–538, 2008.

[97] D. Vazquez-Padin, F. Perez-Gonzalez, and P. Comesana-Alfaro, "A random matrix approach to the forensic analysis of upscaled images," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2115–2130, 2017.

[98] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM SIGGRAPH 2007 papers on - SIGGRAPH'07*, vol. 26, no. 3, pp. 10–16, 2007.

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

[100] Z. Lin, J. He, X. Tang, and C.-K. Tang, "Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis," *Pattern Recognition*, vol. 42, no. 11, pp. 2492–2501, 2009.

[101] Z. Zhou, Y. Li, Y. Zhang, Z. Yin, L. Qi, and R. Ma, "Residual visualization-guided explainable copy-relationship learning for image copy detection in social networks," *Knowledge-Based Systems*, vol. 228, Article ID 107287, 2021.

[102] T. Bianchi and A. Piva, "Detection of nonaligned double JPEG compression based on integer periodicity maps," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 842–848, 2012.

[103] S. Bayram, H. T. Sencar, and N. Memon, "An efficient and robust method for detecting copy-move forgery," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1053–1056, IEEE, Taipei, Taiwan, April 2009.

[104] Q. Liu and Z. Chen, "Improved approaches with calibrated neighboring joint density to steganalysis and seam-carved forgery detection in JPEG images," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 4, pp. 1–30, 2014.

[105] Q. Liu, "An approach to detecting JPEG down-recompression and seam carving forgery under recompression anti-forensics," *Pattern Recognition*, vol. 65, pp. 35–46, 2017.

[106] S. Taspinar, M. Mohanty, and N. Memon, "PRNU-based camera attribution from multiple seam-carved images," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 3065–3080, 2017.

[107] Q. Liu, "An improved approach to exposing JPEG seam carving under recompression," *IEEE Transactions on Circuits*

and Systems for Video Technology, vol. 29, no. 7, pp. 1907–1918, 2019.

[108] J. Ye and Y. Shi, "A local derivative pattern based image forensic framework for seam carving detection," in *Proceedings of the International Workshop on Digital Watermarking*, pp. 172–184, Springer, Guildford UK, September 2016.

[109] D. Zhang, T. Yin, G. Yang, M. Xia, L. Li, and X. Sun, "Detecting image seam carving with low scaling ratio using multi-scale spatial and spectral entropies," *Journal of Visual Communication and Image Representation*, vol. 48, pp. 281–291, 2017.

[110] D. Zhang, G. Yang, F. Li, J. Wang, and A. K. Sangaiah, "Detecting seam carved images using uniform local binary patterns," *Multimedia Tools and Applications*, vol. 79, no. 13-14, pp. 8415–8430, 2020.

[111] D. Zhang, X. Chen, and F. Li, "Seam-carved Image Tampering Detection Based on the Cooccurrence of Adjacent LBPs," *Security and Communication Networks*, vol. 2020, Article ID 8830310, 12 pages, 2020.

[112] M. Lu and S. Niu, "Detection of Image Seam Carving Using a Novel Pattern," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 9492358, 15 pages, 2019.

[113] M. Lu, S. Niu, and Z. Gao, "An efficient detection approach of content aware image resizing," *Computers, Materials & Continua*, vol. 64, no. 2, pp. 887–907, 2020.

[114] P. Comesaña, "Detection and information theoretic measures for quantifying the distinguishability between multimedia operator chains," in *Proceedings of the 2012 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 211–216, IEEE, Costa Adeje, Spain, December 2012.

[115] Z. Chen, Y. Zhao, and R. Ni, "Detection of operation chain: JPEG-Resampling-JPEG," *Signal Processing: Image Communication*, vol. 57, pp. 8–20, 2017.

[116] Z. Chen, B. Tondi, X. Li, R. Ni, Y. Zhao, and M. Barni, "Secure detection of image manipulation by means of random feature selection," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 9, pp. 2454–2469, 2019.

[117] V. Conotter, P. Comesana, and F. Perez-Gonzalez, "Forensic detection of processing operator chains: recovering the history of filtered JPEG images," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2257–2269, 2015.

[118] P. Ferrara, T. Bianchi, and A. De Rosa, "Reverse engineering of double compressed images in the presence of contrast enhancement," in *Proceedings of the 2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 141–146, IEEE, Pula, Italy, September 2013.

[119] T. Bianchi and A. Piva, "Reverse engineering of double JPEG compression in the presence of image resizing," in *Proceedings of the 2012 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 127–132, IEEE, Costa Adeje, Spain, December 2012.

[120] X. Chu, Y. Chen, and K. J. R. Liu, "Detectability of the order of operations: an information theoretic approach," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 4, pp. 823–836, 2016.

[121] D. Baracchi, D. Shullani, M. Iuliani, D. Giani, and A. Piva, "Camera Obscura: exploiting in-camera processing for image counter forensics," *Forensic Science International: Digital Investigation*, vol. 38, Article ID 301213, 2021.

[122] Q. Wang and R. Zhang, "Double JPEG compression forensics based on a convolutional neural network," *EURASIP Journal on Information Security*, vol. 2016, no. 1, pp. 23–12, 2016.

[123] J. Bunk, J. H. Bappy, and T. M. Mohammed, "Detection and localization of image forgeries using resampling features and deep learning," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1881–1889, Costa Adeje, Spain, July 2017.

[124] L. Bondi, S. Lameri, and D. Guera, "Tampering detection and localization through clustering of camera-based CNN features," in *Proceedings of the Computer Vision & Patterm Recognition Workshops*, pp. 1855–1864, Hawaii, USA, July 2017.

[125] C. Chen, S. McCloskey, and J. Yu, "Image splicing detection via camera response function analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition (CVPR)*, pp. 5087–5096, Honolulu, USA, June 2017.

[126] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection," in *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, pp. 159–164, New York, USA, August 2017.

[127] P. Zhou, X. Han, and V. I. Morariu, "Learning rich features for image manipulation detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1053–1061, Salt Lake City, UT, USA, May 2018.

[128] M. Lu and S. Niu, "A detection approach using LSTM-CNN for object removal caused by exemplar-based image inpainting," *Electronics*, vol. 9, no. 5, pp. 858–922, 2020.

[129] J. Chen, X. Kang, Y. Liu, and Z. J. Wang, "Median filtering forensics based on convolutional neural networks," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1849–1853, 2015.

[130] J. H. Bappy, C. Simons, L. Nataraj, B. S. Manjunath, and A. K. Roy-Chowdhury, "Hybrid LSTM and encoder-decoder architecture for detection of image forgeries," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3286–3300, 2019.

[131] Y. Wu, W. Abd-Almageed, and P. Natarajan, "Image copy-move forgery detection via an end-to-end deep neural network," in *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1907–1915, IEEE, Lake Tahoe, NV, USA, March 2018.

[132] Y. Wu, W. Abd-Almageed, and P. Natarajan, "BusterNet: detecting copy-move image forgery with source/target localization," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 168–184, Glasgow, UK, August 2018.

[133] G. Muzaffer and G. Ulutas, "A new deep learning-based method to detection of copy-move forgery in digital images," in *Proceedings of the 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, pp. 1–4, IEEE, Istanbul, Turkey, April 2019.

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

[135] Y. Zhu, C. Chen, G. Yan, Y. Guo, Y. Dong, and Ar-Net, "AR-net: adaptive attention and residual refinement network for copy-move forgery detection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6714–6723, 2020.

[136] X. Zhu, Y. Qian, X. Zhao, B. Sun, and Y. Sun, "A deep learning approach to patch-based image inpainting

forensics," *Signal Processing: Image Communication*, vol. 67, pp. 90–99, 2018.

[137] X. S. Zhu, Y. J. Qian, B. Sun et al., "Image inpainting forensics algorithm based on deep neural network," *Acta Optica Sinica*, vol. 38, no. 11, pp. 1110005–1110113, 2018.

[138] P. Yang, R. Ni, and Y. Zhao, "Recapture image forensics based on Laplacian convolutional neural networks," in *Proceedings of the International Workshop on Digital Watermarking*, pp. 119–128, Springer, Guangzhou China, September 2016.

[139] Y. Liu and Q. Guan, "Image forgery localization based on multi-scale convolutional neural networks," in *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec)*, pp. 85–90, Innsbruck, Austria, June 2018.

[140] Y. Liu, Q. Guan, and X. Zhao, "Copy-move forgery detection based on convolutional kernel network," *Multimedia Tools and Applications*, vol. 77, no. 14, Article ID 18269, 2018.

[141] F. Marra, D. Gragnaniello, L. Verdoliva, and G. Poggi, "A full-image full-resolution end-to-end-trainable CNN framework for image forgery detection," *IEEE Access*, vol. 8, Article ID 133488, 2020.

[142] Q. Bammey, R. G. V. Gioi, and J. M. Morel, "An adaptive neural network for unsupervised mosaic consistency analysis in image forensics," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14194–14204, Seattle, WA, US, June 2020.

[143] C. Yang, H. Li, and F. Lin, "Constrained R-CNN: a general image manipulation detection model," in *Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, London, United Kingdom, July 2020.

[144] P. Zhuang, H. Li, S. Tan, B. Li, and J. Huang, "Image tampering localization using a dense fully convolutional network," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2986–2999, 2021.

[145] H. Li and J. Huang, "Localization of deep inpainting using high-pass fully convolutional network," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8300–8309, Seoul, Korea (South), November 2019.

[146] M. Barni, Q.-T. Phan, and B. Tondi, "Copy move source-target disambiguation through multi-branch CNNs," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1825–1840, 2021.

[147] Y. Zhang, J. Zhang, and S. Xu, "A hybrid convolutional architecture for accurate image manipulation localization at the pixel-level," *Multimedia Tools and Applications*, vol. 80, pp. 1–16, 2021.

[148] Y. Zhang, F. Ding, S. Kwong, and G. Zhu, "Feature pyramid network for diffusion-based image inpainting detection," *Information Sciences*, vol. 572, no. 2021, pp. 29–42, 2021.

[149] Y. Rao, J. Ni, and H. Xie, "Multi-semantic CRF-based attention model for image forgery detection and localization," *Signal Processing*, vol. 183, Article ID 108051, 2021.

[150] J. Hao, Z. Zhang, S. Yang, D. Xie, and S. Pu, "TransForensics: image forgery localization with dense self-attention," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR)*, Article ID 15055, Nashville, TN, USA, June 2021.

[151] X. Wang, H. Wang, and S. Niu, "An intelligent forensics approach for detecting patch-based image inpainting," *Mathematical Problems in Engineering*, vol. 2020, Article ID 8892989, 10 pages, 2020.

[152] H. Li, B. Li, S. Tan, and J. Huang, "Identification of deep network generated images using disparities in color components," *Signal Processing*, vol. 174, Article ID 107616, 2020.

[153] L. Nataraj, T. M. Mohammed, B. S. Manjunath et al., "Detecting GAN generated fake images using Co-occurrence matrices," *Electronic Imaging*, vol. 31, no. 5, pp. 532–541, 2019.

[154] S. McCloskey and M. Albright, "Detecting GAN-generated imagery using color cues," in *Proceedings of the 2019 IEEE International Conference on Image Processing*, pp. 4584–4588, IEEE, Piscataway, January 2019.

[155] F. Marra and D. Gragnaniello, "Do GANs leave artificial fingerprints?" in *Proceedings of the 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pp. 506–511, San Jose, CA, USA, March 2019.

[156] Y. Ning, S. D. Larry, and F. Mario, "Attributing fake images to GANs: learning and analyzing GAN fingerprints," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7556–7566, Seoul, Korea (South), December 2019.

[157] S. Y. Wang and O. Wang, "CNN-generated images are surprisingly easy to spot for now," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8692–8701, Seattle, WA, USA, June 2020.

[158] H. Mo, B. Chen, and W. Luo, "Fake faces identification via convolutional neural network," in *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security*, pp. 43–47, Innsbruck, Austria, June 2018.

[159] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, "Detection of GAN-generated fake images over social networks," in *Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pp. 384–389, Miami, FL, USA, February 2018.

[160] Z. Mi, X. Jiang, T. Sun, and K. Xu, "GAN-generated image detection with self-attention mechanism against GAN generator defect," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 969–981, 2020.

[161] L. Guarnera and O. Giudice, "DeepFake detection by analyzing convolutional traces," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 666-667, Seattle, WA, USA, June 2020.

[162] S. Lee, S. Tariq, Y. Shin, and S. S. Woo, "Detecting handcrafted facial image manipulations and GAN-generated facial images using Shallow-FakeFaceNet," *Applied Soft Computing*, vol. 105, Article ID 107256, 2021.

[163] B. Chen, X. Liu, and Y. Zheng, "A robust GAN-generated face detection method based on dual-color spaces and an improved Xception," *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.

[164] B. Chen, W. Tan, Y. Wang, and G. Zhao, "Distinguishing between natural and GAN-generated face images by combining global and local features," *Chinese Journal of Electronics*, vol. 31, no. 1, pp. 59–67, 2022.

[165] G. Tang, L. Sun, and X. Mao, "Detection of GAN-Synthesized Image Based on Discrete Wavelet Transform," *Security and Communication Networks*, vol. 2021, Article ID 5511435, 10 pages, 2021.

[166] X. Wang, S. Niu, and H. Wang, "Image inpainting detection based on multi-task deep learning network," *IETE Technical Review*, vol. 38, no. 1, pp. 149–157, 2021.

[167] Y. Huang, F. Juefei-Xu, and R. Wang, "FakeLocator: Robust Localization of GAN-based Face Manipulations," 2020, https://arxiv.org/abs/2001.09598#:~:text=FakeLocator%3A%20Robust%20Localization%20of%20GAN-Based%20Face%20Manipulations,-Yihao%20Huang%2C%20Felix&text=F.

[168] D. Gragnaniello, D. Cozzolino, and F. Marra, "Are GAN generated images easy to detect? A critical analysis of the state-of-the-art," in *Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, Shenzhen, China, July 2021.

[169] M. Barni, M. C. Stamm, and B. Tondi, "Adversarial multimedia forensics: overview and challenges ahead," in *Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 962–966, IEEE, Rome, Italy, September 2018.

[170] J. C. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proenca, and J. Fierrez, "GANprintR: improved fakes and evaluation of the state of the art in face manipulation detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 1038–1048, 2020.

[171] X. Zhang, S. Karaman, and S. F. Chang, "Detecting and simulating artifacts in gan fake images," in *Proceedings of the 2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–6, IEEE, Delft, Netherlands, December 2019.

[172] M. Du, S. Pentyala, and Y. Li, "Towards Generalizable Forgery Detection with Locality-Aware Autoencoder," 2019, https://arxiv.org/abs/1909.05999.

[173] D. Cozzolino, J. Thies, and A. Rossler, "Spoc: spoofing camera fingerprints," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 990–1000, Nashville, TN, USA, June 2021.

[174] J. Wu and W. Sun, "Towards multi-operation image anti-forensics with generative adversarial networks," *Computers & Security*, vol. 100, Article ID 102083, 2021.

[175] J. Brockschmidt, J. Shang, and J. Wu, "On the generality of facial forgery detection," in *Proceedings of the 2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW)*, pp. 43–47, IEEE, Monterey, CA, USA, November 2019.

[176] R. Huang, F. Fang, and H. Nguyen, "Security of facial forensics models against adversarial attacks," in *Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP)*, pp. 2236–2240, IEEE, Abu Dhabi, United Arab Emirates, October 2020.

[177] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proceedings of the 3rd International Conference on Leaning Representations (ICLR)*, San Diego, CA, USA, May 2015.

[178] A. Kurakin, I. J. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," in *Proceedings of the 5th International Conference on Leaning Representations Workshop (ICLRW)*, Toulon, France, April 2017.

[179] H. Zeng, J. Liu, J. Yu, X. Kang, Y. Q. Shi, and Z. J. Wang, "A framework of camera source identification Bayesian game," *IEEE Transactions on Cybernetics*, vol. 47, no. 7, pp. 1757–1768, 2017.

[180] M. Barni and B. Tondi, "Adversarial source identification game with corrupted training," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3894–3915, 2018.

[181] B. Tondi, N. Merhav, and M. Barni, "Detection games under fully active adversaries," *Entropy*, vol. 21, no. 1, p. 23, 2018.

[182] F. Ding, G. Zhu, Y. Li, and X. Zhang, "Anti-Forensics for Face Swapping Videos via Adversarial Training," *IEEE Transactions on Multimedia*, 2021.

[183] H. Xie, J. Ni, and Y. Q. Shi, "Dual-Domain Generative Adversarial Network for Digital Image Operation Anti-forensics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, 2021.

[184] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," *Proceedings SPIE Media Forensics and Security*, vol. 7254, 2009.

[185] T. Bianchi and A. Piva, "Image forgery localization via block-grained analysis of JPEG artifacts," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1003–1017, 2012.

[186] J. Park and D. Cho, "Double JPEG detection in mixed JPEG quality factors using deep convolutional neural network," in *Proceedings of the European Conference on Computer Vision*, Munich, Germany, September 2018.

[187] T. T. Ng, S. F. Chang, and Q. Sun, "A Data Set of Authentic and Spliced Image Blocks," Columbia University, ADVENT, Technical Report, , pp. 203–2004, 2004.

[188] Y. F. Hsu and S. F. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency," in *Proceedings of the 2006 IEEE International Conference on Multimedia and Expo*, pp. 549–552, IEEE, Ontario, Canada, July 2006.

[189] J. Dong, W. Wang, and T. Tan, "Casia image tampering detection evaluation database," in *Proceedings of the IEEE China Summit and International Conference on Signal and Information Processing*, pp. 422–426, Chengdu, China, July 2013.

[190] T. J. de Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. de Rezende Rocha, "Exposing digital image forgeries by illumination color classification," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 7, pp. 1182–1194, 2013.

[191] T. Gloe and R. Böhme, "The' Dresden Image Database' for benchmarking digital image forensics," *Proceedings of the 2010 ACM Symposium on Applied Computing - SAC'10*, vol. 3, pp. 150–159, 2010.

[192] A. Novoz´amsk´y, B. Mahdian, and S. Saic, "IMD2020: a large-scale annotated dataset tailored for detecting manipulated images," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (CVPRW)*, pp. 71–80, Long Beach, CA, USA, June 2020.

[193] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1841–1854, 2012.

[194] D. Cozzolino, G. Poggi, and L. Verdoliva, "Efficient dense-field copy-move forgery detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2284–2297, 2015.

[195] D. Tralic, I. Zupancic, and M. Grgic, "CoMoFoD - new database for copy-move forgery detection," in *Proceedings of the 55th International Symposium ELMAR*, pp. 49–54, Zadar, Croatia, September 2013.

[196] B. Wen, Y. Zhu, and R. Subramanian, "COVERAGE-a novel database for copy-move forgery detection," in *Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP)*, pp. 161–165, IEEE, Phoenix, Arizona, September 2016.

[197] H. Guan, M. Kozak, E. Robertson et al., "MFC datasets: large-scale benchmark datasets for media forensic challenge evaluation," in *Proceedings of the 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pp. 63–72, HI, USA, January 2019.

[198] S. Heller, L. Rossetto, and H. Schuldt, "The Ps-Battles Dataset-An Image Collection for Image Manipulation Detection," 2018, https://arxiv.org/abs/1804.04866.

[199] G. Mahfoudi, B. Tajini, and F. Retraint, "DEFACTO: image and face manipulation dataset," in *Proceedings of the 2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5, IEEE, Coruña, Spain, September 2019.