REVIEW

Comprehensive review of watermarking techniques in deep-learning environments

Himanshu Kumar Singh and Amit Kumar Singh*
National Institute of Technology Patna, Department of CSE, Patna, Bihar, India

Abstract. Recently, the demand for the generation, sharing, and storage of massive amounts of multimedia information—especially in the form of images—from different intelligent devices and sensors has increased drastically. This introduces issues including the illegal access and fraudulent usage of this information as well as other security concerns. Watermarking consists of embedding a watermark design in a digital cover and then later extracting it to provide a solution for ownership conflict and copyright violation issues involving the media data. Presently, in watermarking, the use of deep-learning approaches is incredibly beneficial due to their accuracy, superior results and strong learning ability. We present a comprehensive review of watermarking techniques in deep-learning environments. We start with basic concepts of traditional and learning-based digital watermarking; we later review the popular deep-learning model-based digital watermarking methods; then, we summarize and compare the most recent contribution in the literature; finally, we highlight obfuscation challenges and further research directions. © *2022 SPIE and IS&T* [DOI: 10.1117/1.JEI.32.3.031804]

Keywords: digital image watermarking; data hiding; deep learning; convolutional neural network; generative adversarial network; deep neural network.

Paper 220885SS received Aug. 28, 2022; accepted for publication Nov. 8, 2022; published online Nov. 29, 2022.

1 Introduction

In recent years, deep learning has made unprecedented progress in a wide range of image processing tasks, such as classification, segmentation, super-resolution, deblurring, and denoising.[1] It learns the data representation hierarchically from raw images by sidestepping manual feature engineering. Today, digital images are increasingly used in various applications, including healthcare, communications, forensics, education, research and development departments, etc. However, many images involve individuals' sensitive information and even organizational confidentiality; therefore, they should not be visited or viewed by unauthorized persons. Researchers have developed the watermarking scheme to deal with the security and copyright violation issues of digital data. This technique enables us to send and receive personal data from the sender to the receiver via smart devices and unsecured open channels without noticeable distortion of the host data.[2] A few well-known applications[3] of watermarking are shown in Fig. 1.

Apart from the applications mentioned here, watermarking techniques are used for a portion (as a percentage) of special applications, as shown in Fig. 2.

The primary objective of the watermarking technique is to enhance three requirements:[3] imperceptibility, capacity, and robustness. While performing watermarking, the original signal should not be visibly distorted after concealing the hidden data.[1,4] Techniques for watermarking are believed to work with other media and in additional applications; a watermarking trade-off triangle, which is shown in Fig. 3, details the requirements that any watermarking technique must meet.

The capacity requirement is identifiable by how much information (in number of bits) is conveyed by the host mark. This is contrary to the other two requirements: (i) imperceptibility, which refers to the visual quality of the watermark, and (ii) robustness, which is the capability of preserving the mark even when the carrier media experiences certain distortions.

*Address all correspondence to Amit Kumar Singh, amit.singh@nitp.ac.in

Fig. 1 Recent applications of watermarking.

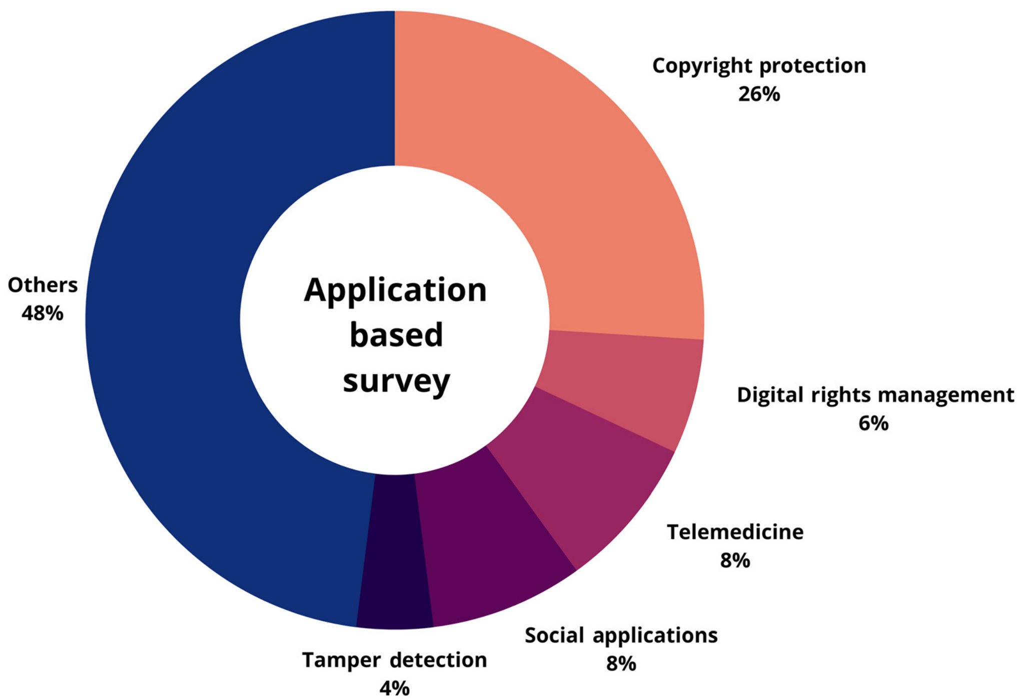

Fig. 2 Application-based survey.

The classical watermarking technique embeds a secret key or authentication code into the image before sending it through a public channel. Verifying the embedded private key or authentication code proves the image's authenticity. The general architecture of the watermarking technique is shown in Fig. 4. It primarily contains two processes: embedding and extraction.[5]

In the embedding process, the secret watermark, cover media, and secret key are given as input to the embedding algorithm, which generates the watermarked by hiding the watermark in the cover media. The embedding algorithm uses either spatial or transform domain techniques. Encryption, encoding, and hashing[6–8] can also be used to improve the security of the watermarking scheme. Generally, the extraction process is the inverse of the embedding process. Let M be

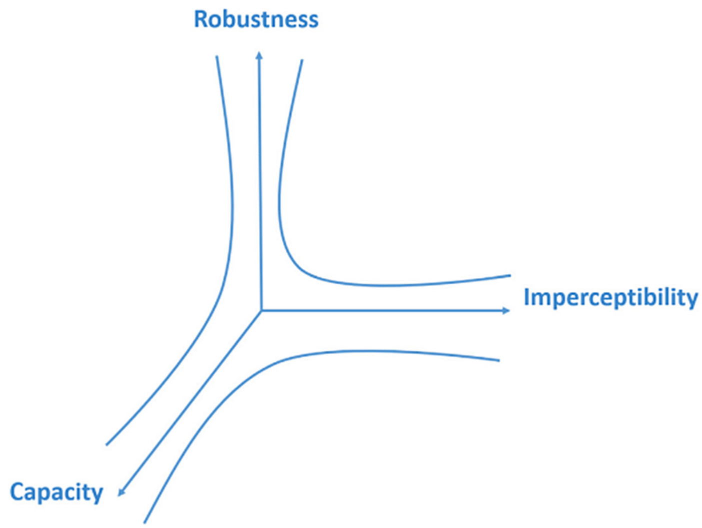

Fig. 3 Watermarking trade-off triangle.

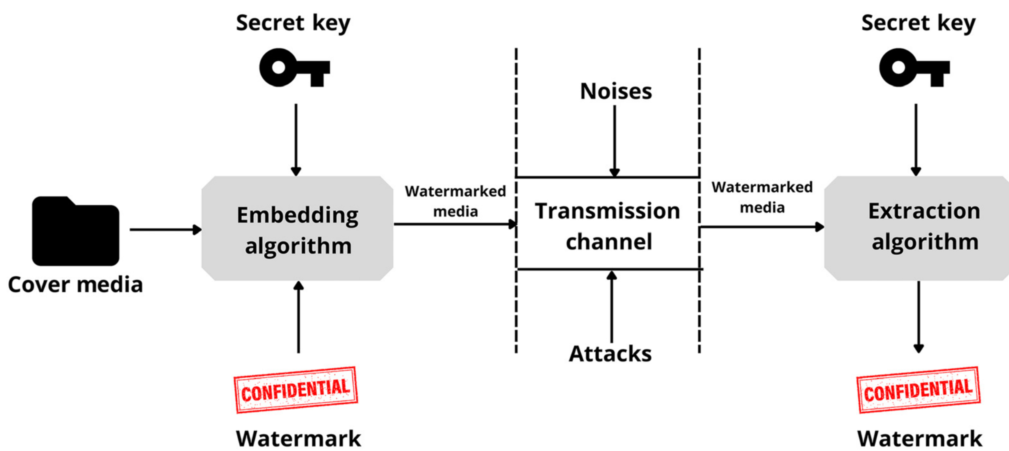

Fig. 4 General architecture of watermarking.

the cover media, W be the watermark, S be the secret key, and Embed () be the embedding function. The watermarked media M' is mathematically obtained by Eq. (1). The three approaches—image-based, linguistic-based, and structure-based—can be used to embed the watermark.[9]

$$Embed(M, W) = M'. \tag{1}$$

When transferred over a public, unprotected network, watermarked media M' is vulnerable to distortion or manipulation. Hence, the media could be attacked intentionally (attacker) or unintentionally (noise) during transmission. To provide copyright protection, even after an attack such as alteration, redistribution, JPEG compression, etc., the embedded watermark must be detectable, and the confidential information should be viable for extraction.

Let Extract () be the extraction function and M'' be the received watermarked media; then the watermarked media W' is mathematically obtained by Eq. (2) as

$$\text{Extract}(M'') = W'. \tag{2}$$

In addition to the classical watermarking schemes, deep-learning-based methods have been widely studied in recent years, achieving an outstanding performance compared with classical methods.[1,10,11] The key advantages to consider when selecting deep learning for watermarking are[1,12] (a) watermark generation for robust watermarking, (b) finding the ideal embedding position in the cover media, (c) identifying the best embedding strength that efficiently offers a balanced trade-off between quality and robustness, (d) offering attack simulation for efficient watermark extraction, and (e) reducing errors and denoising for obtained watermarks. However, the deep-learning model also faces the challenges of model security and privacy.[13] Presently, in deep learning, watermarking approaches are incredibly beneficial for the issues of copyright violation and ownership conflicts of deep-learning models and devices.[1] Here, watermarking techniques embed useful information into deep-learning models and devices and play an essential role in ownership verification.

In the last few years, various papers containing the survey on digital media protection, devices, and artificial model protection using watermarking techniques have been published.[1,10,14,15] In Ref. 1, the authors summarized the roles and usage of deep-learning models in the different phases of the watermarking techniques. In Ref. 10, the authors discussed different watermarking methods in the artificial intelligence domain. A detailed study on the protection of deep-learning models using watermarking is summarized in Ref. 14. Authors in Ref. 15 provide surveys on intellectual property protection using deep learning. In contrast, the primary goal of our work is to provide a thorough analysis of the key aspects of deep-learning models widely used in watermarking for ownership, copyright protection and model protection. Compared with the cited articles, we provide a more thorough discussion on the usage and role of some well-known deep-learning models for watermarking. Table 1 compares the recent existing surveys in the article, including ours, based on parameters that include the deep-learning model-based study, role and usage of deep-learning models, the study of popular models for the data hiding, and the comparison of different state-of-the-art techniques for watermarking in tabular form, pie chart descriptions, and other perspectives.

This article provides a thorough analysis of watermarking methods based on well-known deep-learning models. The article's contributions are as follows:

1. First, we discuss the basic concepts of classical watermarking, recent applications, requirements, and how deep-learning techniques helps in watermarking.
2. Then, we discuss the advantages and roles of using deep learning in watermarking.
3. Next, we review the most popular deep-learning models for digital watermarking techniques and their detailed usage and merits.
4. Finally, we summarize and compare the most recent contribution in the literature, and we highlight the significant challenges with using deep-learning models for watermarking, followed by further research directions.

The organization of this paper is shown in Fig. 5. Section 2 details the deep-learning-based watermarking and their role and usage as well as different popular deep-learning models. Section 3 reviews the most recent works in the deep-learning-based watermarking domain based on the popular model used in the deep-learning field. In Sec. 4, based on the survey, we mention some of the identified issues with deep-learning-based watermarking. Finally, in Sec. 5, we conclude the paper.

2 Learning-based Watermarking

Deep-learning-based frameworks automatically learn from training data to represent the hierarchical data without requiring feature representations.[16] Learning models based on deep-learning methods do not require manual feature representation. To be specific, a deep network takes the content to be processed in a raw format (an image or an audio signal) as input and maps it. Recently, they have been widely used in data hiding and image processing because of their remarkable potential to mimic human brain learning capacities and interact more naturally.[17] Based on the survey (Fig. 6), deep-learning models such as deep neural network (DNN), recurrent neural network (RNN), convolutional neural network (CNN), and generative adversarial network (GAN) are widely used in watermarking techniques. These results show that CNN is more popular than other deep-learning models. A detailed comparison of the widely used deep-learning models is given in Table 2. In the past decades, the role of deep learning in

Table 1 Comparison of recent existing surveys with our survey.

Survey/ Year	Deep-learning-based study	Role of deep-learning for data hiding	Popular model-based study	Requirements for deep models for watermarking discussed	Major issues discussed	Tabular comparison	Pie chart description	Other perspectives
[1]/2022	✓	✓	×	×	✓	✓	✓	Role and usage of deep learning in data hiding techniques
[10]/2022	✓	✓	×	×	✓	✓	✓	Study of watermarking in the artificial intelligence domain
[14]/2021	✓	×	×	×	✓	✓	×	Study of different watermarking DL models
[15]/2021	✓	✓	×	×	×	×	×	Study of different attack simulations and datasets.
This article	✓	✓	✓	✓	✓	✓	✓	Detailed explanation of different deep learning models for watermarking and model protection.

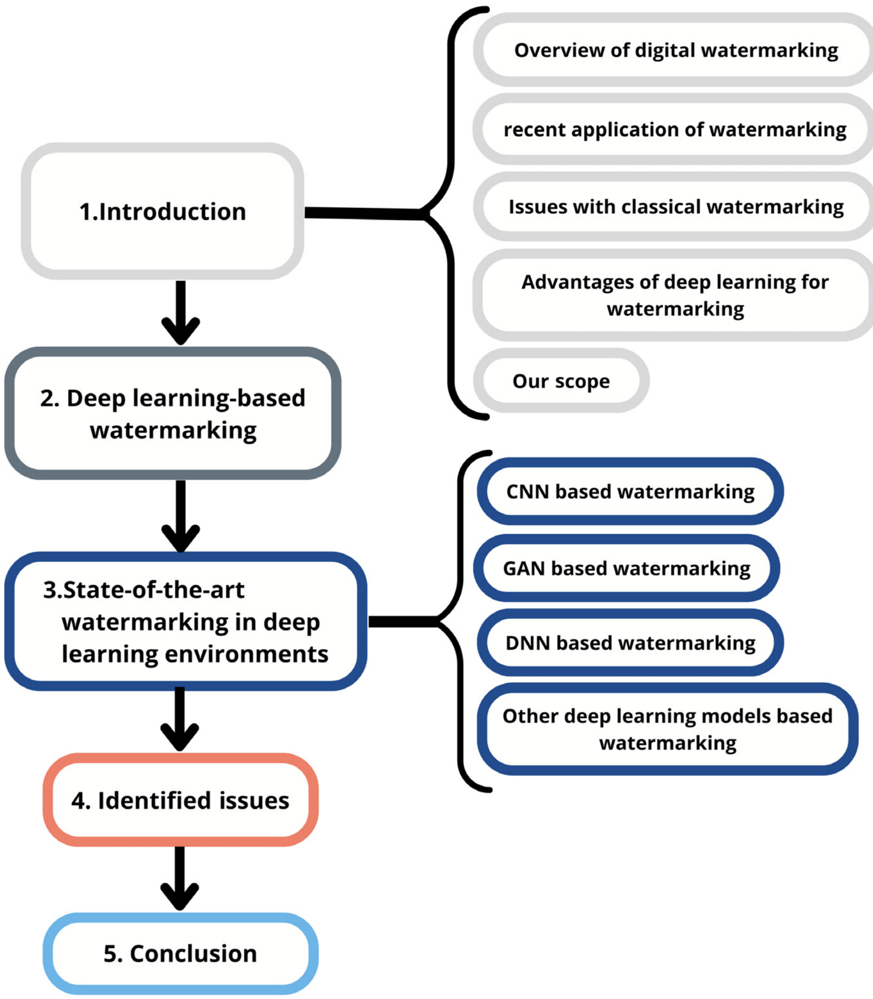

Fig. 5 Organization of the paper.

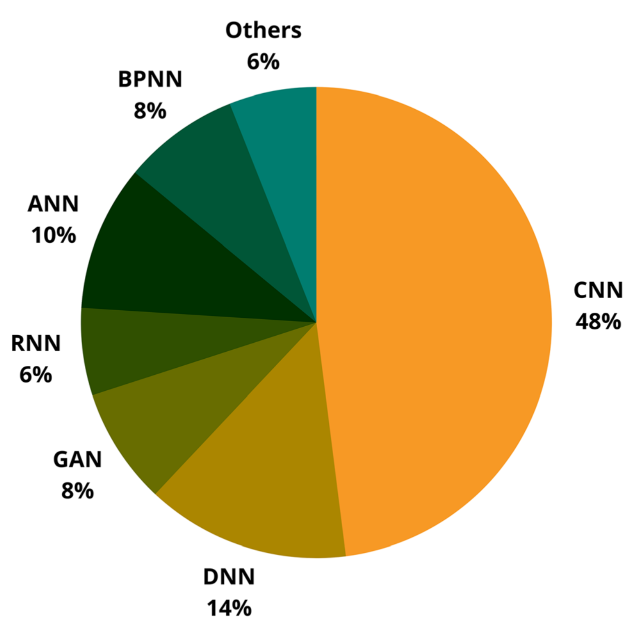

Fig. 6 Popular deep-learning model used for watermarking.

Table 2 Summary of popular deep learning neural networks.

Important aspects	CNN	GAN	DNN	RNN
Neural network family	Feed forward	Feed forward	Feed forward	Recurrent
Layers	Five layers are input, convolution, pooling, fully connected, and output.	Two networks: generator and discriminator	Three layers are input, hidden, and output.	Time series network as input layers uses the output of each layer in the successive epochs.
Data type accepted	Non-sequential	Non-sequential	Non-sequential	Sequential and time series
Recurrent connections	No	No	No	Yes
Input length	Fixed	Fixed	Fixed	Variable
Idle use for	Image, video, and text	Image, video, and text generation	Image and video	Text and speech
Gradient vanishing and exploding problem	Yes	Yes	Yes	Yes
Applications	Image classification, face recognition, image segmentation, NLP, and text analysis	Image generation and classification	Image classification, face recognition, and speech recognition	NLP, speech-to-text and text-to-speech translation, and time-series-based prediction
Architecture	—	—	—	—

watermarking has grown significantly. These networks have been used for many purposes, such as watermark generation, identification of the embedding location, determining an appropriate embedding strength, and watermark embedding and extraction.

The role and usage of deep learning are shown in Fig. 7. Deep-learning models are used for the early stages of watermark generation, pre-processing, and watermark embedding. At later stages, it can be used for watermark extraction and attack simulation.

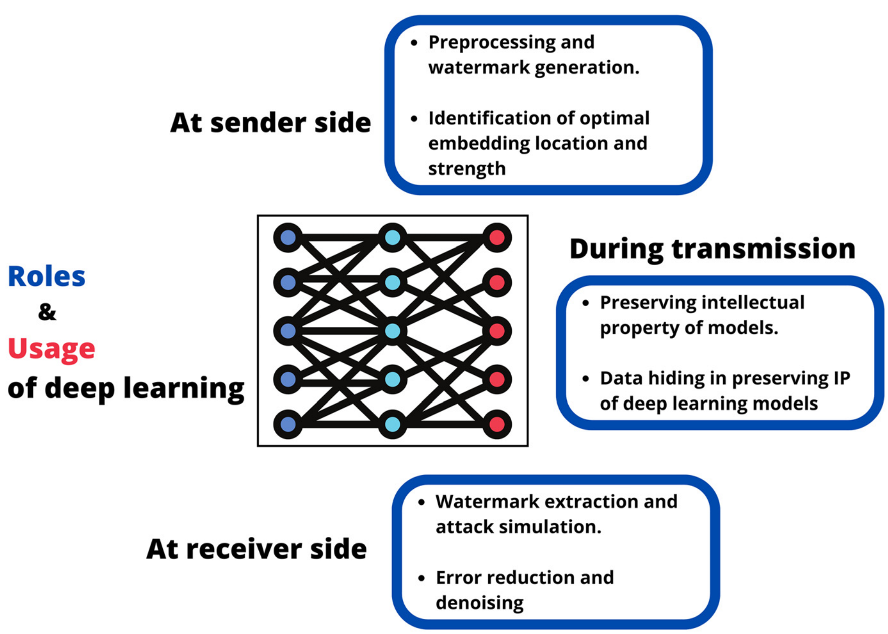

Fig. 7 Roles and usage of deep learning at different phases of watermarking.

3 State-of-the-Art Watermarking in Deep-Learning Environments

Recently, deep-learning models, such as CNN, GAN, and DNN have been widely used in the field of watermarking. In this section, we discuss these deep-learning models in detail and review recent papers based on them.

3.1 *CNN-based Watermarking*

CNN is a deep-learning network that has convolution characteristics. It generally contains five layers: the input layer, the convolution + rectified linear unit (ReLU) layer, the pooling layer, the fully connected layer, and the output layer. CNN is inspired by the visual cortex of the human brain. A neuron is only activated in a receptive field in response to a stimulus. The CNN model takes an input (image) as a vector, and later, the convolution operation is applied to extract the essential features (kernels or filters are used for this). The ReLU activation function is used to make the negative values zero, and later, pooling is applied to reduce the number of parameters. The reduced pooled information is passed to the fully connected network by flattening the input for classification purposes. The CNN model is widely used in image watermarking and data hiding due to its lesser complexity. Though it has many applications, CNN is used primarily in watermarking for embedding the watermark, extracting the watermark, increasing the visual quality of the cover image, and identifying the best embedding location for the watermark. Ingaleshwar and Dharwadkar[18] proposed an optimization-based watermarking technique using deep CNN to embed the watermark image into the cover image. Decomposition on the cover image was performed into the different grids to obtain the related features of gridlines. The interesting region was computed using the deep CNN, which was trained using the optimization techniques. Fitness measures are used for embedding the watermark, and for that, wavelet sub-bands are selected. For recovering the watermark, cover media and fitness function are required. Compared with the methods presented in Refs. 19–23, the suggested method performs better in peak signal-to-noise ratio (PSNR). Even though the technique is robust, there has been no detailed investigation of overall embedding and recovery costs. Also, further study is needed on the robustness of the approach by performing more attacks at varying noise levels.

A blind image watermarking scheme was suggested by Bagheri et al.[19] in which Mask R-CNN was used to compute the embedding strength. The watermark was embedded in the cover by computing the block into the selected discrete cosine transform (DCT) and discrete wavelet transform (DWT) blocks using lower region-of-interest pixels. Even though the watermark is extracted well, the method provides limited embedding capacity and watermark security. A wavelet-based watermarking for digital media was developed by Zheng et al.[24] to investigate the imperceptibility and robustness of the watermark. Initially, the cover media was transformed into different bands using DWT, and then the watermark data singular value was inserted into the high bands of the cover media. Then, the transformation was applied to the low bands by wavelet transformation, in which the watermark sequence was embedded into the selected low bands. The singular vector of the watermark and scrambled sequence were used to obtain the singular and watermark sequence, respectively. Later, using the CNN established the connection between the watermark, cover media, and watermarked media for robust extraction of the watermark. Due to their low embedding cost for different types of watermarks, they extend the suitability for various practical applications. In Ref. 25, a CNN based watermarking technique was developed to embed the scrambled watermark into the DWT cover image. The technique uses a fast region-based CNN model to achieve robust and blind extraction of the embedded watermark. Compared with conventional techniques,[26–29] this method provided better classification accuracy with strong invisibility and lower execution time. However, proper investigation is needed into the robustness of performance against different image processing attacks. Plata and Syga[30] proposed a watermarking technique based on CNN to embed the spatial watermark data into the cover image. A loss function was designed by the authors for the neural network training to improve the robustness and other watermarking trade-offs. Compared with other existing work,[31–33] the scheme provides higher robustness with minimal distortion. The schemes need to be evaluated for different image processing attacks.

A deep CNN network-based blind watermarking technique for copyright protection was proposed by Mun et al.[34] In this technique, first the carrier and watermark media are divided into non-overlapping blocks. Later, the secret is computed using watermark data corresponding to the positions of the block. Subsequently, the watermark is embedded into the carrier using the deep-CNN model, and finally, the model is used to extract the watermark from the cover. The scheme is robust to different salt and pepper noise attacks with good imperceptibility. However, the cost of the method is relatively high. The encryption-compression-based watermarking technique was proposed in Ref. 35. The method uses transform domain watermarking using the lifting wavelet transform (LWT), randomized singular value decomposition (RSVD) and Heisenberg decomposition (HD). Further, a CNN-based denoising network was used to improve the extracted watermark. This technique outperformed the existing methods proposed in Refs. 36–39 with an improvement of 27.84% examined on different attacks.

A transform domain watermarking method was proposed by Hsu and Hu[40] based on quaternion discrete cosine transform (QDCT). The tradeoff between robustness and imperceptibility of the watermarked image was balanced by the grey wolf optimizer. The blind watermark extraction was used, and the denoising CNN model was used to improve the visual quality of the watermark. Based on the experimental evaluation, the watermarking technique performed better than the traditional techniques presented in Refs. 41–44. Another transform domain-based watermarking scheme was proposed by Kandi et al.[45] in using the CNN network to improve the tradeoffs of the watermark. The technique uses the auto-encoder learning capability of a CNN to improve its robustness. It uses the input–output information for embedding using the CNN weights, which are highly prone to attacks. To protect the ownership of the deep CNN network, a watermark-based scheme was proposed by Nagai et al.,[46] which uses the concepts of the technique proposed in Ref. 47. In their technique, the secret is inserted into the different groups of the convolution layer of the original network. The technique is adequate, but the technique is limited to dealing with attacks such as surrogate model attacks and watermark overwriting. Another technique for protection of the ownership of a CNN network was proposed by Guan et al.[48] using the watermarking techniques. The technique uses the model compression's pruning theory to embed the hash using the secure hash algorithm (SHA-256) as watermark data into the convolution layer of the residual neural network (ResNet) 152 network. However, the technique limits the usability of the real-time applications due to their high computational costs.

Some of the recent works using the CNN model are mentioned below are also given in Table 3.

The challenges of CNN models for watermarking systems are as follows:

1. Operations such as max-pool make the CNN models significantly slower.
2. Sometimes the training process takes more time due to the misconfiguration of the network parameters.
3. A CNN model requires a larger dataset for training and processing.
4. Due to its complex nature, sometimes a CNN network runs into problems, such as overfitting or underfitting.

3.2 *Generative Adversarial Network-based Watermarking*

This is a deep-learning-based model with generative properties. It has two networks: the generator network and the discriminator network. The GAN network is widely used in watermarking for watermark verification and generation due to its generative and discriminative ability. GAN is also used to enhance network security by generating a unique watermark for every input. The generator network G takes random noise input Z from distribution $P(Z)$ and generates sample S. The goal is to replicate a particular type of distribution $P(X)$. Later, the discriminator network D discriminates between the generated (fake) samples and the actual sample. During training, the discriminator penalizes itself for misclassifying a real instance as fake or a fake instance as real. The combined generator and discriminator loss function (L) used by the GAN model is obtained by Eq. (3) as

$$L = \min_{G} \max_{D} [\log(D(x)) + \log(1 - D(G(z)))]. \tag{3}$$

Table 3 Summary of CNN-based watermarking.

Ref No.	Objective	Goal	Strategies	Role of CNN	Embedding location	Results	Cover \ mark size	Noticed weakness	Applications oriented
18	Medical images privacy preserving	To Improve visual quality and robustness	WWO, CFOA, and DWT	Optimal embedding region selection	LL and HH coefficient of carrier	45.2157 dB PSNR, NC =1, BER= 0	—	Need to analyze robustness for more attacks.	Medical applications
19	Blind watermarking technique for color images	To obtain acceptable visual quality and high robustness	DWT, DCT	Calculation of embedding strength	8×8 DCT block	PSNR = 49.1052 dB PSNR, SSIM of 0.9985, and NC of 1	512×512 / 4×4	Limited embedding capacity, insufficient security, and complexity	Tested for COCO dataset
24	Secure watermarking technique for digital images	High robustness and high embedding capacity	DWT and SVD	Relation between the cover and watermarked image	Singular matrix	38.5659 dB PSNR with NC of 0.9608	512×512 / 512×512	Time complexity is missing	Gray scale and general images
25	Watermarking with high classification accuracy	Improved robustness with high security and invisibility	DCT, DWT	Detection and adaptive recovery of the watermark	Local neighbors feature points	PSNR = 50.12 dB, Accuracy = 93.75%	—	Analyze robustness against different attacks	Smart city applications
30	Blind watermarking technique	High robustness and capacity	Spatial spreading	Attack simulation to provide high robustness	16×16 blocks	37.81 dB PSNR	256×256 / Text length =3 2	Robustness evaluation can be done.	COCO dataset
34	CNN based watermarking	High visual quality with resistance against different attacks	SGD	Embedding, attack simulation, and extraction of mark	8×8 block	35.9 dB PSNR with, NC of 1	512×512	Higher computational complexity	Grayscale images
35	Encryption-Compression based watermarking scheme	Improve watermark robustness and security	RSVD, Chaotic encryption, LWT, HD, DnCNN SPIHT compression,	Enhance the quality of the recovered watermark image	Singular matrix of the host image	37.6175 dB PSNR, with SSIM of 0.99, and NC of 1	512×512 / 256×256 or 128×128 or 64×64	Limited embedding capacity and less robust for histogram equalization	General images

Table 3 (*Continued*).

Ref No.	Objective	Goal	Strategies	Role of CNN	Embedding location	Results	Cover \ mark size	Noticed weakness	Applications oriented
40	Optimization-based robust watermarking	High resistance with minimum distortion	QDCT, GWO, DnCNN	Embedding and extraction of watermark	8×8 QDCT block	38.1 dB PSNR with MSSIM of 0.947, BER of 0%, and NC of 1	512×512 / 64×64 or 192×64 or 128×128	High computational complexity	CVG-UGR Image Database
45	Watermarking using the encoding functionality of CNN	Robust and secure watermarking	Codebook, a random permutation	Codebook generation for robust embedding and extraction	Cover image	58. 91 dB PSNR with MSE of 0.08 36, BER of 0, NC of 1, and SSIM of 0.998	128×12 8/ 64×64	Robustness for JPEG compression and average filtering can be improved	General Grayscale images
46	Watermarking in a convolutional layer of DCNN	To secure the intellectual properties of DCNN		To protect owner authorization of DCNN	Convolutional layers of the DCNN model	Test Error of 7.69 and BER of 0,	32×32, 300 $\times 200$/ 256-bit	The security of the DNN can be improved	CIFAR-10, Caltech-101
48	CNN model verification using watermarking	To provide high capacity with high robustness	Histogram shifting, SHA256	Integrity authentication of DCNN	Convolutional layers of the DCNN model	Accuracy = 85.9%	—	High computational cost	ImageNet

Wu et al.[49] proposed a deep-learning model to verify the corrupted images caused by dense watermarks. The model consists of a generator and a discriminator for improving the recovered images' quality and verification performance. The generator is an autoencoder that maps densely watermarked, corrupted images to a representation vector and decodes the vector to an red-green-blue (RGB) image. The discriminator controls the contents of the generated images and minimizes the feature loss. The ResNet-46 model is used to extract the recovered images feature and ground truth images. It achieves a verification accuracy of 96.36% at the false positive rate (FPR) of 1%. In Ref. 50, the authors proposed a robust data hiding scheme using the GAN for securing genuine documents. At first, using the geometric correction, the document is adjusted to the required form. After that, the document is generated by the adversarial network, and its security is enhanced by embedding the secret information into the document using the pseudo-random number to generate a watermarked document. The technique is robust and performs better than the work mentioned in Refs. 51–53. A robust watermarking system based on variational autoencoder networks is provided by Wei et al.[54] for copyright protection. Encoder, decoder and detector subnetworks make up the embedder and extractor network. A 1-bit watermark image is embedded in the host image during training, and the encoder and decoder subnetworks build a robust representation model of the cover image. The detector subnetwork acquires the ability to extract the 1-bit mark from the watermarked image. The approach improves the visual quality of the marked image, but further research is needed to determine the robustness of watermarked images. A blind watermarking scheme based on deep learning is proposed in Ref. 55. The technique consists of four components: an encoder, a decoder, two identical noise layers, and an adversarial discriminator. The two identical layers were used to embed and extract the watermark encoder and decoder, respectively, and to make the watermark robust against different attacks. Further, an adversarial discriminator was used to improve the robustness and concealment of the watermark. The technique is robust against various attacks with good imperceptibility. However, the model complexity is very high. To prevent malicious attacks during transmission or from illegal use of diffusion-weighted imaging (DWI) images, a multiscale robust watermarking technique was proposed by Fan et al.[56] The suggested technique includes multiscale characteristics and a generative adversarial. By integrating full-scale features, the DWI images are first rebuilt to mimic the original DWI images. Watermark is embedded into the multiscale reconstructed features. To enhance the visual quality of the reconstructed picture, an optimized boundary equilibrium generative adversarial network| discriminator is suggested. Finally, to learn the watermark distribution feature, pyramid filters and multiscale max-pooling are used. Fang et al.[57] introduced a unique triple-phase watermarking scheme to prevent image distortion in practice. A noise-free initial phase, a mask-guided frequency augmentation phase, and an adversarial-training phase comprise the approach. In the first phase, an encoder–decoder was trained end-to-end using a just-noticeable difference (JND) mask image loss. The encoded characteristics are then subjected to a mask-guided frequency augmentation method in the second step. Later in the final phase, it intends to train a decoder for dealing with non-differentiable distortion through adversarial training. The approach is more robust than the studies reported in Refs. 58–61. In Ref. 62, the authors proposed a semi-fragile watermarking scheme based on deep learning for media authentication. The technique consists of three modules: an encoder network, a decoder network, and an adversarial discriminator network. The encoder network encodes the input images and produces the watermarked images. After that, the marked image goes through two image transformation functions: one from a benign transformation set and the other from a malicious transformation set to produce a benign and malicious watermarked image, respectively. Later, the benign and malicious watermarked images are fed to the decoder network using an adversarial network that discriminates between the benign and malicious images. The technique is robust and provides tamper detection, but the technique needs to be investigated for additional types of image processing attacks.

Some works based on the GAN model are mentioned below and are summarized in Table 4. The challenges of the GAN model for watermarking systems include the following:

1. The irregularity between the generator and discriminator network seeds the problem of overfitting.
2. The network never converges due to the network parameters' oscillation and destabilization.

Table 4 Summary of GAN-based watermarking.

Ref No.	Objective	Goal	Strategies	Role of GAN	Embedding location	Results	Cover\mark size	Noticed weakness	Applications oriented
49	Remove dense watermarking	Recover grayscale image	MSE loss	Denoising the image	—	PSNR = 23.37 dB, TPR@FPR=1% = 96.36%	—	PSNR is not up to the mark	Grayscale images
50	Robust data hiding scheme	High robustness	Image denoising	Improve robustness	Cover image	PSNR = 35.82 dB, SSIM = 0.9988	512 × 512	Need a detailed analysis of robustness	DSSE dataset
54	Robust image watermarking using C-GAN	Copyright and ownership protection	Variational autoencoder	Embedding and extraction of watermark	Cover image	PSNR = 34.97 dB, SSIM = 0.979	—	Limited capacity	Color image
55	Blind watermarking scheme	Robustness and high visual quality	MSE loss	Embedding and extraction	Cover image	PSNR = 41.02 dB in V channel	—	High complexity	General image
56	Multiscale robust watermarking	High security		Watermarking and image denoising	Cover image	Accuracy = 0.999	—	Tested on limited attacks	DWI images
57	Watermarking for distortion-free real images	High watermarked image quality	JND-mask-based loss	Image enhancement	Cover image	PSNR – 36.25 dB, Accuracy = 84.9%	—	High complexity	Medical imaging
62	Watermarking for media authentication	Identify deep fakes images	Encoder-decoder network	Extraction and classification	Encoded cover image	PSNR = 36.38 dB	256 × 256	Limited capacity	General color images

3. Sometimes the discriminator gets too successful, and the generator gradient vanishes and learns nothing.
4. Sometimes the generator network falls in, causing limited variations of the samples.

3.3 *DNN-based Watermarking*

The DNN is the most commonly-used network; in it, multiple hidden layers are present between two layers: the input layer and the output layer. DNNs model resembles the human brain; to be specific, a DNN takes the raw format input (image or audio) $x \in R^n$ and maps it to the output layer using the parametric function $p = F_\theta(x)$, where $p \in [1, n]$, which is based on the network architecture and combined parameters of all of the layers in the network. When training the DNN, the goal is to minimize the loss between the predicted and ground truth labels and optimize the network parameters θ. The back propagation algorithm is the widely used approach for training a DNN; in it, the loss gradient of the output layer is back propagated to update the network parameters. Training and adjusting the parameters of the DNN allows the network to learn, encode, and decode the watermark for embedding, extracting the watermark, and network protection. DNNs are widely-used in the field of watermarking for embedding and extracting the watermark.

Hou et al.[63] proposed a watermarking technique based on the enhanced version of multiple histogram modification. The method follows a particular criterion to embed a watermark into multiple histograms. Initially, the cover image decomposes into two different sets of pixels. Later, using the DNN, multiple histograms are produced by the classification method. For embedding the watermark data, the produced histograms' optimal bins are selected from each. However, the visual quality between the cover and watermarked image is excellent. Compared with the existing technique,[64–66] the technique achieves a higher PSNR, but it is necessary to assess the technique's robustness against various attacks. Further, the overall embedding cost needs to be computed. For copyright protection and ownership confirmation of DNN models, Zhang et al. presented a watermarking technique in Ref. 67. The Uchida et al.[46] method has several drawbacks; however, an upgraded threat model is employed to enable the black-box mode verification and application programming interface (API) access. The method shows good accuracy at the cost of small overhead. However, the authors do not conduct a transparent investigation of the overhead and security of the model. Another method is presented by Wu et al.[68] for deep model copyright protection and ownership verification using the watermarking technique. In this method, the output image from the deep-learning model is used to obtain the watermarked image. The embedded watermark can only be extracted by the extraction network for verification. The method's usefulness for real-time applications is constrained by the fact that it is only evaluated for three different types of attacks and there is no transparent study of the computational time of the method. To preserve intellectual property rights and to protect the ownership of the DNN, Deeba et al.[69] proposed a watermarking technique. This technique generates and embeds the watermark data into the neural network. The ownership verification is done by a specific type of input–output pair. The technique performs better, but the performance is judged based on only two attacks, and the authors do not conduct a transparent investigation of the execution time of the technique. In Ref. 70, the authors proposed a watermarking technique based on a DNN for the ownership verification of the multimedia document. The cover image is divided into 8×8 pixels to embed the binary watermark of 1-bit $\{-1, 1\}$. After that, at the μ'th block, the watermark image is embedded in the DCT coefficient. In the same procedure, the inverse function is applied for extraction. The technique produces a similar original and watermarked image with acceptable PSNR.

Some works based on the DNN model are mentioned below and are summarized in Table 5. The challenges of DNN models for watermarking are as follows:

1. They require lots of training data.
2. They are expensive to train due to the requirements of special processing hardware.
3. They need detailed knowledge about deep learning to train and tune the network parameters.
4. They involve the complexity of the hidden layers, creating black box types of situations.

Table 5 Summary of DNN-based watermarking.

Ref No.	Objective	Goal	Strategies	Role of DNN	Embedding location	Results	Cover\ mark size	Noticed weakness	Applications oriented
63	Multiple histograms modification-based watermarking	Better visual quality with high capacity	PEE, Memo based optimization	Produce multiple histograms for embedding	Histogram of the carrier media	59.97 dB PSNR with NC of 1	512×512 / 10000 bits	Complexity and robustness analysis could be done	Grayscale common images
67	Ownership verification using the watermarking	To support verification like white box and black box	Intrinsic learning-based embedding	To verify the ownership of the deep learning model	DNN model	Accuracy on MNISTT = 100% and on CIFAR 99.93%	28×28, 32×32	No precise study of security and overhead	MNIST and CIFAR
68	Protect ownership of the DNN model	To provide high security and robustness	—	Identify the owner of the host network	Color image	36.63 dB PSNR with SSIM of 0.982	$256 \times 256 \times 3$/ 64×64	Limited robustness analysis	Danbooru2019 dataset
69	Spatial domain-based secure watermarking	High efficiency with high security	LSB	Detection of watermark during extraction	LSB of the cover image	—	10×10 bits	No transparent investigation of performance	General images
70	DNN based watermarking	High robustness	—	Extraction of watermark	Coefficient of block	—	8×8 pixels	Possibility of information loss	General images

3.4 *Others Perspectives*

In addition to the above-mentioned work, some other deep-learning models such as RNNs, back-propagation neural networks (BPNN), and artificial neural networks (ANN) are also used for watermarking techniques. Some of these works are mentioned below and are summarized in Table 6.

For the ownership verification and copyright protection of digital color images, Sinhal et al.[71] proposed a low-cost, blind watermarking scheme. The image is first converted into a YCbCr model using a selection of 4×4 blocks of Y-component randomization. Later, using a low-cost ANN model to implant a binary watermark, the chosen component is dissected using integer wavelet transformation. In the LWT domain, Islam et al.[72] suggested a reliable watermarking method utilizing an ANN. The watermark is embedded using the LWT cover image's randomized coefficient, and it is extracted using an ANN model. The selected sub-band coefficient is first randomized using a key after the cover picture is modified using the LWT. Later, using a different key, the randomized coefficient is used to obtain the randomized blocks. The chosen sub-randomized band's block is then used to incorporate the watermark. ANN is utilized for watermark extraction. Overall, the approach is reliable and blind, but the embedding and extraction costs were not transparently investigated in the study, and the watermark's capacity is insufficient for practical use. The edge detection-based watermarking approach was suggested by Kazemi et al.,[73] and it involves embedding private information in the edge of the color picture. An edge detector is utilized to identify the edge of the color RGB cover media, and the resulting media is subjected to a contourlet transform to calculate the directional components. Using a genetic technique, the logo image is first scrambled to provide better verification before being embedded into the cover image. The hidden watermark is extracted via a combination of differential evolution and multilayer perceptron. The approach, however, exhibits weak resilience to the few attacks. A deep-learning-based watermarking scheme was proposed by Singh et al.[74] in which multiple watermarks are embedded into the DWT-DCT domain single value of the cover image. The three-level DWT is used to first break down the cover image. The low-frequency band (LL3) and low-high bands (LH2) are taken into consideration for embedding the watermark and encoded text data into the cover picture, respectively. Later, selective encryption is used to encrypt the watermarked picture to save cost. Finally, to reduce the distortion effect applied to the extracted watermark image, a BPNN is used. The technique ensures the confidentiality of the data and shows stronger robustness against different attacks compared with Refs. 77–82. However, a transparent investigation of security and cost is required. A data-hiding method utilizing the long short-term memory-based recurrent neural network (LSTM-RNN) was proposed by Singh et al.[75] A particular criterion is used to acquire the electrocardiogram signal (TP) section of the ECG signal, which is then used to encrypt watermark data. The distortion between the original and forecasted signals is reduced using LSTM-RNN. The suggested method outperforms the conventional methods indicated in Refs. 83–92. However, the watermark's capacity is constrained, which limits the method's applicability for real-life use. A blind DCT-SVD-based watermarking technique was proposed by Wang et al.[76] Initially using the median filter, the cover image is enhanced to improve the robustness of the watermark. Later, without altering the cover picture, RCNN is used to map the association between the watermark and cover images. The technique provides better robustness and stability than the works presented in Refs. 93–98.

4 Challenges and Open Research Directions

Due to the strong learning ability with accurate and superior results provided by deep-learning approaches, deep-learning models are exceptionally beneficial. However, the security and privacy of deep-learning models and media data remain challenging tasks. Current studies still try to mitigate security and privacy challenges. A summary of the most recent challenges in deep-learning-based watermarking is shown in Fig. 8. The major issues with deep-learning-based watermarking are as follows:

1. There is always a need to maintain a trade-off between embedding capacity, imperceptibility, and robustness that remains challenging for most watermarking systems.

Table 6 Summary of other deep-learning model-based watermarking.

Ref No.	Objective	Goal	Strategies	Model	Model role	Embedding location	Results	Cover \ mark size	Noticed weakness	Applications
71	Blind watermarking for protection of color media	Low embedding cost	DCT, IWT, and Mersenne Twister random number generator	ANN	Low-cost watermark embedding	4×4 DCT block	40.1304 dB PSNR with BER of 0, NC of 1, and SSIM of 0.9977	512×512 / 32×32	High complexity	Image databases
72	Blind watermarking system for grayscale images	Good visual quality with High robustness	LWT and encryption	ANN	Watermark extraction, improving robustness	Cover image sub-band 3rd level LWT vertical	43.88 dB PSNR with BER of 0 and NC of 0.9922	512×512 / 16×32	Low embedding capacity, Limited robustness analysis	USC-SIPI and CVG-UGR database
73	Watermarking scheme for color images	Low complexity and high robustness	CT, Arnold, Zenzo edge detector	MLP	Efficient extraction of logo mark	Edges of color image	48.81 dB PSNR with NC of 1	512×512	Not robust to few attacks	Color images
74	Securely transmit the digital content	Achieve high security and capacity with high robustness	Selective encryption, DCT, DWT, BCH, and Hamming code	BPNN	Denoising the retrieved watermark for high robustness	DWT coefficient and singular matrix	32.22 dB PSNR with NC of 0.99, and BER of 0	512×512 / 128×128	Concepts of biometrics, turbo code, and other transforms can be included	Color images
75	ECG signal watermarking	Improve visual quality, robustness, and security	XOR, PCA	LSTM RNN	Error reduction between extracted and original ECG signal	TP segment of ECG signal	72.52 dB PSNR	64×64, 128×128 and 256×256	Low embedding capacity and high computational complexity	European ST-T database
76	Non-embedding-based watermarking	Map relation with watermark and host	DCT, SVD, Median filter	RCNN	Non-embedding-based watermarking	Parameters of trained mapping based RCNN model	NC = 0.9906	512×512 / 32×32	Time complexity analysis is missing	General grayscale images

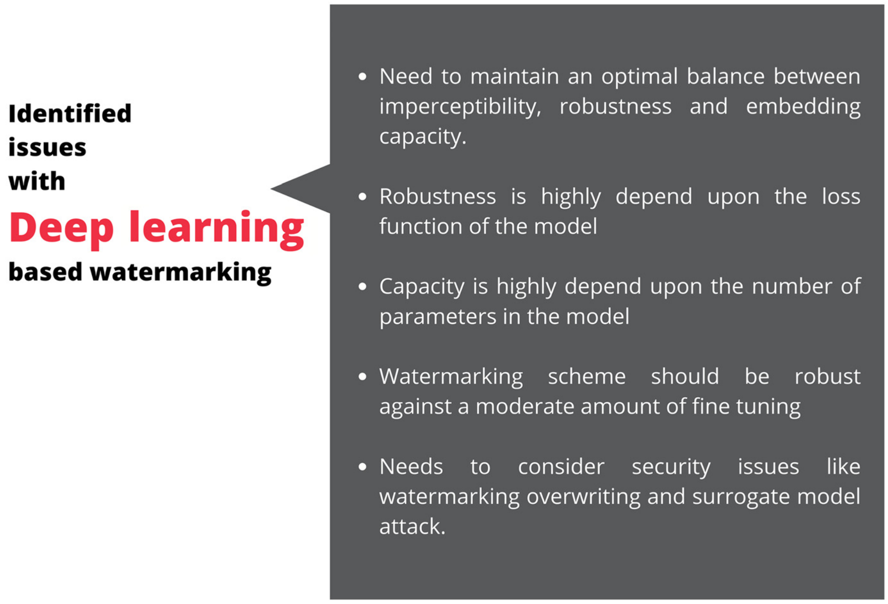

Fig. 8 Summary of issues identified for deep-learning-based watermarking.

2. Most schemes did not significantly provide the solution for data security and complexity issues and a deep-learning model security.
3. Studies have shown that transform domain-based watermarking methods are more robust than spatial domain-based watermarking approaches. So, there is a need to overcome the constraints of a single type of domain method for deep-learning-based watermarking.
4. Watermark security can be improved by combining encryption with watermarking. But the complexity grows as a result.
5. To minimize the model training complexity, pre-trained models are widely used, which leads to the problems of model overwriting and surrogate model attacks.
6. Watermarking robustness highly depends upon the number of samples used during training and the loss function used.
7. Most deep-learning-based watermarking has poor embedding capacity.
8. Watermarking schemes should be robust against network pruning and a moderate amount of fine-tuning.
9. Further investigations and proposals are required to maintain the complete security of digital media, which remains an open challenge.

5 Conclusion

Deep learning has had impressive development in the field of image processing, and more recently, it has also been developing in the field of data concealment to give a reliable watermarking approach with effective performance. This paper provides a comprehensive survey based on the popular deep-learning model used for the watermarking schemes. Starting with the classical watermarking approach, we outlined a few of the identified limitations of the classical approach and how deep-learning techniques can be used to overcome these. Further, we mentioned the requirements that need to be considered in developing a deep-learning-based watermarking. We thoroughly covered the roles and uses of the deep-learning-based model for watermarking, as well as the article's objective, goals, strategies, model function, embedding location, results, limits, and applications, as well as recent problems. Additionally, we outlined some of the issues with deep-learning-based techniques. We hope this survey gives insight into using the deep-learning model for watermarking techniques and provides a valuable information source for potential researchers.

Acknowledgments

This work is supported by research project order No. DLRL/21CR0003/SWCC&ENT/GN/LP dt. 29 August, 2020, DLRL, Hyderabad, India.

References

1. A. Anand and A. K. Singh, "A comprehensive study of deep learning based covert communication," *ACM Trans. Multimedia Comput. Commun. Appl.* **18**(2s), 1–19 (2022).
2. A. K. Singh, "Robust and distortion control dual watermarking in LWT domain using DCT and error correction code for color medical image," *Multimedia Tools Appl.* **78**(21), 30523–30533 (2019).
3. O. P. Singh et al., "Image watermarking using soft computing techniques: a comprehensive survey," *Multimedia Tools Appl.* **80**(20), 30367–30398 (2021).
4. O. P. Singh et al., "SecDH: security of COVID-19 images based on data hiding with PCA," *Comput. Commun.* **191**, 368–377 (2022).
5. S. P. Mohanty et al., "Everything you want to know about watermarking: from paper marks to hardware protection: from paper marks to hardware protection," *IEEE Consum. Electron. Mag.* **6**(3), 83–91 (2017).
6. S. Gull et al., "An efficient watermarking technique for tamper detection and localization of medical images," *J. Ambient Intell. Hum. Comput.* **11**(5), 1799–1808 (2020).
7. D. Singh and S. K. Singh, "DCT based efficient fragile watermarking scheme for image authentication and restoration," *Multimedia Tools Appl.* **76**(1), 953–977 (2017).
8. X. T. Wang et al., "Reversible data hiding for high quality images exploiting interpolation and direction order mechanism," *Digit. Signal Process.* **23**(2), 569–577 (2013).
9. N. S. Kamaruddin et al., "A review of text watermarking: theory, methods, and applications," *IEEE Access* **6**, 8011–8028 (2018).
10. P. Amrit and A. K. Singh, "Survey on watermarking methods in the artificial intelligence domain and beyond," *Comput. Commun.* **188**, 52–65 (2022).
11. Y. Li, H. Wang, and M. Barni, "A survey of deep neural network watermarking techniques," *Neurocomputing* **461**, 171–193 (2021).
12. W. Wan et al., "A comprehensive survey on robust image watermarking," *Neurocomputing* **488**, 226–247 (2022).
13. C. Zhang et al., "A brief survey on deep learning based data hiding, steganography and watermarking," arXiv:2103.01607 (2021).
14. Y. Li, H. Wang, and M. Barni, "A survey of deep neural network watermarking techniques," arXiv:2103.09274 (2021).
15. O. Byrnes et al., "Data hiding with deep learning: a survey unifying digital watermarking and steganography," arXiv:2107.09287 (2021).
16. J. Heaton, "Ian Goodfellow, Yoshua Bengio, and Aaron Courville: deep learning," *Genet. Program Evolvable Mach.* **19**, 305–307 (2018).
17. M. W. Hatoum et al., "Using deep learning for image watermarking attack," *Signal Process Image Commun.* **90**(June), 116019 (2021).
18. S. Ingaleshwar and N. V. Dharwadkar, "Water chaotic fruit fly optimization-based deep convolutional neural network for image watermarking using wavelet transform," *Multimedia Tools Appl.* 1–25 (2021).
19. M. Bagheri et al., "Adaptive control of embedding strength in image watermarking using neural networks," arXiv:2001.03251 (2020).
20. F. Tu et al., "Deep convolutional neural network architecture with reconfigurable computation patterns," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **25**(8), 2220–2233 (2017).
21. B. D. Rouhani, H. Chen, and F. Koushanfar, "Deepsigns: a generic watermarking framework for protecting the ownership of deep learning models," arXiv, pp. 1–13 (2018).
22. V. Sharma and R. N. Mir, "An enhanced time efficient technique for image watermarking using ant colony optimization and light gradient boosting algorithm," *J. King Saud Univ.-Comput. Inf. Sci.* **34**(3), 615–626 (2019).

23. L. Zhang and D. Wei, "Image watermarking based on matrix decomposition and gyrator transform in invariant integer wavelet domain," *Signal Process.* **169**, 107421 (2020).
24. W. Zheng et al., "Robust and high capacity watermarking for image based on DWT-SVD and CNN," in *13th IEEE Conf. Ind. Electron. and Appl.*, May, IEEE, pp. 1233–1237 (2018).
25. D. Li et al., "A novel CNN based security guaranteed image watermarking generation scenario for smart city applications," *Inf. Sci.* **479**, 432–447 (2019).
26. G. Badshah et al., "Watermark compression in medical image watermarking using Lempel–Ziv–Welch (LZW) lossless compression technique," *J. Digit. Imaging* **29**(2), 216–225 (2016).
27. E. Etemad et al., "Robust image watermarking scheme using bit-plane of hadamard coefficients," *Multimedia Tools Appl.* **77**(2), 2033–2055 (2018).
28. Q. Su and B. Chen, "Robust color image watermarking technique in the spatial domain," *Soft Comput.* **22**(1), 91–106 (2018).
29. H. Qian, L. Tian, and C. Li, "Robust blind image watermarking algorithm based on singular value quantization," in *Proc. Int. Conf. Internet Multimedia Comput. and Service*, August, pp. 277–280 (2016).
30. M. Plata and P. Syga, "Robust spatial-spread deep neural image watermarking," in *IEEE 19th Int. Conf. Trust, Security and Privacy in Comput. and Commun. (TrustCom)*, December, IEEE, pp. 62–70 (2020).
31. J. Zhu et al., "Hidden: hiding data with deep networks," *Lect. Notes Comput. Sci.* **11219**, 682–697 (2018).
32. X. Luo et al., "Distortion agnostic deep watermarking," in *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recognit.*, pp. 13548–13557 (2020).
33. M. Ahmadi et al., "ReDMark: framework for residual diffusion watermarking based on deep networks," *Expert Syst. Appl.* **146**, 113157 (2020).
34. S. Mun et al., "A robust blind watermarking using convolutional neural network," arXiv:1704.03248 (2017).
35. O. P. Singh and A. K. Singh, "Data hiding in encryption–compression domain," *Complex Intell. Syst.* 1–14 (2021).
36. A. K. Singh, M. Dave, and A. Mohan, "Hybrid technique for robust and imperceptible multiple watermarking using medical images," *Multimedia Tools Appl.* **75**(14), 8381–8401 (2016).
37. S. Thakur et al., "Improved DWT-SVD-based medical image watermarking through hamming code and chaotic encryption," in *Advances in VLSI, Communication, and Signal Processing*, D. Dutta et al., Eds., pp. 897–905, Springer, Singapore (2020).
38. A. Anand and A. K. Singh, "An improved DWT-SVD domain watermarking for medical information security," *Comput. Commun.* **152**, 72–80 (2020).
39. A. Anand et al., "Compression-then-encryption-based secure watermarking technique for smart healthcare system," *IEEE Multimedia* **27**(4), 133–143 (2020).
40. L. Y. Hsu and H. T. Hu, "QDCT-based blind color image watermarking with aid of GWO and DnCNN for performance improvement," *IEEE Access* **9**, 155138–155152 (2021).
41. S. W. Byun, H. S. Son, and S. P. Lee, "Fast and robust watermarking method based on DCT specific location," *IEEE Access* **7**, 100706–100718 (2019).
42. J. Li et al., "A QDCT-and SVD-based color image watermarking scheme using an optimized encrypted binary computer-generated hologram," *Soft Comput.* **22**(1), 47–65 (2018).
43. B. Chen et al., "Quaternion discrete fractional random transform for color image adaptive watermarking," *Multimedia Tools Appl.* **77**(16), 20809–20837 (2018).
44. M. Moosazadeh and G. Ekbatanifard, "A new DCT-based robust image watermarking method using teaching-learning-based optimization," *J. Inf. Security Appl.* **47**, 28–38 (2019).
45. H. Kandi, D. Mishra, and S. R. S. Gorthi, "Exploring the learning capabilities of convolutional neural networks for robust image watermarking," *Comput. Security* **65**, 247–268 (2017).
46. Y. Nagai et al., "Digital watermarking for deep neural networks," *Int. J. Multimedia Inf. Retrieval* **7**(1), 3–16 (2018).

47. Y. Uchida et al., "Embedding watermarks into deep neural networks," in *Proc. 2017 ACM on Int. Conf. Multimedia Retrieval*, pp. 269–277 (2017).
48. X. Guan et al., "Reversible watermarking in deep convolutional neural networks for integrity authentication," in *Proc. 28th ACM Int. Conf. Multimedia*, pp. 2273–2280 (2020).
49. J. Wu et al., "De-Mark GAN: removing dense watermark with generative adversarial network," in *Int. Conf. Biometrics*, IEEE, pp. 69–74 (2018).
50. V. L. Cu et al., "A robust data hiding scheme using generated content for securing genuine documents," in *Int. Conf. Document Anal. and Recognit.*, IEEE, pp. 787–792 (2019).
51. C. V. Loc, J. C. Burie, and J. M. Ogier, "Stable regions and object fill-based approach for document images watermarking," in *13th IAPR Int. Workshop on Document Anal. Syst.*, April, IEEE, pp. 181–186 (2018).
52. C. V. Loc, J. C. Burie, and J. M. Ogier, "Document images watermarking for security issue using fully convolutional networks," in *24th Int. Conf. Pattern Recognit.*, August, IEEE, pp. 1091–1096 (2018).
53. V. L. Cu, J. C. Burie, and J. M. Ogier, "Watermarking for security issue of handwritten documents with fully convolutional networks," in *16th Int. Conf. Front. in Handwriting Recognit.*, August, IEEE, pp. 303–308 (2018).
54. Q. Wei, H. Wang, and G. Zhang, "A robust image watermarking approach using cycle variational autoencoder," *Security Commun. Netw.* **2020**, 9 (2020).
55. L. Zhang, W. Li, and H. Ye, "A blind watermarking system based on deep learning model," in *IEEE 20th Int. Conf. Trust, Security, and Privacy in Comput. and Commun. (TrustCom)*, pp. 1208–1213 (2021).
56. B. Fan, Z. Li, and J. Gao, "DwiMark: a multiscale robust deep watermarking framework for diffusion-weighted imaging images," *Multimedia Syst.* **28**(1), 295–310 (2022).
57. H. Fang et al., "Encoded feature enhancement in watermarking network for distortion in real scenes," *IEEE Trans. Multimedia*, 1–13 (2022).
58. X. Kang, J. Huang, and W. Zeng, "Efficient general print-scanning resilient data hiding based on uniform log-polar mapping," *IEEE Trans. Inf. Forensics Security* **5**(1), 1–12 (2010).
59. J. Zhu et al., "Hidden: hiding data with deep networks," *Lect. Notes Comput. Sci.* **11219**, 657–672 (2018).
60. Y. Liu et al., "A novel two-stage separable deep learning framework for practical blind watermarking," in *Proc. 27th ACM Int. Conf. Multimedia*, pp. 1509–1517 (2019).
61. Z. Ma et al., "Local geometric distortions resilient watermarking scheme based on symmetry," *IEEE Trans. Circuits Syst. Video Technol.* **31**(12), 4826–4839 (2021).
62. P. Neekhara et al., "FaceSigns: semi-fragile neural watermarks for media authentication and countering deepfakes," arXiv:2204.01960 (2022).
63. J. Hou et al., "Reversible data hiding based on multiple histograms modification and deep neural networks," *Signal Process. Image Commun.* **92**, 116118 (2021).
64. W. He et al., "Efficient PVO-based reversible data hiding using multistage blocking and prediction accuracy matrix," *J. Visual Commun. Image Represent.* **46**, 58–69 (2017).
65. Y. Jia et al., "Reversible data hiding based on reducing invalid shifting of pixels in histogram shifting," *Signal Process.* **163**, 238–246 (2019).
66. X. Li et al., "Efficient reversible data hiding based on multiple histograms modification," *IEEE Trans. Inf. Forensics Security* **10**(9), 1824–1834 (2015).
67. J. Zhang et al., "Protecting intellectual property of deep neural networks with watermarking," in *Proc. 2018 on Asia Conf. Comput. and Commun. Security*, May, pp. 159–172 (2018).
68. H. Wu et al., "Watermarking neural networks with watermarked images," *IEEE Trans. Circuits Syst. Video Technol.* **31**(7), 2591–2601 (2021).
69. F. Deeba et al., "Digital watermarking using deep neural networks," *Int. J. Mach. Learn. Comput.* **10**(2), 277–282 (2020).
70. I. Hamamoto and M. Kawamura, "Image watermarking technique using embedder and extractor neural networks," *IEICE Trans. Inf. Syst.* **E102.D**(1), 19–30 (2019).
71. R. Sinhal, D. K. Jain, and I. A. Ansari, "Machine learning based blind color image watermarking scheme for copyright protection," *Pattern Recognit. Lett.* **145**, 171–177 (2021).

72. M. Islam, A. Roy, and R. H. Laskar, "Neural network based robust image watermarking technique in LWT domain," *J. Intell. Fuzzy Syst.* **34**(3), 1691–1700 (2018).
73. M. Kazemi, M. A. Pourmina, and A. H. Mazinan, "Analysis of watermarking framework for color image through a neural network-based approach," *Complex Intell. Syst.* **6**(1), 213–220 (2020).
74. A. K. Singh et al., "Multiple watermarking technique for securing online social network contents using back propagation neural network," *Future Gen. Comput. Syst.* **86**, 926–939 (2018).
75. S. Banerjee and G. K. Singh, "A new approach of ECG steganography and prediction using deep learning," *Biomed. Signal Process. Control* **64**, 102151 (2021).
76. X. Wang et al., "Mapping based residual convolution neural network for non-embedding and blind image watermarking," *J. Inf. Security Appl.* **59**, 102820 (2021).
77. Y. Xing and J. Tan, "A color image watermarking scheme resistant against geometrical attacks," *Radioengineering* **19**(1), 62–67 (2010).
78. A. Ghafoor and M. Imran, "A non-blind color image watermarking scheme resistent against geometric attacks," *Radioengineering* **21**(4), 1246–1251 (2012).
79. V. Santhi and A. Thangavelu, "DC coefficients based watermarking technique for color images using singular value decomposition," *Int. J. Comput. Electr. Eng.* **3**(1), 1793–8163 (2011).
80. M. Zhao and Y. Dang, "Color image copyright protection digital watermarking algorithm based on DWT & DCT," in *4th Int. Conf. Wireless Commun., Netw. and Mobile Comput.*, October, IEEE, pp. 1–4 (2008).
81. X. Xiong, "A new robust color image watermarking scheme based on 3D-DCT," *World J. Eng. Technol.* **3**(03), 177 (2015).
82. H. Shi, F. Lv, and Y. Cao, "A blind watermarking technique for color image based on SVD with circulation," *J. Softw.* **9**(7), 1749–1756 (2014).
83. M. Urvoy, D. Goudia, and F. Autrusseau, "Perceptual DFT watermarking with improved detection and robustness to geometrical distortions," *IEEE Trans. Inf. Forensics Security* **9**(7), 1108–1119 (2014).
84. A. Ibaida and I. Khalil, "Wavelet-based ECG steganography for protecting patient confidential information in point-of-care systems," *IEEE Trans. Biomed. Eng.* **60**(12), 3322–3330 (2013).
85. S. Edward Jero, P. Ramu, and S. Ramakrishnan, "Discrete wavelet transform and singular value decomposition based ECG steganography for secured patient information transmission," *J. Med. Syst.* **38**(10), 1–11 (2014).
86. B. Lei et al., "Reversible watermarking scheme for medical image based on differential evolution," *Expert Syst. Appl.* **41**(7), 3178–3188 (2014).
87. S. T. Chen et al., "Hiding patients confidential data in the ECG signal via a transform-domain quantization scheme," *J. Med. Syst.* **38**(6), 1–8 (2014).
88. A. Tareef and A. Al-Ani, "A highly secure oblivious sparse coding-based watermarking system for ownership verification," *Expert Syst. Appl.* **42**(4), 2224–2233 (2015).
89. C. A. Liji, K. P. Indiradevi, and K. A. Babu, "Integer-to-integer wavelet transform based ECG steganography for securing patient confidential information," *Proc. Technol.* **24**, 1039–1047 (2016).
90. P. Ramu and R. Swaminathan, "Imperceptibility—robustness tradeoff studies for ECG steganography using continuous ant colony optimization," *Expert Syst. Appl.* **49**, 123–135 (2016).
91. E. Candes et al., "Fast discrete curvelet transforms," *Multiscale Model. Simul.* **5**(3), 861–899 (2006).
92. A. Abuadbba and I. Khalil, "Walsh–Hadamard-based 3-D steganography for protecting sensitive information in point-of-care," *IEEE Trans. Biomed. Eng.* **64**(9), 2186–2195 (2017).
93. F. Ernawan and M. N. Kabir, "A block-based RDWT-SVD image watermarking method using human visual system characteristics," *Visual Comput.* **36**(1), 19–37 (2020).
94. X. B. Kang et al., "A novel hybrid of DCT and SVD in DWT domain for robust and invisible blind image watermarking with optimal embedding strength," *Multimedia Tools Appl.* **77**(11), 13197–13224 (2018).

95. T. K. Araghi, A. Abd Manaf, and S. K. Araghi, "A secure blind discrete wavelet transform based watermarking scheme using two-level singular value decomposition," *Expert Syst. Appl.* **112**, 208–228 (2018).
96. S. H. Soleymani, A. H. Taherinia, and A. H. Mohajerzadeh, "WACA: a new blind robust watermarking method based on Arnold Cat map and amplified pseudo-noise strings with weak correlation," *Multimedia Tools Appl.* **78**(14), 19163–19179 (2019).
97. S. M. Mun et al., "Finding robust domain from attacks: a learning framework for blind watermarking," *Neurocomputing* **337**, 191–202 (2019).
98. V. S. Verma, A. Bhardwaj, and R. K. Jha, "A new scheme for watermark extraction using combined noise-induced resonance and support vector machine with PCA based feature reduction," *Multimedia Tools Appl.* **78**(16), 23203–23224 (2019).

Himanshu Kumar Singh is currently pursuing his PhD in computer science and engineering from National Institute of Technology Patna, Bihar, India. His research interests include watermarking, deep learning, machine learning, and data security.

Amit Kumar Singh is an associate professor in the Computer Science and Engineering Department at the National Institute of Technology Patna, Bihar, India. His research interests include watermarking and image processing.