Circuits, Systems, and Signal Processing (2024) 43:4339–4368
https://doi.org/10.1007/s00034-024-02651-z

Deep Learning-Based Watermarking Techniques Challenges: A Review of Current and Future Trends

Saoussen Ben Jabra[1] · **Mohamed Ben Farah**[2]

Received: 16 February 2023 / Revised: 23 February 2024 / Accepted: 26 February 2024 /
Published online: 2 April 2024

Abstract

The digital revolution places great emphasis on digital media watermarking due to the increased vulnerability of multimedia content to unauthorized alterations. Recently, in the digital boom in the technology of hiding data, research has been tending to perform watermarking with numerous architectures of deep learning, which has explored a variety of problems since its inception. Several watermarking approaches based on deep learning have been proposed, and they have proven their efficiency compared to traditional methods. This paper summarizes recent developments in conventional and deep learning image and video watermarking techniques. It shows that although there are many conventional techniques focused on video watermarking, there are yet to be any deep learning models focusing on this area; however, for image watermarking, different deep learning-based techniques where efficiency in invisibility and robustness depends on the used network architecture are observed. This study has been concluded by discussing possible research directions in deep learning-based video watermarking.

Keywords Watermarking · Deep learning · Neural network · Robustness · Embedding · Extraction

1 Introduction

The rapid and extensive progress of internet and networking technologies has simplified the replication, alteration, reproduction, and distribution of multimedia content through physical transmission media. This occurs during communication, information

✉ Mohamed Ben Farah
mohamed.benfarah@bcu.ac.uk

Saoussen Ben Jabra
saoussen.bj@gmail.com

[1] LimTic Lab, National Engineering School of Sousse, University of Sousse, Sousse, Tunisia

[2] Birmingham City University, Birmingham B4 7XG, UK

Birkhäuser

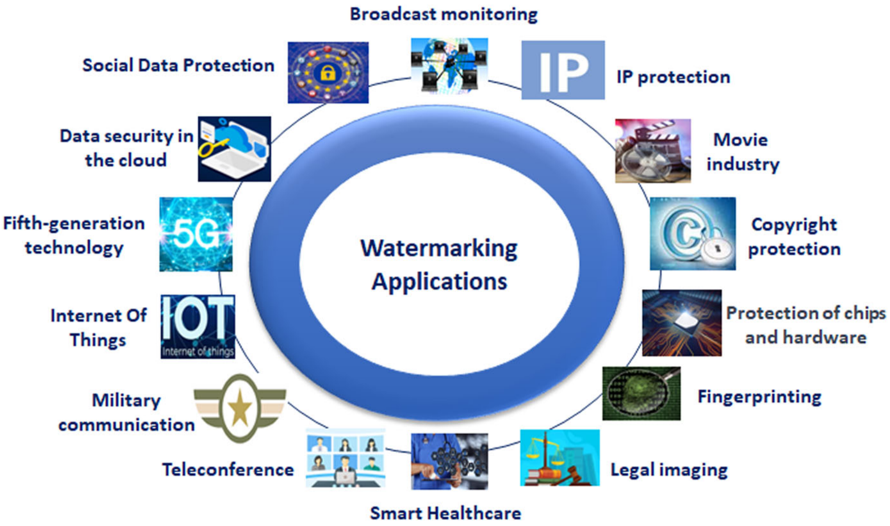

Fig. 1 watermarking applications

processing, and data storage, all at a low cost and without compromising the quality of the content. Therefore, securing data and maintaining digital information from upcoming hackers threats is primordial. Different data hiding techniques were proposed to resolve this problem, such as cryptography, steganography, and digital watermarking. This last one consists in embedding the signature into the original content and then trying to detect it after different manipulations are applied to the marked content. Watermarking is used for several applications, such as content protection, copyright management, content authentication, and tamper detection. Figure 1 illustrates several widely recognized applications of watermarking.

In the last two decades, many traditional watermarking approaches have been proposed to secure different types of media, such as image [84], 2D video [88], 3D models [20], and audio [36]. These traditional approaches are based on embedding signatures into feature regions using spatial or frequency domains, and they approve efficiency in invisibility and robustness against attacks. Recently, watermarking has seen a lot of scientific interest-based artificial intelligence. Deep learning is nowadays the most powerful, time, and cost-efficient machine learning approach. Deep learning has significantly improved in numerous applied research areas such as computer vision, medicine, natural language processing, object detection, face recognition, handwriting recognition, and speech recognition. It is one of the fastest-developing methods with a significant breakthrough performance. Consequently, the high performance of the deep learning models is considered efficient in protecting the intellectual property of any digital multimedia content. Since 2017, several deep learning-based watermarking techniques have been proposed to embed signatures into media content. However, most of these works focused on image content where techniques are generally classified based on their network architecture [17]. Although many traditional watermarking algorithms are proposed for video, 3D models, and audio, deep learning models are yet to focus on these areas. Indeed, to our knowledge, there is only one proposed paper for 3D models [93], two papers for video and no paper exploring deep learning models for audio watermarking.

As deep learning-based watermarking is a relatively recent area of research, current surveys concentrate on traditional algorithms. Since 2020, a few survey papers have been proposed concerning deep learning image watermarking, but no survey paper has focused on deep learning-based video watermarking techniques. Byrnes et al. [17] proposed a comprehensive survey regarding deep data hiding models unifying digital watermarking and steganography. Zhang et al. [95] also presents a brief survey on deep learning-based data hiding, steganography, and watermarking for images. Li et al. [51] provides an overview of watermarking of deep learning models, and [29] gives a brief survey of image watermarking based on deep neural network architecture.

As deep learning-based watermarking continues to expand, and several works were recently proposed for video, it is important to summarize and compare the current methods proposed for image and video. This survey aims to briefly classify traditional watermarking techniques for image and video and discuss existing deep learning models for image and video watermarking. We present future directions in deep learning-based video watermarking areas that research may take. The key contributions of the survey are as follows:

- This survey briefly classifies and compares the existing traditional watermarking techniques proposed for image and video.
- We provide a classification of deep learning-based watermarking techniques based on the network architecture and the embedding domain.
- We discuss and compare the most popular deep learning-based image and video watermarking techniques to give researchers a clear understanding of the practical challenges of deep learning-based watermarking.
- We also present some future directions for deep learning-based video watermarking.

The rest of the paper is organized as follows. In Sect. 2, we introduce a review of the traditional image and video watermarking techniques by classifying them based on the embedding target and the used domain. Section 3 presents a comparison of image watermarking techniques utilizing deep learning methods, focusing on their network architecture. The next section details the small number of existing deep learning-based video watermarking techniques and shows their advantages. Sect. 5 discusses video deep learning-based watermarking challenges and gives some suggestions to researchers in this domain. Eventually, in Sect. 7, we draw some conclusions and highlight some directions for future works.

2 Traditional Image and Video Watermarking Techniques Review

Watermarking is a branch of data hiding technology that hides information in digital content to be transmitted securely in the network. Information-hiding technology mainly includes steganography, covert communication, and watermarking. This technique protects digital content against several security problems, such as illegal data

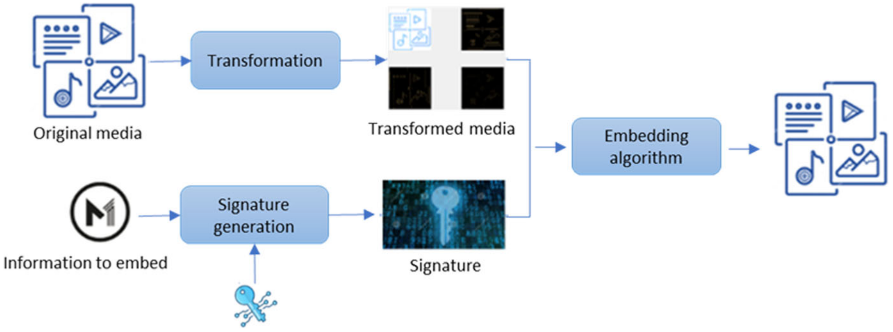

Fig. 2 Embedding stage

distribution, usage, duplication, manipulation, and storage. Indeed, it embeds a signature into the original content and then tries to detect it after different manipulations are applied on the marked content. Usually, a robust watermarking technique should be invisible. Watermarking is an important research area, thanks to its use in several media applications such as copyright protection and owner Identification, copy control and fingerprinting, content authentication and integrity verification, broadcast monitoring, indexing, and medical applications.

2.1 Watermarking Terminology

The watermarking process comprises two main stages: signature embedding and signature detection. Embedding is the stage in which a signature containing the author's information or copyright information is embedded within a hosting multimedia content through a specific embedding method, as shown in Fig. 2. First, the hosting content, an image, a video, or a 3D model, is eventually transformed depending on the chosen embedding target (DWT, DCT, FFT, etc.). Then, the signature is generated by scrambling watermark information randomly by using a secret key to enhance the security of the embedding method. Watermark can also be generated by applying several encryption algorithms as proposed in [14, 19, 53, 69]. The obtained mark is embedded within the selected coefficients, which will then be brought back into the original domain to obtain the marked content.

The signature detection stage tries to extract the embedded watermark and it is usually decomposed in the same steps of the embedding stage as shown in Fig. 3. Given a marked media, the same transformation used at embedding will be applied and the detection algorithm will be applied to the obtained coefficients. The signature detection stage may require knowledge of the original content. In this case, we say that a watermarking algorithm is non-blind. Contrary, if the watermark is recovered without resorting to the comparison between the original media and the marked one, the watermarking algorithm is blind.

If the signature contains a sequence of N bits, it can be read from the marked media. In this case, the watermarking algorithm is called multi-bit watermarking. However, in the 0-bit watermarking, the detector tries to decide whether a known signature is

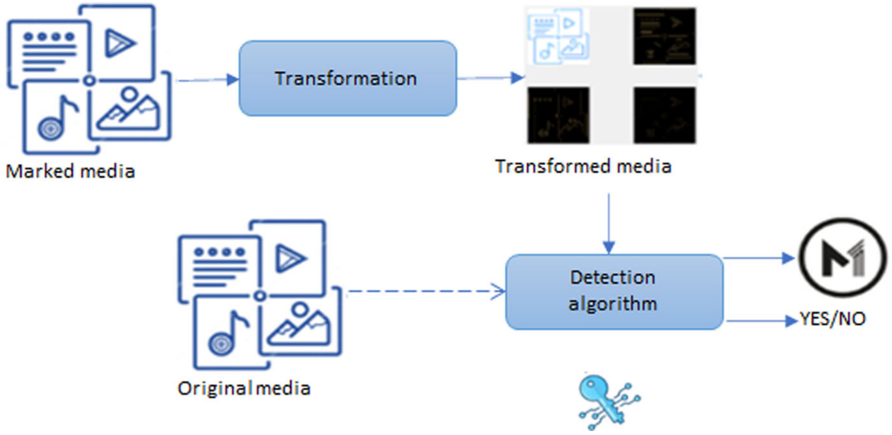

Fig. 3 Detection stage

present in the given media. In several applications, the two types can be required where the detector must verify at the first time the presence of the signature and if so, identify which message is encoded.

Any watermarking technique must satisfy three main requirements: invisibility, capacity, and robustness. Based on applications, these requirements evaluate the performance of watermarking systems. In the case of invisible watermarking, the market and the original content should be perceptually indistinguishable from humans. This fidelity can be evaluated qualitatively by asking a group of people to confirm the visual quality of the marked content or quantitatively by calculating several criteria. The standard criteria used to evaluate the invisibility quantitatively are the mean peak signal-to-noise ratio ($MPSNR$) and mean structural similarity index ($MSSIM$). In the case where the marked content is an image, $PSNR$ is calculated as shown in the following equation:

$$PSNR = 10\log_{10}\frac{255^2}{MSE}(dB) \tag{1}$$

$$MSE = \frac{1}{M \times N}\sum_{m=1}^{M}\sum_{n=1}^{N}[f(m,n) - f_w(m,n)]^2 \tag{2}$$

where M × N is the size of the image, f and f_w are the original and marked images, and MSE is the mean square error between f and f_w. In the case of a video and if the number of marked frames is K, we calculate the Mean $PSNR$ as follows:

$$MPSNR = \frac{1}{K}\sum_{k=1}^{K} PSNR_k \tag{3}$$

Despite its simplicity, $PSNR$ or $MPSNR$ cannot sometimes provide subjective evaluation results, so $SSIM$ or $MSSIM$ are introduced to evaluate visual quality of the marked image or video quality. The $MSSIM$ is defined as follows:

$$MSSIM = \frac{1}{k}\sum_{k=1}^{K} SSIM(f_k, f_{kw}) \tag{4}$$

$$SSIM(f_k, f_{wk}) = \frac{(2\mu_{f_k}\mu_{f_{kw}} + C_1)(2\sigma_{f_k f_{kw}} + C_2)}{(\mu_{f_k}^2 + \mu_{f_{kw}}^2 + C_1)(\sigma_{f_k}^2 + \sigma_{f_{kw}}^2 + C_2)} \tag{5}$$

where μ_{f_k} and $\mu_{f_{kw}}$ are the mean values of the original image and the marked one, respectively; σ_{f_k} and $\sigma_{f_{kw}}$ are the variances of the original image and the marked one. $\sigma_{f_k f_{kw}}$ denotes the covariance of the original image and the marked one; and C_1 and C_2 are two stability constants. We note that there exist watermarking techniques that are visible, but their use is limited to specific applications.

The second requirement is capacity (also called payload) which presents the quantity of embedded information in host media. For several applications, if the watermarking technique needs high invisibility, it is necessary to reduce the signature capacity to avoid too much modification in the host media.

The last requirement is robustness which is the ability to extract the embedded signature even when the marked media undergoes several signal processing manipulations. These manipulations include non-malicious attacks that are unintentional processing that may perturb the embedded signature such as geometric operations (translation, rotation, scaling), noises add, and filtering which can be applied to image or video content and malicious attacks which try to damage or remove the embedded signature. Among these attacks, we distinguish compression attacks and collusion which are specific to video content. Note that, depending on the application, not all watermarking techniques are robust against the same manipulations.

Referring to the robustness level, techniques can be classified into robust, fragile, and semi-fragile watermarking. Robust watermarking requires the watermark to resist noisy operations, as well as geometric or non-geometric manipulations. This class of watermarking is used in different applications such as copyright protection, broadcast monitoring, copy control, and fingerprinting. If the embedded signature is lost or altered after the application of the host content, the watermarking is fragile. This class of watermarking is usually used for integrity verification and content authentication applications. The last type of watermarking is the semi-fragile class that is robust against some attacks, but it fails after malicious manipulations. This class can be used for image authentication applications.

Bit error rate (BER) and normalized correlation (NC) are used to evaluate the robustness of a given watermarking. These two metrics are calculated to compute the dissimilarity between the embedded signature and the extracted one after applying different attacks to the marked content. In fact, the BER provides the percentage of erroneous bits during the transmission, and it is given by the following equation where S is the original signature, S' is the extracted one, $\sum_i Ber_i$ is the number of bit in error, and $\sum_i Btrans_i$ is the total number of transmitted bits:

$$BER(S, S^{'}) = \frac{\sum_i Ber_i}{\sum_i Btrans_i} \tag{6}$$

The NC calculates the similarity between two media. It is a value in the range [0,1] where a higher value proves a better similarity between media. Given an original and an extracted signatures S and $S^{'}$, NC metric is calculated as follows:

$$NC(S, S^{'}) = \frac{1}{WH} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} \delta(S_{i,j}, S^{'}_{i,j}) \tag{7}$$

where

$$\delta(S_{i,j}, S^{'}_{i,j}) = \begin{cases} 1, & \text{if } S = S^{'} \\ 0, & \text{otherwise} \end{cases}$$

The signature capacity, invisibility, and robustness are mutually restricted. Indeed, the most difficult challenge in the research of image and video watermarking area is how to choose embedding target that minimize the visual impact and have a high robustness and an acceptable capacity in the same technique.

2.2 Robust Traditional Image and Video Watermarking Techniques Classification

The main criterion used for image and video watermarking techniques classification is the embedding domain which can be spatial, frequency or hybrid domain.

Spatial watermarking embeds signature by directly modifying the luminance or the chrominance of the original image or video frame pixels. Spatial techniques are characterized by their low complexity and high invisibility. However, they suffer from the lack of robustness against several attacks. The main spatial domain techniques proposed for image and video watermarking include least significant bit (LSB) modification, spread spectrum modulation, and so on.

Concerning image content, LSB is the most used for the spatial domain where the least significant bit of several selected pixels is modified to embed signature [58]. LSB is very simple, but it fails to resist several attacks. For this reason, alternative methods, such as MIDSB (Middle significant bit) [12] and ISB (intermediate significant bit) [62] where the least significant bit was replaced, respectively, by the Middle significant bit and the best pixel value in between the Middle and the edge of the range, have been developed to improve the robustness. Other spatial techniques were proposed [74] to improve robustness while keeping the visual quality level. For video, LSB is also the most classical technique with the same method used for image watermarking while applying LSB for all or some selected frames composed the original video [42]. Despite the simplicity of the LSB technique, its robustness is very poor. The spread spectrum techniques were proposed as an effective spatial watermarking where the original video frames are scanned to obtain a one-dimensional signal and the signature is modulated by spread spectrum technology and inserted in the video [60]. Other spatial video watermarking techniques were also proposed in [8, 48, 82] to improve robustness against attacks. However, the application of these techniques is limited due to their poor robustness, especially with the development of video coding technology.

Frequency domain-based watermarking converts the original content (image or video frames) using a chosen transform and then modify the obtained coefficients to embed the signature. After that, the coefficients are converted back to the spatial domain to obtain the marked content. The most used frequency domain transforms for image watermarking are the discrete cosine transform (DCT) [38, 54, 75, 80, 83], discrete Fourier transform (DFT) [23, 37, 64, 70] discrete wavelet transform (DWT) [5, 32, 46, 85], and singular value decomposition (SVD) [4, 81]. Every frequency transform presents its own advantages and disadvantages where some transforms are robust against several attacks while they fail against others. For example, the spatial domain usually ensures robustness against translation and noises but it does not resist to compression and filtering contrary to DCT which is robust against rotation, filtering, and JPEG compression but it fails to resist noises. To resolve this problem, several image watermarking algorithms are based on the hybrid domain which combines different transforms with spatial domain together to profit from the advantages of these transforms [2, 13, 76]. Note that these algorithms ensure the best trade-off between robustness, capacity, and invisibility.

Concerning video content, like image, the common frequency domain transforms include DCT [18, 34, 49, 89], DWT [15, 30, 72, 79], and SVD which is usually combined with another transform as DWT [33, 77] and DCT [61]. As concluded for the image, the robustness of the video watermarking techniques depends on the characteristics of the chosen transform. However, to better improve performance, many watermarking algorithms use the hybrid domain that combines the advantages of the different transformations. Therefore, different techniques were proposed combining DCT and DWT [39, 73] or combining different transforms with the spatial domain as suggested in [44].

Since video content can be considered as a set of frames, any image watermarking technique can be adopted for video watermarking by embedding the signature into spatial redundancy of all or some selected frames. However, image-based techniques cannot resist video-specific attacks. In fact, video is also defined by temporal information which makes its processing more sensitive and the temporal redundancy in a video gives more chances to hackers to estimate signatures by using malicious attacks such as collusion. This last attack and frame-based attacks such as compression frames dropping and swapping should be considered by researchers when developing watermarking techniques for video. To resist these attacks, different techniques based on temporal information, such as mosaic [45], multi-sprites [11], and Krawtchouk moments [10], have been proposed and they proved their good robustness against malicious attacks, especially against collusion attack.

As video data is nowadays frequently used and transmitted on the Internet, the compression process is usually applied to reduce video size. However, watermarking techniques based on the original video decode the video during signature embedding and detecting stages and can destroy the signature and deteriorate the visual quality. To resolve this problem, a new class of video watermarking algorithms has emerged where the compressed domain is used. These algorithms embed signatures into compressed videos and combine the embedding stage with corresponding video coding standards which include MPEG [21, 22, 90], H.264 [27, 98], and H.265 [24, 55, 71]. Compressed

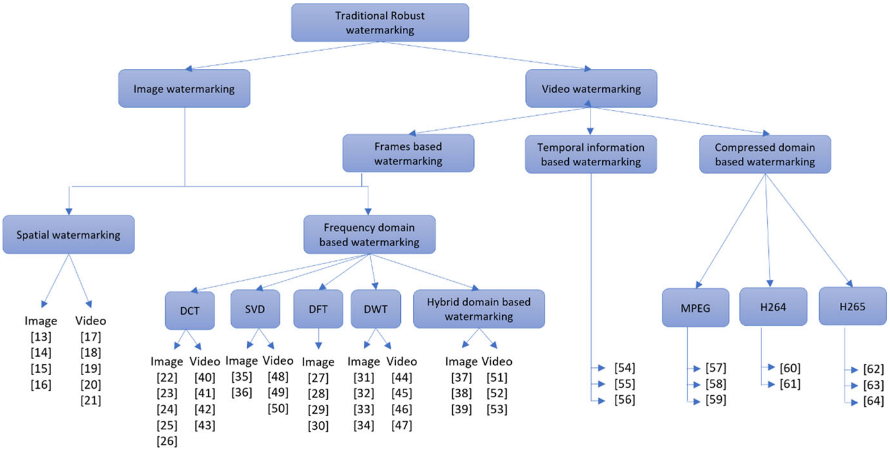

Fig. 4 Classification of the traditional robust image and video watermarking techniques

domain-based watermarking is robust against several attacks such as filtering, noises, and compression.

In summary, the classification of the traditional robust image and video watermarking techniques is illustrated in Fig. 4.

3 Basic Concepts of Deep Learning-Based Watermarking

With the success of deep learning in computer vision and image processing domains, it has been adopted for various tasks. Recently, deep learning models have attracted the attention of researchers in data hiding techniques including steganography [9, 25] and watermarking.

3.1 General Framework of Deep Learning-Based Watermarking Schemes

Deep learning-based watermarking usually uses an encoder–decoder based on convolutional neural networks (CNNs) structure to train models and to embed them in a robust and invisible way the signature. It is more efficient than traditional watermarking thanks to its advantage to be retrained to resist several attacks. In addition, it does not need an expert to develop the embedding method. Finally, the black-box nature of deep learning models allows for improving security.

The deep learning-based watermarking scheme is decomposed into three main stages as shown in Fig. 5. The first stage is the encoder which embeds the signature in the original content. The second stage is attack simulation and finally, a signature is extracted using the decoder network stage. Thanks to the iterative learning process, the embedding is more robust against attacks applied during the second stage, and the extraction network improves the integrity of the extracted signature. The main advantage of deep learning-based watermarking over traditional watermarking is that

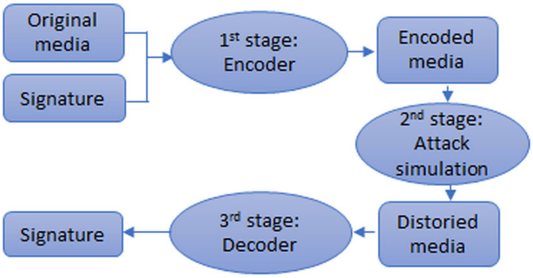

Fig. 5 Encoder–decoder architecture stages for digital watermarking

it can be easily retrained for various applications and different attacks instead of being designed from scratch.

An image or video watermarking scheme based on deep learning works as follows:

1. Training the encoder network to embed input messages to original content where the main goal is to minimize an objective function. This function calculates both the difference between original content and marked content and between the embedded and extracted signatures.
2. Applying different attacks to the marked content through distortion layers. These attacks can include different forms of manipulations such as cropping and compression.
3. Extracting the embedded message from distorted content using the decoder network.

3.2 Neural networks Architectures Used in Watermarking

Deep learning frameworks utilize automatic learning to capture hierarchical information directly from training data, eliminating the need for manual feature representations. Specifically, a deep network takes raw input data, such as an image or audio signal, and performs a mapping operation. Due to their impressive capability to imitate human brain learning abilities and engage in more natural interactions, deep learning techniques have gained widespread usage in data hiding and image processing applications.

Two deep learning models are widely used in watermarking techniques: convolutional neural network (CNN) and generative adversarial network (GAN).

CNNs are well suited for different applications such as classification and recognition, thanks to their efficiency in data representation with limited number of parameters [50]. The CNN algorithm is a specialized multilayer perceptron primarily developed for extracting and recognizing two-dimensional image details. The CNN architecture typically consists of multiple layers, including an input layer, convolutional layers, pooling layers, and an output layers shown in Fig. 6. The CNN initiates by taking an input image and subjecting it to a series of convolutions and subsampling operations. Each convolution layer comprises a collection of filter matrices, which are multiplied with the preceding image matrix to extract significant features referred to as output channel maps. Subsequently, pooling layers are employed to decrease the dimensions of the input map while preserving crucial information. Max pooling, a subsampling

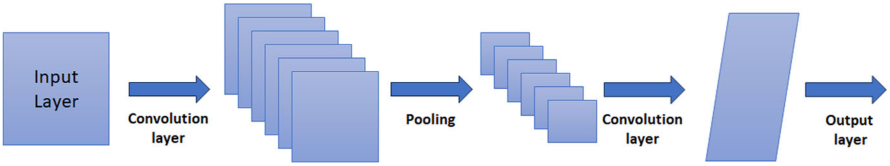

Fig. 6 CNN architecture

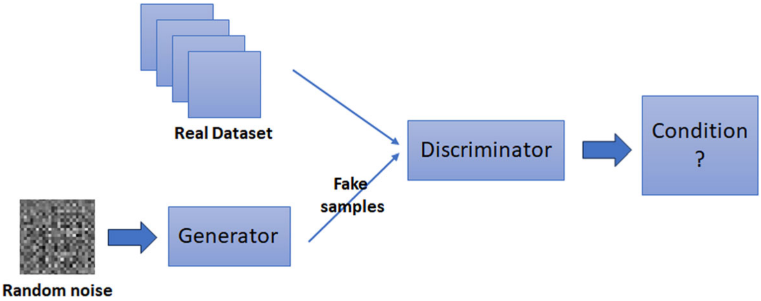

Fig. 7 GAN architecture

technique, selects the maximum value within each block. Nonlinearity is introduced into the network through activation functions like the rectified linear unit (ReLU), which sets negative values to zero. To mitigate overfitting and expedite learning, batch normalization can be employed during network training.

Concerning GAN, it is a type of neural networks widely employed in unsupervised learning. GAN consists of two neural network models that engage in a competitive process, enabling them to examine, grasp, and replicate the diverse patterns present in a given dataset. In fact, GAN is decomposed of two models: generative and discriminative model. It has the same principle as the encoder–decoder described in Fig. 5 with a difference from the discriminator network which classifies the mixture of encoded and unaltered images that are given to it (Fig. 7). The use of these discriminative networks can greatly improve data imperceptibility.

3.3 Examples of Datasets for Watermarking

To assess the performances of a deep learning-based watermarking scheme, different datasets were used in the literature. Among these datasets, we mention:

- ImageNet: ImageNet is a widely used dataset in computer vision research, consisting of millions of labeled images across thousands of categories. While not specifically designed for watermarking, it can be used to evaluate the effectiveness of watermarking techniques on various types of images.
- MS COCO (Microsoft Common Objects in Context): MS COCO is another popular dataset used for object detection and image segmentation tasks. It contains a large collection of images with diverse content, making it suitable for watermarking research.

- BOSSbase (BOWS-2): BOSSbase is a benchmark dataset for digital image watermarking. It contains 10,000 grayscale images with a resolution of 512x512 pixels. The dataset includes both the original images and the corresponding watermarked versions, making it suitable for evaluating the robustness and imperceptibility of watermarking algorithms.
- UCF101: UCF101 is a dataset commonly used for action recognition in videos. It consists of 13,320 videos covering 101 action categories. While primarily used for action recognition, it can be employed to evaluate video watermarking techniques on action-based content.
- The Kinetics dataset is a large-scale video dataset commonly used for action recognition tasks. It consists of approximately 650,000 video clips covering 700 action categories. The dataset is diverse and includes a wide range of human actions captured from YouTube videos. While the Kinetics dataset is not designed specifically for watermarking research, it can still be useful for evaluating certain aspects of watermarking techniques on action-based video content.

4 Deep Learning-Based Image Watermarking Review

While the current research on deep learning-based watermarking predominantly revolves around image watermarking, other forms of media are still in an early stage of development. Only a limited number of works have been proposed for text [1] and 3D images [92]. These approaches offer improved efficiency compared to traditional techniques by leveraging their ability to learn complex insertion patterns that are resilient against various attacks. This robustness is obtained since the networks of deep learning can be easily retrained to become robust to different types of attacks. Moreover, they can target capacity payload or imperceptibility optimization without developing new algorithms for each different application. Deep learning models are characterized by their high nonlinearity which makes the retrieval of the embedded signature impossible by an adversary.

4.1 Classification of Deep Learning-Based Image Watermarking Schemes

Current deep learning-based image watermarking techniques can be categorized into two classes based on the chosen network architecture. The first class uses the encoder–decoder framework including CNNs where we distinguish techniques which are based CNN encoder–decoder (Fig. 5) and those based on the convolutional auto-encoders which are a special case of the encoder–decoder used in unsupervised-learning scenarios.

Two traditional convolutional auto-encoders for watermark embedding and extraction were proposed in [41]. These auto-encoder CNN models allow for obtaining high invisibility of the embedded signature. Moreover, the watermarking proposed in [41] proved its efficiency in terms of robustness and outperforms the traditional watermarking techniques. Another convolutional auto-encoder-based robust and blind watermarking technique was proposed in [63]. This approach is decomposed into three steps: embedding, attack simulation, and updating. In the second step, the CNN simulates the various attacks while in the updating, the loss function is minimized by updating the model's weights.

In [78], the authors present a method of watermarking digital images using CNNs. First, an encoder network is used to extract latent features from the cover and secret images. These features are then concatenated to create a marked image. On the receiving end, a CNN is used to retrieve the secret marked image after removing noise variations from the received image using a denoising auto-encoder network.

Ahmadi et al. [3] presents a new approach called ReDMark which uses two full convolutional neural networks (FCNNs) for embedding and extraction. It contains a differentiable attack layer to simulate different distortions. This technique improves robustness against attacks and maximizes the trade-off between robustness and imperceptibility. Zhong et al. [99] proposes a CNN-based watermarking technique which is robust and blind and can be used for several applications. This technique generalizes the watermarking process by training a deep neural network to learn the general rules of watermark embedding and extraction. This technique outperforms the two auto-encoder CNN methods proposed in [41, 63], and allows obtaining greater robustness. Another watermarking model developed in [47] uses a simple CNN for both embedding and extraction. It contains an image pre-processing network that can adapt images of any resolution for the watermarking process and a watermark pre-processing as well as a strength scaling factor to control the trade-off between robustness and imperceptibility.

Luo et al. [57] improve the CNN-based encoder–decoder framework by adopting trained CNNs for attack simulation instead of using a differentiable attack layer. The addition of adversarial components to model training can improve the robustness of the embedded mark. In fact, in [57], the distortions are generated via adversarial training by a trained CNN.

In [68], an optimized deep fusion convolutional neural network (FCNN)-based digital color image watermarking scheme was proposed for copyright protection. It suggests a deep fusion CNN that uses an optimization method as its basis. The octave convolutional module added by the embedding network reduces spatial redundancy and increases the receptive field. The ECO method can help choose a suitable strength factor with great exploration capabilities.

The second class of deep learning-based image watermarking is based on generative adversarial networks (GAN) [28]. Several variants of the GAN exist, and they include Wasserstein GANs (WGANs) and CycleGANs which are used for image watermarking and provide good results in terms of invisibility and robustness. HiDDeN [100] is the first scheme which uses an adversarial discriminator to improve the performance of the watermarking process. It is decomposed of an encoder network which trains to embed an encoded bit string, a decoder network which tries to extract the information from the encoded image and an adversary network which predicts if the image was encoded or not.

ROMark [87] and [31] improve the HiDDeN technique where the goal of [87] is to minimize decoding loss across a range of attacks, rather than training the model to resist specialized attacks. This technique is more robust than [100] in some specialized attack categories. Concerning [31], it uses a rotation layer and an additive noise layer, allowing the model to learn robustness against geometric rotation attacks. It also uses a noise strength factor to maximize the trade-off robustness/invisibility. Zhang et al. [96] proposed a new GAN-based watermarking technique which uses inverse gradient attention (IGA) to embed signature. This technique identifies pixels that are robust based on an attention mask which provides the values of the gradient of the original image. This allows for improving the capacity and robustness of the marked images compared with other techniques. Another GAN-based watermarking was proposed in [52] where TSDL (two-stage separable deep learning) framework was introduced. This framework can use true non-differentiable noise attacks such as JPEG compression during training. Liu et al. [52] achieves good robustness compared with the previous techniques.

Annadurai et al. [6] presents an approach of digital watermarking based on discrete wavelet transform (DWT) quantization model with convolutional generative adversarial neural networks which are used for segmentation and classification of the processed image. In this technique, the SVD-based discrete wavelet transform quantization model is used for watermarking.

Other watermarking techniques using GAN variants were proposed. The first used variant is Wasserstein GAN (WGAN) which improves the stability during training and the sensitivity of the training of the GAN model [7]. WGANs contain a critic component rather than a discriminator component that returns a score which indicates if the input image is real or not.

Plata et al. [66] proposed a new watermarking based on WGAN where the signature is spread over the spatial domain of the image. The suggested technique uses a new method for differentiable noise approximation of non-differentiable distortions which allows the simulation of subsampling attacks. In [67], the authors improve the previous work by using a double discriminator/detector architecture. The discriminator is placed after the noise layer and learns to distinguish watermarked and non-watermarked images with attacks already applied. Wang et al. [86] proposed a technique which enhances the quality of the encoded image based on texture analysis. The texture of the original image is analyzed using a gray co-occurrence matrix which classifies regions into complex and flat types.

The second variant of GAN used for image watermarking is the CycleGAN [101] which includes two generative and two discriminative models. [94] is the only watermarking technique which uses this framework. This technique uses an attention model to embed data that an attention mask, which represents the attention sensitiveness of each pixel in the cover image. This enhances the embedding process of the encoder network.

4.2 Comparison of Deep Learning-Based Image Watermarking Schemes

Figure 8 illustrates the classification of image watermarking using deep learning, depicting the fluctuation in the number of proposed papers according to the employed architecture. Table 1 summarizes the differences between the various techniques of deep learning-based image watermarking techniques proposed in the literature. In fact, based on the study of the art, we can observe that the GAN is more efficient and

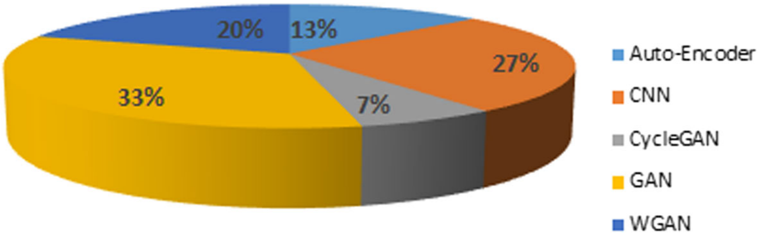

Fig. 8 Number of deep learning-based image watermarking articles per Architecture

promising in terms of robustness, thanks to the inclusion of an adversarial network which greatly improve the invisibility of the watermarking. However, to improve the robustness GAN-based watermarking, it should be combined with other methods as proposed in [94] and [96] where an attention mechanism and IGA method were used. Note that the robustness depends on the attack types used during training. Finally, we note that every class presents robustness against a set of attacks and all techniques except [41] are blind.

Regarding the invisibility criterion, Fig. 11 illustrates the comparison of PSNR values obtained after the application of various existing deep learning-based watermarking techniques. While exploring this figure, we can observe that most deep learning-based image watermarking techniques provide high visual quality for the watermarked image. Nevertheless, we notice that CNN-based techniques offer the best PSNR values, and this is achieved through its capacity to directly align the intricate mapping between low-resolution and high-resolution images. This alignment enhances the recovery of lost high-frequency information, surpassing the performance of numerous conventional methods.

5 Deep Learning-Based Video Watermarking Review

Despite the good number of existing techniques of deep learning-based watermarking proposed for images, video watermarking based on deep learning has only recently begun to be explored and is still an open problem. In fact, as far as we know, there are only a very few number of video watermarking techniques based on deep learning in the literature [16, 26, 35, 40, 43, 56, 59, 65, 91, 97] that appeared since 2019.

5.1 Classification of deep Learning-Based Video Watermarking Techniques

Deep learning-based video watermarking techniques can be classified based on the original video domain which can be the original frames or the compressed domain. In fact, [16, 26, 35, 40, 43, 56, 59, 65, 91, 97] have been proposed for video frames where [16, 35, 43, 56, 65, 91, 97] are proposed for robust multi-bit embedding. [26] is a robust zero watermarking and [40] is proposed recently for compressed videos. Finally, Mansour et al. [59] is based on the mosaic image generated from original video (Fig. 9).

Table 1 Comparison of the existing deep learning-based image watermarking techniques

References	Architecture	Technique	Domain	Capacity	Robustness
[41]	Auto encoder CNN	Uses two convolutional auto-encoders for watermark embedding and extraction	Frequency	128 × 128	JPEG Filtering Geo. attacks Cropping Noise
		Non-blind			
[63]		Uses shallow network and visual mask	Spatial	512 × 512	
		Blind			
[78]		Uses an auto-encoder-based embedder network and a denoising Auto-Encoder network		128 × 128	
		Blind			
[3]	CNN	Uses a circular convolutional embedding network and DCT layer for embedding and extraction networks	Frequency	128 × 128	JPEG Filtering Noise Cropping Rotation
		Blind			
[68]		Uses octave convolutional module (OCM)		512 × 512	
		Enhanced chimp optimization algorithm			
[99]		Uses an invariance layer including a redundancy parameter to tolerate distortions notseen during network training	Spatial	128 × 128	
		Blind			
[47]		Uses a simple CNN as embedding and extraction network, an image pre-processing network, a watermark pre-processing and a strength scaling factor			
		Blind			
[57]		Uses an adversarial training to generate the distortions			
		Blind			
[31]	GAN	Uses a neural network for attack simulation	Spatial	64 × 64	Dropout cropout filtering JPEG mask JPEG drop noise rotation re-encoding attacks
		Blind			
[100]		Uses an adversarial discriminator		128 × 128	

Table 1 continued

References	Architecture	Technique	Domain	Capacity	Robustness
		Blind			
[87]		Uses a min–max formulation for robust optimization Blind			
[52]		Two-stage separable deep learning (TSDL) framework for watermarking Blind			
[96]		Uses inverse gradient attention (IGA) mask to identify robust pixels Blind	Frequency		
[6]		SVD-based discrete wavelet transform quantization model of the processed image The marked image was segmented and classified using convolutional generative adversarial neural networks			
[66]	WGAN	Uses a spatial spread embedding technique Blind	Spatial	256 × 256	
[67]		Uses a double discriminator/detector architecture Blind			
[86]		Uses texture analysis based on a gray co-occurence matrix Blind		128 × 128	
[101]	CycleGAN	Uses an attention mask representing the attention sensitiveness of each pixel Blind	Spatial	512 × 512	

Zhang et al. [97] introduces a new architecture called RIVAGAN, for robust video watermarking composed of two adversaries: a critic and an adversary network. The first one evaluates the quality of the marked video, and the second one tries to remove the watermark. These two components work with the encoder and decoder networks which, respectively, embed and extract the watermark for the video. The proposed architecture is based on an attention-based mechanism which identifies regions that are robust for embedding and generates marked regions with high visual quality. The attention module is composed of two convolutional layers shared between the encoder and decoder. It generates an attention mask from the original frames by applying the two convolutional blocks. This mask contains the data, time, and size dimensions. This

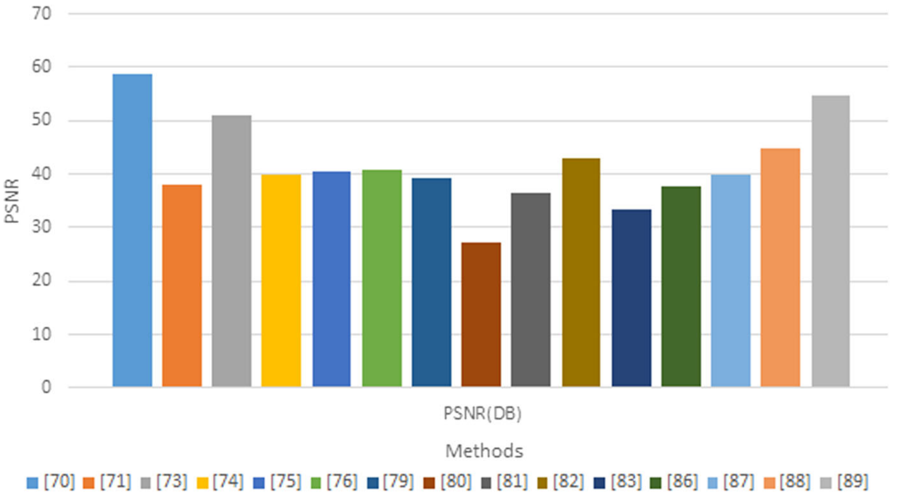

Fig. 9 PSNR comparison of the deep learning-based image watermarking existing techniques

mechanism makes it easy the training and enhances the robustness against different attacks such as scaling and compression.

Luo et al. [56] is another multi-bit robust video watermarking using deep learning called DVMark. It is composed of four stages: an encoder, a decoder, a distortion layer, and a GAN discriminator. In fact, the encoder is a multiscale network that embeds the signature, which is repeated across spatial and temporal dimensions, in the original video on two different spatial–temporal scales. A scalar factor is used to change the signature strength at the time of inference. Concerning the decoder, network is composed of a transform layer and two detector heads which can detect marked frames from unmarked ones. In the distortion layer, different distortions like frame dropping, cropping, and compression are applied to the marked video and the decoder network generates a predicted message from the distorted video. Finally, the multiscale video discriminator's architecture with its 3D ResBlock allows the network to detect both spatial and temporal differences between the cover and watermarked videos. This approach was compared with a 3D-DWT traditional video watermarking and with the HiDDeN method and it was found more efficient in terms of robustness against attacks.

Bistron and Piotrowski introduce in [16] a video watermarking algorithm that merges CNNs with an entropy-driven information mapper. Their approach involves incorporating the watermark into the YUV color space's luminance channel. By utilizing an information mapper, intricate multi-bit binary signatures can be embedded into the watermark of a signal frame. Although the article acknowledges the utilization of CNNs and an entropy-based information mapper to enhance resilience, it overlooks the algorithm's effectiveness when faced with advanced watermarking attacks like geometric transformations, compression, cropping, and collusion.

In [43], the authors aim to create a video watermarking system using curriculum learning approaches and deep neural networks. The attention module is a part of the encoder and decoder component of RivaGAN. Overall, an encoder network that is hidden from the decoder network interrupts the segmented videos' first frame.

The suggested method in [65] uses an Improved Invasive Honey Badger Optimization (IIHBO) algorithm to embed hidden audio components into videos. The process consists of two primary stages: extraction and embedding. Using a Shepard convolutional neural network (ShCNN) trained by IIHBO, the secret audio is incorporated into the predicted object position during the embedding phase. Using the same methods in reverse, the extraction step extracts the hidden audio from the embedded video. For effective ShCNN training, the IIHBO—a hybrid of Improved Invasive Weed Optimization (IIWO) and Honey Badger Optimization (HBO)—is employed.

Reversible medical video watermarking with a Deep CNN based on SCBSA is discussed in [35]. In order to integrate videos, the approach combines the Sine Cosine and Bird Swarm Algorithm (SCBSA), which includes key frame extraction, region identification, and embedding. From gridded video frames, features like CNN, LOOP, neighborhood-based, and histogram features are retrieved. The appropriate area for embedding a hidden message is determined by the Deep CNN classifier, which has been trained using SCBSA. Using a two-level decomposition based on wavelet transform, the SCBSA, a hybrid technique of SCA and BSA, makes message embedding and extraction simpler.

ItoV [91] presents an approach that adapts image watermarking techniques based on deep learning to video watermarking. The main goal is to address issues like computational complexity and temporal interdependence that are present in video data. The authors concentrate on combining the channel and temporal dimensions of videos so that deep neural networks can process videos like images. They investigate how different convolutional blocks affect video watermarking and find that, although depthwise convolutions greatly lower computational costs with no effect on performance, spatial convolution is essential. In watermark embedding, the neural network's task is to understand the cover video's pixel distribution so that messages can be added with the least amount of distortion.

Gao et al. [26] proposed a robust zero-watermarking technique for copyright protection of videos contents. This technique is based on a CNN architecture with a self-organizing map (SOM) in polar complex exponential transform (PCET) space. First, the scheme extracts the feature for the frames composing the original video using CNN. Then, it selects some significant frames by applying SOM clustering and maximum entropy. Given the selected frames, the invariant moments are detected using PCET and the dimensions are reduced by singular value decomposition (SVD). The obtained moments will be used to generate the binary matrix. Finally, the zero-watermark signal is generated by applying a bitwise exclusive-OR operation on the binary matrix and the watermark is encrypted by the chaotic map. The experiments showed that this technique is robust against several attacks such as geometric, compression, and inter-frame attacks and proved a superior efficiency compared with existing video zero-watermarking and traditional video watermarking methods.

The deep learning-based video watermarking technique proposed in [40] is a recent method which works in the compressed domain for protecting encoded videos with H.265/ HEVC codec compression. First, the encoder subsystem generates the watermark by applying the adjustable subsquares properties. This watermark will be introduced to the preliminary network. The encoder DNN takes as input the hidden image with the original one and decomposes the secret image from the preliminary network to the set of features in order to encode the watermark. Then, the deconvolution of the obtained set of secret image features is carried out with the carrier image. During the learning process, the neural network automatically selects the optimal filters by which the image modification can be carried out. The encoding is done with the adjustable subsquares properties algorithm to obtain a bit-encoded image. After the image has passed through the compression channel of the HEVC codec, it will be decoded using a decoder network with return the recovered watermark which is processed by the decoder subsystem to identify the watermark by recognizing the information encoded in the recovered image. This technique presents a high visual quality of the marked video.

Recently, a novel approach to video watermarking utilizing deep learning and employing a mosaic image has been presented in [59]. This method extends the concept of image watermarking to video watermarking. It involves four key steps: pre-processing networks for generating the mosaic image from the original video and for handling the signature, an embedding network, attack simulation, and an extraction network. The primary objective of generating the mosaic image is to construct an image from the original video while ensuring resilience against malicious attacks, particularly collusion attacks. The proposed technique adjusts the resolution of the mosaic image and incorporates signature information, incorporating various CNN training methods like averaging, batch normalization, and rectified linear unit (ReLU). During the attack simulation phase, all attacks, except collusion and MPEG compression, are included in each mini-batch.

5.2 Comparison of Deep Learning-Based Video Watermarking Techniques

Figure 10 illustrates the distribution of existing works based on the type of architecture used. It shows that the CNN architecture is the most commonly employed for video watermarking. The widespread adoption of CNN architecture for video watermarking can be attributed to several reasons. First, CNNs are designed to automatically extract relevant features from input data, which is crucial for video watermarking where specific patterns need to be identified and incorporated imperceptibly. Then, videos contain complex and dynamic information. CNNs, with their deep architecture, can capture complex relationships between successive frames, which is essential for video watermarking. Finally, CNNs are designed to be invariant to translations and deformations, making them robust to minor modifications in an image an important property for video watermarking where the video may undergo alterations.

Table 2 summarizes the advantages and the differences between the existing deep learning-based video watermarking techniques. These techniques use different network architectures and different domains of embedding. Their robustness depends on the used domain and the chosen architecture. Undoubtedly, the specific procedure of deep learning-based video watermarking schema may differ based on the employed technique, but overall, it entails training a deep neural network to comprehend the video's characteristics along with the watermark. Subsequently, this trained network

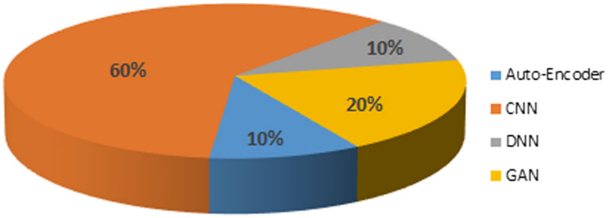

Fig. 10 Distribution of deep learning-based video watermarking articles based on architecture

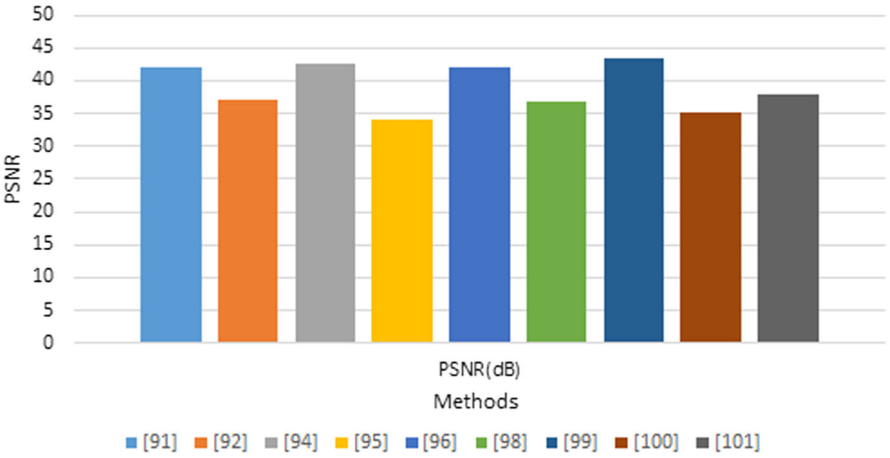

Fig. 11 Invisibility comparison of the deep learning-based video watermarking existing techniques

is utilized to insert the watermark into the video by modifying the video frames in a way imperceptible to the human eye. Likewise, the same neural network can extract the watermark from the watermarked video. One significant advantage of deep learning-based video watermarking methods is their superior ability to effectively address video watermarking challenges, such as motion and compression artifacts, surpassing the capabilities of conventional approaches.

Although numerous papers referenced in this study employ various embedding techniques, it is not possible to definitively determine the optimal solution for the video watermarking process. Notably, certain research indicates that treating video frames as images and employing conventional digital image watermarking methods for embedding may be an option. However, this approach hinders the neural network's ability to acquire valuable video-specific features, thereby posing challenges in effectively countering distortions specifically aimed at videos.

Figure 11 compares the visual quality of existing video watermarking techniques based on deep learning, relying on PSNR values. We can observe that these values are close and range between approximately 34 and 44 dB. These values demonstrate the invisibility guaranteed by these watermarking techniques.

In summary, deep learning-based approaches for video watermarking surpass conventional methods in multiple aspects. These benefits comprise the watermark's

Table 2 Comparison of existing deep learning-based video watermarking techniques

References	Domain	Capacity	Network architecture	Techniques	Robustness
[97]	Video frames	Multi-bits	RIVAGAN	Uses a critic and an adversary network.	Scaling
		32 × 32		Based on an attention module	Cropping
					Compression
[56]	Video frames	Multi-bits 8	3D CNN	Uses multiscale encoder and decoder networks.	Compression
		× 128 × 128		Uses Gan discriminator and a distortion layer	Frame drop
					Frame averaging
					Frame swapping
					Frame blurring
					Cropping
[16]	Video frames	Multi-bits	CNN	Uses YUV color space's luminance channel	Geo. transformations
		256 × 256		Based on entropy-driven information mapper	Cropping
					Collusion
					Compression
[26]	Video frames	Zero-bit	CNN	Uses a combination of convolutional neural network (CNN) and self-organizing map (SOM) in polar complex exponential transform (PCET) space	Rotation
				Uses SVD decomposition	Gaussian noise
					Salt & pepper noise
					Medium filter
					Frame dropping
					Frame swapping
					Frame averaging
[40]	Compressed domain	Multi-bits	DNN auto-encoder	Uses the adjustable subsquares properties encoding data algorithm	Compression
		128 × 128			
[59]	Mosaic Image	Multi-bits	CNN	Uses mosaic image generated from original video as embedding target.	Geo. transformations

Table 2 continued

References	Domain	Capacity	Network architecture	Techniques	Robustness
		128 × 128		Uses attack simulation including various manipulations	Cropping
					Collusion
					Compression
[43]	Video frames	Multi-bits	GAN	Uses an encoder–decoder model with an attention module in the architecture.	Cropping
		256 × 256		Uses a curriculum learning strategy	Scaling
					H.264 Compression
[65]	Video frames	Multi-bits	ShCNN	Uses IIHBO-based ShCNN for video object watermarking	Noise attacks
		–			
[35]	Video frames	Multi-bits	CNN	Training of Deep CNN using developed SCBSA for region selection	Impulse noise
		–			Gaussian noise
					Salt and pepper noise
[91]	Video frames	Multi-bits	CNN	Integrates the temporal dimension into the channel dimension for deep neural networks	H.264 Frame Drop
		128 × 128			- Frame swap
					Random crop
					Gaussian blur
					Random hue

substantial capacity and imperceptibility, resilience against different attacks, versatility across diverse video formats and resolutions, and proficiency in addressing challenges associated with video watermarking, such as motion and compression artifacts.

6 Discussion and Suggestions for Future Research

Deep learning for watermarking is a recent and evolving research domain. As shown in this survey, existing works were all focused on image watermarking, but there are many other important applications for watermarking video with deep learning. As far as we know, only the techniques of video watermarking based on deep learning described in this survey were proposed despite the large number of traditional watermarking techniques proposed for video. In fact, video content continues to present additional challenges, such as temporal coherence, which is a spatial location that cannot be resolved with fixed images. Moreover, video compression is not differentiable, and it is difficult to integrate it into a deep neural network training framework. Furthermore, it is not easy to visualize a robust model that uses temporal correlations in a video while maintaining temporal coherence and perceptual quality. Thus, for videos, deep learning-based watermarking is still in its early stages. However, given the urgent need of protecting videos and the efficiency in terms of invisibility and robustness that can offer deep learning networks, we expect that the above challenges will be the focus of extensive research that will take up most of the academics' time in the coming years.

Based on the state of the art presented in this paper, we note that CNN and GAN are the most used architectures for image and video watermarking. These two architectures present different challenges. In fact, the difficulties encountered by CNN models in watermarking systems are outlined below:

- CNN models experience increased latency due to operations like max-pool.
- Longer training times can be incurred occasionally due to misconfigured network parameters.
- Larger datasets are necessary for the training and processing of CNN models.
- The complexity of CNN networks can lead to issues like overfitting or underfitting at times.
- Applying CNN model to video watermarking, which involves the processing of multiple frames and temporal dependencies, can be time-consuming and resource-intensive.

Concerning GAN model, the posed challenges are as follows :

- Overfitting arises from discrepancies between the generator and discriminator networks.
- The network parameters' oscillation and destabilization prevent convergence.
- In certain cases, the discriminator becomes overly adept, leading to the vanishing of the generator gradient and a lack of learning.
- The generator network occasionally gets stuck, resulting in limited variations of the samples.

However, many other efficient deep learning architectures were developed for other applications such as classification and recognition and we recommend exploring them in watermarking schemes. For an example, RNN (recurrent neural network) was used for many tasks for video content and can provide good results for video watermarking.

Additionally, research findings indicate that transform domain-based watermarking techniques exhibit greater resilience compared to spatial domain-based approaches. Therefore, it is recommended to combine multiple frequency domains in the same image watermarking scheme to achieve enhanced security. Besides, to reduce the complexity of model training, pre-trained models are extensively employed, giving rise to challenges such as model overwriting and surrogate model attacks. In fact, a pre-trained model is a deep learning model that have been trained on large datasets and can be used as a starting point for various tasks without training from scratch as YOLO model.

Moreover, since many deep learning-based image watermarking techniques, that are robust and invisible, have been proposed in the literature, we can profit from their advantages by adapting them for video watermarking. In fact, an image can easily be generated from video with a reversible scheme allowing to return to the original video such as mosaic generation and Krawtchouk Matrix generation. By transforming a video to an image with a reversible algorithm, we can apply image watermarking to the obtained image to embed a signature into a video.

Note that the problem with some proposed deep learning-based video watermarking techniques is that they did not focus on testing the robustness of the method against malicious attacks such as collusion (type I and II) which are very dangerous attacks that should be considered when developing a video watermarking technique.

7 Conclusion

An overview of deep learning techniques used in watermarking and applied in images and video is presented in this survey paper. Firstly, watermarking terminology was presented, and traditional image and video watermarking techniques were classified based on their embedding domain. Then, deep learning-based image watermarking was classified and compared based on their network architecture. The survey also compared the four existing deep learning-based video watermarking proposed recently. Finally, this paper provided possible suggestions for future research in the field of video watermarking based on deep learning. This last one is a promising recent field of research with the potential to revolutionize the protection and security of video communication. In a conclusion, we can confirm that deep learning-based methods for watermarking will greatly surpass the capabilities of any traditional watermarking techniques in all media and greatly enhance digital information security.

Author Contributions S.BJ. had the idea of the presented paper and written it. S.BJ. and M.BF. performed the literature search and data analysis and they critically revised the work.

Funding No funding was received for conducting this study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. S. Abdelnabi, M. Fritz, Adversarial watermarking transformer: towards tracing text provenance with data hiding, in *2021 IEEE Symposium on Security and Privacy (SP)* (IEEE, 2021), pp. 121–140
2. A.K. Abdulrahman, S. Ozturk, A novel hybrid DCT and DWT based robust watermarking algorithm for color images. Multimed. Tools Appl. **78**(12), 17027–17049 (2019)

3. M. Ahmadi, A. Norouzi, N. Karimi, S. Samavi, A. Emami, ReDMark: framework for residual diffusion watermarking based on deep networks. Expert Syst. Appl. **146**, 113157 (2020)
4. M. Ali, C.W. Ahn, M. Pant, P. Siarry, A reliable image watermarking scheme based on redistributed image normalization and SVD. Discrete Dyn. Nat. Soc. (2016). https://doi.org/10.1155/2016/3263587
5. S.P. Ambadekar, J. Jain, J. Khanapuri, Digital image watermarking through encryption and DWT for copyright protection, in *Recent Trends in Signal and Image Processing* (Springer, 2019), pp. 187–195
6. C. Annadurai, I. Nelson, K.N. Devi, R. Manikandan, A.H. Gandomi, Image watermarking based data hiding by discrete wavelet transform quantization model with convolutional generative adversarial architectures. Appl. Sci. **13**(2), 804 (2023)
7. M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in *International Conference on Machine Learning* (PMLR, 2017), pp. 214–223
8. Z. Bahrami, F. Akhlaghian Tab, A new robust video watermarking algorithm based on surf features and block classification. Multimed. Tools Appl. **77**(1), 327–345 (2018)
9. B. Bashir, A. Selwal, Towards deep learning-based image steganalysis: practices and open research issues. Available at SSRN 3883330 (2021)
10. I. Bayoudh, S. Ben Jabra, E. Zagrouba, A robust video watermarking for real-time application, in *International Conference on Advanced Concepts for Intelligent Vision Systems* (Springer, 2017), pp. 493–504
11. I. Bayoudh, S. Ben Jabra, E. Zagrouba, On line video watermarking-a new robust approach of video watermarking based on dynamic multi-sprites generation, in *VISAPP (3)*, pp. 158–165 (2015)
12. I. Bayoudh, S. Ben Jabra, E. Zagrouba, Online multi-sprites based video watermarking robust to collusion and transcoding attacks for emerging applications. Multimed. Tools Appl. **77**(11), 14361–14379 (2018)
13. M. Begum, J. Ferdush, M.S. Uddin, A hybrid robust watermarking system based on discrete cosine transform, discrete wavelet transform, and singular value decomposition. J. King Saud Univ. Comput. Inf. Sci **34**(8), 5856–5867 (2021)
14. M.A. Ben Farah, A. Kachouri, M. Samet, Improvement of cryptosystem based on iterating chaotic map. Commun. Nonlinear Sci. Numer. Simul. **16**(6), 2543–2553 (2011)
15. A. Bhardwaj, V.S. Verma, R.K. Jha, Robust video watermarking using significant frame selection based on coefficient difference of lifting wavelet transform. Multimed. Tools Appl. **77**(15), 19659–19678 (2018)
16. M. Bistroń, Z. Piotrowski, Efficient video watermarking algorithm based on convolutional neural networks with entropy-based information mapper. Entropy **25**(2), 284 (2023)
17. O. Byrnes, W. La, H. Wang, C. Ma, M. Xue, Q. Wu, Data hiding with deep learning: a survey unifying digital watermarking and steganography. arXiv preprint arXiv:2107.09287 (2021)
18. A. Cedillo-Hernandez, M. Cedillo-Hernandez, M.N. Miyatake, H.P. Meana, A spatiotemporal saliency-modulated JND profile applied to video watermarking. J. Vis. Commun. Image Represent. **52**, 106–117 (2018)
19. B.P. Devi, K.M. Singh, S. Roy, New copyright protection scheme for digital images based on visual cryptography. IETE J. Res. **63**(6), 870–880 (2017)
20. D. Dhaou, S. Ben Jabra, E. Zagrouba, A review on anaglyph 3D image and video watermarking. 3D Res. **10**(2), 1–12 (2019)
21. H. Ding, R. Tao, J. Sun, J. Liu, F. Zhang, X. Jiang, J. Li, A compressed-domain robust video watermarking against recompression attack. IEEE Access **9**, 35324–35337 (2021)
22. H. Ding, R. Tao, J. Sun, J. Liu, F. Zhang, X. Jiang, J. Li, A compressed-domain robust video watermarking against recompression attack. IEEE Access **9**, 35324–35337 (2021)
23. M.T. Gaata, An efficient image watermarking approach based on Fourier transform. Int. J. Comput. Appl. **136**(9), 8–11 (2016)
24. S. Gaj, A. Kanetkar, A. Sur, P.K. Bora, Drift-compensated robust watermarking algorithm for H. 265/HEVC video stream. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **13**(1), 1–24 (2017)
25. S. Gaj, A. Sur, P.K. Bora, Prediction mode based H. 265/HEVC video watermarking resisting re-compression attack. Multimed. Tools Appl. **79**(25), 18089–18119 (2020)
26. Y. Gao, X. Kang, Y. Chen, A robust video zero-watermarking based on deep convolutional neural network and self-organizing map in polar complex exponential transform domain. Multimed. Tools Appl. **80**(4), 6019–6039 (2021)

27. M. Ghasempour, M. Ghanbari, A low complexity system for multiple data embedding into H. 264 coded video bit-stream. IEEE Trans. Circuits Syst. Video Technol. **30**(11), 4009–4019 (2019)
28. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in *Advances in Neural Information Processing Systems*, vol. 27 (2014)
29. M. Gupta, R.R. Kishore. A survey of watermarking technique using deep neural network architecture, in *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (IEEE, 2021), pp. 630–635
30. G. Gupta, V.K. Gupta, M. Chandra, An efficient video watermarking based security model. Microsyst. Technol. **24**(6), 2539–2548 (2018)
31. I. Hamamoto, M. Kawamura, Neural watermarking method including an attack simulator against rotation and compression attacks. IEICE Trans. Inf. Syst. **103**(1), 33–41 (2020)
32. K. Hannoun, H. Hamiche, M. Lahdir, M. Laghrouche, S. Kassim, A novel dwt domain watermarking scheme based on a discrete-time chaotic system. IFAC-PapersOnLine **51**(33), 50–55 (2018)
33. Y. Himeur, A. Boukabou, A robust and secure key-frames based video watermarking system using chaotic encryption. Multimed. Tools Appl. **77**(7), 8603–8627 (2018)
34. J.-U. Hou, MPEG and DA-AD resilient DCT–based video watermarking using adaptive frame selection. Electronics **10**(20), 2467 (2021)
35. S.S. Ingaleshwar, D. Jayadevappa, N.V. Dharwadkar, Sine cosine bird swarm algorithm-based deep convolution neural network for reversible medical video watermarking. Multimed. Tools Appl. pp. 1–26 (2023)
36. R. Jain, M.C. Trivedi, S. Tiwari. Digital audio watermarking: a survey, in *Advances in Computer and Computational Sciences*, vol. 2 (Springer, 2018), pp. 433–443
37. S.S. Jamal, M.U. Khan, T. Shah, A watermarking technique with chaotic fractional s-box transformation. Wirel. Pers. Commun. **90**(4), 2033–2049 (2016)
38. M. Jana, B. Jana, A new DCT based robust image watermarking scheme using cellular automata. Inf. Secur. J. A Glob. Perspect. pp. 1–17 (2021)
39. A.M. Joshi, S. Gupta, M. Girdhar, P. Agarwal, R. Sarker. Combined DWT–DCT-based video watermarking algorithm using arnold transform technique, in *Proceedings of the International Conference on Data Engineering and Communication Technology*, vol. 1 (Springer, 2017), pp. 455–463
40. M. Kaczyński, Z. Piotrowski, High-quality video watermarking based on deep neural networks and adjustable subsquares properties algorithm. Sensors **22**(14), 5376 (2022)
41. H. Kandi, D. Mishra, S.R.K.S. Gorthi, Exploring the learning capabilities of convolutional neural networks for robust image watermarking. Comput. Secur. **65**, 247–268 (2017)
42. H. Kaur, V. Kaur, Invisible video multiple watermarking using optimized techniques, in *2016 Online International Conference on Green Engineering and Technologies (IC-GET)* (IEEE, 2016), pp. 1–9
43. Z. Ke, H. Huang, Y. Liang, Y. Ding, X. Cheng, Q. Wu, Robust video watermarking based on deep neural network and curriculum learning, in *2022 IEEE International Conference on e-Business Engineering (ICEBE)* (IEEE, 2022), pp. 80–85
44. A. Kerbiche, S. Ben Jabra, E. Zagrouba, V. Charvillat. Robust video watermarking approach based on crowdsourcing and hybrid insertion, in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)* (IEEE, 2017), pp. 1–8
45. A. Kerbiche, S. Ben Jabra, E. Zagrouba, V. Charvillat, A robust video watermarking based on feature regions and crowdsourcing. Multimed Tools Appl **77**(20), 26769–26791 (2018)
46. S. Kumar, B.K. Singh, Dwt based color image watermarking using maximum entropy. Multimed Tools Appl **80**(10), 15487–15510 (2021)
47. J.-E. Lee, Y.-H. Seo, D.-W. Kim, Convolutional neural network-based digital image watermarking adaptive to the resolution of image and watermark. Appl Sci **10**(19), 6854 (2020)
48. X. Li, X. Wang, W. Yang, X. Wang, A robust video watermarking scheme to scalable recompression and transcoding, in *2016 6th International Conference on Electronics Information and Emergency Communication (ICEIEC)* (IEEE, 2016), pp. 257–260
49. H. Li, X. Guo, Embedding and extracting digital watermark based on DCT algorithm. J. Comput. Commun. **6**(11), 287–298 (2018)
50. Y. Li, Z. Hao, H. Lei, Survey of convolutional neural network. J. Comput. Appl. **36**(9), 2508 (2016)
51. Y. Li, H. Wang, M. Barni, A survey of deep neural network watermarking techniques. Neurocomputing **461**, 171–193 (2021)

52. Y. Liu, M. Guo, J. Zhang, Y. Zhu, X. Xie, A novel two-stage separable deep learning framework for practical blind watermarking, in *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 1509–1517 (2019)
53. R. Liu, An improved logistic chaotic map and self-adaptive model for image encryption. J. Comput. Methods Sci. Eng. **16**(2), 287–301 (2016)
54. S. Liu, Z. Pan, H. Song, Digital image watermarking method based on DCT and fractal encoding. IET Image Proc. **11**(10), 815–821 (2017)
55. Y. Liu, S. Liu, H. Zhao, S. Liu, A new data hiding method for H. 265/HEVC video streams without intra-frame distortion drift. Multimed. Tools Appl. **78**(6), 6459–6486 (2019)
56. X. Luo, Y. Li, H. Chang, C. Liu, P. Milanfar, F. Yang. DVMark: a deep multiscale framework for video watermarking. arXiv preprint arXiv:2104.12734 (2021)
57. X. Luo, R. Zhan, H. Chang, F. Yang, P. Milanfar, Distortion agnostic deep watermarking, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13548–13557 (2020)
58. G.R. Manjula, A. Danti, A novel hash based least significant bit (2-3-3) image steganography in spatial domain. arXiv preprint arXiv:1503.03674 (2015)
59. S. Mansour, S. Ben Jabra, E. Zagrouba, A robust deep learning-based video watermarking using mosaic generation, in *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 5: VISAPP* (NSTICC, SciTePress, 2023), pp. 668–675
60. M. Masoumi, M. Rezaei, A. Ben Hamza, A blind spatio-temporal data hiding for video ownership verification in frequency domain. AEU-Int. J. Electron. Commun. **69**(12), 1868–1879 (2015)
61. K. Meenakshi, K. Swaraja, P. Kora. A robust DCT-SVD based video watermarking using zigzag scanning, in *Soft Computing and Signal Processing* (Springer, 2019), pp. 477–485
62. G.N. Mohammed, A. Yasin, A.M. Zeki. Robust image watermarking based on dual intermediate significant bit (DISB), in *2014 6th International Conference on Computer Science and Information Technology (CSIT)* (IEEE, 2014), pp. 18–22
63. S.-M. Mun, S.-H. Nam, H. Jang, D. Kim, H.-K. Lee, Finding robust domain from attacks: a learning framework for blind watermarking. Neurocomputing **337**, 191–202 (2019)
64. J. Ouyang, G. Coatrieux, H. Shu, Robust hashing for image authentication using quaternion discrete Fourier transform and log-polar transform. Digit. Signal Proc. **41**, 98–109 (2015)
65. A.S. Patil, G. Sundari, Deep learning-based wavelet embedding for covert audio object embedding in video object steganography. Ann. For. Res. **66**(1), 849–869 (2023)
66. M. Plata, P. Syga, Robust spatial-spread deep neural image watermarking, in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (IEEE, 2020), pp. 62–70
67. M. Plata, P. Syga, Robust watermarking with double detector-discriminator approach. arXiv preprint arXiv:2006.03921 (2020)
68. M. Rai, S. Goyal, M. Pawar. An optimized deep fusion convolutional neural network-based digital color image watermarking scheme for copyright protection. Circuits Syst. Signal Process. pp. 1–32 (2023)
69. R.S.R. Rao Channapragada, M.V.N.K. Prasad, Digital watermarking based on magic square and ridgelet transform techniques, in *Intelligent Computing, Networking, and Informatics*, vol. 243 (Springer, 2014), pp. 143–161
70. S.S. Raut, A.R. Mune. A review paper on digital watermarking techniques. Int. J. Eng. Sci. **1**, 10460 (2017)
71. B. Ray, S. Mukhopadhyay, S. Hossain, S.K. Ghosal, R. Sarkar, Image steganography using deep learning based edge detection. Multimed. Tools Appl. **80**(24), 33475–33503 (2021)
72. M.N. Sakib, S.D. Gupta, S.N. Biswas, A robust DWT-based compressed domain video watermarking technique. Int. J. Image Gr. **20**(01), 2050004 (2020)
73. J. Sang, Q. Liu, C.-L. Song, Robust video watermarking using a hybrid DCT-DWT approach. J. Electron. Sci. Technol. **18**(2), 100052 (2020)
74. M. Saqib, S. Naaz, Spatial and frequency domain digital image watermarking techniques for copyright protection. Int. J. Eng. Sci. Technol. (IJEST) **9**(6), 691–699 (2017)
75. Satendra Pal Singh and Gaurav Bhatnagar, A new robust watermarking system in integer DCT domain. J. Vis. Commun. Image Represent. **53**, 86–101 (2018)

76. D.G. Savakar, A. Ghuli, Robust invisible digital image watermarking using hybrid scheme. Arab. J. Sci. Eng. **44**(4), 3995–4008 (2019)
77. M. Shanmugam, A. Chokkalingam, Performance analysis of 2 level DWT-SVD based non blind and blind video watermarking using range conversion method. Microsyst. Technol. **24**(12), 4757–4765 (2018)
78. H.K. Singh, A.K. Singh, Digital image watermarking using deep learning. Multimed. Tools Appl. **83**, 2979–2994 (2023)
79. K. Singh et al., A robust rotation resilient video watermarking scheme based on the sift. Multimed. Tools Appl. **77**(13), 16419–16444 (2018)
80. Soumitra Roy and Arup Kumar Pal, A blind DCT based color watermarking algorithm for embedding multiple watermarks. AEU-Int. J. Electron. Commun. **72**, 149–161 (2017)
81. D. Vaishnavi, T.S. Subashini, Robust and invisible image watermarking in RGB color space using SVD. Procedia Comput. Sci. **46**, 1770–1777 (2015)
82. P.S. Venugopala, H. Sarojadevi, N.N. Chiplunkar, V. Bhat, Video watermarking by adjusting the pixel values and using scene change detection, in *2014 Fifth International Conference on Signal and Image Processing* (IEEE, 2014), pp. 259–264
83. V.P. Vishwakarma, V. Sisaudia, Gray-scale image watermarking based on DE-KELM in DCT domain. Procedia Comput. Sci. **132**, 1012–1020 (2018)
84. W. Wan, J. Wang, Y. Zhang, J. Li, Yu. Hui, J. Sun, A comprehensive survey on robust image watermarking. Neurocomputing **488**, 226–247 (2022)
85. J. Wang, D. Zhiguo, A method of processing color image watermarking based on the Haar wavelet. J. Vis. Commun. Image Represent. **64**, 102627 (2019)
86. K. Wang, L. Li, T. Luo, C.-C. Chang, Deep neural network watermarking based on texture analysis, in *Artificial Intelligence and Security*. ed. by X. Sun, J. Wang, E. Bertino (Springer, Singapore, 2020), pp.558–569
87. B. Wen, S. Aydore, ROMark: a robust watermarking system using adversarial training. arXiv preprint arXiv:1910.01221 (2019)
88. Yu. Xiaoyan, C. Wang, X. Zhou, A survey on robust video watermarking algorithms for copyright protection. Appl. Sci. **8**(10), 1891 (2018)
89. L. Yang, H. Wang, Y. Zhang, J. Li, P. He, S. Meng, A robust DCT-based video watermarking scheme against recompression and synchronization attacks, in *International Workshop on Digital Watermarking* (Springer, 2021), pp. 149–162
90. Y. Yang, Z. Li, W. Xie, Z. Zhang, High capacity and multilevel information hiding algorithm based on pu partition modes for HEVC videos. Multimed. Tools Appl. **78**(7), 8423–8446 (2019)
91. G. Ye, J. Gao, Y. Wang, L. Song, X. Wei, ItoV: efficiently adapting deep learning-based image watermarking to video watermarking. arXiv preprint arXiv:2305.02781 (2023)
92. I. Yoo, H. Chang, X. Luo, O. Stava, C. Liu, P. Milanfar, F. Yang, 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*, pp. 10031–10040 (2022)
93. I. Yoo, H. Chang, X. Luo, O. Stava, C. Liu, P. Milanfar, F. Yang., 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*, pp. 10031–10040 (2022)
94. C. Yu, Attention based data hiding with generative adversarial networks, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 1120–1128 (2020)
95. C. Zhang, C. Lin, P. Benz, K. Chen, W. Zhang, I.S. Kweon, A brief survey on deep learning based data hiding, steganography and watermarking. arXiv preprint arXiv:2103.01607 (2021)
96. H. Zhang, H. Wang, Y. Cao, C. Shen, Y. Li, Robust data hiding using inverse gradient attention. arXiv preprint arXiv:2011.10850 (2020)
97. K.A. Zhang, L. Xu, A. Cuesta-Infante, K. Veeramachaneni, Robust invisible video watermarking with attention. arXiv preprint arXiv:1909.01285 (2019)
98. W. Zhang, X. Li, Y. Zhang, R. Zhang, L. Zheng, Robust video watermarking algorithm for H. 264/AVC based on JND model. KSII Trans. Internet Inf. Syst. (TIIS) **11**(5), 2741–2761 (2017)
99. X. Zhong, P.-C. Huang, S. Mastorakis, F.Y. Shih, An automated and robust image watermarking scheme based on deep neural networks. IEEE Trans. Multimed. **23**, 1951–1961 (2020)
100. J. Zhu, R. Kaplan, J. Johnson, L. Fei-Fei, Hidden: hiding data with deep networks, in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 657–672 (2018)

101. J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232 (2017)