mathematics

Review

Review of Image Forensic Techniques Based on Deep Learning

Chunyin Shi [1,†], Luan Chen [1,†], Chengyou Wang [1,2,*], Xiao Zhou [1,2] and Zhiliang Qin [1,3]

[1] School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China; shicy@mail.sdu.edu.cn (C.S.); chenluan@mail.sdu.edu.cn (L.C.); zhouxiao@sdu.edu.cn (X.Z.); qinzhiliang@beiyang.com (Z.Q.)
[2] Shandong University–Weihai Research Institute of Industrial Technology, Weihai 264209, China
[3] Weihai Beiyang Electric Group Co., Ltd., Weihai 264209, China
* Correspondence: wangchengyou@sdu.edu.cn; Tel.: +86-631-568-8338
† These authors contributed equally to this work.

Abstract: Digital images have become an important carrier for people to access information in the information age. However, with the development of this technology, digital images have become vulnerable to illegal access and tampering, to the extent that they pose a serious threat to personal privacy, social order, and national security. Therefore, image forensic techniques have become an important research topic in the field of multimedia information security. In recent years, deep learning technology has been widely applied in the field of image forensics and the performance achieved has significantly exceeded the conventional forensic algorithms. This survey compares the state-of-the-art image forensic techniques based on deep learning in recent years. The image forensic techniques are divided into passive and active forensics. In passive forensics, forgery detection techniques are reviewed, and the basic framework, evaluation metrics, and commonly used datasets for forgery detection are presented. The performance, advantages, and disadvantages of existing methods are also compared and analyzed according to the different types of detection. In active forensics, robust image watermarking techniques are overviewed, and the evaluation metrics and basic framework of robust watermarking techniques are presented. The technical characteristics and performance of existing methods are analyzed based on the different types of attacks on images. Finally, future research directions and conclusions are presented to provide useful suggestions for people in image forensics and related research fields.

Keywords: image forensics; image forgery detection; robust image watermarking; deep learning

MSC: 94A08; 68U10

check for updates

Citation: Shi, C.; Chen, L.; Wang, C.; Zhou, X.; Qin, Z. Review of Image Forensic Techniques Based on Deep Learning. *Mathematics* **2023**, *11*, 3134. https://doi.org/10.3390/math11143134

Academic Editor: Konstantin Kozlov

Received: 14 June 2023
Revised: 8 July 2023
Accepted: 11 July 2023
Published: 16 July 2023

1. Introduction

Digital images are important information carriers, and with the rapid development of this technology, digital images have gradually been included in all aspects of life. However, data stored or transmitted in digital form is vulnerable to external attacks, and digital images are particularly susceptible to unauthorized access and illegal tampering. As a result, the credibility and security of digital images are under serious threat. If these illegally accessed and tampered images appear in the news media, academic research, and judicial forensics, which require high originality of images, social stability and political security will be seriously threatened. To solve the above problems, digital image forensics has become a hot issue for research and is the main method used to identify whether the images are illegally acquired or tampered with. Digital image forensic technology is a novel technique to determine the authenticity, integrity, and originality of image content by analyzing the statistical characteristics of images, which is of great significance for securing cyberspace and maintaining social order.

Digital image forensics technology is mainly used to detect the authenticity of digital images and realize image copyright protection, which can be divided into active forensics

techniques and passive forensic techniques according to different detection methods, as shown in Figure 1.

The active forensic techniques are used to embed prior information, and then extract the embedded information, such as image watermarking and image signatures, and compare it with the original information to identify whether the image was illegally obtained. Image watermarking is a technique that embeds hidden information into digital images with the aim of protecting the copyright and integrity of the images. Watermarks are usually invisible or difficult to perceive and are used to identify the owner or provide authorization information for the image. They can be used to trace and prevent unauthorized copying, modification, or tampering of the image. Image watermarking techniques typically utilize image processing algorithms and encryption techniques to ensure the robustness and security of the watermark. On the other hand, image signature is a technique used to verify the integrity and authenticity of an image. Image signatures are typically based on digital signatures using encryption and hash functions to ensure that the image has not been tampered with during transmission or storage. With the rapid development of the digital information era, there has been an alarming rise in copyright infringement facilitated by image tampering techniques. Among the aforementioned active forensic techniques, image watermarking, especially robust image watermarking that can recover watermark information intact against intentional or unintentional attacks, is more effective compared to image signatures in authenticating and protecting copyright information when images are subjected to malicious tampering. Therefore, our focus in active forensics is on image watermarking, which aligns with our research interests as well.

Passive forensic techniques, which do not require prior information about the image, determine whether the image has been illegally tampered with by analyzing the structure and statistical characteristics of the image. In passive forensic techniques, even if an image has been forged beyond recognition by the human eyes, the statistical characteristics of the image will change, causing various inconsistencies in the image, and these inconsistencies are used to detect and locate tampered images. With the development of deep learning technology, deep learning technology has been widely used in many fields, such as speech processing, computer vision, natural language processing, and medical applications [1]. Deep learning technology has been widely applied in the field of image forensics, which has promoted the development of image forensics technology. In this survey, we focus on passive image forgery detection technology based on deep learning and robust image watermarking technology based on deep learning.

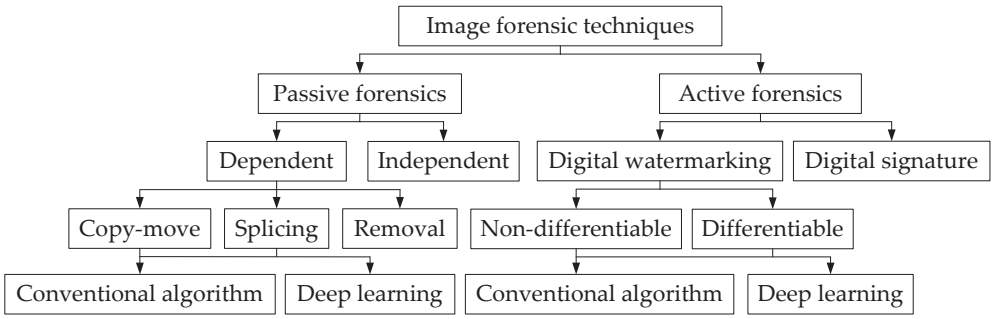

Figure 1. Classification of image forensic techniques.

There have been many reviews of passive forgery detection techniques over the last few years. Kuar et al. [2] created a detailed overview of the process of image tampering and the current problems of tampering detection, and sorted out the conventional algorithms in tampering detection from different aspects, without reviewing the techniques of deep learning. Zhang et al. [3] sorted out and compared the conventional copy-move tampering detection algorithms, giving the advantages and disadvantages of each conventional algorithm in detail, without reviewing the techniques related to deep learning. Zanardelli et al. [4] reviewed deep learning-based tampering algorithms and compared copy-move, splicing,

and deep fake techniques. However, generic tampering detection algorithms were not provided in detail. Nabi et al. [5] reviewed image and video tampering detection algorithms, giving a detailed comparison of tampering detection datasets and algorithms, but did not summarize the evaluation metrics of tampering detection algorithms. Surbhi et al. [6] focused on the universal type-independent techniques for image tampering detection. Some generic methods based on resampling, inconsistency-based detection, and compression are compared and analyzed in terms of three aspects: the methodology used, the dataset and classifier used, and performance. A generic framework for image tampering detection is given, including dataset selection, data preparation, feature selection and extraction, classifier selection, and performance evaluation. The top journals and tampering public datasets in the field of tampering detection are organized. Finally, a reinforcement learning-based model is proposed to provide an outlook on future works. However, in-depth analysis and summary of image tampering detection-based deep learning are not presented. This survey provides an overview of deep learning techniques and reviews the latest generic tamper detection algorithms. In this survey the evaluation metrics commonly used for tampering detection are summarized.

For active forensics, Rakhmawati et al. [7] analyzed a fragile watermarking model for tampering detection. Kumar et al. [8] summarized the existing work on watermarking from the perspective of blind and non-blind watermarking, robust and fragile watermarking, but it did not focus on the methods to improve the robustness. Menendez et al. [9] summarized the reversible watermarking model and analyzed its robustness. Agarwal et al. [10] reviewed the robustness and imperceptibility of the watermarking model from the spatial domain and transform domain perspectives. Amrit and Singh [11] analyzed the watermarking models based on deep watermarking in recent years, but they did not discuss the methods to improve the robustness of the models for different attacks. Wan et al. [12] analyzed robustness enhancement methods for geometric attacks in deep rendering images, motion images, and screen content images. Evsutin and Dzhanashia [13] analyzed the characteristics of removal, geometric and statistical attacks and summarized the corresponding attack robustness enhancement methods, but there was less analysis of watermarking models based on deep learning. Compared to the existing active forensic reviews, we start from one of the fundamentals of model robustness enhancement (i.e., generating attack-specific adversarial samples). According to their compatibility with deep learning-based end-to-end watermarking models, the attacks are initially classified into differentiable attacks and non-differentiable attacks. According to their impact on the watermarking model, the network structure and training methods to improve the robustness of different attacks are further subdivided.

The rest of this survey is organized as follows: Section 2 gives the basic framework for tampering detection and robust watermarking based on deep learning, evaluation metrics, attack types, and tampering datasets. Section 3 presents the state-of-the-art techniques of image forgery detection based on deep learning. Section 4 describes the state-of-the-art techniques of robust image watermarks based on deep learning. Section 5 gives the conclusion and future work.

2. Image Forensic Techniques

2.1. Passive Forensics

Passive image forgery detection techniques can be classified as conventional manual feature-based and deep learning-based. The conventional detection algorithms for copy-move forgery detection (CMFD) are: discrete cosine transform (DCT) [14], discrete wavelet transform (DWT) [15], polar complex exponential transform (PCET) [16], scale-invariant feature transform (SIFT) [17], speeded up robust feature (SURF) [18], etc. The conventional detection algorithms for splicing tamper detection are: inconsistency detection by color filter array (CFA) interpolation [19], and inconsistency detection by noise features [20]. However, these conventional detection methods have the drawbacks of low generalization, poor robustness, and low detection accuracy. Deep learning-based detection algorithms,

which take advantage of autonomous learning features, solve the above problems of conventional algorithms. In this survey, we mainly review forgery detection algorithms based on deep learning.

2.1.1. Basic Framework of Image Forgery Detection

With the development of deep learning techniques, certain advantages have been achieved in areas such as image classification and segmentation. Deep learning is also increasingly being applied to image forgery detection. A basic framework of image forgery detection based on deep learning is shown in Figure 2. First, the tamper detection network is built, and feature extraction, feature classification, and localization are performed by the network model. The weights of the network are saved by learning the parameters within the network through big data training to obtain the weights when optimal. The image to be detected is input to the network and tampering detection is performed using the saved network model.

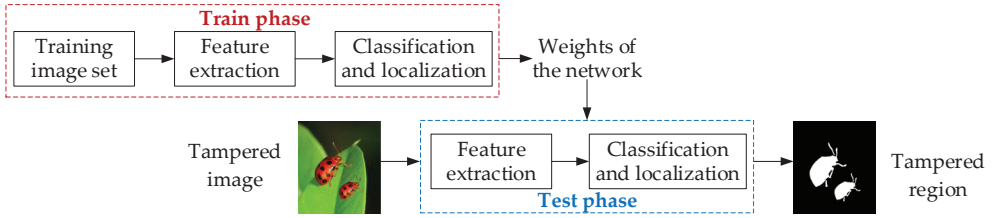

Figure 2. A basic framework of image forgery detection based on deep learning.

2.1.2. Performance Evalution Metrics

The forgery detection task can be considered to be a binary classification task of pixels, i.e., whether a pixel is tampered with or not. Therefore, the evaluation metrics of the forgery detection algorithm need to use the amount of categorization of the samples, including true positive, false positive, true negative, and false negative. The commonly used evaluation metrics for forgery detection are precision p, recall r, and F_1 score [3], which are expressed as Equations (1)–(3), respectively:

$$p = \frac{T_P}{T_P + F_P} \tag{1}$$

$$r = \frac{T_P}{T_P + F_N} \tag{2}$$

$$F_1 = \frac{2pr}{p + r} \tag{3}$$

where T_P denotes the number of tampered pixels detected as tampered; F_P denotes the number of authentic pixels detected as tampered; F_N denotes the number of tampered pixels detected as authentic. The values of p, r, and F_1 are in the range [0, 1]. The larger p, r, and F_1 are, the higher the accuracy of detection results is.

Another important metric is the area under curve (AUC), which is defined as the area under the receiver operating characteristic (ROC) curve and can reflect the classification performance of a binary classifier. Like the F_1 score, AUC can evaluate the precision and recall together. Its value is generally between 0.5 and 1. The closer the value of AUC is to 1, the higher the performance of the algorithm. When it is equal to 0.5, the true value is the lowest and has no application value.

2.1.3. Datasets for Image Forgery Detection

Diverse datasets for forgery detection are described in this section. These datasets contain original images, tampered images, binary labels, and partially post-processed images. Different datasets are used depending on the problem being solved. In order

to validate the results of tamper detection algorithms, different public tampered image datasets are used to test the performance of these algorithms. Table 1 describes the 14 public datasets for image forgery detection, presenting the type of forgery, number of forged images and authentic images, image format, and image resolution.

Table 1. Datasets for image forgery detection.

Dataset	Year	Type of Forgery	Number of Forged Images/Authentic Images	Image Format	Image Resolution
Columbia color [21]	2006	Splicing	183/180	BMP, TIF	757 × 568–1152 × 768
MICC-F220 [22]	2011	Copy-move	110/1100	JPG	480 × 722–1070 × 800
MICC-F600 [22]	2011	Copy-move	160/440	JPG, PNG	722 × 480–800 × 600
MICC-F2000 [22]	2011	Copy-move	700/1300	JPG	2048 × 1536
CASIA V1 [23]	2013	Copy-move, Splicing	921/800	JPG	284 × 256
CASIA V2 [23]	2013	Copy-move, Splicing	5123/7200	JPG, BMP, TIF	320 × 240–800 × 600
Carvalho [24]	2013	Splicing	100/100	PNG	2048 × 1536
CoMoFoD [25]	2013	Copy-move	4800/4800	PNG, JPG	512 × 512–3000 × 2500
COVERAGE [26]	2016	Copy-move	100/100	TIF	2048 × 1536
Korus [27]	2017	Copy-move, Splicing	220/220	TIF	1920 × 1080
USCISI [28]	2018	Copy-move	100,000/-	PNG	320 × 240–640 × 575
MFC 18 [29]	2019	Multiple manipulation	3265/14,156	RAW, PNG, BMP, JPG, TIF	128 × 104–7952 × 5304
DEFACTO [30]	2019	Multiple manipulation	229,000/-	TIF	240 × 320–640 × 6405
IMD 2020 [31]	2020	Multiple manipulation	37,010/37,010	PNG, JPG	193 × 260–4437 × 2958

2.2. Active Forensics

Digital watermarking is one of the most effective active forensics methods to realize piracy tracking and copyright protection. It can be classified into conventional methods based on manually designed features and deep learning-based methods. Early digital watermarking was mostly embedded in the spatial domain, such as least significant bits (LSB) [32], but it lacked robustness and was easily detected by sample pair analysis [33]. To improve robustness, a series of transform domain-based methods were proposed, such as DWT [34], DCT [35], contourlet transform [36], and Hadamard transform [37]. In recent years, with the continuous update and progress of deep learning technology, deep learning has been widely used in image watermarking and achieved remarkable achievements. In this survey, we will focus on the analysis of deep learning-based robust watermarking.

2.2.1. Basic Framework of Robust Image Watermarking Algorithm

The main basic framework of deep learning-based end-to-end robust watermarking is shown in Figure 3. It consists of three components: embedding layer (including image feature extraction network and watermark feature enhancement network), attack layer, and extraction layer. The model includes two stages of forward propagation and reverse gradient updating when iteratively updating parameters. During forward propagation, the original image and the watermarked image pass through the image feature extraction network and the watermark feature enhancement network, respectively, to extract high-order image features and high-order watermark features. The output results are then fed into the watermark embedding network to obtain the watermarked image. The attack simulation layer includes various types of attacks such as noise and filtering, geometric attacks, and JPEG compression. When the watermarked image passes through the attack simulation layer, different adversarial samples after attacks are generated. The extraction layer performs watermark extraction on the adversarial samples or the watermarked images to obtain the watermark authentication information. Above is the model forward propagation training. In backpropagation, the PSNR loss and SSIM, which measures the similarity of two images from three aspects: grayscale, contrast, and structure loss of the original image and the watermarked image are usually set to improve the similarity between the watermarked image and the original image, which are expressed as Equations (4) and (5), respectively.

At the same time, the MSE loss of extracted watermark and original watermark, which are expressed as Equation (6), is set to improve the accuracy of watermark recovery.

$$L_{\mathrm{MSE-I}} = \sum_{i=0}^{P-1} \sum_{j=0}^{Q-1} [I(i,j) - I_{\mathrm{W}}(i,j)]^2 \tag{4}$$

$$L_{\mathrm{SSIM}} = \frac{(2\mu_I \mu_{I_{\mathrm{W}}} + C_1)(\sigma_{I,I_{\mathrm{W}}} + C_2)}{(\mu_I^2 + \mu_{I_{\mathrm{W}}}^2 + C_1)(\sigma_I^2 + \sigma_{I_{\mathrm{W}}}^2 + C_2)} \tag{5}$$

$$L_{\mathrm{MSE-W}} = \sum_{i=0}^{L-1} [W(i,j) - W_{\mathrm{e}}(i,j)]^2 \tag{6}$$

where P denotes the width of the original image; Q denotes the length of the original image; μ_I and $\mu_{I_{\mathrm{W}}}$ denote the mean value of the gray value of the original image and the watermarked image, respectively; σ_I^2 and $\sigma_{I_{\mathrm{W}}}^2$ denote the variance of the gray value of the original image and the watermarked image, respectively; $I(i,j)$ and $I_{\mathrm{W}}(i,j)$ denote the (i,j) original image and watermarked image; $\sigma_{I,I_{\mathrm{W}}}$ denotes the covariance of the original image and the watermark image; C_1 and C_2 are constant in range $[10^{-4}, 9 \times 10^{-4}]$. Then the model calculated the corresponding loss point by point by a gradient from the output end of the model according to the above loss, and the optimizer (usually Adam optimizer, SGD optimizer) was used to update the model parameters, so as to optimize the task of the model (improving the imperceptibility of the watermarked image and the accuracy of watermark extraction after the watermarked image was attacked).

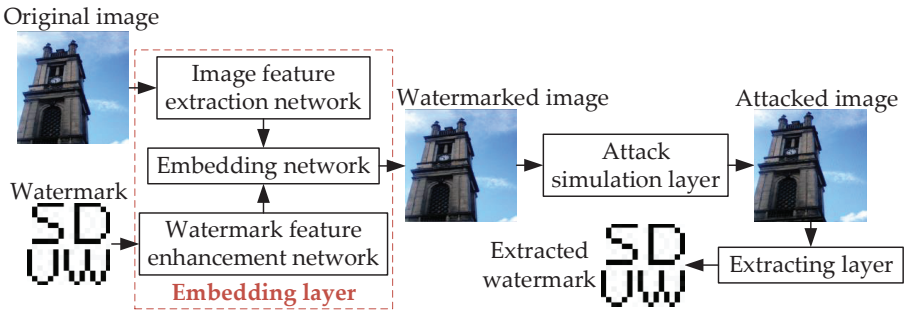

Figure 3. A basic framework of end-to-end robust watermarking based on deep learning.

2.2.2. Performance Evaluation Metrics

In a digital image watermarking algorithm, the most important three evaluation indicators are robustness, imperceptibility, and capacity.

Robustness: Robustness is used to measure the ability of a watermark model to recover the original watermark after an image has been subjected to a series of intentional or unintentional image processing during electronic or non-electronic channel transmission. Bit error rate (BER) and normalized cross-correlation (NCC) are usually used as the objective evaluation metrics, which are expressed as Equations (7) and (8), respectively:

$$E_{\mathrm{BER}}(w, w') = \frac{1}{L} \sum_{i=1}^{L} |w_i - w_i'| \tag{7}$$

$$E_{\mathrm{NCC}} = \frac{\sum_{i=1}^{L} (w_i - \bar{w})(w_i' - \bar{w}')}{\sqrt{\sum_{i=1}^{L} (w_i - \bar{w})^2 \sum_{i=1}^{L} (w_i' - \bar{w}')^2}} \tag{8}$$

where w_i and $\bar{w}$ represent the ith bit of the original watermark and the mean value of the original watermark, respectively; w_i' and $\bar{w}'$ represent the i th bit of the extracted

watermark and the mean value of the extracted watermark; L represents the length of the watermark.

Imperceptibility: Imperceptibility is used to measure the sensory impact of the embedding point on the whole image after the model has completed watermark embedding (i.e., the watermarked image is guaranteed to be indistinguishable from the original image). Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are usually used as the objective evaluation metrics, which are expressed as Equations (9) and (10), respectively:

$$E_{\mathrm{PSNR}}(I, I_{\mathrm{W}}) = 10\log_{10}\frac{W \times H \times G_{\mathrm{MAX}}^2}{\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}[I(i,j) - I_{\mathrm{W}}(i,j)]^2} \tag{9}$$

$$E_{\mathrm{SSIM}}(I, I_{\mathrm{W}}) = \frac{(2\mu_I\mu_{I_{\mathrm{W}}} + C_1)(2\sigma_{I,I_{\mathrm{W}}} + C_1)}{(\mu_I^2 + \mu_{I_{\mathrm{W}}}^2 + C_1)(\sigma_I^2 + \sigma_{I_{\mathrm{W}}}^2 + C_2)} \tag{10}$$

where W denotes the width of the original image; H denotes the length of the original image; G_{MAX} denotes the maximum gray level of the original image; μ_I and $\mu_{I_{\mathrm{W}}}$ denote the mean value of the gray value of the original image and the watermarked image, respectively; σ_I^2 and $\sigma_{I_{\mathrm{W}}}^2$ denote the variance of the gray value of the original image and the watermarked image, respectively; $I(i,j)$ and $I_{\mathrm{W}}(i,j)$ denote the (i,j) original image and watermarked image; σ_{I,I_W} denotes the covariance of the original image and the watermark image; C_1 and C_2 are constant in range $[10^{-4}, 9 \times 10^{-4}]$.

Capacity: The capacity represents the maximum watermark embedding bits of the model while maintaining the established required imperceptibility and robustness metrics. It is mutually constrained with imperceptibility and robustness. Increasing the watermark embedding capacity, imperceptibility and robustness decrease, and vice versa. The number of embedded watermark bits per pixel (bpp) is usually used to measure the capacity metric of the model, which is expressed as Equation (11):

$$E_{\mathrm{bpp}} = \frac{W_{\mathrm{num}}}{I_{\mathrm{num}}} \tag{11}$$

where W_{num} denotes the number of watermark bits, and I_{num} denotes the total number of original image pixels.

2.2.3. Attacks of Robust Watermarking

In the end-to-end watermarking framework, the attack simulation layer plays a decisive role in improving the robustness of the framework. However, when implementing backpropagation updates in parameters, it is necessary to ensure that each node is differentiable. For non-differentiable attacks, the model cannot perform parameter updates during backpropagation. Therefore, here comes a subdivision of whether the attack is differentiable or not.

Differentiable Attacks:

Noise and Filtering Attacks: Noise and filtering attacks refer to some intentional or unintentional attacks on the watermarked image in the electronic channel transmission, such as Gaussian noise, salt and pepper noise, Gaussian filtering, and median filtering. Its robustness can usually be improved by generating adversarial samples in the ASL layer.

Geometric Attacks: Applying geometric attacks to an image can break the synchronization between the watermark decoder and the watermarked image. Geometric attacks can be subdivided into rotation, scaling, translation, cropping, etc.

Non-differentiable Attacks:

Joint Photographic Experts Group (JPEG) Attacks: The JPEG standard has two basic compression methods: the DCT-based method and the lossless compression prediction method. In this survey, we focus on the DCT-based JPEG compression, which can consist of DCT, inverse DCT, quantization, inverse quantization, entropy coding, and entropy

decoding (i.e., color conversion and sampling), as shown in Figure 4. Quantization and dequantization are non-differentiable processes, leading to the incompatibility of the model with simulated attacks.

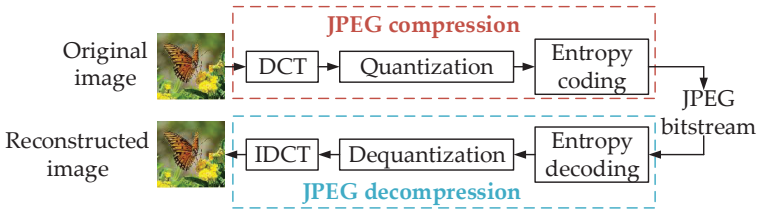

Figure 4. A flowchart of JPEG compression.

Screen-shooting Attacks: In the process of watermarked images through the camera photo processing, it will undergo a series of analog-digital (A/D) conversion and digital-analog (D/A) conversion inevitably, both of which affect the extraction of the watermark seriously.

Agnostic Attacks: Agnostic attacks refer to attacks in which the model parameters are unknown to the attack's prior information. For neural network models, it is difficult for the encoder and decoder to adapt to agnostic attacks without generating attack prior samples.

3. Image Forgery Detection Based on Deep Learning

Image passive forgery detection methods can be divided into a single type of forgery detection and generic forgery detection depending on the detection types. Single forgery detection methods can only detect specific types of tampered images, including image copy-move forgery detection and image splicing forgery detection. Generic forgery detection methods can be applied to different types of tampered images, including copy-move, splicing, and removal. Diverse passive forgery detection methods are described in this section. A table is given to compare the performance of passive forgery detection methods from different aspects at the end of this section. An overview of image forgery detection based on deep learning, as shown in Figure 5.

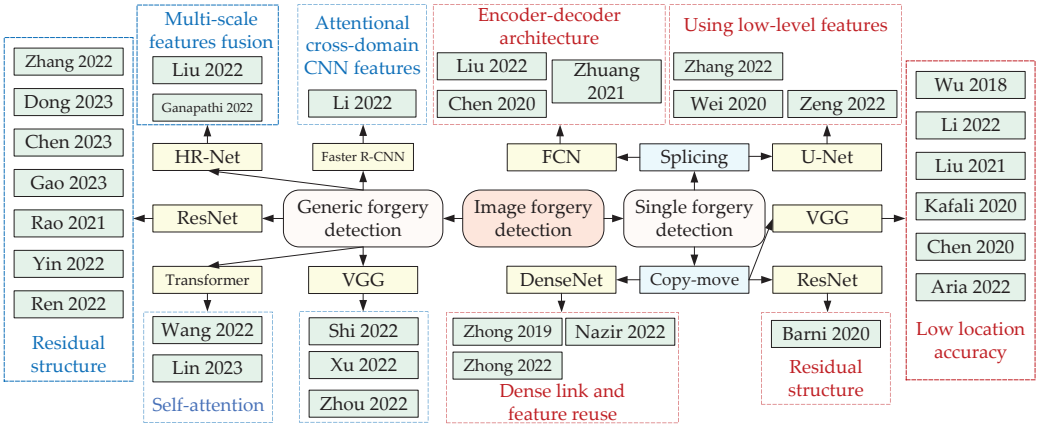

Figure 5. An overview of image forgery detection based on deep learning.

3.1. Image Copy-Move Forgery Detection

The CMFD methods detect tampered images by extracting features associated with tampering. For an image, it is possible to globally capture the features either for the entire image or locally for regions. The choice of feature extraction methods affects the performance of CMFD methods greatly. The CMFD methods are divided into two categories, conventional manual feature-based methods and deep learning-based methods.

The conventional manual feature methods can be divided into block-based and keypoint-based CMFD methods. The block-based CMFD methods can locate tampered

regions accurately, but there are main problems, such as high computational complexity and difficulty in resisting large-scale rotation and scaling. To solve these problems, keypoint-based CMFD methods are proposed. The keypoint-based CMFD methods use the keypoint extraction techniques to extract keypoints of the images and find similar features using feature-matching algorithms. The keypoint-based CMFD methods can accurately locate tampered regions of common tampered images, but it mainly suffers from problems, such as the small number of keypoints in smooth regions leading to undetectable and poor algorithm generalization ability. To solve the problems of conventional manual feature methods, deep learning-based CMFD methods are proposed.

With the rapid development of deep learning techniques, deep learning-based methods have been applied to the field of tampering detection. Deep learning-based CMFD methods have shown great performance improvements. A well-trained model can learn the latent features of images. The difference between the two types of images is found to discriminate the tampered image. Compared with conventional methods, deep learning-based CMFD can provide more accurate and comprehensive feature descriptors.

To detect whether an image was tampered, Rao and Ni [38] proposed a deep convolutional neural network (DCNN)-based model that used the spatial rich models (SRM) with 30 basic high-pass filters to initialize the weights of the first layer, which helped suppress the effects of image semantics and accelerated the network convergence. The image features were extracted by the convolutional layer and classified using a support vector machine (SVM) to discern whether the image was tampered with or not. Kumar and Gupta [39] proposed a convolutional neural network (CNN)-based model to automatically extract image features for image tampering classification with robust to image compression, scaling, rotation, and blurring, but the method also required analysis at the pixel level to locate tampered with regions. In [38,39], the authors only perform the detection of whether tampering has been performed, and cannot localize tampered regions, which has a limited application.

To further improve the application space of the algorithm and achieve the localization of copy and move regions, Li et al. [40] proposed a method combining image segmentation and DCNN, using Super-BPD [41] to segment the image to obtain the edge information of the image. VGG 16 and atrous spatial pyramid pooling (ASPP) [42] networks obtained the multi-scale features of the image to improve the accuracy of the algorithm. The feature matching module was introduced to achieve the localization of tampered regions. But the segmentation module leads to high computational complexity. Liu et al. [43] designed a two-stage detection framework. The first stage introduced atrous convolution with autocorrelation matching based on spatial attention to improve similarity detection. In the second stage, the superglue method was proposed to eliminate false warning regions and repair incomplete regions, thus improving the detection accuracy of the algorithm. Zhong and Pun [44] created an end-to-end deep neural network, referring to the Inception architecture fusing multi-scale convolution to extract multi-scale features. Kafali et al. [45] proposed a nonlinear inception module based on a second-order Volterra kernel by considering the linear and nonlinear interactions among pixels. Nazir et al. [46] created an improved mask region-based convolution network. The network used the DenseNet 41 model to extract deep features, which were classified using the Mask-RCNN [47] to locate tampered regions. Zhong et al. [48] proposed a coarse-fine spatial channel boundary attention network and designed the attention module for boundary refinement to obtain finer forgery details and improve the performance of detection. In a few papers [38–40,43–45], the authors improved the detection performance and robustness of the algorithm but did not distinguish between source and target areas.

To correctly distinguish between source and target regions, some methods have been proposed. Wu et al. [28] proposed the first parallel network for distinguishing between source and target. The manipulation detection branch located potential manipulation regions by visual artifacts, and the similarity detection branch located source and target regions by visual similarity. Chen et al. [49] used a serial structure and added the atrous

convolutional attention module in the similarity detection phase to improve the detection accuracy. The network structure is shown in Figure 6. Aria et al. [50] proposed a quality-independent detection method, which used a generative adversarial network to enhance image quality and a two-branch convolutional network for tampering detection. The network could detect multiple tampered regions simultaneously and distinguish the source and target of tampering. It was resistant to various post-processing attacks and had good detection results in low-quality tampered images. Barni et al. [51] proposed a multi-branch CNN network, which exploited the irreversibility caused by interpolation traces and the inconsistency of target region boundaries to distinguish source and target regions.

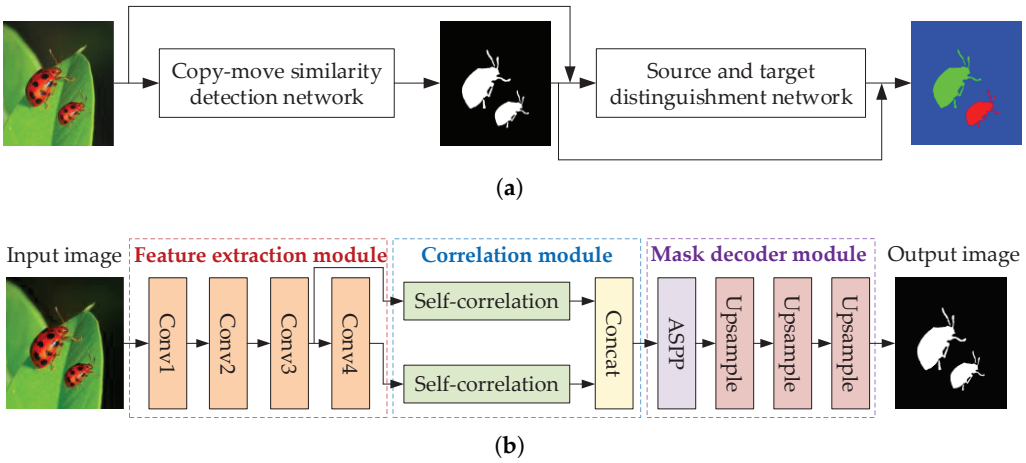

Figure 6. A serial network for CMFD: (**a**) the architecture of the proposed scheme and (**b**) the architecture of copy-move similarity detection network. Adapted from [49].

3.2. Image Splicing Forgery Detection

Image splicing forgery is also one of the most popular ways to manipulate the content of an image. The manipulation operation of splicing is copying an area from one image and pasting it to another image. The tampered images are a serious threat to the security of image information. It is crucial to develop suitable methods to detect image splicing forgery. Image splicing detection techniques are divided into manual feature-based methods and deep learning-based methods.

The current manual feature-based methods can be divided into three categories: textural feature-based techniques, noise-based techniques, and other techniques. The textural feature-based techniques use the difference between the local frequency distribution in tampered images and the local frequency of the real image to detect image splicing. The commonly used textural feature descriptors are local binary pattern (LBP) [52], gray level co-occurrence matrixes (GLCM) [53], local directional pattern (LDP) [54], etc. Since the splicing images come from two different images, the noise distribution of the tampered image is changed. Therefore, the noise-based techniques detect tamper by estimating the noise of the tampered image. The detection of splicing images can also be performed by fusing multiple features. Manual feature-based methods use different descriptors to obtain specific features, and the detection effect is various for different datasets. The generalization of manual feature-based techniques is poor. The deep learning-based methods can automatically learn a large number of features, which improves the accuracy and generalization ability performance of image splicing detection.

Deep learning-based methods can learn and optimize the feature representations for forgery forensics directly. This has inspired researchers to develop different techniques to detect image splicing. In recent years, the U-Net structure has been more widely used in splicing forgery detection. Wei et al. [55] proposed a tamper detection and localization network based on U-Net with multi-scale control. The network used a rotated residual structure to enhance the learning ability of features. Zeng et al. [56] proposed a multitask model for locating splicing tampering in an image, which fused an attention mechanism,

densely connected network, ASPP, and U-Net. The model can capture more multi-scale features while expanding the receptive field and improving the detection accuracy of image splicing tampering. Zhang et al. [57] created a multi-task squeeze and extraction network for splicing localization. The network consisted of a label mask stream and edge-guided stream, which used U-Net architecture. Squeeze and excitation attention modules (SEAMs) were incorporated into the multi-task network to recalibrate the fused features and enhance the feature representation.

Many researchers have also used fully convolutional networks (FCN) commonly used in semantic segmentation for image splicing forgery detection. Chen et al. [58] proposed a residual-based fully convolutional network for image splicing localization. The residual blocks were added to FCN to make the network easier to optimize. Zhuang et al. [59] created a dense fully convolutional network for image tampering localization. This structure comprised dense connections and dilated convolutions to capture subtle tampering traces and obtain finer feature maps for prediction. Liu et al. [60] proposed an FCN with a noise feature. The network enhanced the generalization ability by extracting noise maps in the pre-processing stage to expose the subtle changes in the tampered images and improved the robustness by adding the region proposal network. The technique could accurately locate the tampered regions of images and improve generalization ability and robustness.

In recent years, attention mechanisms have been deeply developed and have gained a great advantage in the field of natural language processing. Many researchers have started to incorporate the attention mechanism in tampering detection. Ren et al. [61] proposed a multi-scale attention context-aware network and designed a multi-scale multi-level attention module, which not only effectively solved the inconsistency of features at different scales, but also automatically adjusted the coefficients of features to obtain a finer feature representation. To address the problem of poor accuracy of splicing boundary, Sun et al. [62] proposed an edge-enhanced transformer network. A two-branch edge-aware transformer was used to generate forgery features and edge features to improve the accuracy of tampering localization.

3.3. Image Generic Forgery Detection

To enable multiple types of tampering detection, the algorithm has been made more generalizable. Zhang et al. [63] proposed a two-branch noise and boundary network that used an improved constrained convolution to extract the noise map. It can effectively solve the problem of training instability. The edge prediction module was added to extract the tampered edge information to improve the accuracy of localization. But the detection performance was poor when the tampered image contained less tampered information. Dong et al. [64] designed a multi-view, multi-scale supervised image forgery detection model. The model combined the boundary and noise features of the tampered regions to learn semantic-independent features with stronger generalization, which improved detection accuracy and generalization. But the detection effect was poor when the tampered region was the background region. Chen et al. [65] proposed a network based on signal noise separation to improve the robustness. The signal noise separation module separated the tampered regions from the complex background regions with post-processing noise, reducing the negative impact of post-processing operations on the image and improving the robustness of the algorithm. Liu et al. [66] proposed a network for learning and enhancing multiple tamper traces, fusing multiple features of global noise, local noise, and detailed artifact features for forgery detection, which enabled the algorithm to have high generalization and detection accuracy. However, when the tampering artifacts are reduced, the lack of effective tampering traces results in tampered regions being undetectable. Wang et al. [67] proposed a multimodal transformer framework, which consisted of three main modules: high-frequency feature extraction, an object encoder, and an image decoder, to address the difficulty of capturing invisible subtle tampering in the RGB domain. The frequency features of the image were first extracted, and the tampered regions were

identified by combining RGB features and frequency features. The effectiveness of the method was shown on different datasets.

To achieve a more refined prediction of tampering masks, a progressive mask-decoding approach is used. Liu et al. [68] proposed a progressive spatio-channel correlation network. The network used two paths, the top-down path acquired local and global features of the image, and the bottom-up path was used to predict the tampered mask. The spatio-channel correlation module was introduced to capture the spatial and channel correlation of features and extract features with global clues to enable the network to cope with various attacks and improve the robustness of the network. To solve the problem of irrelevant semantic information, Shi et al. [69] proposed a progressively-refined neural network. Tampered regions were localized progressively under a coarse-to-fine workflow and rotated residual structure was used to suppress the image content during the generation process. Finally, the refined mask was obtained.

To solve the existing problems of low detection accuracy and poor boundary localization, Gao et al. [70] proposed an end-to-end two-stream boundary-aware network for generic image forgery detection and localization. The network introduced an adaptive frequency selection module to adaptively select appropriate frequencies to mine inconsistent statistical information and eliminate the interference of redundant information. Meanwhile, a boundary artifact localization module was used to improve the boundary localization effect. To address the problem of poor generalization ability to invisible manipulation, Ganapathi et al. [71] proposed a channel attention-based image forgery detection framework. The network introduced a channel attention module to detect and localize forged regions using inter-channel interactions to focus more on tampered regions and achieve accurate localization. To identify forged regions by capturing the connection between foreground and background features, Xu et al. [72] proposed a mutually complementary forgery detection network, which consisted of two encoders for extracting foreground and background features, respectively. A mutual attention module was used to extract complementary information from the features, which consisted of self-feature attention and cross-feature attention. The network significantly improved the localization of forged regions using the complementary information between foreground and background.

To improve the generalization ability of the network model, Rao et al. [73] proposed a multi-semantic conditional random field model to distinguish the tampered boundary from the original boundary for the localization of the forged regions. The attention blocks were used to guide the network to capture more intrinsic features of the boundary transition artifacts. The attention maps with multiple semantics were used to make full use of local and global information, thus improving the generalization ability of the algorithm. Li et al. [74] proposed an end-to-end attentional cross-domain network. The network consisted of three streams that extracted three types of features, including visual perception, resampling, and local inconsistency. The fusion of multiple features improved the generalization ability and localization accuracy of the algorithm effectively. Yin et al. [75] proposed a multi-task network based on contrast learning for the localization of multiple manipulation detection. Contrast learning was used to measure the consistency of statistical properties of different regions to enhance the feature representation and improved the performance of detection and localization.

To solve the problems of low accuracy and insufficient training data, Zhou et al. [76] designed a coarse-to-fine tampering detection network based on a self-adversarial training strategy. A self-adversarial training strategy was used to dynamically extend the training data to achieve higher accuracy. Meanwhile, to solve the problem of the insufficient dataset, Ren et al. [77] designed a novel dataset, called the multi-realistic scene manipulation dataset, which consisted of three kinds of tampering, including copy-move, splicing, and removal, and covered 32 different tampering scenarios in life. A general and efficient search and recognition network was proposed to reduce the computational complexity of forgery detection.

The state-of-the-art deep learning-based CMFD algorithms and performance comparison are described in Table 2. Table 2 describes the methods from four aspects: type of detection, backbone, robustness performance, and dataset.

Table 2. A comparison of deep learning-based passive forgery detection methods.

Ref.	Year	Type of Detection	Backbone	Robustness Performance	Dataset
Li et al. [40]	2022	Copy-move forgery	VGG 16, Atrous convolution	Brightness change, Image blurring, JPEG compression, Color reduction, Contrast adjustments, Noise adding	USCISI, CoMoFoD, CASIA V2
Liu et al. [43]	2022	Copy-move forgery	VGG 16, SuperGlue	Rotation, Scaling, Noise adding, JPEG compression	Self-datasets
Zhong et al. [44]	2020	Copy-move forgery	DenseNet	Rotation, Scaling, Noise adding, JPEG compression	FAU, CoMoFoD, CASIA V2
Kafali et al. [45]	2021	Copy-move forgery	VGG 16, Volterra convolution	Brightness change, Image blurring, JPEG compression, Color reduction, Contrast adjustments, Noise adding	USCISI, CoMoFoD, CASIA
Nazir et al. [46]	2022	Copy-move forgery	DenseNet, RCNN	Brightness change, Image blurring, JPEG compression, Color reduction, Contrast adjustments, Noise adding	CoMoFoD, MICC-F2000, CASIA V2
Zhong et al. [48]	2022	Copy-move forgery	DenseNet	Brightness change, Image blurring, JPEG compression, Color reduction, Contrast adjustments, Noise adding	IMD, CoMoFoD, CMHD [78]
Wu et al. [28]	2018	Copy-move forgery	VGG 16	Brightness change, Image blurring, JPEG compression, Color reduction, Contrast adjustments, Noise adding	USCISI, CoMoFoD, CASIA V2
Chen et al. [49]	2021	Copy-move forgery	VGG 16, Attention module	Brightness change, Image blurring, JPEG compression, Color reduction, Contrast adjustments, Noise adding	USCISI, CoMoFoD, CASIA V2, COVERAGE
Aria et al. [50]	2022	Copy-move forgery	VGG 16	Brightness change, Image blurring, JPEG compression, Color reduction, Contrast adjustments, Noise adding	USCISI, CoMoFoD, CASIA V2
Barni et al. [51]	2021	Copy-move forgery	ResNet 50	JPEG compression, Noise, Scaling	SYN-Ts, USCISI, CASIA, Grip [79]
Wei et al. [55]	2021	Splicing forgery	U-Net, Ringed residual structure	JPEG compression, Gaussian noise, Combined attack, Scaling, Rotation	CASIA, Columbia
Zeng et al. [56]	2022	Splicing forgery	U-Net, ASPP	JPEG compression, Gaussian blurring	CASIA
Zhang et al. [57]	2021	Splicing forgery	U-Net, SEAM	JPEG compression, Scaling, Gaussian filtering, Image sharpening	Columbia, CASIA, Carvalho

Table 2. *Cont.*

Ref.	Year	Type of Detection	Backbone	Robustness Performance	Dataset
Chen et al. [58]	2020	Splicing forgery	FCN	JPEG compression, Gaussian blur, Gaussian noise	DVMM [80], CASIA, NC17, MFC18
Zhuang et al. [59]	2021	Splicing forgery	FCN	JPEG compression, Scaling	PS-scripted dataset, NIST16 [29]
Liu et al. [60]	2022	Splicing forgery	FCN, SRM	Gaussian noise, JPEG compression, Gaussian blurring	CASIA, Columbia
Ren et al. [61]	2022	Splicing forgery	ResNet 50	JPEG compression, Gaussian noise, Scaling	CASIA, IMD2020, DEFACTO, SMI20K
Sun et al. [62]	2022	Splicing forgery	Transformer	JPEG compression, Media blur, Scaling	CASIA, NC2016 [81]
Zhang et al. [63]	2022	Multiple types of tampering detection	ResNet 34, Non-local module	JPEG compression, Gaussian blur	CASIA, COVERAGE, Columbia, NIST16
Dong et al. [64]	2023	Multiple types of tampering detection	ResNet 50	JPEG compression, Gaussian blur	CASIA V2,COVERAGE, Columbia, NIST16
Chen et al. [65]	2023	Multiple types of tampering detection	ResNet 101	JPEG compression, Gaussian blur, Median blur	Self-datasets, NIST16, Columbia, CASIA
Lin et al. [66]	2023	Multiple types of tampering detection	ResNet 50, Swin transformer	JPEG compression, Gaussian blur, Gaussian noise	CASIA, NIST16, Columbia, COVERAGE, CoMoFoD
Wang et al. [67]	2022	Multiple types of tampering detection	Multimodal transformer	JPEG compression, Gaussian blur, Gaussian noise, Scaling	CASIA, Columbia, Carvalho, NIST16, IMD2020
Liu et al. [68]	2022	Multiple types of tampering detection	HR-Net	JPEG compression, Scaling, Gaussian blur, Gaussian noise	Columbia, COVERAGE, CASIA, NIST16, IMD2020
Shi et al. [69]	2022	Multiple types of tampering detection	VGG 19, Rotated residual	JPEG compression, Gaussian blur, Gaussian noise	NIST16, COVERAGE, CASIA, In-The-Wild
Gao et al. [70]	2022	Multiple types of tampering detection	ResNet 101	JPEG compression, Scaling	CASIA, Carvalho, COVERAGE, NIST16, IMD2020
Ganapathi et al. [71]	2022	Multiple types of tampering detection	HR-Net	Flipped horizontally and vertically, Saturation, Brightness	CASIA V2, NIST16, Carvalho, Columbia
Xu et al. [72]	2022	Multiple types of tampering detection	VGG 16	JPEG compression, Scaling, Gaussian blur, Gaussian noise	NIST16, COVERAGE, CASIA, IMD2020
Rao et al. [73]	2022	Multiple types of tampering detection	Residual unit, CRF-based attention, ASPP	JPEG compression, Scaling	COVERAGE, CASIA, Carvalho, IFC
Li et al. [74]	2022	Multiple types of tampering detection	ResNet101, Faster R-CNN	Median filtering, Gaussian noise, Gaussian blur, Resampling	CASIA, Columbia, COVERAGE, NIST16
Yin et al. [75]	2022	Multiple types of tampering detection	Convolution and Residual block	JPEG compression, Gaussian blur, Gaussian noise, Scaling	NIST16, CASIA, COVERAGE, Columbia
Zhou et al. [76]	2022	Multiple types of tampering detection	VGG-style block	JPEG compression, Gaussian noise, Gaussian blur, Scaling	DEFACTO, Columbia, CASIA, COVERAGE, NIST16
Ren et al. [77]	2022	Multiple types of tampering detection	ResNet 50	JPEG compression, Gaussian noise, Scaling	NIST16, CASIA, MSM30K

4. Robust Image Watermarking Based on Deep Learning

According to the gradient updating feature of the attack, robust watermarking can be classified into two categories: robust differentiable attack watermarking and robust non-differentiable attack watermarking. For differentiable attacks, the adversarial simulation layer (ASL) can be introduced into the model to generate attack counterexamples directly, while non-differentiable attacks require other means to improve their robustness, such as differentiable approximation, two-stage training, and network structure improvement, etc.

4.1. Robust Image Watermark against Differentiable Attack

As shown in Figure 7, differential attacks can be further categorized into noise and filtering attacks, as well as geometric attacks.

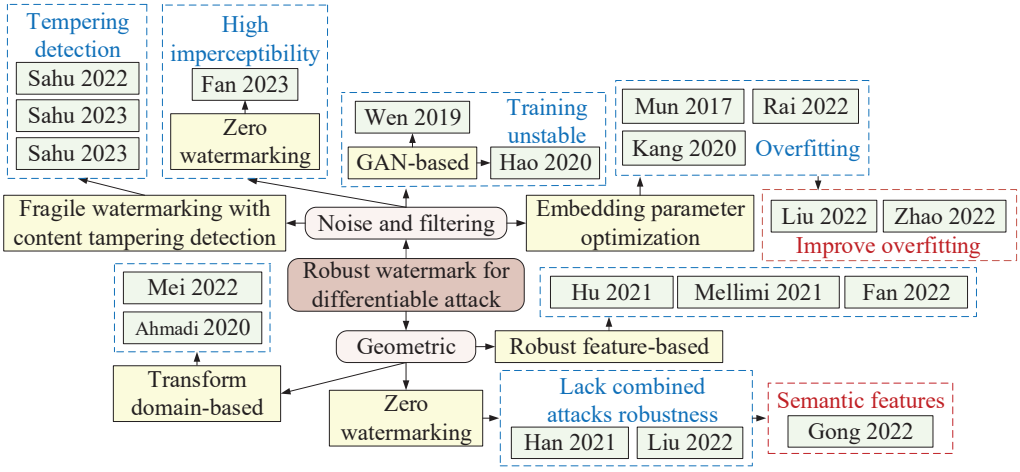

Figure 7. An overview of robustness enhancement methods for differential attacks.

4.1.1. Robust Image Watermark against Noise and Filtering Attack

Noise and filtering attacks apply corresponding noise or filtering to the pixel from the spatial domain and transform domain directly, affecting the amplitude coefficient synchronization of the codec. There are three methods for recovering amplitude synchronization: zero-watermarking, generative adversarial network (GAN) [82]-based, and embedding coefficient optimization.

Gaussian attacks primarily blur the details of an image by adding Gaussian noise, thereby reducing the robustness of the watermark. This attack affects the position and intensity of the embedded watermark, which may result in incorrect extraction or a decrease in the quality of the extracted watermark information. To counter Gaussian attacks, traditional methods typically employ anti-noise and filtering techniques to enhance the robustness of the watermark. For example, adaptive filters [83] and noise estimation [84] techniques can be used to reduce the impact of Gaussian noise. In deep learning approaches, methods such as adversarial training, zero-watermarking, and embedding parameter optimization are utilized to enhance the robustness against Gaussian attacks. For example, Wen and Aydore [85] introduced adversarial training into the watermarking algorithm where the distortion type and the distortion strength were adaptively selected thus minimizing the decoding error.

Zero-watermarking: The conventional zero-watermarking algorithm consists of three steps: original image robust feature extraction, zero-watermark generation, and zero-watermark verification. To extract noise-invariant features, Fan et al. [86] used a pre-trained Inception V3 [87] network to extract the image feature tensor initially, and extracted its low-frequency subbands by DCT transform to generate a robust feature tensor. They dissociated the binary sequence generated by chaos mapping with the watermark information to obtain a noise-invariant zero-watermark.

GAN-based: In the watermarking model, GAN can generate high-quality images more efficiently by competing between the discriminator (i.e., watermarked image discriminator) and the generator (i.e., encoder) under adversarial loss supervision to continuously update each other's parameters and improve the decoding accuracy by supervising the decoder through total loss function indirectly. Since the human visual system focuses more on the central region of the image, Wen and Aydore [85] introduced adversarial training into the watermarking algorithm where the distortion type and the distortion strength were adaptively selected thus minimizing the decoding error. Hao et al. [88] added a high-pass filter before the discriminator so that the watermark tended to be embedded in the mid-frequency region of the image, giving a higher weight to the middle region pixels in the computed loss function. However, it could not resist geometric attacks effectively. Zhang et al. [89] proposed an embedding-guided end-to-end framework for robust watermarking. It uses a prior knowledge extractor to obtain the chrominance and edge saliency of cover images for guiding the watermark embedding. However, it could not be applied to practical scenarios such as printing, camera photography, and geometric attacks. Li et al. [90] designed a single-frame exposure optical image watermarking framework using conditional generative adversarial network (CGAN) [91]. Yu [92] introduced an attention mechanism to generate attention masks to guide the encoder to generate better target images without disturbing the spotlight, and improved the reliability of the model by combining GAN with a circular discriminant model and inconsistency loss. However, refs. [85,88–90,92] did not effectively address the GAN network training instability problem, resulting in none of them being able to further improve the balance of robustness and imperceptibility.

Embedding Parameter Optimization: The position and strength of the embedding parameters determine the algorithm performance directly. Mun et al. [93] performed an iterative simulation of the attack on the watermarking system. But it can only obtain one bit of watermark information from a sub-block. Kang et al. [94] first subjected the host image to DCT transformation to extract the human eye-insensitive LH sub-band and HL sub-band. Particle swarm optimization (PSO) was used to find the best DCT coefficients and the best embedding strength to improve the imperceptibility and robustness of the watermarking algorithm. However, due to the training overfitting of PSO [95], its model generalized and achieved good robust performance only on the experimental dataset. Rai and Goyal [96] combined fuzzy, backpropagation neural networks and shark optimization algorithms. However, refs. [94,96] had the problem of training overfitting. To improve the training overfitting problem, Liu et al. [97] introduced a two-dimensional image information entropy loss to enhance the ability of the model to generate different watermarks, ensuring that the model was always able to assign enough information to a single host image for different watermark inputs and the extractor can extract the watermark information completely, therefore enhancing the dynamic randomness of the watermark embedding. Zhao et al. [98] specifically adopted an end-to-end robust image watermarking algorithm framework, known as the embed-attack-extract paradigm. In the embedding layer of the network, it incorporated the channel spatial attention mechanism. As a result, during training, after forward and backward propagation for parameter updates, the embedding layer's parameters were able to focus on more effective channel and spatial information, which also made the model focus on the optimization of increasing the accuracy of the extracted watermark. To sum up, this optimization of the watermark embedding parameters contributed to enhancing the model's resistance to noise and filtering attacks.

Deep learning-based image watermarking against noise and filtering attack algorithms and performance comparison are described in Table 3. Table 3 describes the methods from five aspects: watermark size (container size), category, method (effect), robustness, and dataset, where σ_b, σ_f, and σ_n represents the variance of Gaussian blur, Gaussian filtering, and Gaussian noise, respectively.

Table 3. The comparison of deep learning-based image watermarking against noise and filtering.

Ref.	Watermark Size (Container Size)	Category	Method (Effect)	Robustness (Attack, Parameter)		Dataset
				BER (%)	NC	
Hao et al. [88]	30 (64 × 64)	GAN-based	GAN (Improving visual quaility)	0.5 (Gaussian blur, 3 × 3, σ_b = 2.0)	–	COCO [99]
Mun et al. [93]	24 (512 × 512)	Embedding parameter optimization	CNN (Feature extraction)	–	0.9625 (Gaussian blur, 3 × 3, σ_b = 1)	MPEG-7 CE Shape-1 [100]
Kang et al. [94]	1024 (1024 × 1024)	Embedding parameter optimization	PSO (Selecting best DCT coefficient)	0 (Gaussian filtering, 3 × 3, σ_f = 0.5)	0.990 (Gaussian filtering, 3 × 3, σ_f = 0.5)	USC-SIPI [101]
Rai et al. [96]	32 (96 × 96)	Embedding parameter optimization	SSO (Gaining ideal embedding parameter)	–	0.8259 (Gaussian noise, σ_n = 0.01)	Self-datasets
Zhao et al. [98]	32 × 32 (512 × 512)	Embedding parameter optimization	Spatial and channel attention mechanism (Improving robustness)	0.09 (Gaussian noise, σ_n = 0.05)	0.9988 (Gaussian noise, σ_n = 0.05)	BOSS Base [102], CIFAR 10 [103]

Fragile Watermark with Content Tampering Detection: The above is based on robust watermarking to achieve the active forensics of noise and filtering attacks, the following is based on fragile watermarking to achieve the copyright protection of noise and filtering attacks and content tampering proof.

Sahu [104] proposed a fragile watermarking scheme based on logical mapping to effectively detect and locate the tampered region of watermarked images. It took advantage of the sensitivity of logical mapping to generate watermark bits by performing a logical XOR operation between the first intermediate significant bit (ISB) and the watermark bit, and embedding the result in the rightmost least significant bits (LSBs). The watermarked image obtained by this method is of high quality and good imperceptibility with an average peak signal-to-noise ratio (PSNR) of 51.14 dB. It can effectively detect and locate the tampering area from the image, and can resist both intentional and unintentional attacks to a certain extent. However, it cannot recover the tampered region.

Sahu et al. [105] studies content tampering in multimedia watermarking. It introduces a feature association model based on a multi-task learning framework that can relatively detect multiple modifications based on location and time. Multi-task learning frameworks leverage the interrelationships between multiple related problems or tasks and can effectively learn from shared representations to provide better models. It also utilizes convolutional neural networks (CNNs) for climate factor prediction tasks, which can be trained and evaluated on large-scale datasets to learn different convolutional models using regression models and distance-based loss functions. The results show that the method using deep learning can achieve high precision and extremely high accuracy. It ensures that image data are collected from reliable sources and manually verifies the authenticity of the data, including images collected from AMOS (archive of many outdoor scenes) and Internet webcams with time stamps, camera ids, and location annotations. This ensures that the information collected is true.

There are still difficulties in the detection of tampered image metadata: the literature points out that the open-source ExifTool is applied to access and modify image metadata between digital workflows, but this method is easy to be tampered with, so additional evidence is needed to prove the authenticity of image content, such as the analysis of other climate factor images. This indicates that there are still challenges in tampering detection of metadata. To sum up, the advantages of this literature are a reliable data acquisition process, multi-task learning framework, and prediction accuracy based on deep learning. However, it is difficult to detect the tampered image metadata and the data set is not evenly distributed. It can also be a baseline in future works to propose a CNN-based content

retrieval and tampering strategy. Sahu et al. [106] proposed an efficient reversible fragile watermarking scheme based on two images (DI-RFWS), which can accurately detect and locate the tamper region in the image. The scheme used a pixel adjustment strategy to embed two secret bits in each host image (HI) pixel to obtain a double-watermarked image (WI). According to the watermark information, the non-boundary pixels of the image are modified to the maximum of ±1. Due to the potency of reversibility, the proposed scheme can be adapted to a wide range of contemporary applications.

4.1.2. Robust Image Watermark against Geometric Attacks

Geometric attacks mainly include rotation, cropping, and translation that change the spatial position relationship of pixels and affect the spatial synchronization of codecs. According to the method of recovering spatial synchronization, it can be divided into transform domain-based, zero-watermarking, and robust feature-based.

For cropping attacks, when the cropping involves the watermark embedding region, the watermark extraction algorithm may not match the watermark features correctly, resulting in the quality of the extracted watermark information being reduced. Multi-location embedding, multi-watermark embedding, and multi-watermark embedding are three main methods to improve the robustness of cropping attacks. Hsu and Tu [107] used the sinusoidal function and the wavelength of the sinusoidal function to design the embedding rule, which gives each watermark bit multiple copies spread across different blocks. Therefore, even under cropping attacks, other watermarks can be saved to achieve copyright authentication.

Transform Domain-based: The transform domain-based watermarking algorithm embeds the watermark into the transform domain coefficients, avoiding geometric attacks from corrupting the synchronization of the codec effectively. Ahmadi [108] implemented watermark embedding and extraction in the transform domain (such as DCT, DWT, Haar, etc.), and introduced circular convolution in the convolution layer used for feature extraction to make the watermark information diffuse in the transform domain, which effectively improved the robustness against cropping. Mei et al. [109] embedded a watermark in the DCT domain and introduced the attention mechanism to calculate the confidence of each image block. In addition, joint source-channel coding was introduced to make the algorithm maintain good robustness and imperceptibility under the gray image watermarking with the background.

Zero-watermarking: The idea of the zero-watermarking is to obtain geometrically invariant features from robust features in images and generate zero-watermark through zero-watermark generation by sequence synthesis operation such as dissimilar operation. Han et al. [110] used pre-trained VGG19 [111] network to extract original image features and selected DFT-transformed low-frequency sub-bands to construct a medical image feature matrix. Liu et al. [112] used neural style transfer (NST) [113] technique combined with a pre-trained VGG19 network to extract the original image style features, fused the original image style with the original watermark content to obtain the style fusion image, and Arnold dislocation [114] to obtain the zero-watermark. Gong et al. [115] used the low-frequency features of the DCT of the original image as labels. Skip connection and loss functions were applied to enhance and extract high-level semantic features. However, none of the authors [110,112,115] could resist the robustness of multiple types of attacks effectively.

Robust Feature-based: Unlike zero-watermarking, the idea of this method is to search for embedded feature coefficients or tensors in the image and embed the watermark in it robustly. Hu et al. [116] embedded the watermark in low-order Zernike moments with rotation and interpolation invariance. Mellimi et al. [117] proposed a robust image watermarking scheme based on lifting wavelet transform (LWT) and deep neural network (DNN). The DNN was trained to identify the changes caused by attacks in different frequency bands and select the best subbands for embedding. However, it was not robust to speckle noise. Fan et al. [118] combined the multiscale features in GAN and used

pyramid filters and multiscale maximum pooling techniques to learn the watermark feature distribution and improve the geometric robustness of watermarking fully.

The state-of-the-art deep learning-based image watermarking against geometric attacks algorithms and performance comparison are described in Table 4. Table 4 describes the methods from five aspects: watermark size (container size), category, method (effect), robustness, and dataset, where s_f represents the scaling factor for the scaling attack; p_c represents the cropping ratio for cropping; Q_R represents the clockwise rotation of the rotation attack.

Table 4. A comparison of deep learning-based image watermarking against geometric attacks.

Ref.	Watermark Size (Container Size)	Category	Method (Effect)	Robustness (Attack, Parameter)		Dataset
				BER (%)	NC	
Ahmadi et al. [108]	1024 (512 × 512)	Transform domain-based	Circular convolution (Diffusing watermark information), Residual connection (Fusing low-level character)	5.9 (Scaling, s_f = 0.5)	–	CIAFAR 10 [103], Pascal VOC [119]
Mei et al. [109]	1024 (512 × 512)	Transform domain-based	DCT, Attention, Joint source-channel coding (Improving robustness)	0.96 (Cropping, p_c = 0.75), 0.34 (Cropping, p_c = 0.5)	–	COCO [99]
Han et al. [110]	–	Zero-watermark	VGG19, DFT (Feature extraction)	–	0.87509 (Q_R = 50)	Self-datasets
Liu et al. [112]	–	Zero-watermark	VGG19 (Feature extraction)	–	0.95 (Q_R = 40), 0.96 (Scaling, s_f = 0.5)	Waterloo Exploration Database [120]
Gong et al. [115]	–	Zero-watermark	Residual-DenseNet (Feature extraction)	–	0.89 (Rotation, Q_R = 45)	Self-datasets
Hu and Xiang [116]	128 (512 × 512)	Robust feature-based	CNN (Feature extraction), GAN (Visual improvement)	0.6 (Scaling, s_f = 2)	–	USC-SIPI [101]
Mellimi et al. [117]	1024 (512 × 512)	Robust feature-based	DNN (Optimal embbeding subband selection)	1.78 (Scaling, s_f = 0.65), 0.2 (Scaling, s_f = 0.75)	0.9353 (Scaling, s_f = 0.65), 0.9930 (Scaling, s_f = 0.75)	USC-SIPI [101]

4.2. Robust Image Watermark against Non-Differentiable Attack

As shown in Figure 8, non-differential attacks can be further categorized into JPEG attacks, screen-shooting attacks, and agnostic attacks.

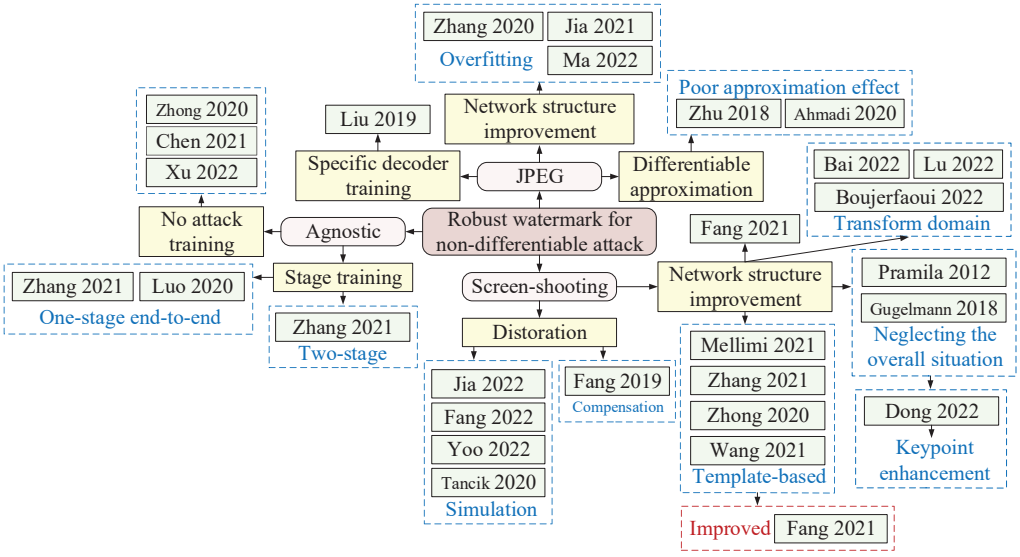

Figure 8. An overview of robustness enhancement methods for non-differential attacks.

4.2.1. Robust Image Watermark against JPEG Attack

In recent years, end-to-end watermarking algorithms based on deep learning have been emerging, and thanks to the powerful feature extraction capabilities of CNN, watermarks can be covertly embedded in low-perception pixel regions of the human eye (such as diagonal line, textured complex regions, high-brightness slow-change regions, etc.) to obtain watermarked images that are very similar to the original images. In the end-to-end training, to improve the robustness of the watermarking algorithm, the watermarked image is added to the differentiable attack by introducing an attack simulation layer to generate the attacked counterexamples, and the decoder parameters are updated by decoding losses such as mean square error (MSE), binary cross entropy (BCE) of the original watermark and decoded watermark, etc. However, due to the introduction of the non-differentiable nature of real JPEG compression, it cannot be introduced into the end-to-end network to implement back-propagation updates of model parameters directly. To address this problem, relevant studies have recently been proposed, which can be subdivided into three directions according to the differences in the methods used to generate the JPEG counterexamples: differentiable approximation, specific decoder training, and network structure improvement.

Differentiable Approximation: Zhu et al. [121] proposed the JPEG-MASK approximation method first, which mainly set the high-frequency DCT coefficients to 0 and retained the 5 × 5 low-frequency coefficients of the Y channel and the 3 × 3 low-frequency coefficients in the U and V channels, which had some simulation effect on JPEG. Based on [121], Ahmadi et al. [108] added a residual network with adjustable weight factors to the watermark embedding network to achieve autonomous adjustment of imperceptibility and robustness. Meanwhile, unlike conventional convolution operation, circular convolution was introduced to achieve watermark information diffusion and redundancy. SSIM was used as the loss function of the encoder network to make the watermarked image more closely resemble the original image in terms of contrast, brightness, and structure. Although differentiable approximation methods effectively solve the back-propagation update problem for training parameters, the two papers [108,121] both suffered from bad simulation approximation, which led to a decoder that cannot perform robust parameter updates more efficiently against real JPEG in turn.

Specific Decoder Training: To avoid the introduction of non-differentiable components in the overall training process of the model. Liu et al. [122] proposed a two-stage separable training watermarking algorithm consisting of noise-free end-to-end adversarial training (FEAT) and a decoder only trained (ADOT) with an attack simulation layer. In FEAT, the encoder and decoder were jointly trained to obtain a redundant encoder, and in ADOT, the encoder parameters were fixed and spectral regularization was used to effectively mitigate the training instability problem of GAN networks. At the same time, corresponding attacks were applied to the watermarked images to obtain the corresponding attack samples, which were then used as the training set to train the dedicated decoder. The disadvantages of the non-gradient nature of JPEG were solved, but the phased training suffered from the problem of training local optima.

Network Structure Improvement: Chen et al. [123] proposed a JPEG simulation network JSNet which could simulate JPEG lossy compression with any quality factors. The three processes of sampling, DCT, and quantization in JPEG were simulated by a maximum pooling layer, a convolution layer, and a 3D noise mask. However, it was found experimentally that the model was still less robust to JPEG compression after the introduction of JSNet (i.e., BER is greater than 30% under both ImageNet [124] and Boss Base [102] dataset tests). This was related to its use of random initialization of parameters in the maximum pooling, convolutional layer, and 3D noise layer during the simulation, which resulted in a poor simulation effect for JPEG compression.

Due to the introduction of the non-differentiable nature of real JPEG compression, it cannot be introduced into the end-to-end network to implement back-propagation updates of model parameters directly. To address this problem, Jia et al. [125] used a mini-batch

of real and simulated (MBRS) JPEG compression to improve the JPEG compression attack robustness. In the attack layer, one of several small mini-batches attacks was selected randomly from the real, simulated, and equivalent sound layers as the noise layer. Please note that the attacks polled by the model in the first iteration are simulated to facilitate cumulative updates based on the differentiable gradient of the first iteration. This was performed thanks to the Adam momentum optimizer with its historical gradient update, which is expressed as Equations (12)–(16).

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \tag{12}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \tag{13}$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{14}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{15}$$

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\hat{v}_t + \varepsilon}} \cdot \hat{m}_t \tag{16}$$

where m_t and v_t denote the first and second moment estimates at the time step t , i.e., the exponential moving average of the gradient and the gradient squared. Where β_1 and β_2 denote the average coefficient, usually set to a value close to 1. g_t denotes the gradient in time step. θ_t denotes the model parameter in time step t. η denotes the learning rate. ε denotes a small constant added for numerical stability to prevent the denominator from being zero. $\hat{m}_t$ and $\hat{v}_t$ denote unbiased estimation modified to the first and second moment estimation, respectively.

Thanks to the Adam momentum optimizer with its historical gradient update, even if the attack layer rotated to the non-differentiable real JPEG compression attack, the internal parameters of the codec network could still be updated by the accumulation of historical differentiable gradients, which avoided the problem of non-differentiable real JPEG compression and achieved a better simulation quality of JPEG compression effectively. However, the model ignored the feature tensor of the image in the spatial and channel directions for image feature extraction, which led to poor robustness in the face of high-intensity salt and pepper noise. Zhang et al. [126] proposed a pseudo-differentiable JPEG method. JPEG pseudo-noise was the difference between the compressed processed image and the original image. Since its backpropagation without going through the pseudo-noise sound path, there was no problem of non-differentiable. However, its robustness to conventional noise was poor due to the lack of noise prior when back propagating. Ma et al. [127] proposed a robust watermarking framework by combining reversible and irreversible mechanisms. In the reversible part, a diffusion extraction module (DEM) (consisting of a series of fully connected layers and a Haar transform) and a fusion separation module (FSM) were designed to implement watermark embedding and extraction in a reversible manner. For the irreversible part, the irreversible attention model which was composed of a series of convolution layers including full-connected layer, squeeze and excitation block, and a dedicated noise selection module (NSM) were introduced to improve the JPEG compression robustness.

The state-of-the-art deep learning-based image watermarking against JPEG attack algorithms and performance comparison are described in Table 5. Table 5 describes the methods from five aspects: watermark size (container size), method, structure, robustness, imperceptibility, and dataset, where Q_F represents the quality factor of the JPEG compression.

Table 5. A comparison of deep learning-based image watermarking against JPEG attack.

Ref.	Watermark Size (Container Size)	Category	Method (Effect)	Robustness (BER (%))	Imperceptibility (PSNR (dB))	Dataset
Ahmadi et al. [108]	1024 (512 × 512)	Differentiable approximation	CNN, Residual connection, Circular convolution	1.2 (50), 0 (70), 0 (90)	35.93	CIFAR 10 [103], Pascal VOC [119]
Liu et al. [122]	30 (128 × 128)	Specific decoder training	CNN, GAN	23.8 (50)	33.5	COCO [99], CIFAR 10 [103]
Chen et al. [123]	1024 (256 × 256)	Network structure improvement	CNN, JSNet	0.097 (90), 32.421(80)	–	ImageNet [124], Boss Base [102]
Jia et al. [125]	64 (128 × 128)	Network structure improvement	CNN, Residual connection	4.14 (50)	39.32	COCO [99]
Zhang et al. [126]	30 (128 × 128)	One-stage end-to-end	Backward ASL	12.64	–	Self-datasets
Ma et al. [127]	30 (128 × 128)	Network structure improvement	DEM, Non-invertible attention module	0.76 (50)	38.51	COCO [99]

4.2.2. Robust Image Watermark against Screen-Shooting Attack

Screen-shooting attacks mainly cause image transmission distortion, brightness distortion, and Moiré distortion [128]. To improve the robustness of the screen-shooting attacks, relevant studies have recently been proposed, which can be subdivided into template-based, distortion compensation-based, decoding based on the attention mechanism, keypoint enhancement, transform domain, and distortion simulation.

Template-based: Template-based watermarking algorithms have a high watermarking capacity. The templates characterizing the watermarking information are embedded in the image in a form similar to additive noise. To ensure robustness, templates usually carry special data distribution features, but conventional template-based watermarking algorithms [117,129–131] were designed with low complexity manually and thus could not cope with sophisticated attacks. Fang et al. [132] designed a template-based watermarking algorithm by exploiting the powerful feature learning capability of DNN. In the watermark embedding phase, the embedding template was designed based on the insensitivity of human eyes to the specific chromatic components, the proximity principle, and the tilt effect. In the watermark extraction phase, a two-stage DNN was designed, containing an auxiliary enhancement sub-network for enhancing the embedded watermark features and classification of the sub-network for extracting the internal information of the watermark.

Distortion Compensation-based: Fang et al. [133] used the method of swapping DCT coefficients to achieve watermark embedding and a distortion compensation extraction algorithm to achieve the robustness of the watermark to photographic processing. Specifically, a line spacing region and a symmetric embedding block were used to reduce the distortion generated by the text.

Decoding Based on Attention Mechanism: Fang et al. [134] designed a transparency, efficiency, robustness, and adaptability coding to effectively mitigate the conflict between transparency, efficiency, robustness, and adaptability. A color decomposition method was used to improve the visual quality of watermarked images, and a super-resolution scheme was used to ensure the embedding efficiency. Bose Chaudhuri Hocquenghem (BCH) coding [135] and an attention decoding network (ADN) were used to further ensure robustness and adaptivity.

Keypoint Enhancement: The feature enhanced keypoints are used to locate the watermark embedding region, but the existing keypoint enhancement methods [136,137] ignore the improvement of the overall algorithm by separating the two steps of keypoint enhancement and watermark embedding. Dong et al. [138] used a convex optimization framework to unify the two steps to improve the accuracy of watermark extraction and blind synchronization of embedding regions effectively.

Transform Domain: Bai et al. [139] introduced a separable noise layer over the DCT domain in the embedding and extraction layers to simulate screen-shooting attacks. SSIM-based loss functions were introduced to improve imperceptibility. Spatial transformation

networks were used to correct the values of pixels on the image formed by the geometric attacks before extracting the watermark. Considering that conventional CNN-based algorithms introduce noise in the convolution operation, Lu et al. [140] used DWT and IDWT instead of down-sampling and up-sampling operations in CNN to enable the network to learn a more stable feature representation from the noisy samples and introduced a residual regularization loss containing the image texture to improve the image quality and watermark capacity. The Fourier transform is invariant to rotation and translation; Boujerfaoui et al. [141] improved the Fourier transform-based watermarking method using a frame-based transmission correction of the captured image in the distortion correction process.

Distortion Simulation: Jia et al. [142] introduced a 3D rendering distortion network to improve the robustness of the model to camera photography and introduced a human visual system-based loss function to supervise the training of the encoder, which mainly contained the just notice difference (JND) loss and learned perceptual image patch similarity (LPIPS) loss of the original and watermarked images to improve the quality of the watermarked images. Fang et al. [143] modeled the three components of distortion with the greatest impacts: transmission distortion, luminance distortion, and Moiré distortion and further differentiated the operation so that the network can be trained end-to-end. The network was trained with end-to-end parameters and the residual noise was simply simulated with Gaussian noise. For imperceptibility, a mask-based edge loss was proposed to limit the embedding region which improved the watermarked image quality. This was performed to address the difficulty of conventional 3D watermarking algorithms to achieve watermark extraction from 2D meshes. Yoo et al. [144] proposed an end-to-end framework containing an encoder, a distortion simulator (i.e., a differentiable rendering layer that simulated the results of a 3D watermarked target after different camera angles), and a decoder to decode from 2D meshes. Tancik et al. [145] proposed the stegastamp steganography model to implement the encoding and decoding of hyperlinks. The encoder used a U-Net-like [146] structure to transform a 400 × 400 tensor with 4 channels (including the input RGB image and watermark information) into a tensor of residual image features. However, the algorithm had a small embedding capacity.

The state-of-the-art deep learning-based image watermarking against screen-shooting attack algorithms and performance comparison are described in Tables 6 and 7. Table 6 describes the methods from four aspects: watermark size (container size), category, robustness with BER metrics, and dataset. Table 7 describes the methods from four aspects: watermark size (container size), category, robustness with other metrics, and dataset.

Table 6. A comparison of deep learning-based image watermarking against screen-shooting attack with BER metrics.

Ref.	Watermark Size (Container Size)	Category	Robustness (BER (%))		Dataset
			Distance (cm)	Angle (°)	
Fang et al. [132]	128 (512 × 512)	Templated-based	1.95 (20), 2.73 (40), 11.72 (60)	4.3 (Up40), 1.17 (Up20), 7.03 (Down20), 7.03 (Down40), 5.47 (Left40), 3.91 (Left20), 2.73 (Right20), 3.52 (Right40)	ImageNet [124], USC-SIPI [101]
Fang et al. [133]	48 (256 × 256)	Decoding based on attention mechanism	5.1 (15), 9.9 (35)	9.4 (Up45) 8.1 (Up30), 8.9 (Down30), 9.45 (Down45), 9.7 (Left45), 8.9 (Left30) , 9.8 (Right30), 9.3 (Right45)	Self-datasets
Fang et al. [134]	32 (512 × 512)	Distortion compensation	2.54 (30), 3.71 (50), 5.27 (70)	6.25 (Up30), 3.13 (Up15), 12.73 (Down15), 14.12 (Down30), 7.05 (Left15), 14.46 (Left40), 5.27 (Right15), 11.52 (Right30)	COCO [99]
Dong et al. [138]	64 (64 × 64)	Keypoint enhancement	0.43 (45), 0.35 (65) , 0.67 (75)	2.0 (Left60), 0.66 (Left30), 0.67 (Right30), 2.68 (Right60)	Self-datasets
Jia et al. [142]	100 (400 × 400)	Distortion simulation	–	1.0 (Left65), 0.7 (Left30), 0.7 (Right30), 5.3 (Right65)	Pascal VOC [119], USC-SIPI [101]
Fang et al. [143]	30 (128 × 128)	Tamper detection	2.08 (40), 1.25 (60), 0.62 (20)	2.92/1.25 (Left/Up40), 1.25/0.93 (Left/Up20), 1.05/1.04 (Right/Down20), 0.62/0.83 (Right/Down40)	USC-SIPI [101]

Table 7. A comparison of deep learning-based image watermarking against screen-shooting attack with other experiment conditions.

Ref.	Watermark Size (Container Size)	Category	Robustness (BER (%))	Dataset
Lu et al. [140]	400 (400 × 400)	Transform domain	11.18	MIR Flickr [147]
Yoo et al. [144]	–	Distortion simulation	9.72 (Q_R = 30°)	ModelNet 40-class [148]
Tancik et al. [145]	100 (400 × 400)	Distortion simulation	0.2	ImageNet [124]

4.2.3. Robust Image Watermark against Agnostic Attack

Agnostic attacks refer to attacks where the attack model cannot access prior information about the attack (i.e., the model cannot generate corresponding adversarial examples precisely to guide the decoder to improve its robustness). To address these problems, relevant studies have recently been proposed, which can be subdivided into two-stage separable training, no-attack training, and one-stage end-to-end.

Two-stage Separable Training: Zhang et al. [129] proposed a two-stage separable watermark training model. The first stage jointly trained the codec as well as an attack classification discriminator which used multivariate cross-entropy loss for convergence to obtain encoder parameters that generated stable image quality and an attack classification discriminator that can accurately classify the type of attacks on the image. In the second stage, a fixed encoder, a multiple classification discriminator, and an attack layer were set up using the obtained watermarked image and attacked image prior to training a specific decoder. However, it still did not solve the local optimal solution of the two-stage separable training.

No Attack Training: Zhong et al. [130] introduced a multiscale fusion convolution module, avoiding the loss of image detail feature information as the number of layers of the network deepens, which triggered the inability of the encoder to find an effective hidden embedding point. The invariance layer was set in the encoder and decoder to reproject the most important information and to disable the neural connections in the independent regions of the watermark. Chen et al. [149] proposed a watermark classification network for implementing copyright authentication of attacked watermarks. In the training phase, the training set was generated by calculating the NC value of each watermarked image and classifying its labels into forged and genuine images according to the set threshold. The training set was fed into the model and supervised by the BCE loss function to obtain the model parameters which can accurately classify the authenticity of the watermark. Under high-intensity attacks, the model can still distinguish the real watermark effectively. However, the classification accuracy was affected by the NC threshold setting of the model itself. Xu et al. [150] proposed a blockchain-based zero-watermarking approach, which alleviates the pressure of authenticating zero-watermarks through third parties effectively.

One-stage End-to-end Training: The encoder trained on a fixed attack layer is prone to model overfitting, which is clearly not applicable to realistic watermarking algorithms that need to resist many different types of attacks, and Luo et al. [151] proposed the use of CNN-based adversarial training and channel encoding which can add redundant information to the encoded watermark to improve the algorithm robustness. Zhang et al. [152] proposed the reverse ASL end-to-end model (i.e., the gradient propagation update of parameters was involved in the forward propagation ASL layer, and the gradient does not pass through the ASL layer in the reverse propagation). Reverse ASL can effectively mitigate model overfitting and improve the robustness against agnostic attacks. Zheng et al. [153] proposed a new Byzantine-robust algorithm WMDefence which detected Byzantine malicious clients by embedding the degree of degradation of the model watermark.

The state-of-the-art deep learning-based image watermarking against agnostic attack algorithms and performance comparison are described in Table 8. Table 8 describes the methods from five aspects: watermark size (container size), category, structure, robustness, and dataset, where p_s represents the ratio of salt and pepper in salt and pepper noise.

Table 8. A comparison of deep learning-based image watermarking against agnostic attack.

Ref.	Watermark Size (Container Size)	Category	Structure	Robustness (Attack, Parameter)	Dataset
Zhang et al. [126]	30 (128 × 128)	One-stage end-to-end	Backward ASL	BER: 12.64 (JPEG, Q_F = 50)	–
Zhang et al. [129]	64 (224 × 224)	Two-stage separable training	CNN, GAN, Attack classification discriminator, Residual network	BER: 18.54 (JPEG, Q_F = 50), 8.47 (Cropping, p_c = 0.7), 11.79 (Rotation, 15°), 1.27 (Salt and pepper noise, p_s = 0.01), 1.9 (Gauss filtering, 3 × 3, σ_f = 2)	Pascal VOC [119]
Zhong et al. [130]	32 × 32 (128 × 128)	One-stage end-to-end training	Multi-scale convolution blocks, Invariance layer	BER: 8.16 (JPEG, Q_F = 10), 6.61 (Cropping, 0.8), 0.97 (Salt and pepper, p_s = 0.05)	ImageNet [124], CIFAR 10 [103]
Chen et al. [149]	64 × 64 (512 × 512)	No attack training	WMNet, CNN	Classification accuracy: 0.978	-
Luo et al. [151]	30 (128 × 128)	No attack training	Channel coding, CNN, GAN	BER: 10.5 (Gaussian noise, σ_n = 0.1), 22.9 (Salt and pepper noise, p_s = 0.15)	COCO [99]

5. Future Research Directions and Conclusions

5.1. Future Research Directions

On the basis of the existing problems in the current research status, this subsection gives future research directions for image forensics.

For passive forensics, although tampering detection algorithm techniques have been developed to some extent, there are still some problems of low generalization ability and poor robustness. Because the performance of deep learning-based tampering localization models depends on the training dataset heavily, the performance usually degrades significantly for the test samples from different datasets. It requires us to analyze the intrinsic relationship among images from different sources more deeply and improve the network architecture to learn more effective features. The network model performance also degrades when the images are subjected to certain post-processing attacks, such as scaling, rotation, and JPEG compression. It requires us to perform data augmentation on the data during training to improve the robustness of the model and push the algorithm into practical applications. The current problem of insufficient tampering and low quality of tampering detection datasets seriously affects the development of deep learning-based tampering detection techniques. It is also very important to construct a dataset that meets the actual forensic requirement. Deep learning techniques continue to evolve, bringing many opportunities and challenges to passive image forensics. We should continuously update tampering detection techniques and use more effective network models and learning strategies to improve the accuracy and robustness of algorithms.

For active forensics, future research on robust image watermarking will use algorithms based on deep learning. In differentiable attacks, for noise enhancement and filtering attacks, choosing a more stable training framework and training methods is the primary method to effectively solve the current training instability, imperceptibility, and robustness tradeoff, such as using the diffusion model [154–156] and training codecs in divided stages. To enhance geometry attacks, designing the structure and transformation of restoring synchronization between watermark information and the decoder is an effective way to solve the problem of lack of synchronization between the decoder and watermark destroyed by geometry attacks. For example, an end-to-end deep learning model combined with the scale-invariant feature transform (SIFT) algorithm can effectively improve the robustness of rotation attacks. In the part of a non-differentiable attack, for a JPEG compression attack, the design of a more effective and realistic simulation of differentiable JPEG compression structure is the primary method to solve the problem that non-differentiable JPEG compression can not achieve model training and the poor effect of differentiable simulation JPEG training, for example, the differentiable analog JPEG combined with the attention mechanism to improve the simulation effect. In digital screen camera attacks, designing an effective analog distortion degradation structure is the primary method to solve the problem of poor robustness due to the difficult prediction of screen camera attack distortion. For example, multiple denoizing and de-denoizing processes in diffusion model [154–156]

are used to simulate the distortion of screen photography to improve the robustness of screen photography attacks. In the unknown attack, it is an effective method to improve the robustness of the unknown attack to design a watermarked image in the attack layer to generate a wider variety of attack adversarial samples after passing through the attack layer. For example, in model training, unsupervised sample types are first enriched, and then supervised watermark recovery accuracy is improved.

5.2. Conclusions

In this review, we synthesize existing research on deep learning-based image forensics from both passive forensics (i.e., tampering detection) and active forensics (i.e., digital watermarking), respectively.

For passive forensics, tampered areas of images are detected and located by analyzing the traces left by image tampering. In this survey, we analyze and review the state-of-the-art techniques for deep learning-based image copy-move, splicing, and generic forgery detection. First, we introduce a framework of image forgery detection based on deep learning, evaluation metrics, and commonly used datasets. According to the different types of tampering detection, image forgery detection methods are classified into three categories: image copy-move forgery detection, image splicing forgery detection, and image generic forgery detection. Then, the state-of-the-art algorithms are compared and analyzed in terms of four aspects: type of detection, backbone, robustness performance, and dataset. Finally, future research directions are analyzed in light of the problems of existing tamper detection algorithms.

For active forensics, we focus on the robustness of digital watermarking. In the beginning, we introduce a classical end-to-end watermarking model. According to whether the attack type is a differentiable attack or not, we subdivide it into five attacks: noise and filtering attacks, geometric attacks, JPEG attacks, screen-shooting attacks, and agnostic attacks. For noise and filtering attacks, most studies introduce an attack simulation layer in the codec to generate attack counterexamples to improve its robustness, but the joint codec training leads to the degradation of the watermarked image generated by the encoder. For geometric attacks, the construction of geometric invariant features from two perspectives, frequency domain coefficients and zero-watermark, ensure the spatial synchronization of the codec to a certain extent, but it does not combine with the robustness of the other attacks. The robustness of JPEG with low-quality factors still needs to be improved. Distortion simulation is the most commonly used method to enhance screen-shooting attacks in existing studies, but it suffers from the complex composition of actual distortion and low simulation accuracy. For agnostic attacks, reverse ASL effectively improves the accuracy of watermark recovery.

Recently, generative AI (for example, chatgpt [157] and DALL-E [158]) technologies are developing rapidly. When it comes to distinguishing between real photos and generated photos, watermarking technology can provide some assistance. However, with the continuous advancement of generative artificial intelligence technology, generated photos are becoming increasingly realistic, making it challenging to rely solely on traditional watermarking technology for accurate differentiation. Generated photos may inherit watermark information from the original photos, making them visually similar to real photos. Additionally, generative AI technology can also generate entirely new photos without any embedded watermarks. Therefore, relying solely on traditional watermarking technology may not be sufficient to differentiate between real photos and generated photos. It may be necessary to combine other image analysis and verification techniques, such as transmembrane state interaction and visual detection algorithms, to improve the accuracy and robustness of identification.

Author Contributions: C.S., L.C. and C.W. chose the topic and designed the structure of the paper. C.S. and C.W. sorted out and analyzed passive forensic techniques. L.C. and C.W. reviewed the robust image watermarking algorithms. C.S., L.C., X.Z. and Z.Q. classified the involved algorithms and analyzed the data. C.W., X.Z. and Z.Q. revised the manuscript. C.W. performed the project administration and supervision. C.W. and X.Z. is funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Shandong Provincial Natural Science Foundation (No. ZR2021MF060), in part by the Joint Fund of Shandong Provincial Natural Science Foundation (No. ZR2021LZH003), in part by the National Natural Science Foundation of China (No. 61702303), in part by the Scientific Research Project of Shandong University–Weihai Research Institute of Industrial Technology (No. 0006202210020011), in part by the Science and Technology Development Plan Project of Weihai Municipality (No. 2022DXGJ13), in part by the Shandong University Graduate Education Quality Curriculum Construction Project (No. 2022038), in part by the Education and Teaching Reform Research Project of Shandong University, Weihai (No. Y2023038), in part by the 17th Student Research Training Program (SRTP) at Shandong University, Weihai (Nos. A22293, A22299, A22086), and in part by the 18th Student Research Training Program (SRTP) at Shandong University, Weihai (Nos. A23246, A23248).

Data Availability Statement: Not applicable.

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

References

1. Dong, S.; Wang, P.; Abbas, K. A survey on deep learning and its applications. *Comput. Sci. Rev.* **2021**, *40*, 100379. [CrossRef]
2. Kaur, G.; Singh, N.; Kumar, M. Image forgery techniques: A review. *Artif. Intell. Rev.* **2023**, *56*, 1577–1625. [CrossRef]
3. Zhang, Z.; Wang, C.; Zhou, X. A survey on passive image copy-move forgery detection. *J. Inf. Process. Syst.* **2018**, *14*, 6–31. [CrossRef]
4. Zanardelli, M.; Guerrini, F.; Leonardi, R.; Adami, N. Image forgery detection: A survey of recent deep-learning approaches. *Multimed. Tools Appl.* **2023**, *82*, 17521–17566. [CrossRef]
5. Nabi, S.T.; Kumar, M.; Singh, P.; Aggarwal, N.; Kumar, K. A comprehensive survey of image and video forgery techniques: Variants, challenges, and future directions. *Multimed. Syst.* **2022**, *28*, 939–992. [CrossRef]
6. Gupta, S.; Mohan, N.; Kaushal, P. Passive image forensics using universal techniques: A review. *Artif. Intell. Rev.* **2022**, *55*, 1629–1679. [CrossRef]
7. Rakhmawati, L.; Wirawan, W.; Suwadi, S. A recent survey of self-embedding fragile watermarking scheme for image authentication with recovery capability. *EURASIP J. Image Video Process.* **2019**, *2019*, 61. [CrossRef]
8. Kumar, C.; Singh, A.K.; Kumar, P. A recent survey on image watermarking techniques and its application in e-governance. *Multimed. Tools Appl.* **2018**, *77*, 3597–3622. [CrossRef]
9. Menendez-Ortiz, A.; Feregrino-Uribe, C.; Hasimoto-Beltran, R.; Garcia-Hernandez, J.J. A survey on reversible watermarking for multimedia content: A robustness overview. *IEEE Access* **2019**, *7*, 132662–132681. [CrossRef]
10. Agarwal, N.; Singh, A.K.; Singh, P.K. Survey of robust and imperceptible watermarking. *Multimed. Tools Appl.* **2019**, *78*, 8603–8633. [CrossRef]
11. Amrit, P.; Singh, A.K. Survey on watermarking methods in the artificial intelligence domain and beyond. *Comput. Commun.* **2022**, *188*, 52–65. [CrossRef]
12. Wan, W.; Wang, J.; Zhang, Y.; Li, J.; Yu, H.; Sun, J. A comprehensive survey on robust image watermarking. *Neurocomputing* **2022**, *488*, 226–247. [CrossRef]
13. Evsutin, O.; Dzhanashia, K. Watermarking schemes for digital images: Robustness overview. *Signal Process. Image Commun.* **2022**, *100*, 116523. [CrossRef]
14. Mahmood, T.; Mehmood, Z.; Shah, M.; Saba, T. A robust technique for copy-move forgery detection and localization in digital images via stationary wavelet and discrete cosine transform. *J. Vis. Commun. Image Represent.* **2018**, *53*, 202–214. [CrossRef]
15. Jaiprakash, S.P.; Desai, M.B.; Prakash, C.S.; Mistry, V.H.; Radadiya, K.L. Low dimensional DCT and DWT feature based model for detection of image splicing and copy-move forgery. *Multimed. Tools Appl.* **2020**, *79*, 29977–30005. [CrossRef]
16. Wo, Y.; Yang, K.; Han, G.; Chen, H.; Wu, W. Copy-move forgery detection based on multi-radius PCET. *IET Image Process.* **2017**, *11*, 99–108. [CrossRef]
17. Park, J.Y.; Kang, T.A.; Moon, Y.H.; Eom, I.K. Copy-move forgery detection using scale invariant feature and reduced local binary pattern histogram. *Symmetry* **2020**, *12*, 492. [CrossRef]
18. Rani, A.; Jain, A.; Kumar, M. Identification of copy-move and splicing based forgeries using advanced SURF and revised template matching. *Multimed. Tools Appl.* **2021**, *80*, 23877–23898. [CrossRef]
19. Singh, G.; Singh, K. Digital image forensic approach based on the second-order statistical analysis of CFA artifacts. *Forens. Sci. Int. Digit. Investig.* **2020**, *32*, 200899. [CrossRef]

20. Zeng, H.; Peng, A.; Lin, X. Exposing image splicing with inconsistent sensor noise levels. *Multimed. Tools Appl.* **2020**, *79*, 26139–26154. [CrossRef]
21. Hsu, Y.F.; Chang, S.F. Detecting image splicing using geometry invariants and camera characteristics consistency. In Proceedings of the IEEE International Conference on Multimedia and Expo, Toronto, ON, Canada, 9–12 July 2006; IEEE: New York, NY, USA, 2006; pp. 549–552. [CrossRef]
22. Amerini, I.; Ballan, L.; Caldelli, R.; Del Bimbo, A.; Serra, G. A SIFT-based forensic method for copy-move attack detection and transformation recovery. *IEEE Trans. Inf. Forensic Secur.* **2011**, *6*, 1099–1110. [CrossRef]
23. Dong, J.; Wang, W.; Tan, T. Casia image tampering detection evaluation database. In Proceedings of the IEEE China Summit and International Conference on Signal and Information Processing, Beijing, China, 6–10 July 2013; IEEE: New York, NY, USA, 2013; pp. 422–426. [CrossRef]
24. De Carvalho, T.J.; Riess, C.; Angelopoulou, E.; Pedrini, H.; de Rezende Rocha, A. Exposing digital image forgeries by illumination color classification. *IEEE Trans. Inf. Forensic Secur.* **2013**, *8*, 1182–1194. [CrossRef]
25. Tralic, D.; Zupancic, I.; Grgic, S.; Grgic, M. CoMoFoD–New database for copy-move forgery detection. In Proceedings of the International Symposium Electronics in Marine, Zadar, Croatia, 13–15 September 2013; IEEE: New York, NY, USA, 2013; pp. 49–54. Available online: http://www.vcl.fer.hr/comofod (accessed on 10 July 2023).
26. Wen, B.; Zhu, Y.; Subramanian, R.; Ng, T.T.; Shen, X.; Winkler, S. COVERAGE–A novel database for copy-move forgery detection. In Proceedings of the IEEE International Conference on Image Processing, Phoenix, AZ, USA, 25–28 September 2016; IEEE: New York, NY, USA, 2016; pp. 161–165. [CrossRef]
27. Korus, P. Digital image integrity–A survey of protection and verification techniques. *Digit. Signal Process.* **2017**, *71*, 1–26. [CrossRef]
28. Wu, Y.; Abd-Almageed, W.; Natarajan, P. Busternet: Detecting copy-move image forgery with source/target localization. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 168–184. [CrossRef]
29. Guan, H.; Kozak, M.; Robertson, E.; Lee, Y.; Yates, A.N.; Delgado, A.; Zhou, D.; Kheyrkhah, T.; Smith, J.; Fiscus, J. MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation. In Proceedings of the IEEE Winter Applications of Computer Vision Workshops, Waikoloa, HI, USA, 7–11 January 2019; IEEE: New York, NY, USA, 2019; pp. 63–72. [CrossRef]
30. Mahfoudi, G.; Tajini, B.; Retraint, F.; Morain-Nicolier, F.; Dugelay, J.L.; Marc, P. DEFACTO: Image and face manipulation dataset. In Proceedings of the 27th European Signal Processing Conference, A Coruna, Spain, 2–6 September 2019; IEEE: New York, NY, USA, 2019; pp. 1–5. [CrossRef]
31. Novozamsky, A.; Mahdian, B.; Saic, S. 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, Snowmass, CO, USA, 1–5 March 2020; IEEE: New York, NY, USA, 2020; pp. 71–80. [CrossRef]
32. Van Schyndel, R.G.; Tirkel, A.Z.; Osborne, C.F. A digital watermark. In Proceedings of the 1st International Conference on Image Processing, Austin, TX, USA, 13–16 November 1994; IEEE: New York, NY, USA, 1994; Volume 2, pp. 86–90. [CrossRef]
33. Dumitrescu, S.; Wu, X.; Wang, Z. Detection of LSB steganography via sample pair analysis. *IEEE Trans. Signal Process.* **2003**, *51*, 1995–2007. [CrossRef]
34. Guo, H.; Georganas, N.D. Digital image watermarking for joint ownership verification without a trusted dealer. In Proceedings of the International Conference on Multimedia and Expo, Baltimore, MD, USA, 6–9 July 2003; IEEE: New York, NY, USA, 2003; Volume 2, pp. 497–500. [CrossRef]
35. Parah, S.A.; Sheikh, J.A.; Loan, N.A.; Bhat, G.M. Robust and blind watermarking technique in DCT domain using inter-block coefficient differencing. *Digit. Signal Process.* **2016**, *53*, 11–24. [CrossRef]
36. Etemad, S.; Amirmazlaghani, M. A new multiplicative watermark detector in the contourlet domain using t location-scale distribution. *Pattern Recognit.* **2018**, *77*, 99–112. [CrossRef]
37. Etemad, E.; Samavi, S.; Reza Soroushmehr, S.; Karimi, N.; Etemad, M.; Shirani, S.; Najarian, K. Robust image watermarking scheme using bit-plane of Hadamard coefficients. *Multimed. Tools Appl.* **2018**, *77*, 2033–2055. [CrossRef]
38. Rao, Y.; Ni, J. A deep learning approach to detection of splicing and copy-move forgeries in images. In Proceedings of the IEEE International Workshop on Information Forensics and Security, Abu Dhabi, United Arab Emirates, 4–7 December 2016; IEEE: New York, NY, USA, 2016; pp. 1–6. [CrossRef]
39. Kumar, S.; Gupta, S.K. A robust copy move forgery classification using end to end convolution neural network. In Proceedings of the 8th International Conference on Reliability, Infocom Technologies and Optimization, Noida, India, 4–5 June 2020; IEEE: New York, NY, USA, 2020; pp. 253–258. [CrossRef]
40. Li, Q.; Wang, C.; Zhou, X.; Qin, Z. Image copy-move forgery detection and localization based on super-BPD segmentation and DCNN. *Sci Rep.* **2022**, *12*, 14987. [CrossRef]
41. Wan, J.; Liu, Y.; Wei, D.; Bai, X.; Xu, Y. Super-BPD: Super boundary-to-pixel direction for fast image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; IEEE: New York, NY, USA, 2020; pp. 9250–9259. [CrossRef]
42. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef] [PubMed]
43. Liu, Y.; Xia, C.; Zhu, X.; Xu, S. Two-stage copy-move forgery detection with self deep matching and proposal superglue. *IEEE Trans. Image Process.* **2021**, *31*, 541–555. [CrossRef]

44. Zhong, J.L.; Pun, C.M. An end-to-end dense-inceptionnet for image copy-move forgery detection. *IEEE Trans. Inf. Forensic Secur.* **2019**, *15*, 2134–2146. [CrossRef]
45. Kafali, E.; Vretos, N.; Semertzidis, T.; Daras, P. RobusterNet: Improving copy-move forgery detection with Volterra-based convolutions. In Proceedings of the 25th International Conference on Pattern Recognition, Milan, Italy, 10–15 January 2021; IEEE: New York, NY, USA, 2021; pp. 1160–1165. [CrossRef]
46. Nazir, T.; Nawaz, M.; Masood, M.; Javed, A. Copy move forgery detection and segmentation using improved mask region-based convolution network (RCNN). *Appl. Soft. Comput.* **2022**, *131*, 109778. [CrossRef]
47. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; IEEE: New York, NY, USA, 2017; pp. 2980–2988. [CrossRef]
48. Zhong, J.L.; Yang, J.X.; Gan, Y.F.; Huang, L.; Zeng, H. Coarse-to-fine spatial-channel-boundary attention network for image copy-move forgery detection. *Soft Comput.* **2022**, *26*, 11461–11478. [CrossRef]
49. Chen, B.; Tan, W.; Coatrieux, G.; Zheng, Y.; Shi, Y.Q. A serial image copy-move forgery localization scheme with source/target distinguishment. *IEEE Trans. Multimed.* **2020**, *23*, 3506–3517. [CrossRef]
50. Aria, M.; Hashemzadeh, M.; Farajzadeh, N. QDL-CMFD: A quality-independent and deep learning-based copy-move image forgery detection method. *Neurocomputing* **2022**, *511*, 213–236. [CrossRef]
51. Barni, M.; Phan, Q.T.; Tondi, B. Copy move source-target disambiguation through multi-branch CNNs. *IEEE Trans. Inf. Forensic Secur.* **2020**, *16*, 1825–1840. [CrossRef]
52. Niyishaka, P.; Bhagvati, C. Image splicing detection technique based on illumination-reflectance model and LBP. *Multimed. Tools Appl.* **2021**, *80*, 2161–2175. [CrossRef]
53. Shen, X.; Shi, Z.; Chen, H. Splicing image forgery detection using textural features based on the grey level co-occurrence matrices. *IET Image Process.* **2017**, *11*, 44–53. [CrossRef]
54. Sharma, S.; Ghanekar, U. Spliced image classification and tampered region localization using local directional pattern. *Int. J. Image, Graph. Signal Process.* **2019**, *11*, 35–42. [CrossRef]
55. Wei, Y.; Wang, Z.; Xiao, B.; Liu, X.; Yan, Z.; Ma, J. Controlling neural learning network with multiple scales for image splicing forgery detection. *ACM Trans. Multimed. Comput. Commun. Appl.* **2020**, *16*, 1–22. [CrossRef]
56. Zeng, P.; Tong, L.; Liang, Y.; Zhou, N.; Wu, J. Multitask image splicing tampering detection based on attention mechanism. *Mathematics* **2022**, *10*, 3852. [CrossRef]
57. Zhang, Y.; Zhu, G.; Wu, L.; Kwong, S.; Zhang, H.; Zhou, Y. Multi-task SE-network for image splicing localization. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 4828–4840. [CrossRef]
58. Chen, B.; Qi, X.; Zhou, Y.; Yang, G.; Zheng, Y.; Xiao, B. Image splicing localization using residual image and residual-based fully convolutional network. *J. Vis. Commun. Image Represent.* **2020**, *73*, 102967. [CrossRef]
59. Zhuang, P.; Li, H.; Tan, S.; Li, B.; Huang, J. Image tampering localization using a dense fully convolutional network. *IEEE Trans. Inf. Forensic Secur.* **2021**, *16*, 2986–2999. [CrossRef]
60. Liu, Q.; Li, H.; Liu, Z. Image forgery localization based on fully convolutional network with noise feature. *Multimed. Tools Appl.* **2022**, *81*, 17919–17935. [CrossRef]
61. Ren, R.; Niu, S.; Jin, J.; Zhang, J.; Ren, H.; Zhao, X. Multi-scale attention context-aware network for detection and localization of image splicing. *Appl. Intell.* **2023**, *53*, 18219–18238. [CrossRef]
62. Sun, Y.; Ni, R.; Zhao, Y. ET: Edge-enhanced transformer for image splicing detection. *IEEE Signal Process. Lett.* **2022**, *29*, 1232–1236. [CrossRef]
63. Zhang, Z.; Qian, Y.; Zhao, Y.; Zhu, L.; Wang, J. Noise and edge based dual branch image manipulation detection. *arXiv* **2022**, arXiv:2207.00724. [CrossRef]
64. Dong, C.; Chen, X.; Hu, R.; Cao, J.; Li, X. MVSS-Net: Multi-view multi-scale supervised networks for image manipulation detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 3539–3553. [CrossRef]
65. Chen, J.; Liao, X.; Wang, W.; Qian, Z.; Qin, Z.; Wang, Y. SNIS: A signal noise separation-based network for post-processed image forgery detection. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 935–951. [CrossRef]
66. Lin, X.; Wang, S.; Deng, J.; Fu, Y.; Bai, X.; Chen, X.; Qu, X.; Tang, W. Image manipulation detection by multiple tampering traces and edge artifact enhancement. *Pattern Recognit.* **2023**, *133*, 109026. [CrossRef]
67. Wang, J.; Wu, Z.; Chen, J.; Han, X.; Shrivastava, A.; Lim, S.N.; Jiang, Y.G. Objectformer for image manipulation detection and localization. In Proceedings of the Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; IEEE: New York, NY, USA, 2022; pp. 2354–2363. [CrossRef]
68. Liu, X.; Liu, Y.; Chen, J.; Liu, X. PSCC-Net: Progressive spatio-channel correlation network for image manipulation detection and localization. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 7505–7517. [CrossRef]
69. Shi, Z.; Chang, C.; Chen, H.; Du, X.; Zhang, H. PR-Net: Progressively-refined neural network for image manipulation localization. *Int. J. Intell. Syst.* **2022**, *37*, 3166–3188. [CrossRef]
70. Gao, Z.; Sun, C.; Cheng, Z.; Guan, W.; Liu, A.; Wang, M. TBNet: A two-stream boundary-aware network for generic image manipulation localization. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 7541–7556. [CrossRef]
71. Ganapathi, I.I.; Javed, S.; Ali, S.S.; Mahmood, A.; Vu, N.S.; Werghi, N. Learning to localize image forgery using end-to-end attention network. *Neurocomputing* **2022**, *512*, 25–39. [CrossRef]

72. Xu, D.; Shen, X.; Lyu, Y.; Du, X.; Feng, F. MC-Net: Learning mutually-complementary features for image manipulation localization. *Int. J. Intell. Syst.* **2022**, *37*, 3072–3089. [CrossRef]
73. Rao, Y.; Ni, J.; Xie, H. Multi-semantic CRF-based attention model for image forgery detection and localization. *Signal Process.* **2021**, *183*, 108051. [CrossRef]
74. Li, S.; Xu, S.; Ma, W.; Zong, Q. Image manipulation localization using attentional cross-domain CNN features. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–15. [CrossRef]
75. Yin, Q.; Wang, J.; Lu, W.; Luo, X. Contrastive learning based multi-task network for image manipulation detection. *Signal Process.* **2022**, *201*, 108709. [CrossRef]
76. Zhuo, L.; Tan, S.; Li, B.; Huang, J. Self-adversarial training incorporating forgery attention for image forgery localization. *IEEE Trans. Inf. Forensic Secur.* **2022**, *17*, 819–834. [CrossRef]
77. Ren, R.; Niu, S.; Ren, H.; Zhang, S.; Han, T.; Tong, X. ESRNet: Efficient search and recognition network for image manipulation detection. *ACM Trans. Multimed. Comput. Commun. Appl.* **2022**, *18*, 1–23. [CrossRef]
78. Silva, E.; Carvalho, T.; Ferreira, A.; Rocha, A. Going deeper into copy-move forgery detection: Exploring image telltales via multi-scale analysis and voting processes. *J. Vis. Commun. Image Represent.* **2015**, *29*, 16–32. [CrossRef]
79. Cozzolino, D.; Poggi, G.; Verdoliva, L. Copy-move forgery detection based on patchmatch. In Proceedings of the IEEE International Conference on Image Processing, Paris, France, 27–30 October 2014; IEEE: New York, NY, USA, 2014; pp. 5312–5316. [CrossRef]
80. Ng, T.T.; Chang, S.F.; Sun, Q. *A Data Set of Authentic and Spliced Image Blocks*; Columbia University, ADVENT Technical Report; Columbia University: New York, NY, USA, 2004; Volume 4, pp. 1–9.
81. Shi, Z.; Shen, X.; Chen, H.; Lyu, Y. Global semantic consistency network for image manipulation detection. *IEEE Signal Process. Lett.* **2020**, *27*, 1755–1759. [CrossRef]
82. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. Acm* **2020**, *63*, 139–144. [CrossRef]
83. Kang, X.; Huang, J.; Zeng, W. Improving robustness of quantization-based image watermarking via adaptive receiver. *IEEE Trans. Multimed.* **2008**, *10*, 953–959. [CrossRef]
84. Goléa, N.E.H.; Seghir, R.; Benzid, R. A bind RGB color image watermarking based on singular value decomposition. In Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications, Hammamet, Tunisia, 16–19 May 2010; IEEE: New York, NY, USA, 2010; pp. 1–5. [CrossRef]
85. Wen, B.; Aydore, S. Romark: A robust watermarking system using adversarial training. *arXiv* **2019**, arXiv:1910.01221. [CrossRef]
86. Fan, Y.; Li, J.; Bhatti, U.A.; Shao, C.; Gong, C.; Cheng, J.; Chen, Y. A multi-watermarking algorithm for medical images using Inception V3 and DCT. *CMC-Comput. Mat. Contin.* **2023**, *74*, 1279–1302. [CrossRef]
87. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; IEEE: New York, NY, USA, 2016; pp. 2818–2826. [CrossRef]
88. Hao, K.; Feng, G.; Zhang, X. Robust image watermarking based on generative adversarial network. *China Commun.* **2020**, *17*, 131–140. [CrossRef]
89. Zhang, B.; Wu, Y.; Chen, B. Embedding guided end-to-end framework for robust image watermarking. *Secur. Commun. Netw.* **2022**, *2022*, 1–11. [CrossRef]
90. Li, J.; Li, Y.; Li, J.; Zhang, Q.; Yang, G.; Chen, S.; Wang, C.; Li, J. Single exposure optical image watermarking using a cGAN network. *IEEE Photonics J.* **2021**, *13*, 6900111. [CrossRef]
91. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784. [CrossRef]
92. Yu, C. Attention based data hiding with generative adversarial networks. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; AAAI: Washington, DC, USA, 2020; Volume 34, pp. 1120–1128. [CrossRef]
93. Mun, S.M.; Nam, S.H.; Jang, H.U.; Kim, D.; Lee, H.K. A robust blind watermarking using convolutional neural network. *arXiv* **2017**, arXiv:1704.03248. [CrossRef]
94. Kang, X.; Chen, Y.; Zhao, F.; Lin, G. Multi-dimensional particle swarm optimization for robust blind image watermarking using intertwining logistic map and hybrid domain. *Soft Comput.* **2020**, *24*, 10561–10584. [CrossRef]
95. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; IEEE: New York, NY, USA, 1995; Volume 4, pp. 1942–1948. [CrossRef]
96. Rai, M.; Goyal, S. A hybrid digital image watermarking technique based on fuzzy-BPNN and shark smell optimization. *Multimed. Tools Appl.* **2022**, *81*, 39471–39489. [CrossRef]
97. Liu, C.; Zhong, D.; Shao, H. Data protection in palmprint recognition via dynamic random invisible watermark embedding. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 6927–6940. [CrossRef]
98. Zhao, Y.; Wang, C.; Zhou, X.; Qin, Z. DARI-Mark: Deep learning and attention network for robust image watermarking. *Mathematics* **2023**, *11*, 209. [CrossRef]
99. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 740–755. [CrossRef]

100. Latecki, L. Shape Data for the MPEG-7 Core Experiment Ce-Shape-1. 2002, 7. Available online: https://www.researchgate.net/figure/The-MPEG-7-Core-Experiment-CE-Shape-1-dataset_fig4_245251999 (accessed on 10 July 2023).
101. Weber, A.G. The USC-SIPI Image Database: Version 5. 2006. Available online: http://sipi.usc.edu/database/ (accessed on 10 July 2023)
102. Bas, P.; Filler, T.; Pevnỳ, T. Break our steganographic system: The ins and outs of organizing BOSS. In Proceedings of the 13th International Conference on Information Hiding, Prague, Czech Republic, 18–20 May 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 59–70. [CrossRef]
103. Krizhevsky, A.; Nair, V.; Hinton, G. Cifar-10 and Cifar-100 Datasets. 2009. Available online: https://www.cs.toronto.edu/~kriz/cifar.htm (accessed on 10 July 2023)
104. Sahu, A.K. A logistic map based blind and fragile watermarking for tamper detection and localization in images. *J. Ambient Intell. Humaniz. Comput.* **2022**, *13*, 3869–3881. [CrossRef]
105. Sahu, A.K.; Umachandran, K.; Biradar, V.D.; Comfort, O.; Sri Vigna Hema, V.; Odimegwu, F.; Saifullah, M. A study on content tampering in multimedia watermarking. *SN Comput. Sci.* **2023**, *4*, 222. [CrossRef]
106. Sahu, A.K.; Sahu, M.; Patro, P.; Sahu, G.; Nayak, S.R. Dual image-based reversible fragile watermarking scheme for tamper detection and localization. *Pattern Anal. Appl.* **2023**, *26*, 571–590. [CrossRef]
107. Hsu, C.S.; Tu, S.F. Enhancing the robustness of image watermarking against cropping attacks with dual watermarks. *Multimed. Tools Appl.* **2020**, *79*, 11297–11323. [CrossRef]
108. Ahmadi, M.; Norouzi, A.; Karimi, N.; Samavi, S.; Emami, A. ReDMark: Framework for residual diffusion watermarking based on deep networks. *Expert Syst. Appl.* **2020**, *146*, 113157. [CrossRef]
109. Mei, Y.; Wu, G.; Yu, X.; Liu, B. A robust blind watermarking scheme based on attention mechanism and neural joint source-channel coding. In Proceedings of the IEEE 24th International Workshop on Multimedia Signal Processing, Shanghai, China, 26–28 September 2022; IEEE: New York, NY, USA, 2022; pp. 1–6. [CrossRef]
110. Han, B.; Du, J.; Jia, Y.; Zhu, H. Zero-watermarking algorithm for medical image based on VGG19 deep convolution neural network. *J. Healthc. Eng.* **2021**, *2021*, 5551520. [CrossRef]
111. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]
112. Liu, G.; Xiang, R.; Liu, J.; Pan, R.; Zhang, Z. An invisible and robust watermarking scheme using convolutional neural networks. *Expert Syst. Appl.* **2022**, *210*, 118529. [CrossRef]
113. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; IEEE: New York, NY, USA, 2016; pp. 2414–2423. [CrossRef]
114. Sun, L.; Xu, J.; Zhang, X.; Dong, W.; Tian, Y. A novel generalized Arnold transform-based zero-watermarking scheme. *Appl. Math. Inf. Sci.* **2015**, *4*, 2023–2035.
115. Gong, C.; Liu, J.; Gong, M.; Li, J.; Bhatti, U.A.; Ma, J. Robust medical zero-watermarking algorithm based on residual-DenseNet. *IET Biom.* **2022**, *11*, 547–556. [CrossRef]
116. Hu, R.; Xiang, S. Cover-lossless robust image watermarking against geometric deformations. *IEEE Trans. Image Process.* **2021**, *30*, 318–331. [CrossRef]
117. Mellimi, S.; Rajput, V.; Ansari, I.A.; Ahn, C.W. A fast and efficient image watermarking scheme based on deep neural network. *Pattern Recognit. Lett.* **2021**, *151*, 222–228. [CrossRef]
118. Fan, B.; Li, Z.; Gao, J. DwiMark: A multiscale robust deep watermarking framework for diffusion-weighted imaging images. *Multimed. Syst.* **2022**, *28*, 295–310. [CrossRef]
119. Everingham, M.; Eslami, S.A.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes challenge: A retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [CrossRef]
120. Ma, K.; Duanmu, Z.; Wu, Q.; Wang, Z.; Yong, H.; Li, H.; Zhang, L. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Trans. Image Process.* **2017**, *26*, 1004–1016. [CrossRef] [PubMed]
121. Zhu, J.; Kaplan, R.; Johnson, J.; Fei-Fei, L. Hidden: Hiding data with deep networks. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; Springer: Cham, Switzerland, 2018; pp. 682–697. [CrossRef]
122. Liu, Y.; Guo, M.; Zhang, J.; Zhu, Y.; Xie, X. A novel two-stage separable deep learning framework for practical blind watermarking. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; .ACM: New York, NY, USA, 2019; pp. 1509–1517. [CrossRef]
123. Chen, B.; Wu, Y.; Coatrieux, G.; Chen, X.; Zheng, Y. JSNet: A simulation network of JPEG lossy compression and restoration for robust image watermarking against JPEG attack. *Comput. Vis. Image Underst.* **2020**, *197*, 103015. [CrossRef]
124. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 15–20 June 2009; IEEE: New York, NY, USA, 2009; pp. 248–255. [CrossRef]
125. Jia, Z.; Fang, H.; Zhang, W. MBRS: Enhancing robustness of DNN-based watermarking by mini-batch of real and simulated JPEG compression. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event, 20–24 October 2021; ACM: New York, NY, USA, 2021; pp. 41–49. [CrossRef]

126. Zhang, C.; Karjauv, A.; Benz, P.; Kweon, I.S. Towards robust data hiding against (JPEG) compression: A pseudo-differentiable deep learning approach. *arXiv* **2020**, arXiv:2101.00973. [CrossRef]
127. Ma, R.; Guo, M.; Hou, Y.; Yang, F.; Li, Y.; Jia, H.; Xie, X. Towards blind watermarking: Combining invertible and non-invertible mechanisms. In Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal, 10–14 October 2022; ACM: New York, NY, USA, 2022; pp. 1532–1542. [CrossRef]
128. Tsai, P.H.; Chuang, Y.Y. Target-driven moire pattern synthesis by phase modulation. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; IEEE: New York, NY, USA, 2013; pp. 1912–1919. [CrossRef]
129. Zhang, L.; Li, W.; Ye, H. A blind watermarking system based on deep learning model. In Proceedings of the IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications, Shenyang, China, 20–22 October 2021; IEEE: New York, NY, USA, 2021; pp. 1208–1213. [CrossRef]
130. Zhong, X.; Huang, P.C.; Mastorakis, S.; Shih, F.Y. An automated and robust image watermarking scheme based on deep neural networks. *IEEE Trans. Multimed.* **2020**, *23*, 1951–1961. [CrossRef]
131. Wang, R.; Lin, C.; Zhao, Q.; Zhu, F. Watermark faker: Towards forgery of digital image watermarking. In Proceedings of the IEEE International Conference on Multimedia and Expo, Shenyang, China, 5–9 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–6. [CrossRef]
132. Fang, H.; Chen, D.; Huang, Q.; Zhang, J.; Ma, Z.; Zhang, W.; Yu, N. Deep template-based watermarking. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 1436–1451. [CrossRef]
133. Fang, H.; Zhang, W.; Ma, Z.; Zhou, H.; Sun, S.; Cui, H.; Yu, N. A camera shooting resilient watermarking scheme for underpainting documents. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 4075–4089. [CrossRef]
134. Fang, H.; Chen, D.; Wang, F.; Ma, Z.; Liu, H.; Zhou, W.; Zhang, W.; Yu, N. Tera: Screen-to-camera image code with transparency, efficiency, robustness and adaptability. *IEEE Trans. Multimed.* **2021**, *24*, 955–967. [CrossRef]
135. Bose, R.C.; Ray-Chaudhuri, D.K. On a class of error correcting binary group codes. *Inf. Control* **1960**, *3*, 68–79. [CrossRef]
136. Pramila, A.; Keskinarkaus, A.; Seppänen, T. Toward an interactive poster using digital watermarking and a mobile phone camera. *Signal Image Video Process.* **2012**, *6*, 211–222. [CrossRef]
137. Gugelmann, D.; Sommer, D.; Lenders, V.; Happe, M.; Vanbever, L. Screen watermarking for data theft investigation and attribution. In Proceedings of the 10th International Conference on Cyber Conflict, Tallinn, Estonia, 29 May–1 June 2018; IEEE: New York, NY, USA, 2018; pp. 391–408. [CrossRef]
138. Dong, L.; Chen, J.; Peng, C.; Li, Y.; Sun, W. Watermark-preserving keypoint enhancement for screen-shooting resilient watermarking. In Proceedings of the IEEE International Conference on Multimedia and Expo, Taipei, Taiwan, 18–22 July 2022; IEEE: New York, NY, USA, 2022; pp. 1–6. [CrossRef]
139. Bai, R.; Li, L.; Zhang, S.; Lu, J.; Chang, C.C. SSDeN: Framework for screen-shooting resilient watermarking via deep networks in the frequency domain. *Appl. Sci.* **2022**, *12*, 9780. [CrossRef]
140. Lu, J.; Ni, J.; Su, W.; Xie, H. Wavelet-based CNN for robust and high-capacity image watermarking. In Proceedings of the IEEE International Conference on Multimedia and Expo, Taipei, Taiwan, 18–22 July 2022; IEEE: New York, NY, USA, 2022; pp. 1–6. [CrossRef]
141. Boujerfaoui, S.; Douzi, H.; Harba, R.; Gourrame, K. Robust Fourier watermarking for print-cam process using convolutional neural networks. In Proceedings of the 7th International Conference on Signal and Image Processing, Suzhou, China, 20–22 July 2022; IEEE: New York, NY, USA, 2022; pp. 347–351. [CrossRef]
142. Jia, J.; Gao, Z.; Chen, K.; Hu, M.; Min, X.; Zhai, G.; Yang, X. RIHOOP: Robust invisible hyperlinks in offline and online photographs. *IEEE T. Cybern.* **2022**, *52*, 7094–7106. [CrossRef]
143. Fang, H.; Jia, Z.; Ma, Z.; Chang, E.C.; Zhang, W. PIMoG: An effective screen-shooting noise-layer simulation for deep-learning-based watermarking network. In Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal, 10–14 October 2022; ACM: New York, NY, USA, 2022; pp. 2267–2275. [CrossRef]
144. Yoo, I.; Chang, H.; Luo, X.; Stava, O.; Liu, C.; Milanfar, P.; Yang, F. Deep 3D-to-2D watermarking: Embedding messages in 3D meshes and extracting them from 2D renderings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; IEEE: New York, NY, USA, 2022; pp. 10021–10030. [CrossRef]
145. Tancik, M.; Mildenhall, B.; Ng, R. Stegastamp: Invisible hyperlinks in physical photographs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; IEEE: New York, NY, USA, 2020; pp. 2114–2123. [CrossRef]
146. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th International Coference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015, pp. 234–241. [CrossRef]
147. Huiskes, M.J.; Lew, M.S. The mir flickr retrieval evaluation. In Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval, Vancouver, BC, Canada, 30–31 October 2008; ACM: New York, NY, USA, 2008; pp. 39–43. [CrossRef]
148. Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; Xiao, J. 3D shapenets: A deep representation for volumetric shapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; IEEE: New York, NY, USA, 2015; pp. 1912–1920. [CrossRef]

149. Chen, Y.P.; Fan, T.Y.; Chao, H.C. Wmnet: A lossless watermarking technique using deep learning for medical image authentication. *Electronics* **2021**, *10*, 932. [CrossRef]
150. Xu, D.; Zhu, C.; Ren, N. A zero-watermark algorithm for copyright protection of remote sensing image based on blockchain. In Proceedings of the International Conference on Blockchain Technology and Information Security, Huaihua City, China, 15–17 July 2022; IEEE: New York, NY, USA, 2022; pp. 111–116. [CrossRef]
151. Luo, X.; Zhan, R.; Chang, H.; Yang, F.; Milanfar, P. Distortion agnostic deep watermarking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; IEEE: New York, NY, USA, 2020; pp. 13545–13554. [CrossRef]
152. Zhang, C.; Karjauv, A.; Benz, P.; Kweon, I.S. Towards robust deep hiding under non-differentiable distortions for practical blind watermarking. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event, 20–24 October 2021; ACM: New York, NY, USA, 2021; pp. 5158–5166. [CrossRef]
153. Zheng, X.; Dong, Q.; Fu, A. WMDefense: Using watermark to defense byzantine attacks in federated learning. In Proceedings of the IEEE Conference on Computer Communications Workshops, New York, NY, USA, 2–5 May 2022; IEEE: New York, NY, USA, 2022; pp. 1–6. [CrossRef]
154. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851. [CrossRef]
155. Song, J.; Meng, C.; Ermon, S. Denoising diffusion implicit models. *arXiv* **2020**, arXiv:2010.02502. [CrossRef]
156. Dhariwal, P.; Nichol, A. Diffusion models beat GANs on image synthesis. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 8780–8794. [CrossRef]
157. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901. [CrossRef]
158. Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; Chen, M. Hierarchical text-conditional image generation with clip latents. *arXiv* **2022**, arXiv:2204.06125. [CrossRef]