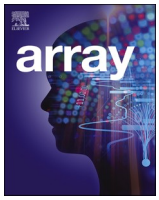
Array 22 (2024) 100353

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array



Comprehensive survey on image steganalysis using deep learning

Ntivuguruzwa Jean De La Croix [a,b], Tohari Ahmad [a,*], Fengling Han [c]

[a] Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
[b] College of Science and Technology, University of Rwanda, Kigali, 3900, Rwanda
[c] School of Computing Technologies, RMIT University, Melbourne, VIC, 3000, Australia

ARTICLE INFO

Keywords:
Data hiding
Information security
National security
Network infrastructure
Steganalysis

ABSTRACT

Steganalysis, a field devoted to detecting concealed information in various forms of digital media, including text, images, audio, and video files, has evolved significantly over time. This evolution aims to improve the accuracy of revealing potential hidden data. Traditional machine learning approaches, such as support vector machines (SVM) and ensemble classifiers (ECs), were previously employed in steganalysis. However, they demonstrated ineffective against contemporary and prevalent steganographic methods. The field of steganalysis has experienced noteworthy advancements by transitioning from traditional machine learning methods to deep learning techniques, resulting in superior outcomes. More specifically, deep learning-based steganalysis approaches exhibit rapid detection of steganographic payloads and demonstrate remarkable accuracy and efficiency across a spectrum of modern steganographic algorithms. This paper is dedicated to investigating recent developments in deep learning-based steganalysis schemes, exploring their evolution, and conducting a thorough analysis of the techniques incorporated in these schemes. Furthermore, the research aims to delve into the current trends in steganalysis, explicitly focusing on digital image steganography. By examining the latest techniques and methodologies, this work contributes to an enhanced understanding of the evolving landscape of steganalysis, shedding light on the advancements achieved through deep learning-based approaches.

1. .Introduction

Technological advancement has made digital media pervasive in public networks, notably images, videos, and audio. The ubiquitousness of digital photos has made them preferred by communication agents to use these images/photos as covers to host confidential data for transmission in the public network [1,2]. To protect confidential data, practitioners apply different methods to hide confidential data from unwanted or unauthorized access. Cryptography and steganography [3–5] have been used. Although they are used similarly, steganography and cryptography are different algorithms for covering data during transmission. Steganography, an approach of hiding secret data in the content of digital media of various types; typical applications can be found in Refs. [6–8], is mainly used for protecting the hidden data from unwanted access in the public networks during transmission [9]. Steganography techniques have been employed in the industrial domain to prevent the illegal copying of digital content where the copyright societies modify digital content in an invisible way to the human eye, thereby embedding information that can be used to establish the ownership of the image or to track its sale or transmission, as reported in Refs. [10,11]. In military applications, steganography has been utilized to transmit sensitive messages without being detected by the adversary. However, there are concerns that illegal groups and terrorists might also employ steganography to communicate covertly, as mentioned in Ref. [12].

$$Cover + Secret = Stego \tag{1}$$

Steganalysis involves the detection of concealed messages within digital media, such as images. It systematically applies analytical techniques, statistical methods, and machine learning models to identify and detect hidden information within digital media, focusing on maintaining a balance between accuracy and adaptability. Feature extraction and binary classification [13] are two critical steps for Steganalysis. With advances in deep learning (DL) and graphic processing units (GPUs) [14], research in steganalysis has yielded promising results in detecting the concealed steganographic payload within digital images. DL has played a crucial role not only in the steganalysis of digital images but also in other advanced tasks related to digital images, as evidenced in the recent research works in Refs. [15,16]. That highlights the superior

* Corresponding author.
E-mail address: tohari@its.ac.id (T. Ahmad).

https://doi.org/10.1016/j.array.2024.100353
Received 29 December 2023; Received in revised form 30 May 2024; Accepted 4 June 2024
Available online 4 June 2024

capability of the DL to tackle the complexity issues related to digital images and their applicability to real-life problem-solving. Referring to Fig. 1, the DL-based algorithm unifies the feature extraction and classification stages under the same architecture, optimizing the parameters simultaneously and reducing the complexity and dimensionality opposite to machine learning (ML). The conventional methods for ML involve a series of techniques that rely on human-guided feature extraction [17]. Subsequently, these techniques train a computer to identify various classes within the provided training data autonomously.

However, achieving accurate classification in steganalysis requires a deep understanding of both steganography and steganalysis. This understanding is gained through a time-consuming process of comprehending the fundamental principles of available steganographic algorithms and tools and identifying patterns associated with payload embedding. One major challenge encountered in this domain is the issue of dimensionality for classifiers. Recent steganalysis techniques, such as spatial rich models (SRM) [18] and maxSRM [19], subtractive pixel adjacency matrix (SPAM) [20], Phase Aware Projection Model (PHARM), Gabor Filter bank (GFR) [21] and discrete cosine transform residual (DCTR) [22], exhibit varying feature dimensionality, specifically 34,671, 686, 12,600, 17,000 and 8000, respectively. Consequently, ML-based steganalysis algorithms require many carefully crafted steganalysis features to capture modifications in embedding accurately, irrespective of variations in image content across different covers. This ML-based arduous process can be effectively addressed through DL-based techniques, which currently represent the cutting-edge approach for tasks involving image classification and recognition. Under the ML paradigm, DL involves training models to perform direct classification based on input data. Prominent DL models include feedforward neural networks (FNN), Shallow neural networks (networks characterized by one or two layers of perceptron known as artificial neurons), and deep neural networks (DNNs) made of numerous layers. Moreover, various DL-based models include autoencoders, convolutional neural networks (CNNs), and recursive and recurrent neural networks (RNNs).

Referring to Fig. 1, it is identified that ML-based steganalysis methods lack backward communication between the feature extraction and classification phases, with information flowing only in the forward direction with feature extraction as the source and the classification as the destination. In contrast, DL, a subset of machine learning, involves training neural networks with multiple layers to automatically learn and extract features from data. DL-based steganalysis methods integrate these two phases, forming a unified phase characterized by a back-forward communication style. DL schemes leverage the decisions made by the classification phase to update the extracted features in the feature extraction phase. Fig. 1(b) demonstrates how DL-based steganalysis techniques extract features directly from the raw input image autonomously.

In references [23,24], the authors introduce an alternative taxonomy for steganalysis, distinguishing between specific and statistical approaches. Moreover, the research work in Ref. [25] proposed a general survey for steganography and steganalysis in spatial and JPEG domains; the surveys in Refs. [26,27] availed critical analysis and classification of the steganalysis method, focusing on blind steganalysis. Based on the fact that the five previously identified surveys seem to be outdated, we also consulted four recent literature surveys with [28,29] focusing on steganography and steganalysis techniques for digital forensics with both limiting their content to the techniques [30], focuses on demonstrating the paradigm shift from ML to DL in general steganalysis tasks, and [31] presented a survey focused on the analysis of the deep CNNs for images steganography combined with steganalysis. To complete the existing literature and avail a new perspective of DL-based steganalysis, in our comprehensive review, we include a wide range of steganalysis categories, not limited to specific ones such as JPEG or universal detection but focusing on digging deep into the DL-based steganalysis methods. Moreover, our review is up to date, offering detailed references spanning from earlier steganalysis methods to state-of-the-art.

In this article, we provide an extensive survey focusing on contemporary techniques for image steganalysis using DL techniques. The evolution of these techniques is thoroughly examined, covering developments from their inception to the present day. To enhance the study, both the advantages and limitations of each technique are thoroughly investigated. The key contributions of this paper can be summarized as follows.

- Providing a thorough of the advancement of DL-based steganalysis techniques, charting their development from traditional methodologies based on ML and their classification. Additionally, it delves deeper into the subject by investigating various DL-based steganalysis models and emerging trends in the field of steganalysis.
- Diverging form the existing surveys, this review comprehensively gives specific attention to DL-based steganalysis techniques in spatial domain images from a new scientific perspective. Adopting a novel scientific perspective, this study delves into unexplored realms to elucidate the progress, challenges, and potential avenues for future research in this specialized domain.
- It is expected that this survey can provide valuable guidance to researchers in developing steganalysis techniques for both traditional and advanced steganographic methods. Moreover, it is a valuable resource for individuals working in DL-based steganalysis and information security, offering insights and support for their endeavours.

The subsequent sections of this paper are organized as follows: Section 2 provides a general overview of the steganalysis techniques in digital images. Section 3 delves into the shift from ML to DL-based steganalysis paradigms. Section 4 explores DL-based techniques, most of which are CNN-based schemes. Section 5 presents the key emerging challenges in digital image steganalysis, which reflect the status of the race between steganography and steganalysis. Finally, Section 6 concludes this review work, summarizing the key points and implications.

2. .Steganalysis techniques for digital images

Based on the paradigm of a steganalysis task, the steganalysis techniques for digital images are classified into several categories. Some

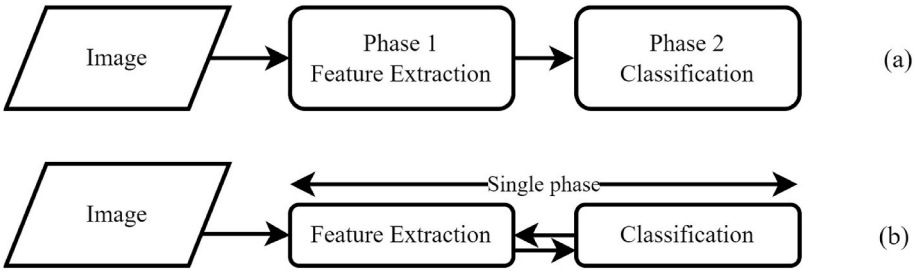


Fig. 1. A) ML-based steganalysis system; b) DL-based steganalysis system.

main categories of steganalysis techniques include visual inspection, statistical analysis, machine learning, feature-based analysis, and model-specific analysis. Visual inspection involves manually examining the image for any visual anomalies or artefacts that may indicate steganographic content. This technique relies on the human observer's ability to notice subtle changes in the image that may not be easily detectable by automated methods [30]. Statistical analysis involves examining various statistical properties of the image to identify any deviations from the expected patterns. For example, steganographic techniques often introduce slight changes in pixel values or distribution, which can be detected through statistical analysis of the image's histogram, correlation, or higher-order moments [32]. Machine learning techniques have been increasingly applied in steganalysis to train models that can automatically detect steganographic content. This involves training a classifier on a large dataset of known cover images and steganographic images, allowing the model to learn patterns and features that distinguish between them. The trained model can then classify new images as cover or stego [30]. The feature-based analysis involves extracting specific features from the image and analyzing them to identify steganographic content. Features could include texture descriptors [33,34], colour moments [35], edge histograms, or wavelet coefficients [36]. Deviations can be identified by comparing these features with expected values for cover images, suggesting the presence of steganography [37]. Model-specific steganalysis consists of some techniques tailored to specific steganographic algorithms or software. These techniques exploit known weaknesses or vulnerabilities of steganographic methods to detect their presence in digital images. This approach requires prior knowledge or suspicion of the specific steganographic algorithm [38].

Based on the targeted result of a steganalysis method, the steganalysis techniques are classified into four main categories, namely, blind steganalysis, locative steganalysis, quantitative steganalysis, and forensic steganalysis. Blind steganalysis focuses on identifying possible confidential data without prior knowledge of the steganographic scheme or the carrying media [39]. Among the recent advancements in blind steganalysis are machine learning approaches, feature extraction techniques, and deep learning-based methods [27]. Locative steganalysis aims to locate the specific regions within digital media where hidden data may be embedded [18,40]. The recent advancements in locative steganalysis we can cite include spatial domain analysis, frequency domain analysis, and content-aware methods [40]. Quantitative steganalysis is interested in demonstrating the quantity or amount of the secret data concealed in the vital content of the cover image [41,42]. Recent schemes in quantitative steganalysis include statistical approaches, payload estimation techniques, distortion analysis, chi-square analysis, Markov feature extraction, and cover source estimation [42]. Forensic steganalysis involves detecting and analyzing steganographic schemes techniques utilized in real-world scenarios to conceal confidential data, primarily for legal or criminal investigations [29,38]. Recent advancements in forensic steganalysis include image tampering,

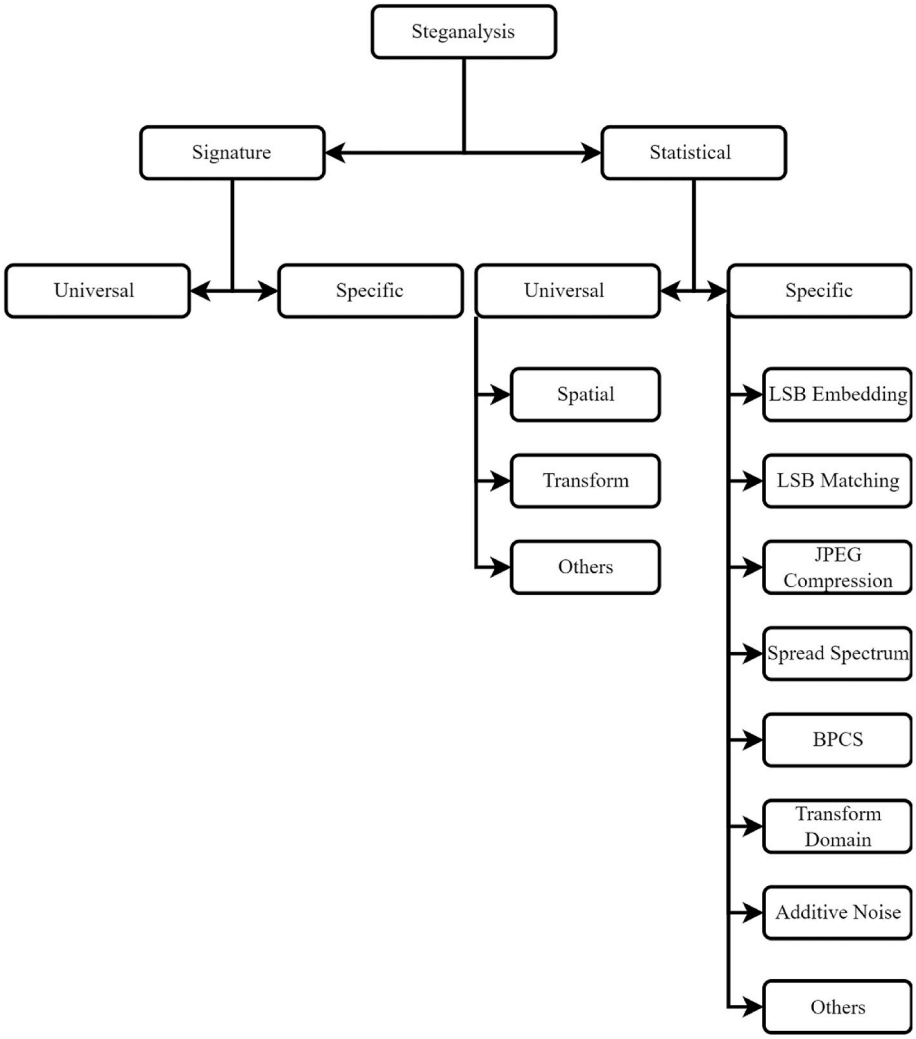


Fig. 2. Taxonomy of image steganalysis techniques.

multimedia forensics, data recovery, resampling detection, double-compressed image analysis, and error-level analysis [29].

To classify the steganalysis techniques for image steganalysis, the techniques above fall into two main categories: signature and statistical steganalysis, as of the taxonomy illustrated in Fig. 2. Both the signature steganalysis and statistical steganalysis have two sub-divisions per each, namely the specific steganalysis that consists of attacking a specific type of steganography, and the universal steganalysis that consists of targeting any steganographic algorithm in general.

2.1. Signature steganalysis

The main objective of any steganographic algorithm is to hide data in an image without violating its visibility quality so that the changes committed remain imperceptible to the human eye [43,44]. However, any steganographic algorithm manipulates and alters the cover media's properties, which damages the media's originality with other patterns [45]. Moreover, these alterations and added patterns are considered broadcasting signatures of the concealed data. In the signature steganalysis schemes, the detection of the possible hidden data in a suspected image is based on inspecting the pattern signatures of a specific steganographic tool [17]. The signatures related to specific steganographic tools are applied in steganalysis to exploit the embedding algorithm. These approaches are concerned explicitly with examining the palette tables in Graphic Interchange Format (GIF) images and any abnormalities resulting from using steganographic tools [3]. The steganalysis tools are applied to the images' palettes within the LSB embedding for indices to palettes. Though these steganalysis approaches present promising performance, they have disadvantages in adaptation to the dynamism of the steganographers, and the reliability is much lower [24].

2.1.1. Specific signature steganalysis

This study classifies and examines four key steganalysis tools designed to discern specific steganographic techniques. These tools include Masker, which employs pattern analysis; Hide and Seek, focused on signature-based methodologies; Bit Plane Complexity Segmentation (BPCS), emphasizing spatial pattern complexity through bit plane segmentation; and Jpegx, a specialized steganalysis tool tailored for JPEG images. Our investigation delves into the unique approaches of each tool, contributing to a comprehensive understanding of their roles in detecting hidden information [46]. Hide and Seek is a software with 4.1 and 5.0 versions, and the steganography algorithm depends on the version selected to create stego images [24]. However, the two versions of hide and seek have some properties in common for the palette information of the stego. Applying hide and seek in grey-scale images with any of the versions 4.1 or 5.0 implies 256 triples with a range of 0–252 in four triples' sets, incrementing by 4 per step from 0 to 252. The detection key for the steganography with this tool is based on the 252 values of the image, which is the whitest part of the image [47].

Steganalysis is based on signature attacks on the steganography data embedded with BPCS, where the data are hidden by swapping the blocks with noise, such as on bit planes. Therefore, the bit plane block schemes for steganalysis are classified into the noisy region and the formative region based on the image complexity for the binary feature image complexity, a metric that reflects the black or white pattern density [48]. With this steganalysis tool's algorithm, a valley is formed in the stego image histogram's complexity, which is considered as a signature for BPCS steganography, and the signature is used for signature steganalytically identify the presence of any possible alteration in an inquiry image's components [49]. Signature steganalysis in Joint Photographic Experts Group (JPEG) is performed by detecting the secret message hidden with the Jpegx steganography tool and any other scheme that manipulates the JPEG data in this image [50]. Upon detecting the steganographic signature, a steganalyst automatically classifies and image as a stego, resulting in a Jpegs steganographic algorithm [51].

2.1.2. Universal signature steganalysis

Universal signature steganalysis aims to create detection approaches based on signatures for a broad spectrum of steganographic schemes, encompassing diverse embedding techniques. Unlike traditional steganalysis approaches tailored for well-known steganographic schemes, universal signature steganalysis focuses on creating a unified framework able to detect different types of confidential information regardless of the specific algorithm or method used for data concealment [17]. Recent advancements in ML and DL techniques have improved the performance and versatility of universal signature steganalysis [52]. However, the Universal Signature Steganalysis achieves poor performance in JPEG carries for the spatial domain steganography approaches because of the quantization resulting from compressing the JPEG that works as an identic fingerprint to detect any modification of the cover image by benchmarking the stego with the original cover image in the JPEG format [52]. For this steganalysis approach, the stego image is first split into blocks of size 8×8, and secondly, the matrix for quantization is generated by analyzing the Discrete Cosine Transform (DCT) coefficient values for all blocks in the first step. Finally, the table in the second step is benchmarked with the original quantization table of the JPEG to check the originality. If all the blocks are incompatible, the image under operation is classified as an image altered with steganography. This method is believed to be the most efficient way to detect any change in a JPEG image, but it presents a weakness in that it cannot detect any steganography when any modification happens on an image because any slight modification wipes out the signature [53].

2.2. Statistical steganalysis

Statistical steganalysis refers to techniques used to detect hidden steganographic content by examining the statistical properties of the cover media. These properties include pixel values, frequency distributions, correlations, and higher-order statistical characteristics. By comparing these properties to the expected values for the cover media, steganalysis algorithms can identify anomalies indicating concealed data presence [54]. The fundamental assumption in statistical steganalysis is that embedding hidden information using steganography introduces statistical irregularities in the cover media. The modifications made to the cover media to accommodate the hidden data disrupt the natural statistical patterns expected in the original, unaltered media [30]. Statistical steganalysis algorithms often leverage machine learning and pattern recognition techniques to identify these statistical anomalies. They are trained using extensive datasets of known cover media and steganographic content. The algorithms learn the statistical patterns of both data types through this training. Subsequently, these algorithms can analyze new media to assess the likelihood of steganographic content being present [55].

Mathematics signature-based steganalysis relies on visual perception, which is considered more reliable. Mathematical approaches provide a stronger foundation for detection, as they are based on quantifiable and objective principles. Statistical steganalysis is categorized into two main types: specific statistical steganalysis and universal statistical steganalysis. The former focuses on detecting specific steganography techniques or algorithms, while the latter aims to identify steganographic content in a broader and more general sense.

2.2.1. Specific statistical steganalysis

A specific statistical steganalysis is a specialized approach used in digital image analysis to detect specific steganography schemes for hiding information in cover images. This method involves examining the statistical properties of images to identify distinctive patterns and anomalies associated with a well-known steganographic scheme [32]. In specific statistical steganalysis, custom algorithms are developed to target and recognize the unique statistical characteristics introduced by steganographic algorithms. These algorithms are trained using datasets comprising cover images containing known instances of embedded

steganographic content and their corresponding statistical properties [24]. The approaches learn to distinguish between cover and stego images throughout training. They extract statistical features from the cover images and compare them against predefined thresholds or models specific to the targeted steganography methods. By detecting significant deviations from the expected statistical properties, specific statistical steganalysis algorithms can infer the presence of confidential data and determine the specific steganographic algorithm employed [24]. Specific statistical steganalysis proves particularly effective when the steganography techniques are well-known or when there is prior knowledge regarding the types of steganography likely to be used. It enables targeted analysis and detection of specific steganographic methods, offering valuable insights for digital forensics, security investigations, and countermeasures against known threats [14].

This review focuses on the statistical steganalysis specific to seven basic steganography algorithms, namely the LSB embedding, the LSB matching, the spread-spectrum steganography, the BPCS-steganography, the JPEG-compression steganography, the transform domain steganography, the additive noise steganography, and others.

2.2.1.1. A. steganalysis of the LSB embedding. LSB is considered the most used steganographic algorithm. This data-hiding technique involves embedding the bits of a secret message in the least significant bits of the pixels of the cover image. The pixel selection depends on the key shared among the communicating bodies. The general logic of the LSB steganography is based on the example below adapted from Ref. [56].

For x and y, the pixels of the same pairs where $x = 206$ and $y = 201$ and $b = 1$, a small amount of confidential data. Before embedding any data, we first calculate two main values, namely the average of the pixels x and y labelled as l and the difference between the pixel labelled as h as per (2) and (3), with the values of l and h are respectively 203 and 5 and the operator. $\lfloor . \rfloor$ stands for the floor operation, which means "the integer part less or equal to," and ignores the decimal part of the obtained result.

$$l = \left\lfloor \frac{x+y}{2} \right\rfloor \tag{2}$$

$$h = x - y \tag{3}$$

To identify the least significant bit in h, its values are converted to the binary system and becomes 101_2. Therefore, the secret bit b is then appended after the least significant bit of h to become h'. Thus, $h' = hb_2$ and the value becomes $h' = 1011_2 = 11_{10}$, by only handling the operations above in a decimal system, the value of h' is got by (4), and after getting the value of h'. The data hiding operation is then proceeded following (4) and (5) to get the values of the new pixels, including the hidden bit x' and y'. The new values of the pixels are $x' = 209$ and $y' = 198$. The pixel values must belong to the set P, such as P is the set of values p, with $0 \leq p \leq 255$. If the value of one of the pixels is less than zero, it is called the underflow problem; if the of one is greater than 255, it is called the overflow problem. To avoid the underflow and overflow problems, for the pixel values out of the domain of definition of the set P, the pairs to which they belong are not used for data hiding with the formulae in (5) and (6), which makes it complicated for a steganalyst to localize the altered pixels in the image.

$$h' = 2h + b \tag{4}$$

$$x' = l + \left\lfloor \frac{h'+1}{2} \right\rfloor \tag{5}$$

$$y' = l - \left\lfloor \frac{h'}{2} \right\rfloor \tag{6}$$

2.2.1.2. B. steganalysis of LSB matching. LSB matching steganalysis is a technique used to detect the presence of hidden information within digital using an LSB matching steganographic algorithm. In LSB matching steganalysis, algorithms are designed to analyze the statistical properties and patterns present in the LSBs of the image pixels [57]. These algorithms aim to identify deviations or irregularities indicative of LSB matching steganography. During the analysis, the algorithms examine the neighbouring pixels and compare their LSB values to determine if they match or differ. In LSB matching steganography, the LSB values are adjusted to match a predetermined sequence based on the secret data, while in regular cover images, the LSB values are typically uncorrelated or random. By examining the correlations between the LSB values of neighbouring pixels, LSB matching steganalysis algorithms can detect deviations from the expected statistical patterns. These deviations suggest the presence of confidential data that has been embedded using LSB-matching steganography [58].

LSB-matching steganalysis is crucial in detecting and combating LSB-matching steganography for covert communication or illicit purposes. Identifying the presence of hidden information assists in digital forensic investigations, security assessments, and preserving the integrity of digital media. The main difference between LSB matching and LSB embedding is that LSB matching is more robust than ordinary LSB replacement. A steganalysis specific to LSB matching has been proposed in Ref. [59] and later improved in Ref. [57] to focus on grayscale images. The research in Ref. [58] proposed a characteristic histogram function for steganalysis in colour images adapted for steganalysis in grayscale images to generate two algorithms related to histogram characteristic function.

2.2.1.3. C. spread spectrum steganalysis. Spread spectrum steganalysis involves detecting and analyzing confidential data within a digital image concealed using spread spectrum steganography. Spread spectrum steganography operates by embedding data within the Gaussian noise of the stego image, concealing it within the cover image [60]. Spread spectrum steganography algorithms possess strong resilience and resistance against adversarial attacks, making them challenging to detect through steganalysis techniques. However, like other steganographic methods, spread spectrum steganography can still raise suspicion due to the potential noise it introduces to images. In spread spectrum steganalysis, specific algorithms are developed to scrutinize the statistical properties and patterns inherent in the images, aiming to identify possible indicators of spread spectrum steganography. These algorithms primarily concentrate on analyzing the frequency domain characteristics of the image [61]. Spread spectrum steganography encompasses concealing confidential data by modulating it with a carrier signal and distributing it over a wide frequency range. Spread spectrum steganalysis algorithms detect potential indications of hidden data by examining the frequency distribution and correlations within the media. These algorithms analyze deviations or anomalies that could imply concealed information [50]. Spread spectrum steganalysis algorithms employ various techniques to reveal concealed information, including spectral analysis, autocorrelation analysis, and statistical modelling. Additionally, these algorithms can integrate machine learning and pattern recognition methodologies. They are trained on datasets comprising known instances of spread spectrum steganography and cover media, enabling them to learn the statistical patterns associated with both data types [62].

Several research works [26,63,64] have been proposed for spread spectrum steganography detection approach by exploiting the histogram characteristic function's center of mass properties; this method sets the center of mass to the first order momentum, and the characteristic histogram function is from the image's histogram of Fourier transform function. Sullivan et al., in Ref. [65], proposed a machine learning-based spread-spectrum steganalysis approach in grayscale images. Later, other algorithms [66–68] illuminated by this work were developed for this specific type of steganalysis.

2.2.1.4. D. BPCS steganalysis. BPCS (Bit-Plane Complexity

Segmentation) steganalysis is a technique used to detect and analyze the presence of hidden data within digital images that have been concealed using the BPCS steganography method. BPCS steganography involves embedding secret data by exploiting the complexity variations in the bit-planes of the image [69]. In BPCS steganalysis, algorithms are designed to examine the statistical properties and patterns present in the bit-planes of the image. The algorithm analyzes the complex variations and irregularities in the bit-plane distribution, looking for hidden data indications. Several methods have been proposed [55,69,70]. The BPCS steganalysis algorithm may involve segmenting the image into different complex regions or extracting features related to complexity variations. By comparing the observed complexity patterns to the expected statistical properties of the cover image, BPCS steganalysis algorithms can detect deviations that suggest the presence of concealed data [70]. ML and pattern recognition techniques can be incorporated into BPCS steganalysis algorithms to enhance their detection capabilities. These algorithms can be trained on datasets that include known instances of BPCS steganography and cover images, enabling them to learn the statistical patterns associated with both data types. This training enables the algorithm to accurately identify the presence of BPCS steganography in new images [25]. The work [71] proposed a steganalysis method to detect BPCS steganography in both the spatial and the transform domains. From a respective sequence of hypothesis testing and the Chi-square test, this approach decides on detecting the presence of possible hidden messages in an image's content. Based on the symmetry of the distribution around zero for the images' histograms for the coefficients of the quantized wavelet sub bands, the experimental results of this method show that changes caused by the addition of data in histograms of the images make this vulnerable to a Chi-Square test to detect the presence of hidden data [72].

2.2.1.5. E. JPEG-compression steganalysis. JPEG-compression steganalysis is a technique used to detect and analyze the presence of confidential data within JPEG-compressed images that has been concealed using steganography methods. Colour images in the JPEG format are considered the best cover for secret message embedding due to their high availability and use over the public network. Due to their ubiquitousness, steganographers have targeted JPEG images to transmit hidden bits of personal data [50]. JPEG-compression steganalysis involves using algorithms that analyze the statistical properties and artefacts introduced during the compression process of JPEG images. These algorithms scrutinize quantization errors, Discrete Cosine Transform (DCT) coefficients, and other characteristics to detect potential indications of hidden data. Detection focuses on identifying anomalies in the statistical distribution of DCT coefficients, alterations in frequency domain properties, and inconsistencies in quantization tables. Bycomparing these properties to the expected statistical characteristics of uncompressed images, suspicious patterns suggestive of concealed data can be identified [22].

ML techniques have been integrated into JPEG-compression steganalysis algorithms to enhance detection capabilities [73–75]. These algorithms can be trained using datasets that include known instances of steganographic content hidden within JPEG-compressed images and cover images without any hidden data. This training enables the algorithms to learn the statistical patterns associated with both types of images, enabling accurate identification of steganography in new JPEG-compressed images. This method achieved an adequate performance in JPEG-compression steganography steganalysis. Based on the JPEG-compressed algorithm, in Ref. [76], a steganalysis algorithm specifically detects the JSteg steganography for stego images in JPEG. Yu applies a generalized Cauchy distribution to model the DCT coefficients distribution of JPEG images. With his algorithm, the original image's histogram of the DCT coefficients of the stego is estimated. Based on the experimental results, Yu's algorithm outperforms the algorithm by Fridrich [77], which combines feature-based steganography and the calibration concept of JPEG image metadata.

In line with the JPEG-compression steganography attack, the works in Refs. [78,79] have been proposed by modelling the original cover images with hyper-dimensional approaches. Their methods created geometric models in the attribute's spaces with hyper-ellipsoid or convex polytopes. However, the hyper-spheres are sometimes used to create the geometric model in the attribute space. The detection of the altered regions of the stego images is identified by comparing it to the file model of the cover image; moreover, departing from Markov empirical matrices to detect the stego images generated with JPEG steganography algorithms. By inspecting the change in the intra- and inter-block dependencies among the DCT coefficient blocks, any alteration caused by data embedding is identified; hence, the steganography is detected in the JPEG image.

2.2.1.6. F. transform domain steganalysis. Transform domain steganalysis refers to detecting and analyzing confidential data within digital media by examining the transformed representation. The transform domain steganography consists of adding the personal data in the high-frequency bands' coefficients of an image because they represent the edge areas of an image and are surrounded by excellent coefficients [80]. Transform domain steganalysis algorithms are specifically developed to examine the statistical characteristics and patterns inherent in the transformed coefficients of digital media. The primary objective of these algorithms is to identify deviations or irregularities that could potentially suggest the existence of confidential data. The selection of an appropriate transform for analysis is contingent upon the steganography technique under investigation. Prominent transforms employed in such analyses encompass the Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), and a range of wavelet transforms. By analyzing the smoothness of the histogram shape for both the cover and the stego images, wavelet domain quantization and modulation are applied to detect the presence of hidden data [81].

Transforming domain steganalysis is integral in identifying and mitigating potential risks linked to covert information embedded in digital images. Its significance extends to diverse domains, including digital forensics, security assessments, and the formulation of effective countermeasures against transform domain steganographic techniques. In line with transform steganalysis detection, the neural network has been introduced to classify images into cover or stego [82–84].

2.2.1.7. G. additive noise steganalysis. Additive noise steganalysis is a technique employed to detect and analyze the presence of hidden data within the statistical histograms of a digital image by examining the characteristics of the additive noise introduced during the embedding process. In additive noise steganography, secret information is concealed by adding imperceptible noise to the cover media. Hiding data in the statistical histograms or DCT coefficients creates noise on the residuals in the DCT [85,86]. Implementing three stages [87] presents a steganalysis scheme that first transforms an image with confidential data into the transform domain and then models the histograms of the DCT coefficients with generalized Gaussian Distribution. Secondly, MAP estimation is applied to predict the secret message, and finally, proceeds by predicting the location and size of the secret message by using the division operation of dissimilar Signal Noise Ratios (SNR) between the stego and the cover images with the method of segmentation based on results extrapolation and local variation. The privilege of this method is that it can be applied to both the colour and grayscale images.

Moreover [88], proposed a steganalysis scheme for additive noise steganography in a file with binary content. His steganalysis algorithm, known as boundary-based steganography detection, identifies hidden data in the characters' boundaries by combining the digitization noise and quantization from the small pixel's disturbances. This approach achieved better detection accuracy results.

2.2.2. Universal statistical steganalysis

Universal steganalysis is concerned with steganalysis methods not designed for specific steganography detection. This type of steganalysis aims to identify concealed data in digital media without relying on specific knowledge about the steganography technique employed [89]. Distinguishing from specific statistical steganalysis, which targets known steganographic methods, universal steganalysis focuses on detecting statistical irregularities that may indicate the presence of hidden information, irrespective of the specific steganography technique employed. To achieve this, algorithms are devised to examine the statistical properties and patterns within the media to reveal potential indicators of hidden data. These algorithms commonly employ mathematical models, statistical analysis techniques, and machine learning approaches to detect deviations from the expected statistical properties of the original media [32]. Universal statistical steganalysis proves particularly beneficial in scenarios where the specific steganography techniques are unknown or when dealing with emerging or novel steganographic methods [18]. Its versatile nature allows for application across diverse image types and steganography techniques, rendering it a valuable tool in digital forensics, security assessments, and identifying potential threats related to confidential information.

The work [90] proposed the first universal steganalysis algorithm that was improved in late research [91–93] for improved accuracy in detecting any possible hidden message in digital images regardless of the steganography algorithm. Universal statistical steganalysis can also be seen in two subdivisions: spatial and transform universal statistical steganalysis.

2.2.2.1. A. spatial universal statistical steganalysis. Spatial universal statistical steganalysis is a technique that aims to detect hidden data within digital media by analyzing their spatial characteristics and properties [9]. Distinguishing from specific spatial steganalysis methods that target known spatial steganography techniques, this approach identifies statistical irregularities or anomalies indicative of concealed information, regardless of the specific spatial steganography technique employed. The algorithms utilized in spatial universal statistical steganalysis are designed to analyze the statistical properties and patterns in the spatial domain of the media [94]. They integrate a range of mathematical models, statistical analysis techniques, and ML approaches to identify deviations from the anticipated statistical properties of the cover media [47]. The adaptability and versatility of spatial universal statistical steganalysis render it invaluable, particularly when encountering unknown or emerging spatial steganography techniques [95].

2.2.2.2. B. transform universal statistical steganalysis. Transform universal statistical steganalysis encompasses a steganalysis approach to identify concealed information within digital media by scrutinizing transformed coefficients' statistical properties and patterns [96]. In contrast to specific transform steganalysis techniques targeting known transform-based steganography methods, the objective of transforming universal statistical steganalysis is to detect statistical deviations indicative of hidden data, irrespective of the specific transform-based steganography technique employed. To achieve this, algorithms [97–99] are devised to scrutinize the statistical characteristics and patterns exhibited by the transformed coefficients of the media. These algorithms integrate mathematical models, statistical analysis techniques, and machine learning approaches to detect deviations from the anticipated statistical properties of the cover media. The training phase involves utilizing extensive datasets encompassing their cover media and instances of steganographic content [97]. By understanding the statistical patterns associated with concealed data, algorithms become proficient in detecting and identifying confidential information within new media [98]. Transform universal statistical steganalysis proves highly valuable in scenarios where the specific transform-based steganography techniques are unknown or when confronting emerging or novel steganographic methods within the transform domain. Its versatility and generality make it an invaluable tool in digital forensics, security assessments, and identifying potential threats connected to hidden information within the transformed coefficients of digital media.

3. .Shift from ML-based to DL-based steganalysis paradigm

The evolution of technology that recently gained great consideration results in several broad research fields, including artificial intelligence (A.I.), plays a considerable part. AI encompasses techniques that target producing artificial systems to mimic human behaviours with various cognitive capacities. With the introduction of AI techniques, it has been possible to design intelligent systems, building smart cities with humanoids to perform tasks in smart production systems [100]. Researchers have recently used complex programming programs such as expert systems and searching trees to develop intelligent systems [101]. However, the existing systems that apply the handcrafted methods showed an inability to handle complex tasks such as natural language processing and other digital media recognition. The limitations of the handcrafted approaches made the introduction of ML imperative to handle the metadata and data of complex systems.

Referring to Fig. 3, ML is a subfield within artificial intelligence. It provides smart systems with intelligence for self-learning based on the data to imitate human behaviours and data analysis to model and manage the patterns of a specific task [102]. Even though ML systems attain better efficiency in tackling complex problems, they have some drawbacks. Building an effective feature extractor with these systems seems impossible because it imposes the intelligence and expertise of a specific system with knowledge in the aimed field. In addition, an extension of the feature extraction output has also been challenging, and the introduction of DL, a subset of machine learning, revolves around models that iteratively learn to perform classification directly from input data. Feedforward neural networks, known as the fundamental deep learning models [103], consist of multiple layers of artificial neurons called perceptron, enabling them to exhibit greater depth than external neural networks comprising only one or two layers. The nature of the ML system's data set is that the system's learning is either supervised or unsupervised.

Supervised learning: In line with the concept of ML, supervised learning focuses on learning the features that best reflect the relationship between input $x^{(i)}$ and the related label $y^{(i)}$ supervised regression. The supervised learning approach and classification are basic examples [104]. Considering D, a dataset, s the number of samples in D, $x^{(i)}$ is the data sample, and $y^{(i)}$ is the data label, supervised learning is mathematically expressed as (7).

$$D = \left\{ \left(x^{(1)}, y^{(1)}\right), \ldots, \left(x^{(s)}, y^{(s)}\right) \right\} \tag{7}$$

Unsupervised learning: Unlike supervised learning, unsupervised learning is an ML-based approach in which the data do not have

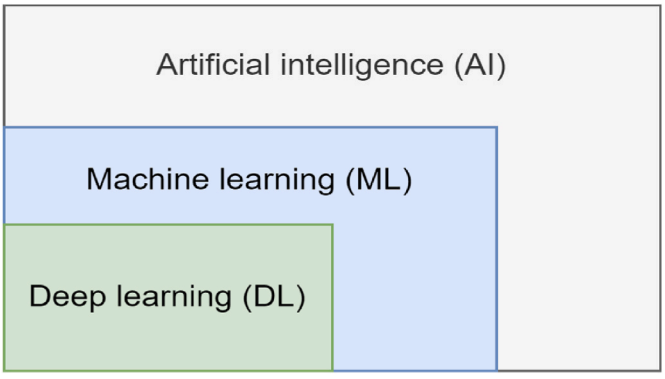


Fig. 3. The relationship between AI, ML, and DL.

associated labels, and the dataset is only enough. Unsupervised learning seeks the data representation that is the most appropriate; for example, the K-mean [104] Unsupervised learning method is based on finding the dependencies and structures between the samples within the dataset D (see (8)).

$$D = \{x^{(1)}, \ldots, x^{(s)}\} \tag{8}$$

Throughout the last two decades, a plethora of steganalysis techniques have emerged. Conventionally, these techniques have relied on a two-stage process to tackle the problem of steganography detection. The initial phase involves extracting manually crafted features to effectively capture the embedding distortions induced by any data-hiding technique applied to an image. The next phase involves classification, where binary SVM or EC [105] is employed to determine hidden information or predict the specific steganographic algorithms or tools for embedding the secret data. A notable challenge these techniques face is dimensionality, commonly called the curse of dimensionality, which impacts the classifiers' performance [30]. Therefore, ML techniques steganalysis involve the painstaking process of crafting a vast number of steganalysis features that are highly sensitive to changes in embedding, irrespective of variations in image content across different cover images. DL-based techniques, which have emerged as the state-of-the-art approach for image classification and recognition tasks, are applied to alleviate this arduous task. In ML-based steganalysis methods, the feature extraction and classification stages lack backward communication, with information flowing directly from the feature extraction stage to the classification stage. In contrast, with the DL-based method, these two stages are integrated into a single cohesive stage, where the decision made by the classification stage is utilized to update the features being extracted in the feature extraction stage.

4. .DL and CNN steganalysis

4.1. General paradigm for DL and CNN

The introduction of the DL in the steganalysis domain reduced the complexity of the traditional ML-based methods because the DL-based approaches unify the feature extraction and classification stages in one phase, and the features are learned directly from the input [106]. Implementing steganalysis with DL-based algorithms enables backward communication between the feature extraction and the classification stages; this improves a steganalysis system's performance [107]. DL-based algorithms include shallow neural networks that are made of at least one perceptron, otherwise known as the layer of the neurons, and deep neural networks capable of containing hundreds of layers. DL models are mainly classified as RNNs, autoencoders, and CNNs.

The CNNs are the DL models specialized in manipulating two-dimensional data in a grid fashion, as in the case of image pixels [100]. CNNs are based on the convolutional layer, a critical feature that makes a convolutional block. The purpose of a convolutional layer is to show the existence of crucial components in an image using the convolutional operation. The convolutional process consists of using the weighted sum of the neighbours of a pixel instead of the pixel $I_{(i,j)}$. When the weights are put into a set, a kernel K is derived, such as $K \in \mathbb{R}^{(2k_1+1)\times(2k_2+1)}$ with k_1 the horizontal size of the neighbourhood, and k_2 the vertical size of the neighbourhood. The convolutional operation scenario is mathematically represented by $O_{(i,j)}$ as of (9).

$$O_{(i,j)} = (I * K)_{(i,j)} = \sum_{u=-k_1}^{k_1} \sum_{v=-k_2}^{k_2} I_{(i+uj+v)} k_{(u,v)} \tag{9}$$

Moreover, an activation operation also takes place in the convolutional layer. Activation is a non-linear function that comes after a convolutional process in the convolutional block. This operation introduces non-linearity with the advantage of making possible stacked layers in a network. Since an activation function outputs feature maps that are identical to the input features map in terms of dimension, it is classified as an operation of the elements-wise level. Sigmoid, tanh, and relu are the three main activation functions. Fig. 4 shows the graphs of the primary activation functions. The sigmoid function is mathematically expressed as (10) and is referred to as a logistic function by default.

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{10}$$

The sigmoid function aims to address the non-derivability problem of the Heaviside function. However, this function presents a couple of issues, namely, nullifying the derivative results out of the interval [−3, +3], which can be a drawback to learning [74]. This problem is called the vanishing gradient problem concerning backpropagation because the slope of early layers results from multiplying the last layers' gradients. Thus, for the gradients of later layers to be mathematically negligible, their products will also be insignificant, leading to the vanishing slopes of earlier layers. Tangent hyperbolic (*tanh*) is another activation function that is best known due to its ability to model the inputs with any value, either positive or negative, based on its results that are zero-centered. The drawback of this function is somehow the same as that of the sigmoid function to vanish the gradients. The tanh is mathematically expressed in (11).

$$tan\, h(x) = \frac{e^{2x}-1}{e^{2x}+1} \tag{11}$$

Rectified linear units (ReLu) represented by (12) and (13) are known for their non-linearity in assisting in the normalization of small values of the inputs. Other benefits of this function are that its derivative is easy to compute and the learning time is low. This function is mainly used by default based on its advantages. However, the output gradients of ReLu also get nullified if the inputs are nearly zero or negative, making the learning and backpropagation operations impossible.

$$relu(x) = \max(0, x) \tag{12}$$

$$relu(z) = \begin{cases} 0 \; if \; x < 0 \\ x \; else \geq 0 \end{cases} \tag{13}$$

Moreover, there is a need to normalize input data before training a CNN; hence, a normalization layer in the block is needed. The normalization function is used to avoid any saturation from the activation functions, such as the sigmoid function, by normalizing the data in the same interval of values [108]. The most used normalization layer is known as batch normalization (BN) and is expressed in (14) with $\Re^{[l]}$ represents a normalized output (activation maps); hence, the activation map is maintained.

$$\Re^{[l]} = BN(A^{[l]}) \tag{14}$$

Toward the classification stage, a convolutional network needs a pooling layer; the pooling layer consists of downsampling the size of the activation maps. The basic operations performed in this layer are the reduction of the parameter number, the requirements for calculation, the compaction of the activation maps for a more efficient network, and finally, the minimization of the likelihood of overfitting [108]. Commonly, the pooling layer follows other layers because it does not have weights for updating. Two parameters, namely the pooling stride $P_s^{[l]}$, and the pooling window $W^{[l]}$ are necessary to define a pooling layer.

Fig. 5 shows the preferred pooling methods: average and maximum. Maximum pooling, or max pool, selects only the highest value and rejects the remaining in the window. Max pool mainly ensures the translational covariance to make CNN focus on extracting features and reduce the sensitivity for the spatial location of the features. Differently, the average pooling, notated as average pool, considers all the feature map values and uses the mean of all the window region's values. The main benefit of the average pool is to achieve generalized results by

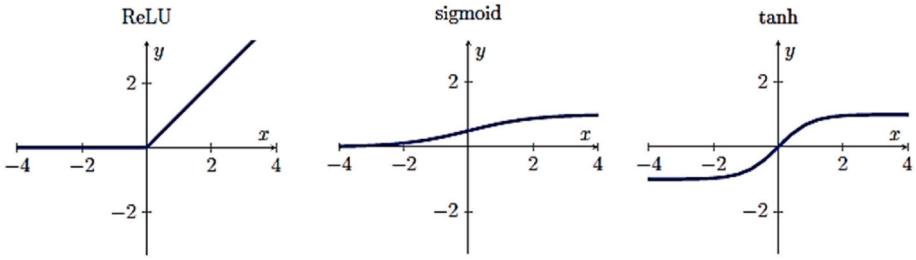


Fig. 4. The graphs of three main activation functions.

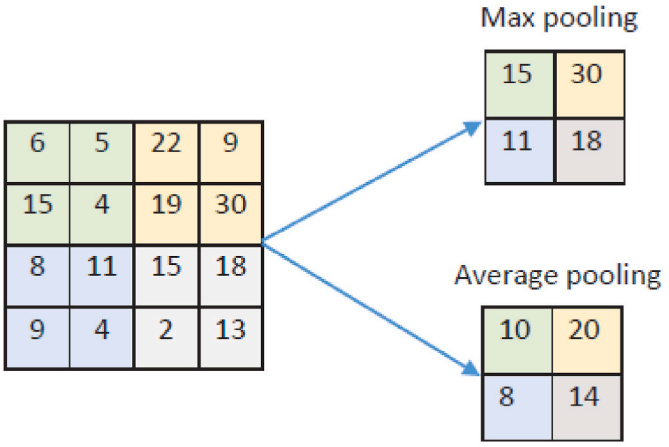


Fig. 5. Maximum and average pooling.

combining various values in one value. This reduces the overfitting issue due to a combination of various features. In addition to the max and average pooling, there is another method called spatial pyramid pooling (SPP) that generates a fixed-size output regardless of the input size, which benefits the network because the highest connected layer must be fixed.

Moreover, SPP admits images of varying sizes because the output features are independent of the input size. After different convolutional blocks for feature extraction, the classification operations follow. These are performed through a classification module that predicts the category or class of the image, which is made of one or more sequential fully connected layers before a loss function.

The fully connected layer is made of neurons in a complete connection fashion with the neurons of the preceding layer. The fully connected side of the CNN is applied to associate the features of an input image to reduce the classification error. The ending part of the fully connected layer must be made of a loss function such as softmax function, sigmoid/softmax cross-entropy, and mean squared error to specify the rate of penalization of the training over the deviation between the assumed and the correct labels when applying the stochastic gradient descent.

The performance of the steganalysis algorithms based on DL is evaluated based on their detection level, which is mainly shown from the existing literature by detection accuracy and prediction error. However, other metrics such as sensitivity, specificity, and cross-entropy loss (L) are of capital importance in evaluating the performance of a steganalysis algorithm applying DL. Considering four classification results, false negative (FN) which represents a class of stego images predicted to be the cover images; true negative (TN), which stands for a class of cover images predicted to be cover images; true positive (TP) which stands for a class of stego images predicted to be stego images, and false positive (FP) which represents a class of cover images predicted to be stego images, we derive P_{FA} as of (15) to represent the false alarm probabilities and P_{MD} to represent the miss-detection probabilities as of (16). To compute the sensitivity, specificity, and accuracy, we use the relations in (17), (18), and (19), respectively. The prediction error is calculated based on the values of the false alarm probabilities P_{FA} and miss-detection probabilities P_{MD} as of (20). We calculate the cross-entropy loss from y_i which represents the label of the sample x_i with $\delta(.)$ an impulse function, N which represents the number of the training samples with K representing the number of all labels, and $O_{ik}(x_i, \theta)$ which represents the output of the i^{th} sample x_i at the k^{th} label as of (21).

$$P_{FA} = \frac{FP}{FP + TP} \tag{15}$$

$$P_{MD} = \frac{FN}{TP + FN} \tag{16}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{17}$$

$$Specificity = \frac{TN}{TN + FP} \tag{18}$$

$$DACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$PE = \frac{1}{2} min_{P_{FA}} (P_{FA} + P_{MD}) \tag{20}$$

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \delta(y_i = k) \log \left(\frac{e^{O_{ik}(x_{i,\theta})}}{\sum_{k}^{K} e^{O_{ik}(x_{i,\theta})}} \right) \tag{21}$$

4.2. Systematic review of the state-of-the-art in DL-based steganalysis schemes

4.2.1. Method for literature selection

To systematically choose the literature to be used in the next sections of this paper, we avail a systematic review of 24 state-of-the-art articles selected following the PRISMA scheme [109] to assess the present state of DL in steganalysis; it is essential to gather information on scientific literature systematically. Various databases have been explored to locate pertinent articles in line with the state-of-the-art. Articles in highly reliable academic digital libraries include various topics that may be divergent if they are all included in this SLR. Therefore, we follow PRISMA frameworks to set a list of relevant keywords. We retrieved the papers from six highly reliable academic libraries considered quality academic databases, as recorded in Table 1, with many articles retrieved from IEEE Xplore and Springer. The search keywords for relevant articles on steganalysis in spatial domain images using DL are in two categories concatenated with an "AND" operator to form appropriate search queries. The first category of DL-related keywords includes DL OR DNN OR CNN. The second category consists of the words associated with steganalysis. These words include steganalysis OR steganalysis system OR passive steganalysis OR universal steganalysis. To form the search query, we use the AND operator to link the keywords of category 2.

For literature exploration, a methodological background description of the methods in digital image steganalysis is conducted based on the

Table 1
Considered research works for this survey.

Title	Year of publication	Library	Proposed Algorithm
Stacked Convolutional Auto-Encoders for Steganalysis of Digital Images [33]	2014	IEEE Xplore	SCAE
Deep learning for steganalysis via convolutional neural Networks [131]	2015	Springer	QianNet-V1
Structural Design of Convolutional Neural Networks for Steganalysis [119]	2016	IEEE Xplore	XuNet
Learning and Transferring Representations for Image Steganalysis using Convolutional Neural Network [111]	2016	IEEE Xplore	QianNet-V2
JPEG-Phase-Aware Convolutional Neural Network for Steganalysis of JPEG Images [118]	2017	ACM	PHASE-AWARE CNN
Deep Convolutional Neural Network to Detect J-UNIWARD [115]	2017	ACM	J-UNIWARD-DETECTION-CNN
Deep Learning Hierarchical Representations for Image Steganalysis [112]	2017	IEEE Xplore	YeNet
YEDROUDJ-NET: An Efficient CNN for Spatial Steganalysis [113]	2018	IEEE Xplore	YedroudjNet
Steganalyzing Images of Arbitrary Size with CNNs [140]	2018	Elsevier	ModifiedYeNet
Efficient feature learning and multi-size image steganalysis based on CNN [136]	2018	Cornell University Library	ZhangNet
Deep Residual Network for Steganalysis of Digital Images [114]	2019	IEEE Xplore	SRNet
CIS-Net: A Novel CNN Model for Spatial Image Steganalysis via Cover Image Suppression [141]	2019	IEEE Xplore	CIS-Net
A novel steganalysis method with deep learning for different texture complexity images [142]	2019	Springer	TCI-Net
Joint multi-domain feature learning for image steganalysis based on CNN [143]	2020	Springer	Wang-Net
A Steganalysis framework based on CNN using the filter subset selection method [144]	2020	Springer	Wu-Net
CNN-based image steganalysis using additional data embedding [145]	2020	Springer	Kim-Net
Steganalysis of convolutional neural network based on neural architecture search [146]	2021	Springer	Pan-Net
Steganalysis using learned denoising kernels [147]	2021	Springer	Singh-Net
GBRAS-Net: A Convolutional Neural Network Architecture for Spatial Image Steganalysis [137]	2021	IEEE Xplore	GBRAS-Net
Spatial Steganalysis Based on Non-Local Block and Multi-Channel Convolutional Networks [95]	2022	IEEE Xplore	Han-Net
CCNet: CNN model with channel attention and convolutional pooling mechanism for spatial image steganalysis [47]	2022	Elsevier	CCNet
Image steganalysis based on attention augmented convolution [148]	2022	Springer	SAA-Net
A convolutional neural network to detect possible hidden data in spatial domain images [96]	2023	Springer	Jean-Net
Self-attention enhanced deep residual network for spatial image steganalysis [93]	2023	Elsevier	Xie-Net

purpose of this research, which is to add significant value to the existing literature. This literature review also demonstrates the evolution of steganalysis through the implementation of DL in recently published articles. This review provides a novel point of view by emphasizing noteworthy achievements from the existing published research report in scientific research articles and potential areas for further research. The systematic review of the literature included 24 papers that were selected from high-impact academic databases. Table 1 chronologically displays the title, year of publication, digital library, and proposed algorithm (using its original representing abbreviation or given an abbreviation to ease its representability). The table demonstrates that this topic has gained significant attention in the last nine years based on the frequency of publications. It is also identified that the primary contributions of the research works achieved the development of various CNNs built upon their predecessors' successes.

4.2.2. State-of-the-art in steganalysis

To identify the progress and the state-of-the-art in steganalysis through the application of DL, mainly the CNN-based techniques, we thoroughly analyze some of the existing works in the steganalysis of digital images by focusing on experimentation and the primary contribution. In line with the implementation of the state-of-the-art steganalysis schemes, it is worth noting that concerning the experimentation and the used datasets, all the explored existing schemes used various datasets based on the domains and steganographic algorithms they pose their steganalysis attack. Referring to the schemes presented in Table 2, the Spatial Color Autoencoder (SCAE) employs a nine-layer convolutional architecture for spatial domain steganalysis, focusing on the BOSSBase 1.01 dataset with the HUGO algorithm [33]. QianNet-V1 [131] and XuNet [119] also operate in the spatial domain, utilizing the BOSSBase 1.01 dataset with steganographic algorithms such as HUGO, WOW, S-UNIWARD, and HILL. QianNet-V2 [110] follows a similar pattern but with different pre-processing layers. The Phase-

Aware [118], CNN introduces two architectures (PNet and VNet) for JPEG steganalysis using BOSSBase 1.01 and focusing on the J-UNIWARD algorithm [115]. BOWS2 and J-UNIWARD-DETECTION CNN also target JPEG steganalysis with specific convolutional structures. YeNet [112] and YedroudjNet [113], JeanNet [96], and XieNet [93], and concentrate on spatial domain steganalysis, utilizing BOSSBase 1.01 with algorithms like WOW, S-UNIWARD, HILL, and YedroudjNet extending its use to ImageNet with WOW and S-UNIWARD. Each algorithm's unique architecture and domain specialization contribute to a diverse and well-rounded choice of experimental datasets.

Moreover, in Table 2, the researchers explored various methods of using CNNs in their experiments. These include manipulating the network height by applying the fully connected layers [110]. They have

Table 2
Experimental setup analysis of considered algorithms.

Algorithm	Architecture	Domain	Experimentation steganographic algorithm
SCAE	−9 convolutional layers - Max pooling −1 FC - Softmax function	Spatial	HUGO
QianNet-V1	−5 convolutional layers −3 FC - Softmax function	Spatial	- HUGO - WOW - S-UNIWARD
XuNet	−1 Pre-processing layer −5 convolutional layers −2 FC - Softmax function	Spatial	- S-UNIWARD - HILL
QianNet-V2	−1 Pre-processing layer −5 convolutional layers −2 FC - Softmax function	Spatial	- S-UNIWARD - WOW
PHASE-AWARE CNN	Proposed two CNNs 1. PNet: Architecture is: −3 convolutional layers −3 FC - Softmax function 2. VNet: Architecture is: −2 pre-processing layers −5 convolutional layers - 1FC - Softmax function	JPEG	J-UNIWARD UED-JC
J-UNIWARD-DETECTION CNN	−16 fixed DCT filters −20 convolutional layers −1 FC - Softmax function	JPEG	J-UNIWARD
YeNet	−8 convolutional layers −1 FC -Softmax function	Spatial	WOW S-UNIWARD HILL
YedroudjNet	−30 SRM filter banks in the Pre-processing phase −5 Convolutional layers −3 FC -Softmax function	Spatial	WOW S-UNIWARD
ModifiedYeNet	−30 SRM filter banks in the Pre-processing phase −8 Convolutional layers −1 FC -Softmax function	Spatial	WOW LSBM
ZhangNet	−30 SRM filter banks in the Pre-processing phase −9 Convolutional layers −1 FC -Softmax function	Spatial	WOW S-UNIWARD
SR-Net	−8 convolutional layers −2 FC	Spatial JPEG	WOW HILL S-UNIWARD J-UNIWARD UED-JC
CIS-Net	−5 convolutional layers −1 FC	Spatial	WOW HILL S-UNIWARD
TCI-Net	−5 convolutional layers −1 FC - Softmax function	Spatial	WOW S-UNIWARD
Wu-Net	−50 SRM filter banks −5 Convolutional layers −1 FC -Softmax function	Spatial	WOW HUGO S-UNIWARD
Kim-Net	Dual networks with Net 1: −5 Convolutional layers −1 FC Net 2: −5 Convolutional layers −2 FC	Spatial	S-UNIWARD
Pan-Net	−12 Convolutional layers −1 FC -Softmax function	Spatial	WOW S-UNIWARD
GBRAS-Net	−30 SRM filter banks −13 Convolutional layers -Softmax function	Spatial	WOW HUGO S-UNIWARD HILL MiPOD
Han-Net	−30 SRM filter banks −6 Convolutional layers −3 FC -Softmax function	Spatial	WOW S-UNIWARD HILL
CCNet	−4 Convolutional layers −2 FC -Sigmoid function	Spatial	WOW S-UNIWARD HUGO
SAA-Net	−6 Convolutional layers −1 FC -Softmax function	Transform (JPEG)	J-UNIWARD JC-UED
Wang-Net	−8 Convolutional layer −1 FC -HPF from SRM - DCTR	Spatial	WOW S-UNIWARD
Singh-Net	−5 Convolutional layers - HPF and Denoising channels −3 FC - Softmax function	Spatial	WOW S-UNIWARD HUGO
Jean-Net	−8 Convolutuional layers −4 DepthwiseSeparable layers −3 FC -Softmax function	Spatial	WOW S-UNIWARD
Xie-Net	- Architecture not specified - Weighted histogram method - HPF	JPEG	J-UNIWARD SI-UNIWARD UERD

also experimented with custom activation features to enhance steganographic payload detection performance and yield a converging network [111–113]. Another approach has been to use CNNs with jumps among the layers of the CNN, such as Residual Networks or Dense Networks, to design deep CNNs that achieve network convergence and improve detection rates [140–148]. In addition, researchers have transferred learned parameters from training sets of CNNs to networks where convergence is complex or detection ability is low [111,114, 136-137] To identify the efficiency of the proposed networks (robustness against cover-source Mismatch), the researchers trained their models with a database different from the one used to test the model's ability to generalize the results [110–118].

Improving statistical modelling through the implementation of an absolute value function (in the ABS layer) has been explored in various studies [113,119]. Another approach involves placing two CNNs in competition, with the first network performing steganography and the second performing steganalysis, to achieve an automatic steganographic process by learning from the characteristics of the two processes [62,95, 120–123]. Researchers have also trained a network to classify images with high resolutions from low-resolution images and predict the payload capacity under a quantitative steganalysis paradigm applying DL in both the spatial and JPEG domains [124]. To increase the size of the database, researchers have considered trimming, rotating, and interpolating operations and using cameras with the same or different

characteristics for image taking while carefully resizing [108,112,113,125,126]. In another approach, three CNNs are placed to work in parallel, with each network using different filters in the pre-processing layer, inspired by Gabo Filters [127] and SRM, and activation functions (ReLU, Sigmoid, and TanH) [128]. Finally, researchers have used a similar approach to the previous one by using colour images [129].

Most CNNs for the steganalysis in digital images apply high-pass filter banks as in (22). These filter banks were initially developed in Ref. [130] and applied for steganalysis tasks in Ref. [131]; high-pass filter parameters were not optimized during the training phase. By doing so, the filter aids in achieving convergence during the CNN training, which can be slow or non-existent if convergence is not employed. However, the most recent CNN designs do not incorporate this filter; instead, they use a filter bank recommended by SRM to obtain residual characteristic maps.

$$F=\frac{1}{12}\begin{pmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{pmatrix} \tag{22}$$

To express the general operation performed in the CNN as shown in (23), we consider M^n as the characteristic map of the $n-th$ layer, M_i^{n-1} as the i^{th} characteristic map of the preceding layer. We also consider H_i^n as the $i-th$ kernel of the $n-th$ layer, b^n as the bias variable (parameter) for the layer at the position n. We use operator $(\times)$ to express the convolutional operation and $f(.)$ as the non-linear operation, the activation function. For the pooling operation, we use $pool\,(.)$ and $norm\,(.)$ for the normalization operator. The convolutional layers in CNNs perform operations in a specific order: first convolution, followed by normalization, activation function, and pooling. Once the final layer generates the feature maps, they are fed into the classification module. This module comprises one or more fully connected layers of neurons and a Softmax layer. The final layer plays a crucial role in normalizing the CNN's values between 0 and 1, representing the probability of an image being classified as either cover or stego. The CNNs investigated in the study use various non-linear activation functions, including Rectification Linear Unit (ReLU), Tangent Hyperbolic (TanH) [132], Gaussian, and TLU. TLU is a unique activation function used specifically for steganalysis in deep learning. Its purpose is to restrict the range of values and prevent the network from modelling large values. TanH is typically used in the initial layers, while ReLU is used in the later ones.

$$M^n=norm\left(pool\left(f\left(\sum_{i=1}^{k}\left(M_i^{n-1}\times H_i^n\right)+b^n\right)\right)\right) \tag{23}$$

Data normalization is performed using BN, as described in equation (4) taken from Ref. [133]. BN normalizes each feature map's distribution to have a zero mean and unit variance. It may also rescale and retranslate the distribution if required. To express the BN operation mathematically, we consider a random variable X depending on the realization of a feature map x such as $x\in\mathbb{R}$. Its mathematical representation is given in (24), considering $E[X]$ as the expectation $Var[X]$ as the variance. The variables γ and β represent the re-scaling and re-translation operations.

$$BN(x,\gamma,\beta)=\beta+\gamma\frac{x-E[X]}{\sqrt{Var[X]+\varepsilon}} \tag{24}$$

During each batch, the expectation ($E[X]$) as well as the variance ($Var[X]$) of the data are updated while the parameters γ and β are learned using the back-propagation paradigm. Implementing BN reduces learning sensitivity to parameter initialization, enables higher learning rates that accelerate the learning process, and improves classification accuracy [118]. The first CNN proposals did not incorporate BN. In CNNs, Average Pooling [134] is widely employed for pooling because the steganographic noise introduced during the embedding process is generally weak. Average Pooling tends to propagate and retain this noise, which is not valid with Max Pooling [134]. The pooling operation used is usually a local computation involving neighbouring pixels.

Based on the existing works, it is noteworthy that traditional CNNs were originally designed to process spatial domain input. However, some CNN methods have been modified to accommodate JPEG images in the transform domain; these CNNs include Qian-Net [131,135], Xu-Net [119], Ye-Net [112], Yedroudj-Net [113], Zhu-Net [136], and GBRAS-Net [137]. Qian-Net uses five convolutional layers with Gaussian function for activation and Average Pooling after each layer, followed by two fully connected layers and a Softmax layer. Xu-Net, as of Qian-Net, comprises five convolutional layers with an ABS layer that follows the first layer, tangent hyperbolic as activation functions in the two first layers, Rectified Linear Unit activation functions for the three ending layers, and Batch Normalization in each layer for convolutional operation. It also has two fully connected layers and a Softmax layer for classification operation. Ye-Net applied the SRM filter bank for steganographic noise extraction and has eight convolutional layers with a truncated linear unit activation function after the first layer and tangent hyperbolic activation for the remaining layers, followed by one fully connected layer and one Softmax layer for the classification operation. Yedroudj-Net as of Ye-Net, used the SRM filter bank with five convolutional layers, one ABS layer after the first layer, truncated linear unit activation function for the first two layers, Rectified Linear Unit activation for the three ending layers, use the average Pooling for layers from 2 to 5, and two fully connected layers followed by a Softmax layer in the classification stage. The architecture of this CNN combines and integrates the most effective characteristics of Xu-Net and Ye-Net. Zhu-Net is distinguished by its utilization of an SRM-based filter bank to initialize the weights of the preprocessing layer. These weights are then refined during the training phase to enhance the impact of noise caused by steganography while reducing the loss of image content. Zhu-Net enhances the feature extraction process by utilizing different convolutions.

Additionally, it employs a multi-level Average Pooling technique called Spatial Pyramid Pooling (SPP) [138]. This enables the network to analyze images of varying sizes, resulting in superior performance compared to Xu-Net, Ye-Net, Yedroudj-Net, and SRM + EC. GBRAS-Net is a CNN architecture that introduces a new approach to steganalysis. The architecture includes a preprocessing stage that employs filter banks to improve the detection of steganographic noise, a feature extraction stage using depthwise and separable convolutional layers and skip connections.

The architecture utilizes the 30 filters from YE-Net for image preprocessing, resulting in superior feature extraction capabilities through preprocessing. Each filter is normalized by its maximum absolute value. This set of filters consists of 8 class 1 filters, 4 class 2 filters, 8 class 3 filters, 1 square 3×3 filter, 4 edges 3×3 filters, 1 square 5×5 filter, and 4 edges 5×5 filters. In Figs. 6 and 7, we present the architectures of the CNNs considered crucial in this review. It is worth noting that the content in the boxes indicates the size of the kernels used for convolutional layers, represented as the number of kernels multiplied by their height, width, and number of feature maps. The content outside the boxes indicates the size of the resulting feature maps after convolution, represented as the number of feature maps multiplied by their height and width. If not specified, a default Stride value of 1 and a Padding value of 0 are assumed.

Table 3 provides a detailed analysis of the outcomes produced by the considered algorithms, shedding light on their primary contributions and best results in terms of error percentage, benefits, and weaknesses. This structured examination offers a comprehensive overview of each algorithm's performance, enabling a nuanced understanding of its strengths and limitations. Fig. 8 shows the average probability error (percentage) for detecting stego images under the S-UNIWARD steganographic algorithm with BOSSBase 1.01. The state-of-the-art CNN algorithms' performance is measured at various payload capacities

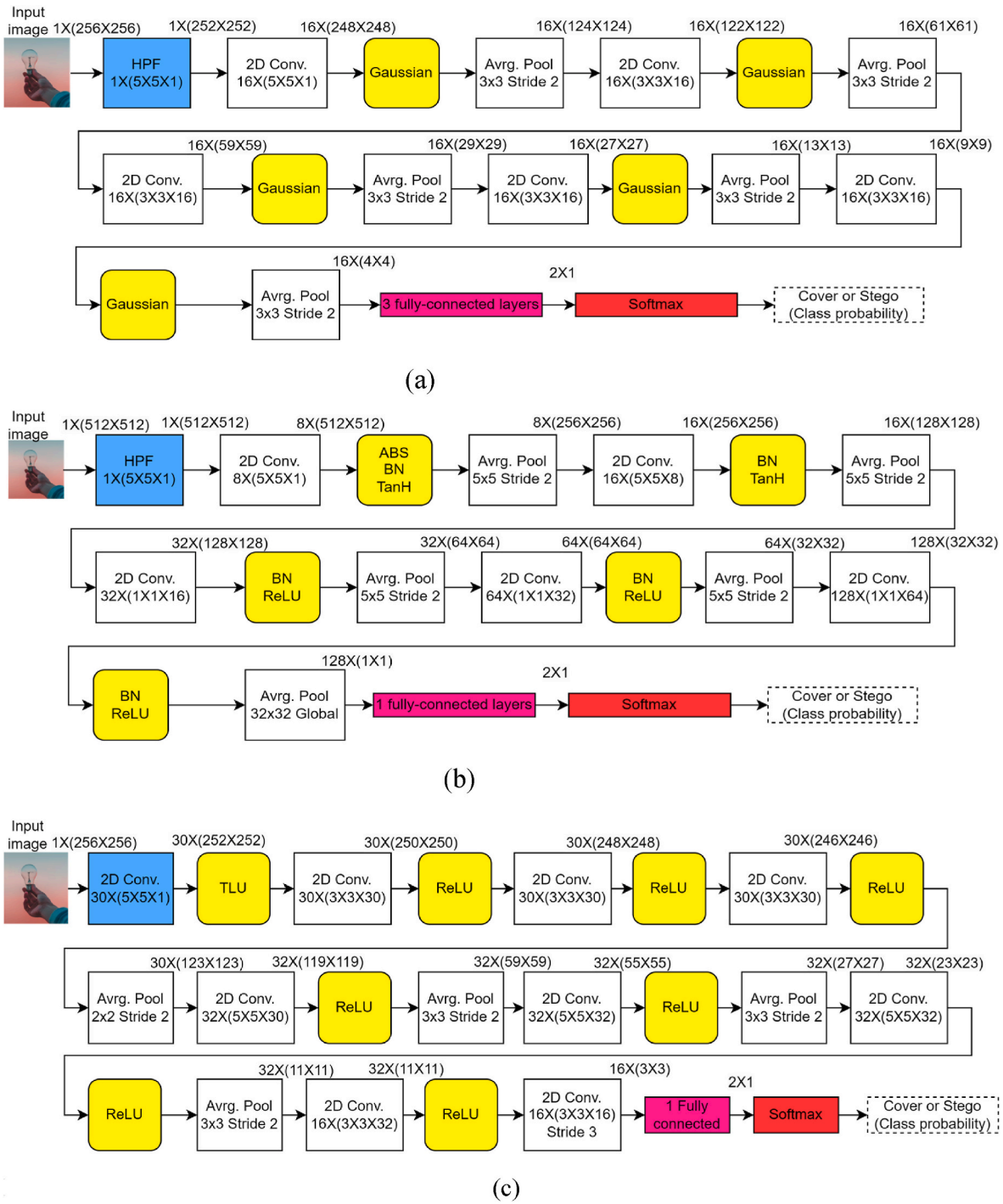


Fig. 6. (A) Qian-Net (b) Xu-Net (c) Ye-Net. (c) Legend: 1. The notation in the boxes of the form $a \times (b \times c \times d)$ is used to show the structure of the layer with a: Number of kernels, b: height, c: width, d: number of input feature maps (The sign $\times$ serves as a link between the four elements). 2. The notation outside the boxes of the form $a \times (b \times c)$ is used to show the structure features feed to the next layer with a: Number of feature maps, b: height, c: width. 3. The meanings for the box colors are such as, white = convolutional layers, blue = preprocessing layer, yellow = auxiliary functions, green = pooling layer, pink = fully connected layer, red = softmax function. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

ranging from 0.05 *bpp* to 0.4 *bpp*. The results show that the average probability error of all algorithms decreases as the payload capacity increases. At 0.05 *bpp*, Ye-Net has the lowest average probability error of 49.0%, followed by QIAN-Net at 48.5%. However, the performance of the other algorithms becomes more competitive at higher payload capacities. At 0.1 *bpp*, Ye-Net and QIAN-Net have the lowest average probability error of 46.0%. At 0.2 *bpp*, Ye-Net still has the lowest average probability error at 40.1%, followed by the lowest average probability error at 15.06%. At 0.4 *bpp*, GBRAS-Net remains the top-performing algorithm with a moderate probability error of 11.2%. The results suggest that GBRAS-Net is the most effective algorithm for detecting stego images under the S-UNIWARD steganographic algorithm with BOSSBase 1.01, especially at higher payload capacities. However, Ye-Net and QIAN-Net are highly competitive, especially at lower payload capacities.

In Figs. 8 and 9, we illustrate the average probability errors yielded by using S-UNIWARD (see Fig. 8) and WOW (see Fig. 9); the payload

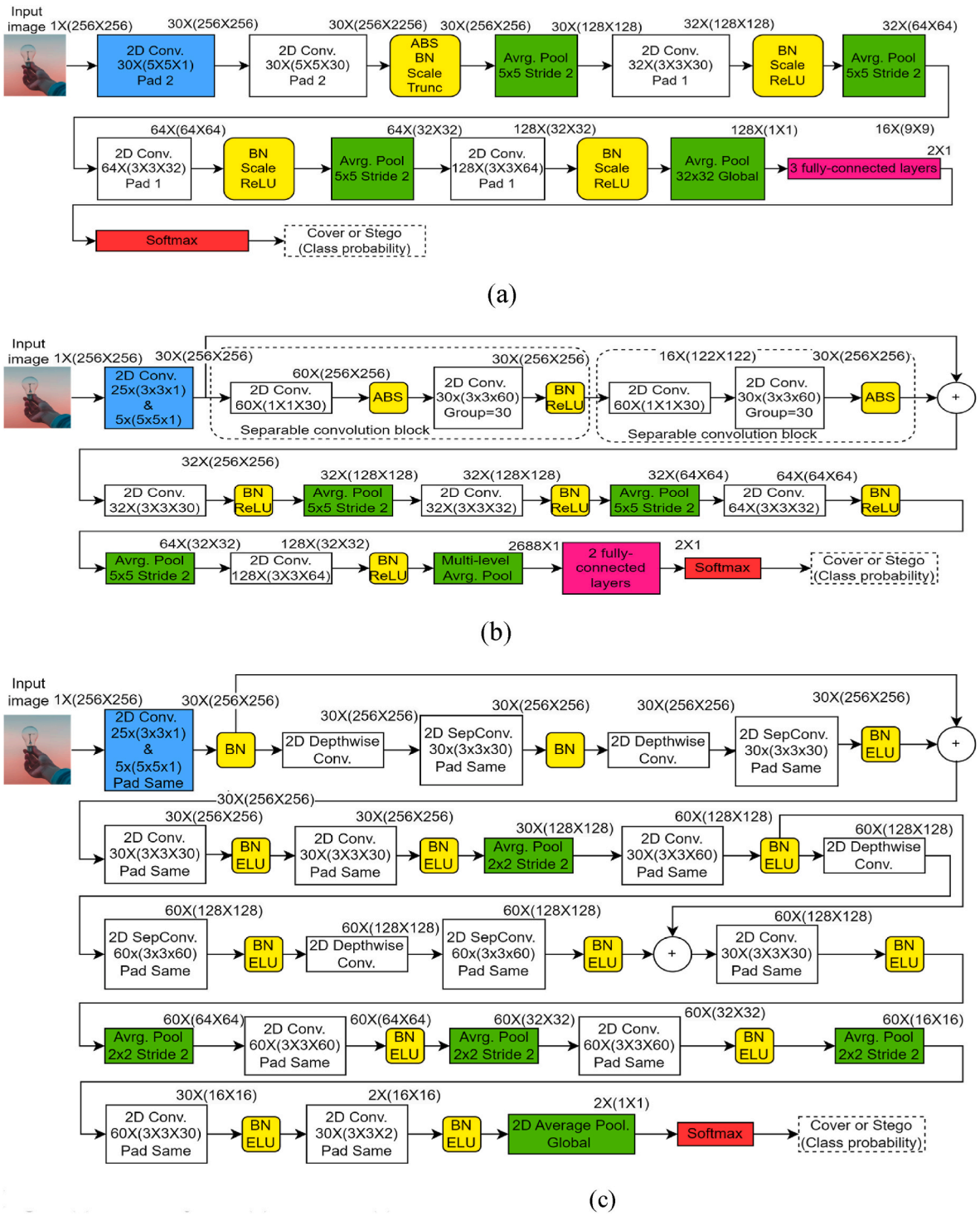


Fig. 7. (A) Yedroudj-Net (b) Zhu-Net (c) GBRAS-Net (c) Legend: 1. The notation in the boxes of the form $a \times (b \times c \times d)$ is used to show the structure of the layer with a: Number of kernels, *b*: height, *c*: width, *d*: number of input feature maps (The sign $\times$ serves as a link between the four elements). 2. The notation outside the boxes of the form $a \times (b \times c)$ is used to show the structure features feed to the next layer with a: Number of feature maps, *b*: height, *c*: width. 3. The meanings for the box colors are such as, white = convolutional layers, blue = preprocessing layer, yellow = auxiliary functions, green = pooling layer, pink = fully connected layer, red = softmax function. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

sizes from 0.05 *bpp* to 0.4 *bpp*. Fig. 9 shows the average probability error (in percentage) for detecting stego images using state-of-the-art CNN algorithms under WOW with BOSSBase 1.01. The payload capacity varies between 0.05 *bpp* to 0.4 *bpp*. The results indicate that as the payload capacity increases, the average probability error for detecting stego images decreases for all the algorithms. At a payload capacity of 0.05 *bpp*, GBRAS-Net has the lowest average probability error of 36.2 %, followed by QIAN-Net at 48.1 %. At 0.1 bpp, GBRAS-Net still has the lowest average probability error of 26.8 %, followed by Ye-Net at 42 %. At 0.2 *bpp*, GBRAS-Net remains the top-performing algorithm with an average probability error of 21.2 %, followed by Ye-Net at 34.1 %. At 0.3 bpp and 0.4 *bpp*, GBRAS-Net outperforms other algorithms with an average probability error of 13.8% and 9.3%, respectively. The results show that GBRAS-Net is the most effective algorithm for detecting stego

Table 3
Analysis of the results of the considered algorithms.

Algorithm	Primary contribution	Best results (Error percentage)	Benefit	Weakness
SCAE	The authors introduced a pioneering steganalysis approach using CNN and convolutional Auto-Encoders for pre-training, marking a departure from manual feature extraction in earlier methods like SRM and SPAM which signifies a notable shift toward automated and more efficient feature learning in steganalysis.	With 0.4 bpp - CNN = 31 % - SPAM = 42 % - SRM = 14 %	The obtained results are better than the ones achieved with SPAM	- Not enough deep and slow because of the use of fully connected (FC) layers
QianNet-V1	The authors employed high-pass filter banks to reduce image metadata and enhance steganographic noise, leading to promising results surpassing SRM and SPAM. Notably, this model represents the first steganalysis CNN using a supervised learning approach.	With 0.4 bpp under BOSSBase - HUGO = 28.29 % - WOW = 29.3 % - S-UNIWARD = 30.29 % Under ImageNet - HUGO = 33.6 % - WOW = 34.1 % - S-UNIWARD = 34.7 %	The obtained results were competitive to the state-of-the-art algorithms	- Slow due to the use of FC layers and the average pooling operation which takes longer.
XuNet	Deviating from the QianNet model, the authors introduced an innovative CNN with enhanced depth, a shallower architecture, and the integration of Rectified Linear Unit (ReLU), marking a significant advancement in network design.	With 0.4 bpp under BOSSBase - S-UNIWARD = 19.76 % - HILL = 20.76 %	Based on noise residuals to increase the performance of the CNNs to detect the secret bits.	- Less performing due to the not being deeper enough. - Takes long time for training.
QianNet-V2	The authors introduced a novel contribution involving a CNN that initially undergoes training with high payload capacity through a combination of regular and fully connected (FC) layers. Subsequently, the model is transferred to train with low payload capacity within the same algorithm, presenting a distinctive advancement in payload capacity management.	With 0.4 bpp under BOSSBase 1.01 - WOW = 21.95 % - S-UNIWARD = 22.05 %	Outperforming the SPAM and applies the Gaussian activation function.	Not enough deeper and achieving the results less than those of SRM
PHASE-AWARE CNN	The authors significantly contributed to the field by extending the application of XuNet to work seamlessly with JPEG domain images. This adaptation yielded two new CNNs, namely PNet and VNet, thereby expanding the versatility of the original model and offering tailored solutions for enhanced performance within the JPEG domain. This contribution reflects an innovative approach towards adapting existing architectures for specific image domains, fostering advancements in image analysis and processing.	With 0.1 bpnzAC QF 75 under BOSSBase (train and test), BOWS2 (test). - PNet: J-UNIWARD = 35.75 % UED-JC = 17.77 % - VNet: J-UNIWARD = 36.15 % UED-JC = 18.97 % With 0.4 bpnzAC QF 75 under BOSSBase (train and test), BOWS2 (test). - PNet: J-UNIWARD = 6.56 % UED-JC = 2.34 % - VNet: J-UNIWARD = 7.05 % UED-JC = 2.32 %	The overall detection accuracy is promising, and the VNet show a reduced computational complexity	The PNet presented a high computing complexity.
J-UNIWARD-DETECTION CNN	The authors achieved a noteworthy advancement by surpassing traditional manual feature extraction methods and successfully addressing vanishing gradients, ensuring the effective implementation of increased depth in the CNN. This accomplishment underscores the significant contribution of their work in enhancing feature extraction processes within neural networks.	With 0.4 bpnzAC QF 75 under BOSSBase: J-UNIWARD = 6.41 % With 0.4 bpnzAC QF 75 under ImageNet: J-UNIWARD = 16.8 %	Outperforms the state-of-the-art methods to detect J-UNIWARD	Narrow and increases the training errors when added the number of convolutional layers.
YeNet	The authors made a significant contribution by creatively employing high-pass filter banks for SRM residual map computation and subsequent trainable filters initialization. Additionally, their	With 0.4 bpp under BOSSBase 1.01 + BOWS 2 WOW = 9.59 %	The introduction of TLU in the pre-processing phase increased the ability of the network to detect the presence of a steganographic payload	Widely used and slow

(*continued on next page*)

Table 3 (*continued*)

Algorithm	Primary contribution	Best results (Error percentage)	Benefit	Weakness
	introduction of the TLU activation function in YeNet stands out as a notable contribution, strategically enhancing the signal-to-noise ratio and improving the efficacy of low payload detection in steganalysis.	S-UNIWARD = 12.81 % HILL = 17.08 %		
YedroudjNet	Authors made a significant contribution by integrating the effective features of XuNet and YeNet, resulting in an enhanced detection accuracy. Furthermore, their scheme incorporated an extended database, further contributing to an overall improvement in the detection performance.	With 0.4 bpp under BOSSBase 1.01 WOW = 14.10 % S-UNIWARD = 22.80 %	Achieved better detection accuracy compared to the previously proposed CNNs	The network was not convergent enough, and the overall CNN was slow due to its complexity
ModifiedYeNet	The authors' key contribution lies in the adaptation of the superior features from YeNet to create a versatile, generalized model. This innovative approach allowed the training of the model with small-resolution images, enhancing its capability to effectively detect high-resolution images.	With 0.4 under BOSSBase 1.01 - Image size 256 × 256 LSBM = 11.77 % WOW = 11.68 % - Image size 1024 × 1024 LSBM = 9.40 % WOW = 14.45 %	Proved a promising ability to detect the steganographic payload in images of arbitrary size	High computational cost in the training phase when dealing with high resolution images.
ZhangNet	The authors made significant contributions through strategic optimizations. They replaced larger 5x5 convolution kernels with more efficient 3x3 kernels in the preprocessing layer, showcasing a thoughtful enhancement. Additionally, their use of separable convolutions not only leveraged channel correlation but also efficiently compressed image content, collectively improving the overall effectiveness of the model.	With 0.4 under BOSSBase 1.01 WOW = 11.8 % S-UNIWARD = 15.3 %	Used data augmentation techniques enhanced the overall performance which led to better accuracy in detecting hidden information.	Though the proposed CNN was fast due to few parameters, it was vulnerable to overfitting when trained with images from a different database
Wang-Net	The authors introduced a novel nonlinear detection method based on SRM, which incorporates a unique approach for acquiring maximum and minimum nonlinear residual features. This adaptation is designed to address the nonlinear distribution characteristics of steganography data, contributing to the improved effectiveness of the detection method.	With 0.4 bpp under BOSSBase - WOW = 14 % -S-UNIWARD = 18.1 %	The joint domain detection mechanism incorporates high pass filters from SRM, and the model's performance is enhanced.	The proposed model is not suitable to colour images steganalysis.
Wu-Net	The authors' innovative approach entails the careful selection of a subset of filters for the pre-processing layer within a CNN-based steganalysis framework. This selection is informed by considerations of the convolution operation mechanism and pixel correlations, leading to the exclusion of several high-pass and derivative filters.	With 0.4 bpp under BOSSBase - WOW = 18.5 % - HUGO = 15.4 % -S-UNIWARD = 16.5 %	It was found that the suggested approach not only ensures precise detection but also enhances the model's training efficiency.	According to the current research, there is a lack of investigation into the blind detection technique for image steganalysis.
Kim-Net	The authors introduce a novel CNN-based steganalysis method that incorporates the embedding of additional data into an input image while using two images as input. They put forward two variations of conventional CNNs, namely the dual channel CNN and dual network CNN, facilitating the utilization of two images as input in this innovative approach.	With 0.4 bpp under BOSSBase -S-UNIWARD = 20.44 %	Overall, the authors propose an innovative approach to CNN-based image steganalysis using two images as input.	The classification accuracy is still low with high time cost for the dual model.
Pan-Net	The article presents a neural architecture search algorithm employed in crafting a deep residual network with a focus on minimizing artificial design of network elements. The authors introduced a long-span residual structure to the traditional residual structure layer, aiming to enhance the signal from hidden messages and capture complex statistical information within digital images.	With 0.4 bpp under BOSSBase - WOW = 8.12 % -S-UNIWARD = 9.16 %	The network is useful for distinguishing between cover and stego images. Better outcomes can be achieved by further exploring additional steganalysis attributes.	When dealing with a low embedding rate of WOW, the improvement in performance is minimal, less than one percent.
Singh-Net	The authors' work primarily centers on developing a steganalytic detector that enhances detection accuracy by utilizing a denoising kernel and noise residual. The detector was trained with a specific emphasis on leveraging the noise residual to achieve	With 0.4 bpp under BOSSBase - WOW = 7.02 % - HUGO = 8.42 % -S-UNIWARD = 7.64 %	The study employs a CNN to learn the denoising kernel used in the pre-processing stage, which achieves better results.	The study did not demonstrate the model's resilience to overfitting.

(*continued on next page*)

Table 3 (*continued*)

Algorithm	Primary contribution	Best results (Error percentage)	Benefit	Weakness
	increased precision in steganography detection.			
GBRAS-Net	The authors contribute significantly by introducing a novel CNN architecture with a distinctive pre-processing step utilizing filter banks to enhance steganographic noise. In the subsequent feature extraction stage, the incorporation of depthwise and separable convolutional layers, along with skip connections, marks a notable advancement in effective feature extraction.	With 0.4 bpp under BOSSBase - WOW = 10.2 % - HUGO = 15.5 % -S-UNIWARD = 12.9 % - HILL = 18.1 % - MiPOD = 18.6 %	The study highlights the potential of CNN architectures, which have proven to be highly effective in classification tasks.	The authors did not investigate how sensitive Deep Learning is when used for this classification problem.
Han-Net	The authors present a novel network model for spatial grayscale image steganalysis, featuring multi-channel convolution and non-local operations in the basic block of feature extraction. Their method outperforms existing CNN-based techniques, establishing superior results in the detection of steganography in grayscale images.	With 0.4 bpp under BOSSBase - WOW = 8.02 % - HILL = 15.91 % -S-UNIWARD = 10.07 %	The article presents a novel CNN-based approach for steganalysis of grayscale images that utilizes non-local operations and multi-channel convolution.	The proposed method did not incorporate non-local operations to analyze colour images.
CCNet	The authors' significant contribution lies in presenting a pioneering network model for spatial grayscale image steganalysis. This model incorporates multi-channel convolution and non-local operations in the basic feature extraction block, resulting in superior performance compared to existing CNN-based techniques. Their work stands out as an impactful advancement in the detection of steganography in grayscale images.	With 0.4 bpp under BOSSBase - WOW = 3.55 % - HUGO = 9.10 % -S-UNIWARD = 3.45 %	Their model demonstrates significantly higher detection accuracy compared to existing models like SRNet, Zhu-Net, and GBRAS-Net.	Despite the model's accuracy, there is still a need to optimize its architecture and reduce the number of parameters while maintaining high accuracy.
SAA-Net	The authors propose a model that introduces attention augmented convolution, assigning greater weights to the steganographic area to facilitate feature learning advantageous for steganalysis. This innovative approach stands as a noteworthy contribution, enhancing the model's effectiveness in discerning steganographic content.	With QF 75, 0.4 bpp under BOSSBase - J-UNIWARD = 10.03 % - JC-UED = 3.32 %	The model incorporates attention augmented convolution to guide the network's focus towards the steganographic region, which leads to improved training outcomes and detection accuracy.	The SAANet model requires further improvement to enhance its capability to detect images with low embedding rates.
SR-Net	The proposed model's most crucial component is an extensively expanded front detector section, featuring the calculation of "noise residuals," and the deactivation of the pooling process to prevent the suppression of stego signals. This adaptation represents a key contribution, emphasizing the importance of preserving stego signal information in the detection process.	In Spatial domain. With 0.4 bpp under BOSSBase + BOWS -WOW = 8.93 % -HILL = 14.14 % -S-UNIWARD = 10.23 % In JPEG with QF 75 at 0.4 bpnzac -J-UNIWARD = 6.70 % -UED-JC = 1.88 %	The deep residual architecture minimizes the usage of externally enforced elements and heuristics. This approach works effectively for both JPEG and spatial domains. Additionally, the second channel is chosen for channel selection.	The dense connection does not deliver satisfactory results, as the network runs significantly slower compared to the Xu-Net.
CIS-Net	The authors introduced the Cover Image Suppression Network (CIS-Net), a novel model aiming to enhance spatial image steganalysis efficiency by minimizing cover image content during learning. Their innovation involves the introduction of two key layers, the Single-value Truncation Layer (STL) and the Sub-linear Pooling Layer (SPL).	With 0.4 bpp under BOSSBase - WOW = 12.13 % - HILL = 18.10 % -S-UNIWARD = 14.62	This method involved calculating the bias for each convolutional layer during initialization by utilizing input cover-stego pairs which improves the performance in the detection of steganographic algorithm.	The proposed CNN was not tested against the compressed images, which currently trending in the steganalysis domain
TCI-Net	The authors proposed an approach involving segmenting the dataset based on image texture complexity and training specific steganalysis models utilizing the Most Effective Region and Inception techniques. This innovative strategy contributes to the model's adaptability to diverse image characteristics, enhancing its overall performance.	With 0.4 bpp under BOSSBase - WOW = 7.4 % -S-UNIWARD = 10.01 %	The texture complexity of an image is calculated to identify the appropriate model for detection, and ensemble learning enhances the precision of the framework.	The model takes long and the experimental results though promising are a bit not different from the existing results.
Jean-Net	The authors proposed a new CNN architecture using depth-wise separable convolutions combined with the conventional convolutional layers to enhance the feature learning. The use of multi-scale pooling makes the proposed	With 0.4 bpp under BOSSBase - WOW = 9.6 % -S-UNIWARD = 10.2 %	Achieving the best performance over the previously proposed models and maintaining a stable training phase without over/underfitting.	The model was only designed for the grayscale images.

(*continued on next page*)

Table 3 (*continued*)

Algorithm	Primary contribution	Best results (Error percentage)	Benefit	Weakness
	model to perform well with images of arbitrary sizes.			
Xie-Net	The authors proposed a method with flexibility to adaptively scale the features of the extraction process based on the filters. The method is also based on the weighted histogram without the conventional rounding.	With 0.4 bpp under BOSSBase QF = 95 J-UNIWARD = 25.2 % SI-UNIWARD = 41.4 % UERD = 19.8 %	Ability to turne the scale of the filters in the feature extraction based on the quality factor (QF) of the colour image.	The model is not flexible to be applied to other domains of digital images.

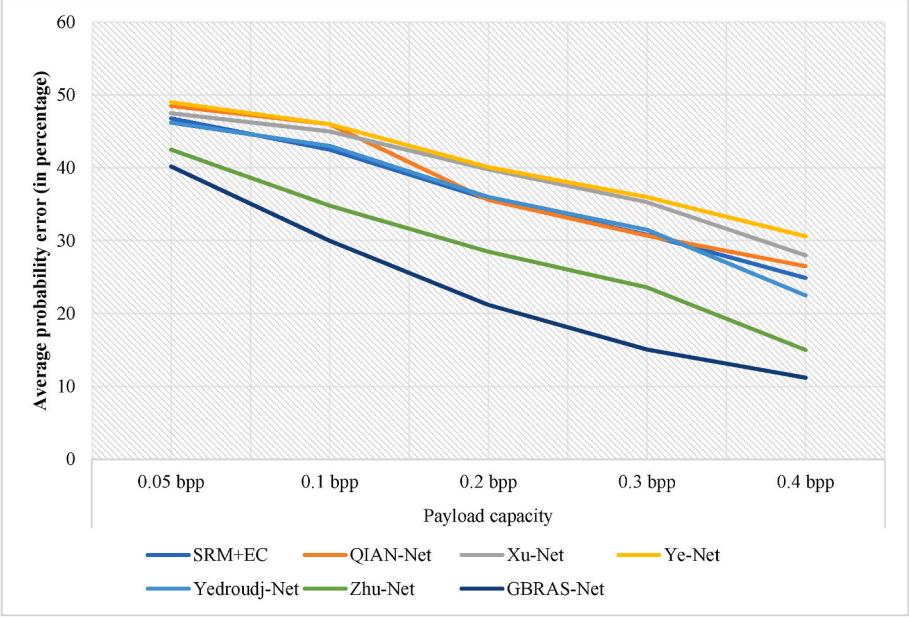


Fig. 8. Average probability error (in percentage) to detect the stego images (steganalysis task) for the main state-of-the-art CNNs under S-UNIWARD with BOSSBase 1.01, and payload capacity varying between 0.05 *bpp* and 0.4 *bpp*

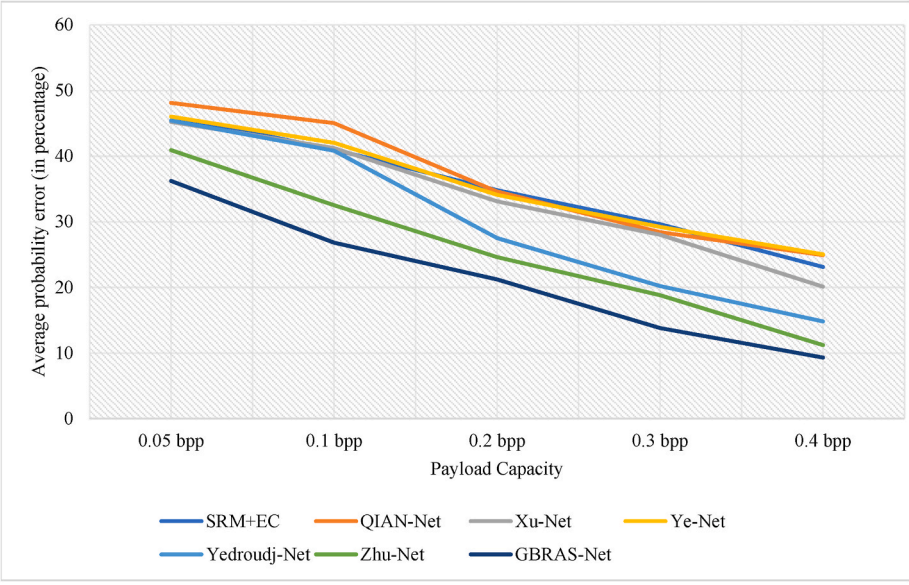


Fig. 9. Average probability error (in percentage) to detect the stego images (steganalysis task) for the main state-of-the-art CNNs under WOW with BOSSBase 1.01, and payload capacity varying between 0.05 *bpp* and 0.4 *bpp* of the proposed CNNs are compared to these traditional algorithms to determine their effectiveness in detecting steganographic images. Initially, the performance of early CNNs was found to be inferior to that of conventional algorithms.

images under WOW with BOSSBase 1.01, especially with higher payload capacities. Ye-Net and QIAN-Net are also competitive, especially at lower payload capacities. However, Yedroudj-Net and Zhu-Net perform relatively poorly compared to other algorithms.

It is essential to highlight that the examined articles document the utilization of various steganographic algorithms within distinct domains. Specifically, in the spatial domain, the reported algorithms encompass S-UNIWARD, HUGO, HILL, and WOW, while in the JPEG domain, the usage of J-UNIWARD, UED, and UERD is documented. This diversity in algorithmic selection underscores the comprehensive nature of the studies considered, providing a foundation for the synthesis and evaluation of state-of-the-art methods. TBits algorithms are typically tested with various payloads, with the most used payload being 0.4 *bpp* in the spatial domain and 0.4 *bpnzAC* (bits per non-zero cover AC DCT coefficient) in the transform domain. The proposed CNNs are compared to traditional algorithms in their ability to detect steganographic images.

Conventional algorithms typically rely on manual extraction of complex features, with some of the essential algorithms being SRM [130], SPAM [20], and variants of Selection-Channel-Aware [110] for the spatial domain. For the frequency domain, traditional algorithms include Selection-Channel-Aware Gabor Filter Residuals [36,127], Discrete Cosine Transform Residuals [22], JPEG Rich Models [139], and PHase Aware pRojection Model [21]. The results.

5. The takeaway message from this comprehensive survey

5.1. Current emerging challenges

The field of image steganalysis faces several challenges when utilizing DL techniques. One of these challenges arises from the restricted availability of training data, which hampers the creation of diverse and comprehensive data for practical training. Furthermore, DL models trained on specific steganographic techniques encounter difficulties in detecting unfamiliar methods; this highlights the need for enhanced generalization capabilities across a broader range of techniques. The presence of adversarial attacks adds another hurdle, as attackers can manipulate stego images to mislead the models. Therefore, developing robust models that can resist such attacks becomes crucial. Another issue lies in the lack of interpretability and explainability exhibited by DL models, which hinders understanding and trust in their decision-making processes. Addressing this challenge requires the development of interpretable models tailored explicitly for image steganalysis. The following seven points clarify the current emerging challenges in steganalysis.

1) Adversarial Attacks Break the Neural Networks: Neural Networks do not present an efficient resistance to adversarial examples, which means that input with high similarity to natural information confuses the classifiers and gets them wrongly classified with high confidence. The existence of adversarial examples makes neural networks vulnerable to attacks, but they are crucial approaches for data-hiding frameworks. To address the problems of adversarial issues, researchers should work on making secure ML algorithms that are able to identify false flags and make available outlines.
2) Dataset quality: Steganalysis schemes yielded less performance in detecting adaptive steganography when working with datasets containing images with heterogeneous content complexities. DL-based approaches best work based on the data extracted from the datasets; hence, the performance of a model's training phase depends on the dataset's quality. Most of the state-of-the-art DL algorithms for steganalysis, as of Fig. 9, use BOSSBase as a testing dataset, making most of the existing results unfair. Therefore, there is a need for standardized datasets to avoid unfair results from almost a single dataset, which also sometimes results in overoptimized results and is remarkably less optimal with other sources of images.
3) Underperformance for images of random sizes: DL-based steganalysis systems impose images with arbitrary dimensions, which makes the existing steganalysis schemes using DL not admissible to pictures of various sizes because they worsen the system's performance. Therefore, the research community in this field should work on designing DL-based steganalysis systems independent of the images' measures because the steganographers use several images with various sizes. Before admitting the input image, existing systems convert it to an arbitrary size, affecting performance due to its residuals under use.
4) Inefficiency in low payload detection: When embedding the secret message in an image with a low capacity of less than 2 %, the steganalysis encounters a challenge to detect images altered with steganalysis; embedding with low payload is still a challenging task for steganalysis because the steganography signal is not easily noticeable. The samples used in the training phase and the learning schemes play a crucial role in the steganalysis performance because the selection of training samples and learning approaches play a significant role in low payload detection and the performance of the steganalyzer in detecting the steganographic signal.
5) Cover Source or Stego Mismatch: These mismatch challenges happen when the training and the testing datasets differ for the systems steganalyzer. Suppose the detector for steganalysis is trained with images from one source and tested with images from another source. In that case, the mismatch problem is likely to happen. Getting a dataset with pictures from the same source in real-life practices seems impossible. To overcome the mismatch problem completely, the overfitting and underfitting problems are still challenging in the DL steganalysis frameworks. The mismatch problems result from several causes, such as various steps in image down-sampling, images from different photo sensors or camera models, and varied processing. The stego mismatch results from the embedding bits number, embedding concealment approaches, and cover images from various sources cause the cover source mismatch.
6) Feature identification and learning: The steganalysis systems using convolutional schemes in DL frameworks take advantage of capturing the similarities among image pixels in the same neighbourhood. However, due to a pooling or convolutional layer scaling operation, the local area layer information fuses for the global CNN. Therefore, there is a need to develop a feature learning approach for steganalysis that ensures global information stability.
7) A significant number of training samples: The available training samples, which play an essential role in learning the small samples in-depth, are still insufficient for steganalysis DL-based CNNs to achieve outperforming results in the detection. Though the dig training datasets increase the training time, they are beneficial in classification for better detection performance. Departing from the fact that many datasets for spatial steganalysis systems training are challenging to find and that sometimes they are impossible to find, the steganalysis frameworks should be designed with the ability to accept a small number of training samples with efficient performance.

5.2. 5.2 Recommendations and possible future directions

The introduction of DL-based steganalysis schemes has provided many efficient performing steganalysis systems that have been able to solve several problems that the traditional methods have not solved. However, these newly introduced DL-based schemes do not efficiently address the challenges mentioned in the previous section. Based on this SLR, which consists of inclusive work to highlight the situation with the current research works and challenges at a glance, we recommend that the researcher in this domain address the highlighted challenges for future research works. The potential points that we recommend as directions for future research are the following.

1) To enhance the accuracy of steganalysis in both the spatial and frequency domains, researchers propose the creation of novel CNNs by combining the strengths of existing networks or developing entirely new architectures with varying depths (such as dense, shallow, or deep architectures).
2) To conduct more comprehensive experiments and further study the impact of Cover-Source Mismatch, it is recommended to utilize diverse digital image databases with varying cameras and other relevant factors. This approach would enable more thorough testing and analysis, yielding more accurate and reliable results.
3) One could conduct steganalysis by experimenting with additional steganographic approaches in the transform domain.
4) Explore the potential of using Generative Adversarial Networks for steganalysis in the spatial domain and investigate their effectiveness for automated steganography in the transform domain.
5) Modify the CNNs used for quantitative steganalysis to enhance the accuracy of their payload prediction outcomes.
6) Utilize deep learning techniques for quantitative steganalysis to accurately predict the payload of steganographic images in the JPEG domain.
7) To enhance the performance of current CNNs, it is imperative to train them using extensive databases and higher-resolution image inputs. This requires utilizing a cluster architecture comprising both a CPU and GPU to meet the processing and memory requirements of the training process.
8) To evaluate the feasibility of transferring knowledge from one steganographic algorithm to another, CNNs can be trained on a specific algorithm and tested on a different one. By conducting such experiments, researchers can investigate how knowledge acquired during training with one algorithm can be effectively applied to another.
9) Incorporating advanced deep learning architectures can significantly enhance the security performance of the proposed ASDL-GAN framework when applied to the transform domain, which contains abundant images. By leveraging the vast amount of available data, researchers can explore the potential of further improving the framework's capabilities.
10) To better represent characteristics, enable the classification of images in the spatial or frequency domain, and process arbitrary images more efficiently, new CNNs can be generated, and novel computational elements can be designed. It is essential to accomplish these goals without resorting to any tricks and with maximum automation by obtaining the noise produced by the steganography process in a more streamlined manner.
11) Analyzing the computational efficiency of current CNNs relative to conventional techniques is worthwhile. This investigation can reveal insights into the strengths and limitations of each approach and offer valuable guidance for selecting the most appropriate method for a given application.
12) Assessing the efficacy of filters employed in the pre-processing stage, specifically high-pass filters (HPF), relative to activation functions used in deep learning (DL) for steganalysis, is worth pursuing. By examining the performance of each technique, researchers can determine their strengths and limitations and make informed decisions about which approach to adopt for a given application.

6. Conclusion

This paper provides a chronological overview of the progression of DL in steganalysis through a comprehensive review of the existing literature. By examining the evolution of this field over time, researchers gain valuable insights into its development, achievements, and areas that require further investigation. In 2014, Tan and Li introduced the first convolutional neural network (CNN) for steganalysis, employing a series of auto-encoders for unsupervised learning. Their approach yielded results similar to those achieved with the Spatial Rich Model (SRM) and outperformed outcomes obtained with the Spatial Pixel Adjacency Matrix (SPAM). The successful application of auto-encoders in this context laid a promising foundation for subsequent research in steganalysis, providing a benchmark for further improvement. Since then, researchers have explored various aspects, including training CNNs with diverse datasets, investigating parameter transfer between networks, conducting quantitative steganalysis, analyzing steganography in images of arbitrary sizes, enhancing image databases, and addressing the Cover-Source Mismatch effect in experiments. This collective effort signifies the continuous refinement and expansion of steganalysis techniques to address evolving challenges and opportunities in the field. This survey also serves as a comprehensive mapping of the obtained results, which demonstrated that several CNN architectures had been introduced for steganalysis, including Qian-Net, Yu-Net, Ye-Net, Yedroudj-Net, Zhu-Net, and GBRAS-Net, all operating in the spatial domain, and a modified version of Yu-Net that uses ResNet has also been adapted for steganalysis in the JPEG domain. The CNN architecture Zhu-Net has demonstrated the most effective detection results in the spatial domain, outperforming even SRM. Furthermore, SRNet is a CNN architecture designed to minimize the use of tricks or heuristics to extract steganographic noise. It operates in both the spatial domain and the JPEG domain.

Numerous potential avenues for future research in this area offer compelling reasons for existing researchers' continued efforts while attracting new ones to explore DL applications in steganalysis. The wide range of possibilities highlights the significance of this field and underscores the need for ongoing research to advance the state-of-the-art.

Funding

This research was supported by Institut Teknologi Sepuluh Nopember, Indonesia.

CRediT authorship contribution statement

Ntivuguruzwa Jean De La Croix: Writing – original draft, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Tohari Ahmad:** Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Fengling Han:** Writing – review & editing, Validation, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Agustinus JT, Mafazy MM, Haq TI, La Croix NJD, Ahmad T. A data hiding scheme via reduced difference expansion to improve the stego quality. In: 2023 3rd international conference on mobile networks and wireless communications (ICMNWC). IEEE; Dec. 2023. p. 1–6. https://doi.org/10.1109/ICMNWC60182.2023.10435709.

[2] Aminy MRH, De La Croix NJ, Ahmad T. A reversible data hiding approach in medical images using difference expansion. In: 2023 IEEE 15th international conference on computational intelligence and communication networks (CICN). IEEE; Dec. 2023. p. 358–62. https://doi.org/10.1109/CICN59264.2023.10402139.

[3] Hussain M, Wahab AWA, Bin Idris YI, Ho ATS, Jung KH. Image steganography in spatial domain: a survey. Signal Process Image Commun Jul. 2018;65:46–66. https://doi.org/10.1016/j.image.2018.03.012.

[4] Chanda D'Layla AW, Nevin M, Sunardi Putra GG, de La Croix NJ, Ahmad T. Steganography in grayscale images: improving the quality of a stego image. In: 2023 3rd international conference on smart generation computing, communication and networking (SMART GENCON). IEEE; Dec. 2023. p. 1–6. https://doi.org/10.1109/SMARTGENCON60755.2023.10442310.

[5] Zou Y, Zhang G, Liu L. Research on image steganography analysis based on deep learning. J Vis Commun Image Represent Apr. 2019;60:266–75. https://doi.org/10.1016/j.jvcir.2019.02.034.

[6] Zaini AFR, De La Croix NJ, Ahmad T. A steganographic approach based on pixel blocks differencing to enhance the quality of the stego image. In: 2024 conference on information communications technology and society (ICTAS). IEEE; Mar. 2024. p. 63–8. https://doi.org/10.1109/ICTAS59620.2024.10507107.

[7] Amrulloh MM, Ahmad T, De La Croix NJ. Analysis of the smoothing and payload distribution method on reversible audio steganography. In: 2023 14th international conference on computing communication and networking technologies (ICCCNT); 2023. p. 1–5. https://doi.org/10.1109/ICCCNT56998.2023.10307703. Delhi, India.

[8] Sunardi Putra GG, Nevin M, De La Croix NJ, Ahmad T. Improving the imperceptibility of hidden data in a digital image using difference expansion. In: 2023 IEEE 2nd industrial electronics society annual on-line conference (ONCON). IEEE; Dec. 2023. p. 1–6. https://doi.org/10.1109/ONCON60463.2023.10430422.

[9] De La Croix NJ, Aminy MRH, Anandha DA, Arsyad H, Nevin M, Ahmad T. Towards a high-capacity data concealment for spatial domain image-steganography. In: 2023 3rd international conference on mobile networks and wireless communications (ICMNWC). IEEE; Dec. 2023. p. 1–6. https://doi.org/10.1109/ICMNWC60182.2023.10435744.

[10] Simmons GJ. The prisoners' problem and the subliminal channel. In: Chaum D, editor. Advances in cryptology. Boston, MA: Springer; 1984. https://doi.org/10.1007/978-1-4684-4730-9_5.

[11] You W, Zhang H, Zhao X. A siamese CNN for image steganalysis. IEEE Trans Inf Forensics Secur 2021;16:291–306. https://doi.org/10.1109/TIFS.2020.3013204.

[12] Yigit Y, Karabatak M. A stenography application for hiding student information into an image. In: 7th international symposium on digital forensics and security, ISDFS 2019. Institute of Electrical and Electronics Engineers Inc.; Jun. 2019. https://doi.org/10.1109/ISDFS.2019.8757516.

[13] Croix NJDL, Ahmad T, Han F. Enhancing secret data detection using convolutional neural networks with fuzzy edge detection. IEEE Access 2023;11:131001–16. https://doi.org/10.1109/ACCESS.2023.3334650.

[14] Castillo Camacho I, Wang K. Convolutional neural network initialization approaches for image manipulation detection. Digit Signal Process: A Review Journal 2022;122(Apr). https://doi.org/10.1016/j.dsp.2021.103376.

[15] K. Wang, X. Song, S. Sun, J. Zhao, C. Xu and H. Song, "Efficient Multi-object Detection for Complexity Spatio-Temporal Scenes,". In: X. Song, R. Feng, Y. Chen, J. Li, G. Min (eds) Web and Big Data. APWeb-WAIM 2023. Lecture Notes in Computer Science, vol vol. 14334. Springer, Singapore. https://doi.org/10.1007/978-981-97-2421-5_13.

[16] Zhao W, Xu C, Guan Z, Wu X, Zhao W, Miao Q, He X, Wang Q. TelecomNet: tag-based weakly-supervised modally cooperative hashing network for image retrieval. IEEE Trans Pattern Anal Mach Intell Nov. 2022;44(11):7940–54. https://doi.org/10.1109/TPAMI.2021.3114089.

[17] Arivazhagan S, Amrutha E, Jebarani WSL. Universal steganalysis of spatial content-independent and content-adaptive steganographic algorithms using normalized feature derived from empirical mode decomposed components. Signal Process Image Commun Feb. 2022;101. https://doi.org/10.1016/j.image.2021.116567.

[18] De La Croix NJ, Ahmad T. Toward secret data location via fuzzy logic and convolutional neural network. Egyptian Informatics Journal Sep. 2023;24(3):100385. https://doi.org/10.1016/j.eij.2023.05.010.

[19] Denemark T, Sedighi V, Holub V, Cogranne R, Fridrich J. Selection-channel-aware rich model for Steganalysis of digital images. In: 2014 IEEE international workshop on information forensics and security (WIFS). IEEE; Dec. 2014. p. 48–53. https://doi.org/10.1109/WIFS.2014.7084302.

[20] Pevný T, Bas P, Fridrich J. Steganalysis by subtractive pixel adjacency matrix. IEEE Trans Inf Forensics Secur Jun. 2010;5(2):215–24. https://doi.org/10.1109/TIFS.2010.2045842.

[21] Holub Vojtěch, Fridrich Jessica. Phase-aware projection model for steganalysis of JPEG images. Proc. SPIE 9409, Media Watermarking, Security, and Forensics 2015:94090T. https://doi.org/10.1117/12.2075239. 4 March 2015.

[22] Holub V, Fridrich J. Low-complexity features for JPEG steganalysis using undecimated DCT. IEEE Trans Inf Forensics Secur Feb. 2015;10(2):219–28. https://doi.org/10.1109/TIFS.2014.2364918.

[23] Chanu YJ, Singh K, Tuithung T. Image steganography and steganalysis: a survey. International Journal of Computer Applications Aug 2012;52(2):1–11. https://doi.org/10.5120/8171-1484.

[24] Nissar A, Mir AH. Classification of steganalysis techniques: a study. Digit Signal Process: A Review Journal 2010;20(6):1758–70. https://doi.org/10.1016/j.dsp.2010.02.003.

[25] Bachrach M, Shih FY. Survey of image steganography and steganalysis. In: Multimedia security. CRC Press; 2017. p. 201–14. https://doi.org/10.1201/b12697-11.

[26] Chandramouli R, Subbalakshmi KP. Current trends in steganalysis: a critical survey. In: 2004 8th international conference on control, automation, robotics and vision (ICARCV); 2004. p. 964–7. https://doi.org/10.1109/icarcv.2004.1468971.

[27] Luo XY, Wang DS, Wang P, Liu FL. A review on blind detection for image steganography. Signal Process Sep. 2008;88(9):2138–57. https://doi.org/10.1016/j.sigpro.2008.03.016.

[28] Karampidis K, Kavallieratou E, Papadourakis G. A review of image steganalysis techniques for digital forensics. J Inf Secur Appl Jun. 2018;40:217–35. https://doi.org/10.1016/j.jisa.2018.04.005.

[29] Dalal M, Juneja M. Steganography and Steganalysis (in digital forensics): a Cybersecurity guide. Multimed Tool Appl Feb. 2021;80(4):5723–71. https://doi.org/10.1007/s11042-020-09929-9.

[30] Selvaraj A, Ezhilarasan A, Wellington SLJ, Sam AR. Digital image steganalysis: a survey on the paradigm shift from machine learning to deep learning based techniques. IET Image Process Feb. 2021;15(2):504–22. https://doi.org/10.1049/ipr2.12043.

[31] Hussain I, Zeng J, Xinhong, Tan S. A survey on deep convolutional neural networks for image steganography and steganalysis. KSII Transactions on Internet and Information Systems 2020;14(3):1228–48. https://doi.org/10.3837/tiis.2020.03.017. Korean Society for Internet Information.

[32] Guo L, Ni J, Su W, Tang C, Shi Y-Q. Using statistical image model for JPEG steganography: uniform embedding revisited. IEEE Trans Inf Forensics Secur Dec. 2015;10(12):2669–80. https://doi.org/10.1109/TIFS.2015.2473815.

[33] Tan S, Li B. Stacked convolutional auto-encoders for steganalysis of digital images. In: 2014 asia-pacific signal and information processing association annual summit and conference, APSIPA 2014. Institute of Electrical and Electronics Engineers Inc.; Feb. 2014. https://doi.org/10.1109/APSIPA.2014.7041565.

[34] Zhong S, Jia C, Chen K, Dai P. A novel steganalysis method with deep learning for different texture complexity images. Multimed Tool Appl Apr. 2019;78(7):8017–39. https://doi.org/10.1007/s11042-018-6573-5.

[35] Yang C, Kang Y, Liu F, Song X, Wang J, Luo X. Color image steganalysis based on embedding change probabilities in differential channels. Int J Distrib Sens Netw May 2020;16(5). https://doi.org/10.1177/1550147720917826.

[36] Alobaidi T, Mikhael W. An adaptive steganography insertion technique based on wavelet transform. J Eng Appl Sci Dec. 2023;70(1):144. https://doi.org/10.1186/s44147-023-00300-x.

[37] Denemark TD, Boroumand M, Fridrich J. Steganalysis features for content-adaptive JPEG steganography. IEEE Trans Inf Forensics Secur Aug. 2016;11(8):1736–46. https://doi.org/10.1109/TIFS.2016.2555281.

[38] Chutani S, Goyal A. A review of forensic approaches to digital image Steganalysis. Multimed Tool Appl Jul. 2019;78(13):18169–204. https://doi.org/10.1007/s11042-019-7217-0.

[39] de La Croix NJ, Ahmad T. Toward hidden data detection via local features optimization in spatial domain images. In: 2023 conference on information communications technology and society (ICTAS). IEEE; Mar. 2023. p. 1–6. https://doi.org/10.1109/ICTAS56421.2023.10082736.

[40] Qiao T, Luo X, Pan B, Chen Y, Wu X. Toward steganographic payload location via neighboring weight algorithm. Secur Commun Network 2022;2022. https://doi.org/10.1155/2022/1400708.

[41] Rupa C, Shaikh S, Chinta M. Multimedia concealed data detection using quantitative steganalysis. Int J Digital Crime Forensics (IJDCF) Sep. 2021;13(5):101–13. https://doi.org/10.4018/IJDCF.20210901.oa6.

[42] De La Croix NJ, Ahmad T, Ijtihadie RM. Pixel-block-based steganalysis method for hidden data location in digital images. International Journal of Intelligent Engineering and Systems 2023;16(6):375–85. https://doi.org/10.22266/ijies2023.1231.31.

[43] Théophile I, De La Croix NJ, Ahmad T. Fuzzy logic-based steganographic scheme for high payload capacity with high imperceptibility. In: 2023 11th international symposium on digital forensics and security (ISDFS). IEEE; May 2023. p. 1–6. https://doi.org/10.1109/ISDFS58141.2023.10131727.

[44] Jahbel AKS, Ahmad T, De La Croix NJ. Reduced difference expansion based on cover image bisection for a quality stego image. In: 2024 conference on information communications technology and society (ICTAS). IEEE; Mar. 2024. p. 51–6. https://doi.org/10.1109/ICTAS59620.2024.10507134.

[45] La Croix NJD, Ahmad T. FuzConvSteganalysis: steganalysis via fuzzy logic and convolutional neural network. SoftwareX May 2024;26:101713. https://doi.org/10.1016/j.softx.2024.101713.

[46] Mandal PC, Mukherjee I, Paul G, Chatterji BN. Digital image steganography: a literature survey. Inf Sci Sep. 2022;609:1451–88. https://doi.org/10.1016/j.ins.2022.07.120.

[47] Fu T, Chen L, Fu Z, Yu K, Wang Y. CCNet: CNN model with channel attention and convolutional pooling mechanism for spatial image steganalysis. J Vis Commun Image Represent Oct. 2022;88:103633. https://doi.org/10.1016/j.jvcir.2022.103633.

[48] Niimi M, Eason RO, Noda H, Kawaguchi E. Intensity histogram steganalysis in BPCS-steganography. In: Proc. SPIE 4314, security and watermarking of multimedia contents III; 2001. https://doi.org/10.1117/12.435440.

[49] Karampidis K, Kavallieratou E, Papadourakis G. A review of image steganalysis techniques for digital forensics. J Inf Secur Appl Jun. 2018;40:217–35. https://doi.org/10.1016/j.jisa.2018.04.005.

[50] Wang J, Yang C, Zhu M, Song X, Liu Y, Lian Y. JPEG image steganography payload location based on optimal estimation of cover co-frequency sub-image. EURASIP J Image Video Process Dec. 2021;2021(1). https://doi.org/10.1186/s13640-020-00542-2.

[51] Chen J, Lu W, Fang Y, Liu X, Yeung Y, Xue Y. Binary image steganalysis based on local texture pattern. J Vis Commun Image Represent Aug. 2018;55:149–56. https://doi.org/10.1016/j.jvcir.2018.06.004.

[52] Bedi P, Singhal A. Estimating cover image for universal payload region detection in stego images. Journal of King Saud University - Computer and Information Sciences Sep. 2022;34(8):5797–809. https://doi.org/10.1016/j.jksuci.2022.01.010.

[53] Qiao T, Wang S, Luo X, Zhu Z. Robust steganography resisting JPEG compression by improving the selection of the cover element. Signal Process 2021;183(Jun). https://doi.org/10.1016/j.sigpro.2021.108048.

[54] Xiang L, Guo G, Yu J, Sheng VS, Yang P. A convolutional neural network-based linguistic steganalysis for synonym substitution steganography. Math Biosci Eng 2020;17(2):1041–58. https://doi.org/10.3934/mbe.2020055.

[55] Lopez-Hernandez J, Martinez-Noriega R, Nakano-Miyatake M, Yamaguchi K. Detection of BPCS-steganography using SMWCF steganalysis and SVM. In: 2008 international symposium on information theory and its applications. IEEE; Dec. 2008. p. 1–5. https://doi.org/10.1109/ISITA.2008.4895497.

[56] Tian J. Reversible data embedding using a difference expansion. IEEE Trans Circ Syst Video Technol Aug. 2003;13(8):890–6. https://doi.org/10.1109/TCSVT.2003.815962.

[57] Luo Weiqi, Huang Fangjun, Huang Jiwu. Edge adaptive image steganography based on LSB matching revisited. IEEE Trans Inf Forensics Secur Jun. 2010;5(2):201–14. https://doi.org/10.1109/TIFS.2010.2041812.

[58] Luo Y, Li X, Yang B. Locating steganographic payload for LSB matching embedding. In: Proceedings - IEEE international conference on multimedia and expo; 2011. https://doi.org/10.1109/ICME.2011.6011833.

[59] Ker AD. Steganalysis of LSB matching in grayscale images. IEEE Signal Process Lett Jun. 2005;12(6):441–4. https://doi.org/10.1109/LSP.2005.847889.

[60] Marvel LM, Boncelet CG, Retter CT. Spread spectrum image steganography. IEEE Trans Image Process Aug. 1999;8(8):1075–83. https://doi.org/10.1109/83.777088.

[61] Wang J, Yang C, Wang P, Song X, Lu J. Payload location for JPEG image steganography based on co-frequency sub-image filtering. Int J Distrib Sens Netw Jan. 2020;16(1). https://doi.org/10.1177/1550147719899569.

[62] Wang J, Yang C, Wang P, Song X, Lu J. Payload location for JPEG image steganography based on co-frequency sub-image filtering. Int J Distrib Sens Netw Jan. 2020;16(1). https://doi.org/10.1177/1550147719899569.

[63] Liu J, Zhang W, Zhang Y, Hou D, Liu Y, Zha H, Yu N. Detection-based defense against adversarial examples from the steganalysis point of view [Online]. Available: http://arxiv.org/abs/1806.09186; Jun. 2018.

[64] Pan B, Qiao T, Li J, Chen Y, Yang C. Novel hidden bit location method towards JPEG steganography. Secur Commun Network 2022;2022. https://doi.org/10.1155/2022/8230263.

[65] Sullivan K, Madhow U, Chandrasekaran S, Manjunath BS. Steganalysis of spread spectrum data hiding exploiting cover memory. Proc. SPIE 5681, Security, Steganography, and Watermarking of Multimedia Contents VII 21 March 2005. https://doi.org/10.1117/12.588121.

[66] Li M, Kulhandjian MK, Pados DA, Batalama SN, Medley MJ. Extracting spread-spectrum hidden data from digital media. IEEE Trans Inf Forensics Secur 2013;8(7):1201–10. https://doi.org/10.1109/TIFS.2013.2264462.

[67] Li M, Kulhandjian M, Pados DA, Batalama SN, Medley MJ, Matyjas JD. On the extraction of spread-spectrum hidden data in digital media. In: 2012 IEEE international conference on communications (ICC); 2012. p. 1031–5. https://doi.org/10.1109/ICC.2012.6364055. Ottawa, ON, Canada.

[68] Ji R, Yao H, Liu S, Wang L, Sun J. A new steganalysis method for adaptive spread spectrum steganography. In: 2006 international conference on intelligent information hiding and multimedia; 2006. p. 365–8. https://doi.org/10.1109/IIH-MSP.2006.265018. Pasadena, CA, USA.

[69] Shi P, Li Z. An improved BPCS steganography based on dynamic threshold. In: 2010 international conference on multimedia information networking and security. IEEE; 2010. p. 388–91. https://doi.org/10.1109/MINES.2010.87.

[70] Tan S, Huang J, Shi YQ. Steganalysis of enhanced BPCS steganography using the hilbert-huang transform based sequential analysis. 2008. p. 112–26. https://doi.org/10.1007/978-3-540-92238-4_10.

[71] Yu Xiaoyi, Tan Tieniu, Wang Yunhong. Reliable detection of BPCS-steganography in natural images. In: Third international conference on image and graphics (ICIG'04); 2004. p. 333–6. https://doi.org/10.1109/ICIG.2004.123. Hong Kong, China.

[72] Bansal R, Nagpal CK, Gupta S. An efficient hybrid security mechanism based on chaos and improved BPCS. Multimed Tool Appl 2018;77:6799–835. https://doi.org/10.1007/s11042-017-4600-6.

[73] Guo Linjie, Ni Jiangqun, Shi Yun Qing. Uniform embedding for efficient JPEG steganography. IEEE Trans Inf Forensics Secur May 2014;9(5):814–25. https://doi.org/10.1109/TIFS.2014.2312817.

[74] Holub V, Fridrich J. Designing steganographic distortion using directional filters. In: 2012 IEEE international workshop on information forensics and security (WIFS). IEEE; Dec. 2012. p. 234–9. https://doi.org/10.1109/WIFS.2012.6412655.

[75] Xu G, Wu HZ, Shi YQ. Ensemble of CNNs for steganalysis: an empirical study. In: IH and MMSec 2016 - proceedings of the 2016 ACM information hiding and multimedia security workshop. Association for Computing Machinery, Inc; 2016. p. 103–7. https://doi.org/10.1145/2909827.2930798.

[76] Yu Xiaoyi, Wang Yunhong, Tan Tieniu. On the estimation of secret message length in JSteg-like steganography. In: Proceedings of the 17th international conference on pattern recognition, 2004. ICPR 2004, vol. 4; 2004. p. 673–6. https://doi.org/10.1109/ICPR.2004.1333862. Cambridge, UK.

[77] Fridrich J. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. In: Fridrich J, editor. Information hiding. IH 2004. Lecture notes in computer science, vol. 3200. Berlin, Heidelberg: Springer; 2004. https://doi.org/10.1007/978-3-540-30114-1_6.

[78] Miri A, Faez K. An image steganography method based on integer wavelet transform. Multimed Tool Appl Jun. 2018;77(11):13133–44. https://doi.org/10.1007/s11042-017-4935-z.

[79] McBride BT, Peterson GL, Gustafson SC. A new blind method for detecting novel steganography. Digit Invest Feb. 2005;2(1):50–70. https://doi.org/10.1016/j.diin.2005.01.003.

[80] Han S, Lv M, Cheng Z. Dual-color blind image watermarking algorithm using the graph-based transform in the stationary wavelet transform domain. Optik Oct. 2022;268:169832. https://doi.org/10.1016/j.ijleo.2022.169832.

[81] Liu S, Yao H, Gao W. Steganalysis of data hiding techniques in the wavelet domain. In: International conference on information technology: coding and computing, 2004. Proceedings. ITCC 2004, vol. 1; 2004. p. 751–4. https://doi.org/10.1109/ITCC.2004.1286558. Las Vegas, NV, USA.

[82] Yu XY, Wang A. Detection of quantization data hiding. In: 2009 international conference on multimedia information networking and security. IEEE; 2009. p. 45–7. https://doi.org/10.1109/MINES.2009.272.

[83] Wu Q, Li W, Yu XY. Revisit steganalysis on QIM-based data hiding. In: 2009 fifth international conference on intelligent information hiding and multimedia signal processing. IEEE; Sep. 2009. p. 929–32. https://doi.org/10.1109/IIH-MSP.2009.316.

[84] Malik H, Subbalakshmi KP, Chandramouli R. Nonparametric steganalysis of QIM steganography using approximate entropy. IEEE Trans Inf Forensics Secur Apr. 2012;7(2):418–31. https://doi.org/10.1109/TIFS.2011.2169058.

[85] Li X, Zhang T, Li K, Ping X. A blind detection method for additive noise steganography in JPEG decompressed images. In: Proceedings - 3rd international conference on multimedia information networking and security. MINES; 2011. p. 489–93. https://doi.org/10.1109/MINES.2011.68. 2011.

[86] Yang J, Fan G, Xie K, Chen Q, Wang A. Additive noise model structure learning based on rank correlation. Inf Sci Sep. 2021;571:499–526. https://doi.org/10.1016/j.ins.2021.05.061.

[87] Holotyak Taras, Fridrich Jessica, Soukal David. Stochastic approach to secret message length estimation in ±k embedding steganography. In: Proc. SPIE 5681, security, steganography, and watermarking of multimedia contents VII; 21 March 2005. https://doi.org/10.1117/12.584201.

[88] Jiang M, Wu X, Wong EK, Memon A. Steganalysis of boundary-based steganography using autoregressive model of digital boundaries. In: 2004 IEEE international conference on multimedia and expo (ICME) (IEEE cat. No.04TH8763). vol. 2; 2004. p. 883–6. https://doi.org/10.1109/ICME.2004.1394342. Taipei, Taiwan.

[89] De La Croix NJ, Ahmad T, Ijtihadie RM. Convolutional neural network with multi-scale pooling for the efficient steganalysis in images of arbitrary sizes. In: 2023 14th international conference on information & communication technology and system (ICTS). IEEE; Oct. 2023. p. 141–6. https://doi.org/10.1109/ICTS58770.2023.10330880.

[90] Avcibas I, Memon ND, Sankur B. Steganalysis of watermarking techniques using image quality metrics. Proc. SPIE 4314, Security and Watermarking of Multimedia Contents III 2001. https://doi.org/10.1117/12.435436.

[91] Vilkhovskiy DE. Steganalysis for LSB inserts in low stego-payload artificial color images. In: Journal of physics: conference series. IOP Publishing Ltd; Mar. 2022. https://doi.org/10.1088/1742-6596/2182/1/012102.

[92] Rahman CR, et al. Identification and recognition of rice diseases and pests using convolutional neural networks. Biosyst Eng Jun. 2020;194:112–20. https://doi.org/10.1016/j.biosystemseng.2020.03.020.

[93] Jin Z, Feng G, Ren Y, Zhang X. Feature extraction optimization of JPEG steganalysis based on residual images. Signal Process May 2020;170. https://doi.org/10.1016/j.sigpro.2020.107455.

[94] Xie G, Ren J, Marshall S, Zhao H, Li R, Chen R. Self-attention enhanced deep residual network for spatial image steganalysis. Digit Signal Process Jul. 2023;139:104063. https://doi.org/10.1016/j.dsp.2023.104063.

[95] Han X, Zhang T. Spatial steganalysis based on non-local block and multi-channel convolutional networks. IEEE Access 2022;10:87241–53. https://doi.org/10.1109/ACCESS.2022.3199351.

[96] Ntivuguruzwa JDLC, Ahmad T. A convolutional neural network to detect possible hidden data in spatial domain images. Cybersecurity Sep. 2023;6(1):23. https://doi.org/10.1186/s42400-023-00156-x.

[97] Sabnis SK, Awale RN. Statistical steganalysis of high-capacity image steganography with cryptography. Procedia Comput Sci 2016;79:321–7. https://doi.org/10.1016/j.procs.2016.03.042.

[98] Lou D-C, Liu C-L, Lin C-L. Message estimation for universal steganalysis using multi-classification support vector machine. Comput Stand Interfac Feb. 2009;31(2):420–7. https://doi.org/10.1016/j.csi.2008.05.017.

[99] Hou X, Zhang T, Xu C. New framework for unsupervised universal steganalysis via SRISP-aided outlier detection. Signal Process Image Commun Sep. 2016;47:72–85. https://doi.org/10.1016/j.image.2016.05.011.

[100] Lerch-Hostalot D, Megías D. Unsupervised steganalysis based on artificial training sets. Eng Appl Artif Intell Apr. 2016;50:45–59. https://doi.org/10.1016/j.engappai.2015.12.013.

[101] Gupta S, Mohan N, Kaushal P. Passive image forensics using universal techniques: a review. Artif Intell Rev Mar. 2022;55(3):1629–79. https://doi.org/10.1007/s10462-021-10046-8.

[102] Selvaraj A, Ezhilarasan A, Wellington SLJ, Sam AR. Digital image steganalysis: a survey on the paradigm shift from machine learning to deep learning-based

techniques. IET Image Process Feb. 2021;15(2):504–22. https://doi.org/10.1049/ipr2.12043.
[103] Bashir B, Selwal A. Towards deep learning-based image steganalysis: practices and open research issues. SSRN Electron J 2021. https://doi.org/10.2139/ssrn.3883330.
[104] Kadhim IJ, Premaratne P, Vial PJ, Halloran B. Comprehensive survey of image steganography: techniques, Evaluations, and trends in future research. Neurocomputing Mar. 2019;335:299–326. https://doi.org/10.1016/j.neucom.2018.06.075.
[105] Płachta M, Krzemień M, Szczypiorski K, Janicki A. Detection of image steganography using deep learning and ensemble classifiers. Electronics (Basel) May 2022;11(10):1565. https://doi.org/10.3390/electronics11101565.
[106] Tabares-Soto R, et al. Strategy to improve the accuracy of convolutional neural network architectures applied to digital image steganalysis in the spatial domain. PeerJ Comput Sci Apr. 2021;7:e451. https://doi.org/10.7717/peerj-cs.451.
[107] Tabares-Soto R, Raúl RP, Gustavo I. Deep learning applied to steganalysis of digital images: a systematic review. IEEE Access 2019;7:68970–90. https://doi.org/10.1109/ACCESS.2019.2918086.
[108] Zhang R, Zhu F, Liu J, Liu G. Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis. IEEE Trans Inf Forensics Secur 2020;15:1138–50. https://doi.org/10.1109/TIFS.2019.2936913.
[109] Moher D, Altman DG, Liberati A, Tetzlaff J. PRISMA statement. Epidemiology Jan. 2011;22(1):128. https://doi.org/10.1097/EDE.0b013e3181fe7825.
[110] Pibre L, Jérôme P, Ienco D, Chaumont M. Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source-mismatch [Online]. Available: http://arxiv.org/abs/1511.04855; Nov. 2015.
[111] Qian Y, Dong J, Wang W, Tan T. Learning and transferring representations for image steganalysis using convolutional neural network. In: Proceedings - international conference on image processing, ICIP. IEEE Computer Society; Aug. 2016. p. 2752–6. https://doi.org/10.1109/ICIP.2016.7532860.
[112] Ye J, Ni J, Yi Y. Deep learning hierarchical representations for image steganalysis. IEEE Trans Inf Forensics Secur Nov. 2017;12(11):2545–57. https://doi.org/10.1109/TIFS.2017.2710946.
[113] Yedroudj M, Comby F, Chaumont M. Yedroudj-net: an efficient CNN for spatial steganalysis. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP); 2018. p. 2092–6. https://doi.org/10.1109/ICASSP.2018.8461438. Calgary, AB, Canada.
[114] Boroumand M, Chen M, Fridrich J. Deep residual network for steganalysis of digital images. IEEE Trans Inf Forensics Secur May 2019;14(5):1181–93. https://doi.org/10.1109/TIFS.2018.2871749.
[115] Xu G. Deep convolutional neural network to detect J-UNIWARD. In: IH and MMSec 2017 - proceedings of the 2017 ACM workshop on information hiding and multimedia security. Association for Computing Machinery, Inc; Jun. 2017. p. 67–73. https://doi.org/10.1145/3082031.3083236.
[116] Wu S, Zhong S, Liu Y. Deep residual learning for image steganalysis. Multimed Tool Appl May 2018;77(9):10437–53. https://doi.org/10.1007/s11042-017-4440-4.
[117] Yang J, Kang X, Wong EK, Shi YQ. JPEG steganalysis with combined dense connected CNNs and SCA-GFR. Multimed Tool Appl Apr. 2019;78(7):8481–95. https://doi.org/10.1007/s11042-018-6878-4.
[118] Chen M, Sedighi V, Boroumand M, Fridrich J. JPEG-phase-aware convolutional neural network for steganalysis of JPEG images. In: IH and MMSec 2017 - proceedings of the 2017 ACM workshop on information hiding and multimedia security. Association for Computing Machinery, Inc; Jun. 2017. p. 75–84. https://doi.org/10.1145/3082031.3083248.
[119] Xu G, Wu HZ, Shi YQ. Structural design of convolutional neural networks for steganalysis. IEEE Signal Process Lett May 2016;23(5):708–12. https://doi.org/10.1109/LSP.2016.2548421.
[120] Hu D, Wang L, Jiang W, Zheng S, Li B. A novel image steganography method via deep convolutional generative adversarial networks. IEEE Access Jul. 2018;6:38303–14. https://doi.org/10.1109/ACCESS.2018.2852771.
[121] Tang W, Tan S, Li B, Huang J. Automatic steganographic distortion learning using a generative adversarial network. IEEE Signal Process Lett Oct. 2017;24(10):1547–51. https://doi.org/10.1109/LSP.2017.2745572.
[122] Hayes J, Danezis G. Generating steganographic images via adversarial training. ArXiv 2017. https://arxiv.org/abs/1703.00371v3.
[123] Jan Kodovský, Fridrich Jessica. Quantitative steganalysis using rich models. Proc. SPIE 8665, Media Watermarking, Security, and Forensics 22 March 2013;2013:86650O. https://doi.org/10.1117/12.2001563.
[124] Chen M, Boroumand M, Fridrich J. Deep learning regressors for quantitative steganalysis. In: IS and T International symposium on electronic imaging science and technology. Society for Imaging Science and Technology; 2018. https://doi.org/10.2352/ISSN.2470-1173.2018.07.MWSF-160.
[125] Zeng J, Tan S, Li B, Huang J. Large-scale JPEG steganalysis using the hybrid deep-learning framework. Nov. 2016. https://doi.org/10.1109/TIFS.2017.2779446.
[126] Zhang Y, Zhang W, Chen K, Liu J, Liu Y, Yu N. Adversarial examples against deep neural network based steganalysis. In: IH and MMSec 2018 - proceedings of the 6th ACM workshop on information hiding and multimedia security. Association for Computing Machinery, Inc; Jun. 2018. p. 67–72. https://doi.org/10.1145/3206004.3206012.
[127] Song X, Liu F, Yang C, Luo X, Zhang Y. Steganalysis of adaptive JPEG steganography using 2D Gabor filters. In: IH and MMSec 2015 - proceedings of the 2015 ACM workshop on information hiding and multimedia security. Association for Computing Machinery, Inc; Jun. 2015. p. 15–23. https://doi.org/10.1145/2756601.2756608.
[128] Li B, Wei W, Ferreira A, Tan S. ReST-net: diverse activation modules and parallel subnets-based CNN for spatial image steganalysis. IEEE Signal Process Lett May 2018;25(5):650–4. https://doi.org/10.1109/LSP.2018.2816569.
[129] Zeng J, Tan S, Liu G, Li B, Huang J. WISERNet: wider separate-then-reunion network for steganalysis of color images. Mar. 2018. https://doi.org/10.1109/TIFS.2019.2904413.
[130] Fridrich J, Kodovsky J. Rich models for steganalysis of digital images. IEEE Trans Inf Forensics Secur Jun. 2012;7(3):868–82. https://doi.org/10.1109/TIFS.2012.2190402.
[131] Qian Y, Dong J, Wang W, Tan T. Deep learning for steganalysis via convolutional neural networks. In: *Media watermarking, security, and forensics 2015*, SPIE; Mar. 2015. 94090J. https://doi.org/10.1117/12.2083479.
[132] Nair V, Hinton G. Rectified linear units improve restricted Boltzmann machines. Proceedings of ICML 2010;27:807–14. https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf.
[133] Tang W, Li B, Tan S, Barni M, Huang J. CNN based adversarial embedding with minimum alteration for image steganography. Mar. 2018. https://doi.org/10.1109/TIFS.2019.2891237.
[134] Boureau Y-Lan, Ponce J, LeCun Y. A theoretical analysis of feature pooling in visual recognition. Icml 2010 - proceedings, 27th international conference on machine learning. 2010. p. 111–8. https://dl.acm.org/doi/10.5555/3104322.3104338.
[135] Qian Y, Dong J, Wang W, Tan T. Feature learning for steganalysis using convolutional neural networks. Multimed Tool Appl Aug. 2018;77(15):19633–57. https://doi.org/10.1007/s11042-017-5326-1.
[136] Zhang R, Zhu F, Liu J, Liu G. Efficient feature learning and multi-size image steganalysis based on CNN. https://arxiv.org/abs/1807.11428; 2018.
[137] Reinel Tabares-Soto, Brayan Arteaga-Arteaga Harold, Alejandro Bravo-Ortiz Mario, Alejandro Mora-Rubio, Daniel Arias-Garzón, Alejandro Alzate-Grisales Jesús, AlejandroBuenaventura Burbano-Jacome, Simon Orozco-Arias, Gustavo Isaza, Raúl Ramos-Pollán. GBRAS-net: a convolutional neural network architecture for spatial image steganalysis. IEEE Access 2021;9:14340–50. https://doi.org/10.1109/ACCESS.2021.3052494.
[138] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer vision – eccv 2014. Lecture notes in computer science, vol. 8691. Cham: Springer; 2014. https://doi.org/10.1007/978-3-319-10578-9_23. ECCV 2014.
[139] Jan Kodovský, Fridrich Jessica. Steganalysis of JPEG images using rich models. In: Proc. SPIE 8303, media watermarking, security, and forensics 2012, 83030A; 2012. https://doi.org/10.1117/12.907495.
[140] Tsang CF, Fridrich J. Steganalyzing images of arbitrary size with CNNs. Electron Imag Jan. 2018;30(7). https://doi.org/10.2352/ISSN.2470-1173.2018.07.MWSF-121. 121-1-121–8.
[141] Wu S, Zhong S, Liu Y, Liu M. CIS-net: a novel CNN model for spatial image steganalysis via cover image suppression [Online]. Available, http://arxiv.org/abs/1912.06540; Dec. 2019.
[142] Zhong S, Jia C, Chen K, Dai P. A novel steganalysis method with deep learning for different texture complexity images. Multimed Tool Appl Apr. 2019;78(7):8017–39. https://doi.org/10.1007/s11042-018-6573-5.
[143] Wang Z, Chen M, Yang Y, Lei M, Dong Z. Joint multi-domain feature learning for image steganalysis based on CNN. EURASIP J Image Video Process Dec. 2020;2020(1). https://doi.org/10.1186/s13640-020-00513-7.
[144] Wu L, Han X, Wen C, Li B. A Steganalysis framework based on CNN using the filter subset selection method. Multimed Tool Appl Jul. 2020;79(27–28):19875–92. https://doi.org/10.1007/s11042-020-08831-8.
[145] Kim J, Park H, Il Park J. CNN-based image steganalysis using additional data embedding. Multimed Tool Appl Jan. 2020;79(1–2):1355–72. https://doi.org/10.1007/s11042-019-08251-3.
[146] Wang H, Pan X, Fan L, Zhao S. Steganalysis of convolutional neural network based on neural architecture search. In: Multimedia systems. Springer Science and Business Media Deutschland GmbH; Jun. 2021. p. 379–87. https://doi.org/10.1007/s00530-021-00779-5.
[147] Singh B, Chhajed M, Sur A, Mitra P. Steganalysis using learned denoising kernels. Multimed Tool Appl Feb. 2021;80(4):4903–17. https://doi.org/10.1007/s11042-020-09960-w.
[148] Huang S, Zhang M, Ke Y, Bi X, Kong Y. Image steganalysis based on attention augmented convolution. Multimed Tool Appl Jun. 2022;81(14):19471–90. https://doi.org/10.1007/s11042-021-11862-4.