Hindawi
Security and Communication Networks
Volume 2023, Article ID 9764791, 1 page
https://doi.org/10.1155/2023/9764791

Retraction

Retracted: Cryptospace Invertible Steganography with Conditional Generative Adversarial Networks

Security and Communication Networks

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] C.-C. Chang, "Cryptospace Invertible Steganography with Conditional Generative Adversarial Networks," *Security and Communication Networks*, vol. 2021, Article ID 5538720, 14 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2021, Article ID 5538720, 14 pages
https://doi.org/10.1155/2021/5538720

Research Article

Cryptospace Invertible Steganography with Conditional Generative Adversarial Networks

Ching-Chun Chang

Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK

Correspondence should be addressed to Ching-Chun Chang; ching-chun.chang@warwickgrad.net

Received 28 January 2021; Revised 17 February 2021; Accepted 23 February 2021; Published 15 March 2021

Academic Editor: Chi-Hua Chen

Deep neural networks have become the foundation of many modern intelligent systems. Recently, the author has explored adversarial learning for invertible steganography (ALIS) and demonstrated the potential of deep neural networks to reinvigorate an obsolete invertible steganographic method. With the worldwide popularisation of the Internet of things and cloud computing, invertible steganography can be recognised as a favourable way of facilitating data management and authentication due to the ability to embed information without causing permanent distortion. In light of growing concerns over cybersecurity, it is important to take a step forwards to investigate invertible steganography for encrypted data. Indeed, the multidisciplinary research in invertible steganography and cryptospace computing has received considerable attention. In this paper, we extend previous work and address the problem of cryptospace invertible steganography with deep neural networks. Specifically, we revisit a seminal work on cryptospace invertible steganography in which the problem of message decoding and image recovery is viewed as a type of binary classification. We formulate a general expression encompassing spatial, spectral, and structural analyses towards this particular classification problem and propose a novel discrimination function based on a recurrent conditional generative adversarial network (RCGAN) which predicts bit-planes with stacked neural networks in a top-down manner. Experimental results evaluate the performance of various discrimination functions and validate the superiority of neural-network-aided discrimination function in terms of classification accuracy.

1. Introduction

Cybersecurity has become an urgent priority for governments, businesses, and individuals all over the globe as an exponentially growing amount of data is communicated and stored in the cyberspace [1]. It is arguably more vital than ever to take positive steps to prevent cyber criminals getting hold of private data. While encryption affords an effective protection of privacy, it may limit functionality as a large number of algorithms are not compatible with encrypted data. In view of this issue, scientists have carried out studies on signal processing and data analysis in the cryptospace [2–6].

As an established discipline closely associated with cybersecurity, steganography concerns the methodologies and applications of hiding information [7–9]. A typical steganographic approach is to modify the cover objects in an imperceptible manner in order to represent messages while simultaneously preserving the content of cover objects. It has been used for a wide range of applications including covert communication [10], copyright protection [11], integrity verification [12], and traitor tracing [13], to name a few. The possibility of carrying additional information within a cover object is bought at the expense of introducing some degree of distortion. Even though this distortion is often minimal and invisible, it might not be permissible when data integrity and high resolution are required. This gave birth to the research on invertible steganography [14], also known as erasable watermarking, lossless data embedding, or reversible information hiding.

With the advent of the Internet of things [15–17], it is believed that invertible steganography will occupy a crucial

role because data communication is a fundamental part in the big data era. It can be utilised to verify authenticity when distributing and archiving data through embedding digital object identifiers, digital signatures, or metadata. Meanwhile, it can remove the modifications and recover a clean copy of data. Recent studies have shown that a type of wilfully crafted noise called adversarial perturbations will cause the output of deep learning models to change substantially [18–20], as illustrated in Figure 1. Whilst no claim has been made that those models will be equally susceptible to the steganographic noise, the characteristic of invertible steganography is desirable since it reduces the risk of dataset contamination to a minimum.

The essence of invertible steganography is to find a set of features that are losslessly compressible and of which randomisation has little impact on the cover object [21–26]. In order to exploit data redundancy, it is necessary to access and analyse the cover object. Redundancy analysis is, however, hardly achievable in the crytospace because an ideal cryptosystem that offers perfect secrecy will output a purely random and uniformly distributed encrypted object. Thus, most invertible steganographic methods cannot be applied directly to the cryptospace.

Cryptospace invertible steganography has come to prominence as a new and promising research paradigm [27]. It inherits the merits of invertible steganography, and on top of that a more secure environment is ensured with more promising applications to be developed. An example of how to apply this technology is illustrated in Figure 2 and narrated as follows. Suppose that a client, Alice, wants to send an image, or a batch of images, to a data scientist, Bob, for analysis purposes. Bob requests all clients to embed messages such as service order numbers and authentication codes into images for facilitating management. Messages are preferred to be embedded in an invertible manner in order to minimise the uncontrollable risks of erroneous analytical results posed by steganographic distortion. Due to limited computational resources and restricted access to steganographic software, Alice resorts to cloud computing. The cloud server, by contrast, has an enormous capacity and a licence for the software. However, Alice has concerns about privacy and wishes not to reveal the content of images to the cloud server. Therefore, Alice encrypts and uploads a batch of cover images along with the messages to the cloud server which then performs the steganographic algorithm in the cryptospace. The resultant images may be returned to Alice or downloaded directly by Bob. A preshared cryptographic key between Alice and Bob is required if the stego images are presented in a state of encryption to Bob. In either case, Bob will receive the stego images, decode the messages, remove the distortion, and then carry out analysis. We would like to note that the workflows and applications of cryptospace invertible steganography are by no means limited to this particular example.

The problem of cryptospace invertible steganography is challenging and there are diverse approaches towards this problem. A possible strategy is to make compromises on security by utilising bespoke encryption schemes in exchange for the redundancy and compressibility of encrypted objects [28–32]. Another strategy is to preprocess the cover objects prior to encryption in order to create space for the subsequent data embedding in the cryptospace [33–39]. From our perspective, both strategies have limitations and their practicality might be open to dispute. The former by no means guarantees security, as conditions for perfect secrecy may not be satisfied when employing dedicated cryptosystems. The latter is unfavourable for expansion as preprocessing prior to encryption is unavoidable and could also be criticised for evading the problem and challenge of cryptospace signal processing altogether.

There is, in addition, one further strategy for cryptospace invertible steganography. Compared with the aforementioned strategies, it suggests analysing and exploiting data redundancy after decryption rather than before or during a state of encryption [40–44]. This methodology usually adopts a standard encryption scheme and has practically no need for preprocessing prior to encryption. In general, it embeds messages by disturbing the encrypted objects, and the ability to recover the original content relies on a discrimination function that acts on the decrypted objects. A drawback of this methodology is that a perfect recovery of unaltered content might not be guaranteed. For a given cover object, the upper bound of recovery accuracy depends on the amplitude and period of perturbations, which are, in practice, factors of steganographic distortion and capacity. The question of how well the bound can be approached is connected to the design of the discrimination function.

In this paper, we address the problem of cryptospace invertible steganography for digital images. In particular, we study the discrimination function from different perspectives and formulate a general framework. We follow a classic cryptospace invertible steganographic methodology denominated as *associative tri-LSB flipping* [45] and carry out spatial, spectral, and structural analyses for discriminating the perturbations. The majority of prior discrimination functions are based on spatial analysis and can be more or less represented by the discrete Laplacian operator that calculates fluctuations in local regions. It is also worthwhile investigating discrimination mechanisms based on spectral analysis. To this interest, we convert the image patches to the frequency domain by the discrete Fourier transform and apply the Butterworth filter in an attempt to detect perturbations. Both spatial and spectral analyses are valid approaches, but there is still room for improvement in terms of recovery accuracy.

Deep learning has revolutionised the academia and industry in an unprecedented manner and has served to promote the development of data-driven intelligent systems [46–52]. The outlook of integrating neural networks with invertible steganography is also positive. Recently, the author conducted an exploratory study on adversarial learning for invertible steganography (ALIS) [53] and demonstrated the potential of deep neural networks to bring an obsolete invertible steganographic method, the regular-singular (RS) method [54], forward into the modern generation. As an extension of it, this paper proposes to *neuralise* cryptospace invertible steganography: we name the project *ALIS in Cryptoland*. In order to be compatible with the associative

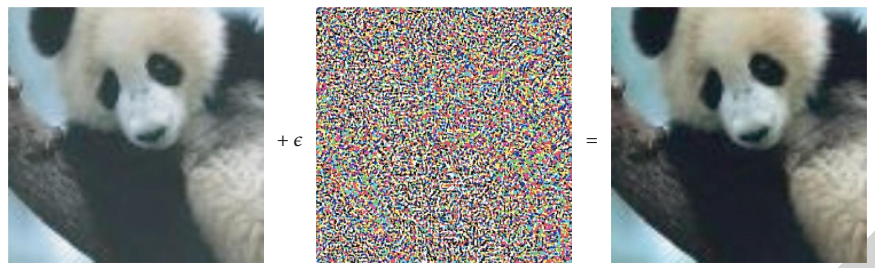


Figure 1: An example of how imperceptible adversarial perturbations may mislead the classification output of a neural network model.

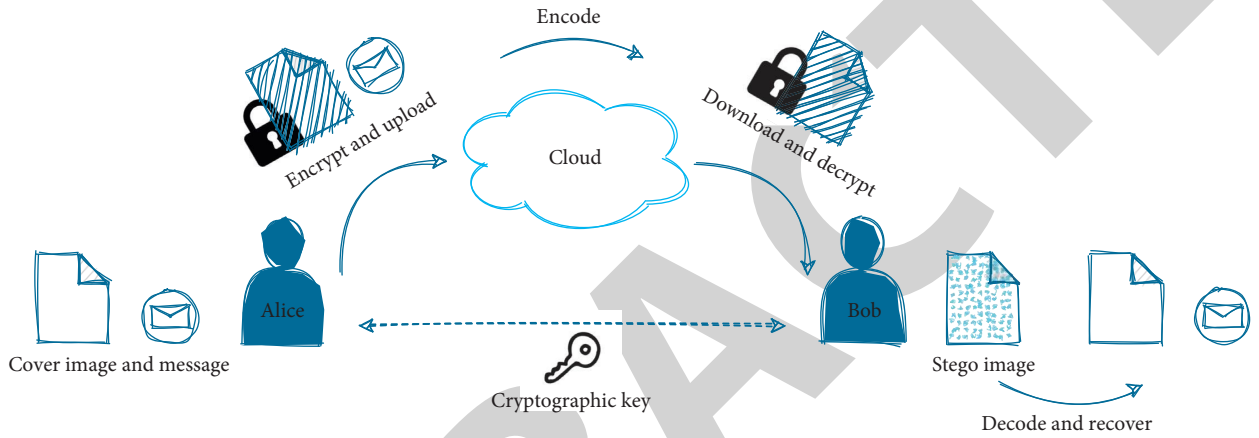


Figure 2: A workflow of cryptospace invertible steganography as privacy-preserving cloud computing.

tri-LSB flipping, we adjust the prior art networks and propose a recurrent conditional generative adversarial network (RCGAN). The discrimination function powered by the RCGAN can be viewed as a form of structural analysis because the RCGAN learns to make a structured prediction of the original state of the flipped bits. As in the famous dictum by Richard Feynman, 'What I cannot create, I do not understand,' accurate prediction could imply good comprehension of structures of natural images and, thus, a good ability to detect abnormality and identify perturbations. The experimental results from large-scale statistical assessment showed that the structural analysis via RCGAN outperforms the aforementioned spatial and spectral analyses. The main contributions of this paper are summarised as follows:

(i) Introduction of deep neural networks to the research of cryptospace invertible steganography

(ii) Formulation of a general framework encompassing spatial, spectral, and structural analyses

(iii) Invention of the RCGAN that learns to generate reference bits in a progressive manner

The remainder of this paper is organised as follows. Section 2 revisits the associative tri-LSB flipping method and formulates some principal concepts. Section 3 presents different strategies for constructing the discrimination function. Section 4 evaluates the performance experimentally. The paper is concluded in Section 5.

2. Cyptospace Invertible Steganography

The associative tri-LSB flipping method was first proposed by Zhang [45]. It marked a significant milestone and has driven considerable research on cryptospace invertible steganography over the past decade. In this section, we reinterpret this fundamental method with a slight simplification, point out some principal concepts, and make an association with the RS method [54].

To recapitulate and give an overview of the associative tri-LSB flipping method, a workflow is outlined as follows. Consider a local client with limited computational resources and restricted access to steganographic software and, by contrast, a cloud server with an enormous capacity and a license for the software. In this scenario, outsourcing, or cloud computing, is a feasible solution for the client to entrust the task of invertible steganography to the cloud server. Due to privacy concerns, the client encrypts and uploads a cover image or a batch of images, along with an intended (compressed and encrypted) message to a cloud server, which then embeds the message into the encrypted image through the addition of invertible noise, resulting in an encrypted stego image. The client, or another authorised party, downloads, decrypts, and obtains the stego image, from which the message can be extracted and the original image can be recovered with the aid of a discrimination function.

Let us consider an 8 bit greyscale image and divide it into nonoverlapping blocks. We define a *tri-bit* as a three-bit

aggregation and abbreviate the least significant tri-bit of a pixel as *tri-LSB*. The associative tri-LSB flipping utilises a synchronous stream cipher as the encryption scheme and realises invertible noise adding by disturbing the tri-LSBs on a block basis. The synchronous stream cipher encrypts an input data by performing the XOR logical operation with a key vector generated independently of the input data. It can be viewed as an approximation of a provably secure cipher, the one-time pad [55]. The result of flipping the cipher bits, when deciphered, matches the result of flipping the plain bits since XOR is associative:

$$\begin{aligned}\text{Flip}(\text{Encrypt}(x)) &= 1 \oplus (x \oplus k) \\ &= k \oplus (x \oplus 1) \\ &= \text{Encrypt}(\text{Flip}(x)),\end{aligned} \tag{1}$$

where x denotes a tri-LSB, k is a 3 bit key vector, 1 is an all-ones vector exerting the effect of flipping, and $\oplus$ is the XOR logical operation.

Let X be a disjoint block of $n \times n$ pixels, which is written as

$$X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,n} \end{pmatrix}. \tag{2}$$

Given a secret random seed, we pseudorandomly generate a block K representing a cryptographic key for encrypting a given block X. Then, we encipher the readable pixels into an unintelligible form by

$$C = X \oplus K. \tag{3}$$

The original description of the associative tri-LSB flipping involves a steganographic key that determines which set of pixels in a block should be flipped when encoding different message bits. For simplicity but without loss of generality, we discard the notion of the optional steganographic key and simply flip all the pixels if the intended message bit is 1 and keep all the pixels unchanged if the intended message bit is 0, as expressed symbolically by

$$C' = \begin{cases} C, & \text{if } m = 0, \\ \overline{C}, & \text{if } m = 1, \end{cases} \tag{4}$$

where C represents a block of enciphered pixels, $\overline{C}$ is the flipped counterpart, and m is a message bit. After decryption, we obtain the stego block of pixels, as given by

$$X' = C' \oplus K. \tag{5}$$

As aforementioned, flipping cipher bits is equivalent to flipping plain bits when applying the associative property. Therefore, decoding the message bit, coupled with recovering the pixels, is equivalent to resolving the problem of whether the present block has been flipped. From a statistical point of view, it can be modelled as to estimate the probability of X' having been flipped. Therefore, the message can be decoded by

$$\widehat{m} = \begin{cases} 0, & \text{if } p(\text{flipped}|X') < 0.5, \\ 1, & \text{otherwise}, \end{cases} \tag{6}$$

and the pixels can be recovered by

$$\widehat{X} = \begin{cases} X', & \text{if } p(\text{flipped}|X') < 0.5, \\ \overline{X}' & \text{otherwise}. \end{cases} \tag{7}$$

For a natural image block, the estimated probability of having been flipped ought to be low if it is in its original condition and high if altered. Borrowing from the RS method, we identify a block of pixels as the regular, singular, or indeterminate class by

$$X \in \begin{cases} \text{Regular}(\mathscr{R}), & \text{if } p(\text{flipped}|X) < 0.5, \\ \text{Singular}(\mathscr{S}), & \text{if } p(\text{flipped}|X) > 0.5, \\ \text{Indeterminate}(\mathscr{I}), & \text{if } p(\text{flipped}|X) = 0.5. \end{cases} \tag{8}$$

Let us denote by $N_{\mathscr{R}}$, $N_{\mathscr{S}}$, and $N_{\mathscr{I}}$ the number of regular, singular, and indeterminate blocks, respectively. When facing an indeterminate block, there will be no alternative but to guess and the chance of being correct or wrong is equal. Thus, the accuracy of decoding and recovery can be computed by

$$\text{Accuracy: } \frac{N_{\mathscr{R}} + (N_{\mathscr{I}}/2)}{N_{\mathscr{R}} + N_{\mathscr{S}} + N_{\mathscr{I}}}. \tag{9}$$

Our objective is to construct a well-behaved discrimination function that maximises $N_{\mathscr{R}}$ while simultaneously minimising $N_{\mathscr{S}}$ and $N_{\mathscr{I}}$.

Before we move towards the construction of the discrimination function, we would like to provide a brief discussion on distortion and capacity, which are the primary concerns of most, if not all, steganographic methods. In the literature, the steganographic distortion is usually assessed by peak signal-to-noise ratio (PSNR) in decibel (dB) and the steganographic capacity is measured by embedding rate (ER), or relative payload, in bits per pixel (bpp). As can be observed from Table 1, a tri-LSB and its flipped version are always summed to 7 in the decimal numeral system, and thus, the average mean squared error (MSE) when flipping occurs can be estimated by

$$\text{MSE}_{\text{flip}}: \frac{1}{8}\sum_{t=0}^{7}[t - (7-t)]^2 = 21. \tag{10}$$

Providing that the probability of flipping is 1/2, the expected PSNR can be approximated by

$$\text{PSNR: } 10 \cdot \log_{10}\frac{255^2}{\text{MSE}_{\text{flip}}/2} = 37.92\,(\text{dB}). \tag{11}$$

The maximum ER is deterministic, as given by

$$\text{Maximum ER: } \frac{\text{total number of blocks}}{\text{total number of pixels}}. \tag{12}$$

It can be lifted by dividing the image into smaller blocks. In order to ensure reliable message decoding, we may apply

Table 1: Behaviour of tri-LSB flipping.

Binary	Decimal
000↔111	0↔7
001↔110	1↔6
010↔101	2↔5
011↔100	3↔4

error-correction code on the message in advance. Suppose that the message is encoded by the Reed–Solomon codes [56] which offer the following guarantee. Given a message of length k, a Reed–Solomon code adds check bits to the message and results in an encoded message of length n such that up to $(n-k)/2$ erroneous bits can be detected and corrected. Therefore, the expected number of erroneous message bits should not exceed the error-correction capability:

$$p \cdot n \leq \frac{n-k}{2}, \tag{13}$$

where p is the probability of erroneous message bits (i.e., inverse accuracy). To estimate the effective number of bits that can be reliably conveyed per pixel, we refer to the Reed–Solomon embedding rate (R–S ER) [57], a tailored metric for evaluating the capacity of deep-learning-based steganographic algorithms, as given by

$$\text{R} - \text{SER: } \frac{k}{\text{total number of pixels}}, \tag{14}$$

where

$$k = n(1-2p). \tag{15}$$

3. Discrimination Functions

The purpose of the discrimination function is to compute a score reflecting to what degree a given block A may have been flipped, written symbolically as

$$f\colon A \longrightarrow \mathbb{R}. \tag{16}$$

By computing further the score for its flipped counterpart $\overline{A}$, the expression can be normalised as probability:

$$p(\text{flipped}|A) = \frac{f(A)}{f(A) + f(\overline{A})}. \tag{17}$$

In this section, we unveil different ways and perspectives towards constructing the discrimination function. Specifically, we explore spatial, spectral, and structural analyses for computing the score. When not causing ambiguity and affecting reproducibility, we shall not delve into complete mathematical details of the basics of image-processing techniques; rather, we will focus on high-level description of strategies.

3.1. Spatial Analysis. Spatial analysis is the most straightforward and common way to detect abnormality of digital images. Typically, this process exploits the correlations between neighbouring pixels and measures local fluctuations. The Laplacian-based spatial discrimination functions can be considered reasonably representative. The Laplacian operator is a second-order differential operator that measures the divergence of image gradient and is sensitive to noise. To determine to what extent an observed block may have been flipped, we can estimate the noise level of it by the sum of the Laplacian, as given by

$$f_{\text{spat}}(A) = \sum |\mathscr{P}(A) * \mathscr{L}|, \tag{18}$$

where $*$ denotes the convolution operation, $\mathscr{P}$ is an optional padding mechanism, and $\mathscr{L}$ is the discrete Laplacian operator which is convolved over the block. A discrete approximation of the Laplacian operator can be realised by

$$\mathscr{L} = \frac{1}{(\alpha+1)} \begin{pmatrix} \alpha & (1-\alpha) & \alpha \\ (1-\alpha) & -4 & (1-\alpha) \\ \alpha & (1-\alpha) & \alpha \end{pmatrix}, \tag{19}$$

where $\alpha \in [0,1]$ is a parameter that controls the element values of the operator. In Section 3.2, we will examine the performance of discrete Laplacian operators with different settings of α, as shown in Figure 3.

3.2. Spectral Analysis. Spectral analysis is of utmost importance in signal processing that shows how the energy of a signal is distributed over a range of frequencies. It may be interesting to see if the minute distortion left by flipping can be traced in the frequency domain. To perform spectral analysis, we first apply the discrete Fourier transform to convert a spatial description of image data into a spectrum of frequency components. We hypothesise that high-frequency components could have more involvement with the flipping distortion than low-frequency components because flipping usually causes rapid fluctuations in pixel intensities. Under the assumption that the noise of flipping is dominant at high frequencies, we may attenuate low frequencies and retain high frequencies through a high-pass filter. The Butterworth filter is a classic signal processing filter operated in the frequency domain. It is designed to have a frequency response that is maximally flat in the passband and rolls off gradually towards zero in the stopband. We can estimate the noise level by aggregating the amplitudes of preserved frequencies, as given by

$$f_{\text{spec}}(A) = \sum |\mathscr{F}(A) \circ \mathscr{B}|, \tag{20}$$

where $\circ$ denotes the Hadamard (or elementwise) product, $\mathscr{F}$ is the discrete Fourier transform, and $\mathscr{B}$ the Butterworth filter. The filter is specified by two parameters: the cutoff frequency and the filter order. The low-pass filter is formulated by

$$\mathscr{B}_{\text{LPF}}(u,v) = \frac{1}{1 + \left[\sqrt{u^2+v^2}/\omega_c\right]^{2n}}, \tag{21}$$

where u and v are coordinates centred at zero and normalised to ± 0.5, $\sqrt{u^2+v^2}$ represents the radius relative to

0.00	1.00	0.00
1.00	−4.00	1.00
0.00	1.00	0.00

(a)

0.25	0.50	0.25
0.50	−3.00	0.50
0.25	0.50	0.25

(b)

0.33	0.33	0.33
0.33	−2.67	0.33
0.33	0.33	0.33

(c)

0.40	0.20	0.40
0.20	−2.40	0.20
0.40	0.20	0.40

(d)

0.50	0.00	0.50
0.00	−2.00	0.00
0.50	0.00	0.50

(e)

Figure 3: Pictured from left to right are the Laplacian operators with different parameter settings. (a)$\mathscr{L}(\alpha = 0)$. (b)$\mathscr{L}(\alpha = 1/3)$. (c)$\mathscr{L}(\alpha = 1/2)$. (d)$\mathscr{L}(\alpha = 2/3)$. (e)$\mathscr{L}(\alpha = 1)$.

centre, ω_c is the cutoff frequency ranging from 0 to 0.5, and n is the filter order. By contrast, the high-pass filter is constructed by

$$\mathscr{B}_{\mathrm{HPF}}(u, v) = 1 - \mathscr{B}_{\mathrm{LPF}}(u, v). \tag{22}$$

In Section 3.3, we will test our hypothesis that the noise of flipping is dominant at high frequencies rather than low frequencies by using the filters shown in Figure 4.

3.3. Structural Analysis. Both spatial and spectral strategies are implemented on a block basis. While blockwise approaches are workable when the block size is sufficiently large, they may suffer from a relatively restricted receptive field when the block size is small. Due to the smoothness nature, it is often the case that a small block is composed of pixels having identical values. In this case, both flipped and unflipped blocks will probably be assigned the same score and become indistinguishable. The underlying cause is that context information outside the block is entirely ruled out and ignored. Hence, we propose the question: how can we effectively and efficiently incorporate context information beyond the local area, or even make full use of all the credible information?

For the associative tri-LSB flipping method, it is reasonable to think of the unchanged five bit-planes as the credible information. We can, therefore, exploit the five upper planes to predict the remaining three lower planes and then use the output as reference. For a query image block A, we may compare it with the corresponding reference block $\widetilde{A}$ extracted from the prediction output to obtain a score indicating the distance. The remaining task is to devise a suitable prediction mechanism.

To this end, we construct RCGAN by stacking up multiple conditional generative adversarial networks (CGANs), as illustrated in Figure 5. Each CGAN is trained to synthesise a lower bit-plane conditioned on the five upper bit-planes and the output planes from the previous CGANs. The synthesis of bit-planes is processed in a top-down manner. We would like to note that, during the training stage, the input to each CGAN is the real bit-planes instead of the synthetic bit-planes from the former CGANs, and each individual CGAN is trained independently. While there are many ways to realise the CGANs, we adopt the pix2pix model [58], a seminal model for various image-to-image translation tasks. This model is composed of a U-Net generator [59] and a Markovian discriminator. We do not lay out the details regarding the pix2pix since there are many available resources and tutorials on the specifics. Further implementation details of the pix2pix for bit-plane synthesis can be found in the author's previous work [53]. It seems possible that the RCGAN can learn the structure of bit-planes and generate realistic ones. Thus, we suggest calculating the distance between a query image block A and a synthetic reference block $\widetilde{A}$ by structural similarity index measure (SSIM):

$$f_{\mathrm{struc}}(A) = \mathrm{SSIM}(A, \widetilde{A}). \tag{23}$$

We will validate the effectiveness of this approach in the following section.

4. Experimental Results

In this section, we evaluate the performance of cryptospace invertible steganography using different discrimination functions. In our experiments, we use a random bit stream to simulate the intended message which is assumed to have been compressed and encrypted. First and foremost, we would like to evaluate the effectiveness of the RCGAN for generating accurate reference images. We begin by evaluating the error rate of synthetic bit-planes and the structural similarity of the synthetic reference images. Then, we move from the effectiveness of the RCGAN to how it may benefit invertible steganography. In particular, we evaluate the accuracy of decoding and recovery, as well as the Reed–Solomon steganographic capacity. Furthermore, we are interested in the superiority of the structural discrimination function based on deep neural networks over the spatial and spectral ones based on common image-processing tools and handcrafted features. We compare their average accuracies of decoding and recovery, as well as their average percentages of regular, singular, and indeterminate cases. Last but not least, we analyse the security of encryption by showing a uniform distribution over the cryptospace.

4.1. Datasets. The image samples for training and testing are from the BOSSbase [60]. This database originated from an academic competition for steganography and has been recognised as one of the most prestigious ones since. It contains a collection of 10000 greyscale photographs

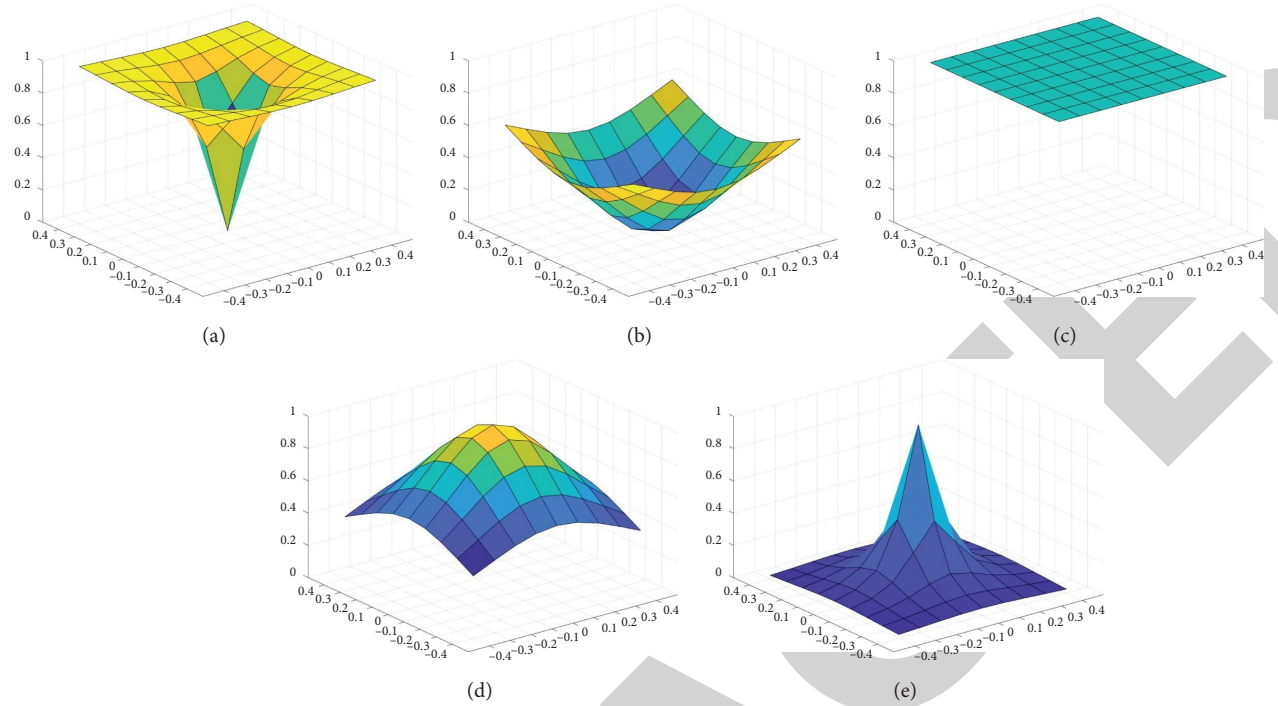


FIGURE 4: Pictured from left to right are the Butterworth high-pass, all-pass, and low-pass filters with fixed order ($n = 1$) and different cutoff frequency settings. (a)$\mathscr{B}_{\mathrm{HPF}}(\omega_c = 0.1)$. (b)$\mathscr{B}_{\mathrm{HPF}}(\omega_c = 0.5)$. (c)$\mathscr{B}_{\mathrm{APF}}$. (d)$\mathscr{B}_{\mathrm{LPF}}(\omega_c = 0.5)$. (e)$\mathscr{B}_{\mathrm{LPF}}(\omega_c = 0.1)$.

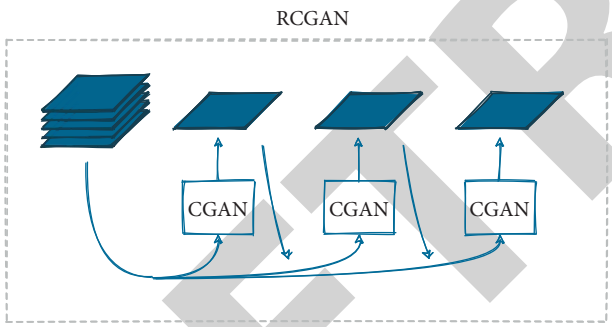


FIGURE 5: The RCGAN constructed by stacking independently trained CGANs for synthesising bit-planes in a top-down manner.

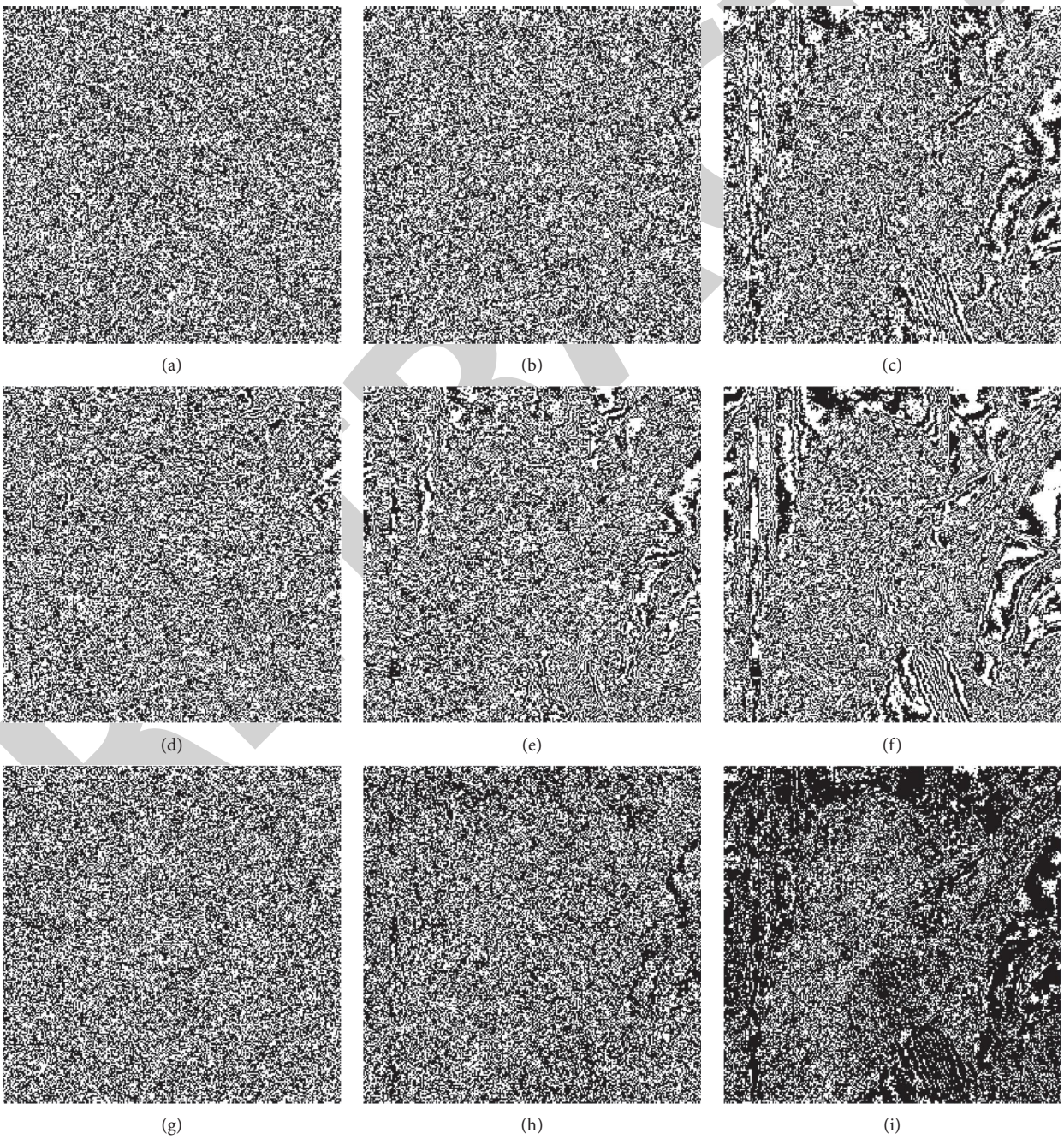
covering a wide variety of topics and scenes. In our experiments, we use 8000 samples for training the neural network model and the other 2000 samples for performance evaluations and analyses. For future reference, we also include experimental results on some commonly used test images selected from the USC-SIPI dataset [61], as shown in Figure 6. Throughout the experiments, all the images were converted to 8 bit greyscale and resampled to 256×256 pixels.

4.2. Evaluations. Starting from Figure 7, we can catch a glimpse of how the synthetic bit-planes look and how much difference there might be between the real and the synthetic ones. A quantitative assessment based on a large amount of data is provided in Figure 8, showing the bit error rate (BER) of synthetic bit-planes of differing order. As expected, bit-planes of a higher order can be generated with fewer errors. It is notable that even for the least significant bit-plane, the error rate on average is, though only slightly, better than random guessing. Accurate prediction of the least significant bit-plane is challenging due to error propagation from the synthetic upper bit-planes. Figure 9 shows the SSIM of reference images created by merging the synthetic lower bit-planes with the intact upper bit-planes. It suggests that the quality of reference images is generally high in terms of structural information.

Turning to the heart of cryptospace invertible steganography, the accuracy of message decoding and image recovery is reported in Figure 10. By viewing the problem of decoding and recovery as that of binary classification, we can interpret the performance with the receiver operating characteristic (ROC) curve by plotting the true positive rate (TPR) against false positive rate (FPR) at various thresholds. The diagonal corresponds to the performance of random guessing and the further from the diagonal the better performance achieved. It can be observed that accuracy is directly proportional to the block size. While a larger block size could yield a gain of more correctly decoded message bits, the block size itself puts a ceiling on the maximum capacity, as the message is embedded at one bit per block. It is therefore interesting to analyse the R–S ER at different settings of block size. Figure 11 suggests that a much greater number of bits can be effectively conveyed with a smaller size of blocks. We can also observe a relatively varied distribution of capacity for a small block size in contrast to a fairly consistent distribution of capacity for a large block size. The underlying explanation is that a near-

(a) (b) (c) (d)

Figure 6: Standard test images selected from the USC-SIPI dataset. (a) Aeroplane. (b) Lena. (c) Mandrill. (d) Peppers.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 7: Examples of real and synthetic bit-planes and their residuals. The bit-plane order is represented by subscript. (a) Real_1. (b) Real_2. (c) Real_3. (d) Synthetic_1. (e) Synthetic_2. (f) Synthetic_3. (g) Residual_1. (h) Residual_2. (i) Residual_3.

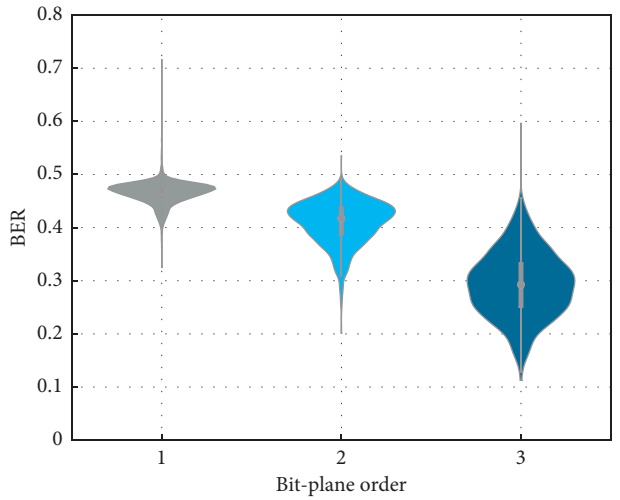


Figure 8: Bit error rates of synthetic bit-planes.

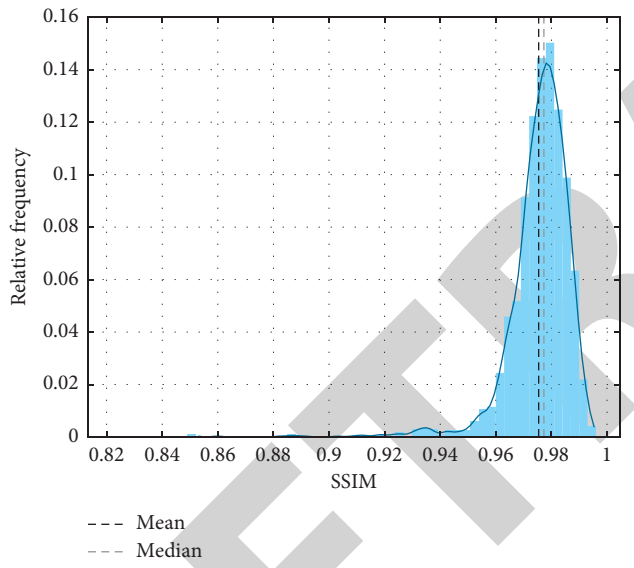


Figure 9: Structural similarity index measures of synthetic images.

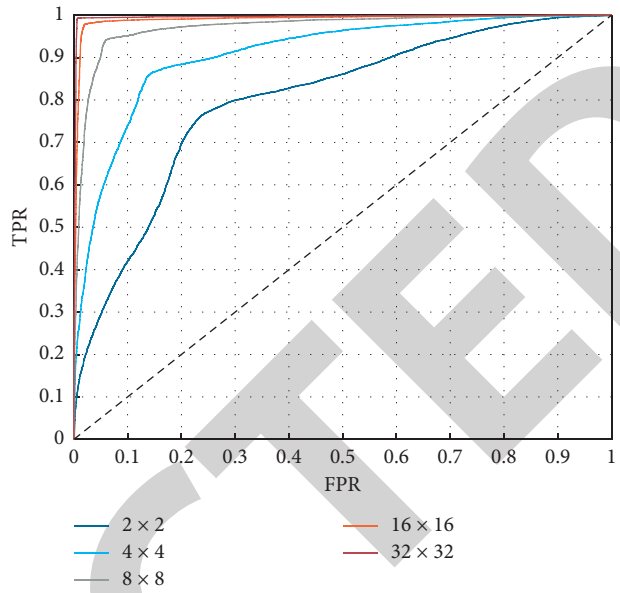


Figure 10: Receiver operating characteristic curves of structural analysis with regard to different block sizes.

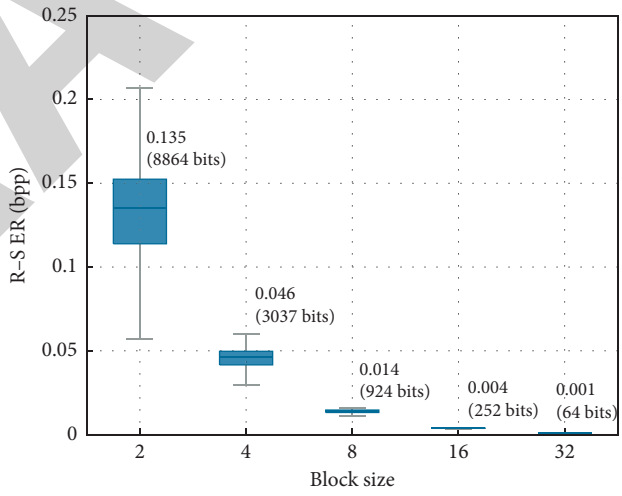


Figure 11: Reed–Solomon embedding rates with respect to different block sizes.

perfect decoding accuracy is achieved with a large block size for most of the images, while the accuracy with a small block size is much more dependent on the content of images. A summary of the performance on selected test images is reported in Table 2.

4.3. Comparisons. Before comparing against spatial and spectral discrimination functions, we deliver analyses on their parameter configurations. Figure 12 demonstrates the accuracy of decoding and recovery by using different Laplacian operators. While there seems no significant gap between the performances of different operators, the best results were achieved by configuring $\alpha = 0$, which is in fact equivalent to the discrimination function originally described in the literature of associative tri-LSB flipping scheme [45]. Figure 13 shows accuracy when using high-pass, all-pass, and low-pass Butterworth filters. The best results were obtained by using a high-pass filter with $\omega_c = 0.1$, which validated our hypothesis that steganographic distortion caused by tri-LSB flipping is primarily concentrated at high frequencies. By applying the best configurations, Figure 14 compares the structural discrimination function which uses neural networks as the backbone against the Laplacian-based spatial approach and the Butterworth-based spectral approach. The results suggest that although the three strategies all converged to a near-perfect accuracy with a large block size, the structural strategy outperformed the others significantly when a small block size was used.

It would be helpful to have a more in-depth analysis of how the three strategies discriminate image blocks, and hence, we provide the statistics on relative frequencies of the

Table 2: A summary of accuracy, capacity, and distortion on standard test images.

Block size	2 × 2		4 × 4		8 × 8		16 × 16		32 × 32	
	Accuracy	R–S ER	Accuracy	R–S ER	Accuracy	R–S ER	Accuracy	R–S ER	Accuracy	R–S ER
Aeroplane	0.769	0.134	0.885	0.048	0.960	0.014	0.984	0.004	1.000	0.001
Lena	0.800	0.150	0.914	0.052	0.987	0.015	1.000	0.004	1.000	0.001
Mandrill	0.641	0.070	0.740	0.030	0.862	0.011	0.961	0.004	1.000	0.001
Peppers	0.822	0.161	0.948	0.056	0.995	0.016	1.000	0.004	1.000	0.001
Average PSNR	38.02 (dB)		38.04 (dB)		38.09 (dB)		37.75 (dB)		38.13 (dB)	
Maximum ER	0.250 (bpp)		0.063 (bpp)		0.016 (bpp)		0.004 (bpp)		0.001 (bpp)	

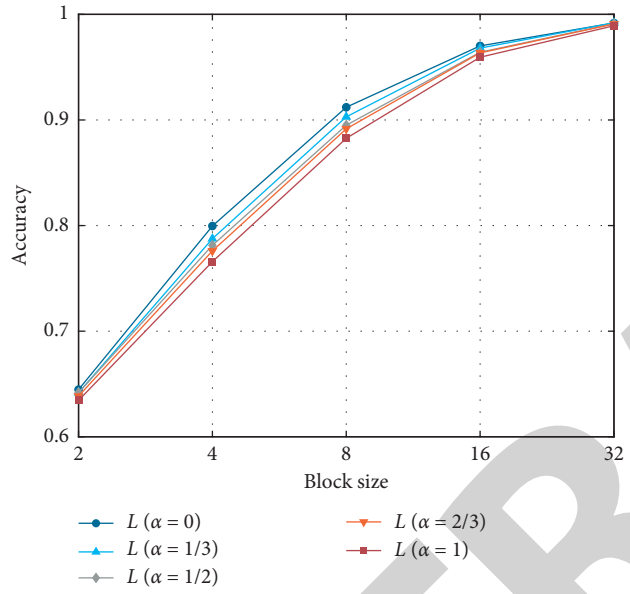


Figure 12: Accuracy evaluation by applying different Laplacian spatial filters.

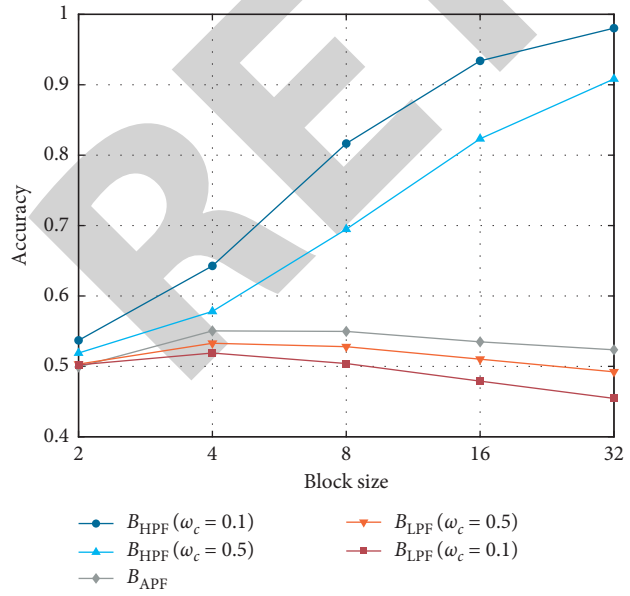


Figure 13: Accuracy evaluation by applying different Butterworth spectral filters.

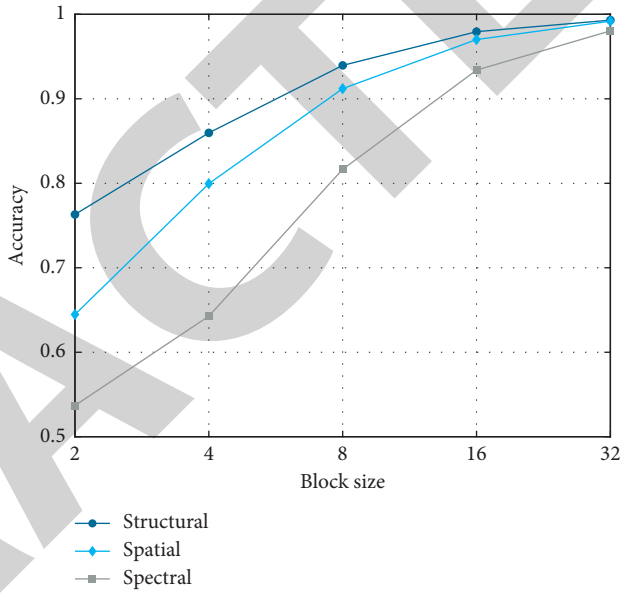


Figure 14: Accuracy comparison amongst structural, spatial, and spectral analyses.

regular, singular, and indeterminate cases. Figure 15 illustrates some examples of RSI maps produced by using spatial, spectral, and structural discrimination functions, and Figure 16 presents the average percentages of regular, singular, and indeterminate cases based on a large number of test samples. It is evident that the spatial and spectral strategies are much more likely to make an indeterminate decision due to the problem of restricted receptive field, which conforms with our presumption. The percentage of regular cases increases monotonically with the block size as expected.

4.4. Security Analysis. We close our experiments with a security analysis. It can be observed from Figure 17 that semantic secrecy is preserved because the image in a state of encryption is visually random and semantically uninterpretable. By examining the histogram of the encrypted image, the occurrence of each intensity value is virtually even, suggesting a uniform distribution and thus statistical secrecy.

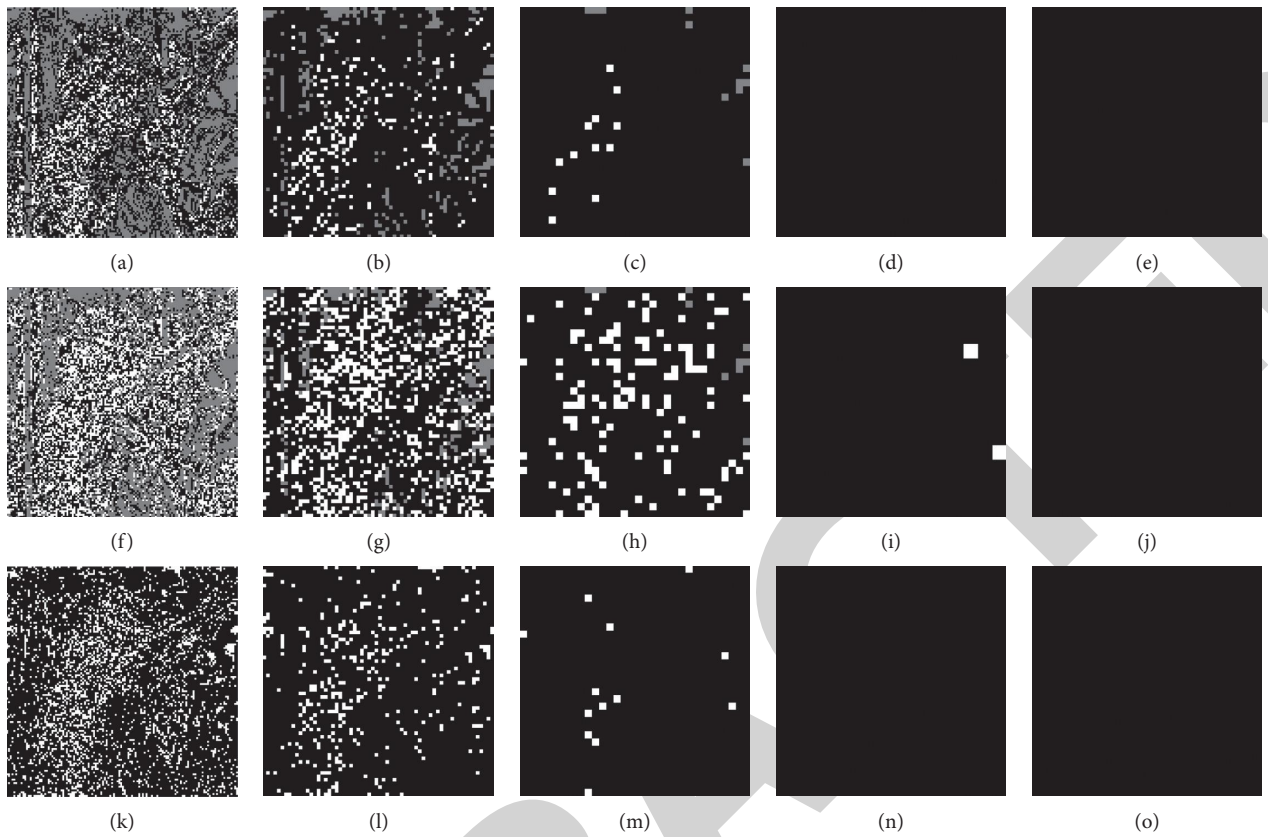


Figure 15: Examples of RSI maps by spatial, spectral, and structural analyses with respect to different block sizes. Regular, singular, and indeterminate blocks are coloured in black, white, and grey, respectively. (a) Spatial_2. (b) Spatial_4. (c) Spatial_8. (d) Spatial_{16}. (e) Spatial_{32}. (f) Spectral_2. (g) Spectral_4. (h) Spectral_8. (i) Spectral_{16}. (j) Spectral_{32}. (k) Structural_2. (l) Structural_4. (m) Structural_8. (n) Structural_{16}. (o) Structural_{32}.

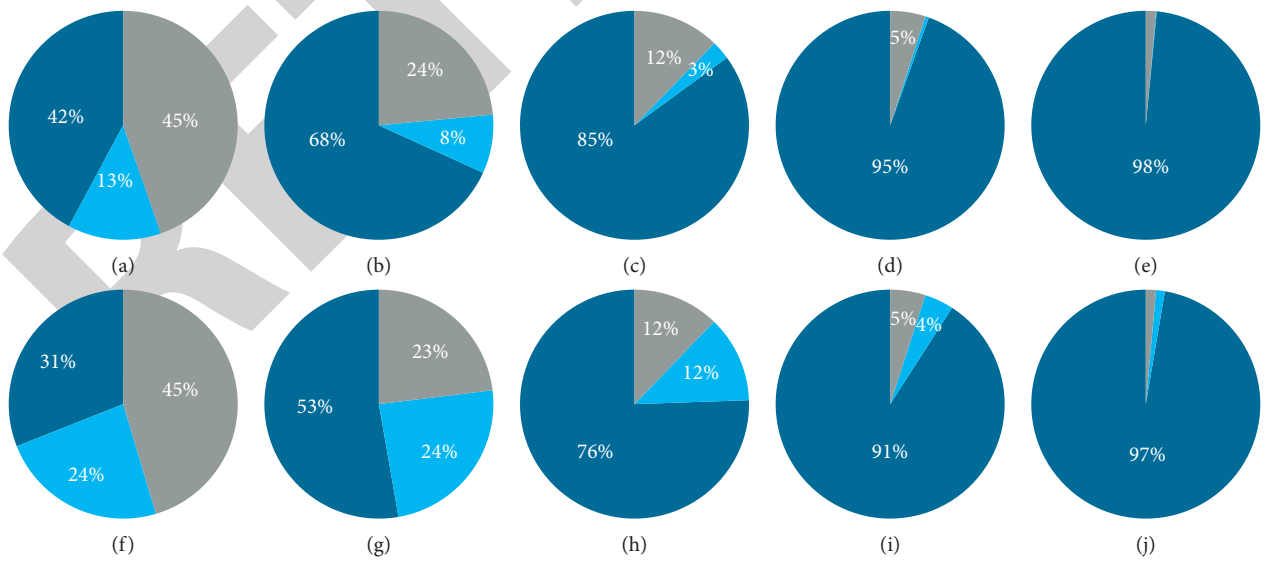


Figure 16: Continued.

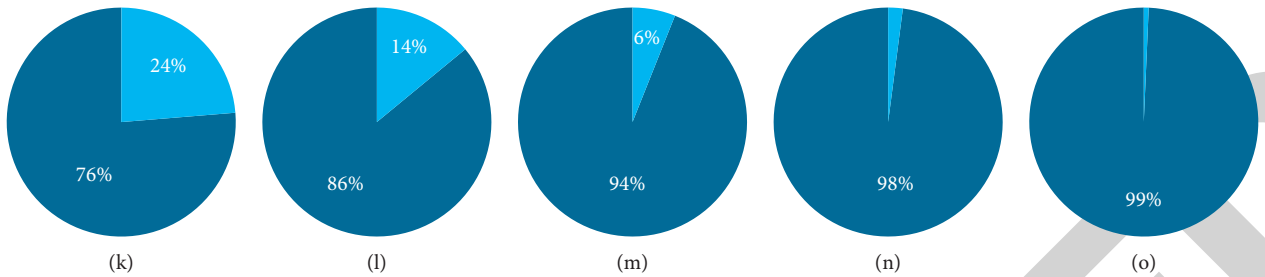


Figure 16: Average percentages of regular (dark blue), singular (light blue), and indeterminate (grey) blocks by spatial, spectral, and structural analyses with respect to different block sizes. (a) Spatial_2. (b) Spatial_4. (c) Spatial_8. (d) Spatial_{16}. (e) Spatial_{32}. (f) Spectral_2. (g) Spectral_4. (h) Spectral_8. (i) Spectral_{16}. (j) Spectral_{32}. (k) Structural_2. (l) Structural_4. (m) Structural_8. (n) Structural_{16}. (o) Structural_{32}.

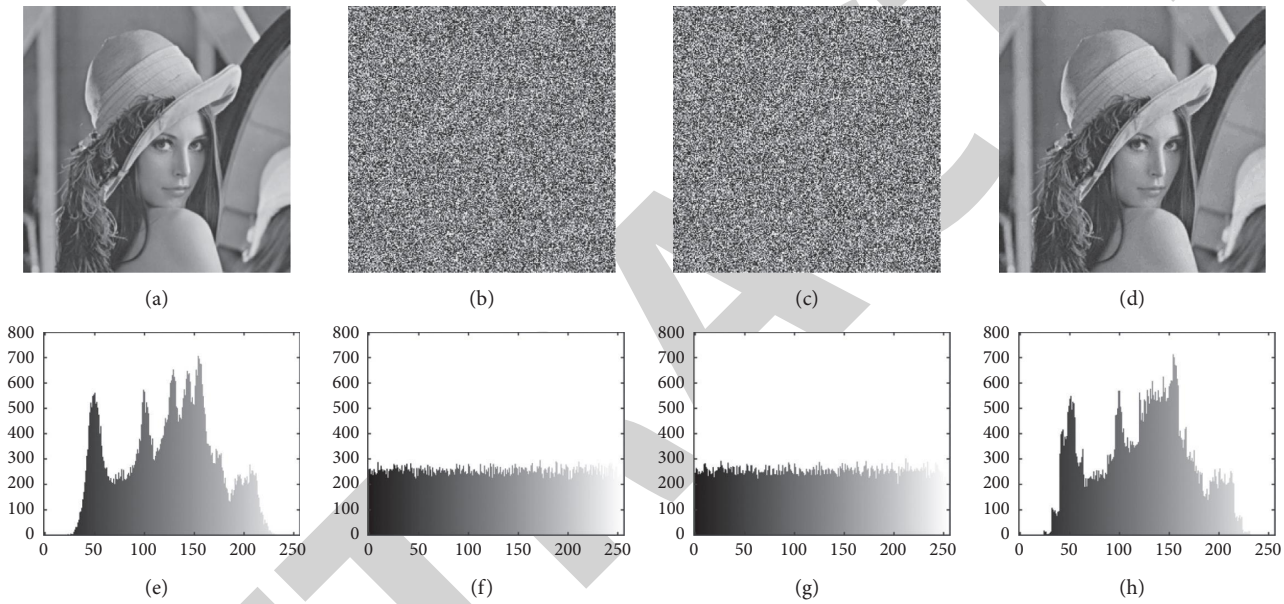


Figure 17: Security analysis in terms of semantic and statistical secrecy. Top row: resultant images at different stages. Bottom row: corresponding image histograms. (a) Cover image. (b) Encrypted image. (c) Encrypted stego image. (d) Stego image. (e) Cover image. (f) Encrypted image. (g) Encrypted stego image. (h) Stego image.

5. Conclusion

In this paper, we neuralised a classic method of cryptospace invertible steganography by introducing generative adversarial networks. We validated the effectiveness of the RCGAN for learning structural information of bit-planes and generating realistic ones in a top-down manner. In addition, we analysed the performance of spatial, spectral, and structural discrimination functions and demonstrated the superiority of deep neural networks over traditional handcrafted analytics. Furthermore, we showed that the applied encryption scheme for digital images satisfies semantic and statistical perfect secrecy. We envision that, by exploring the potential of deep neural networks, the accuracy and capacity can be further improved. It is also interesting to investigate the possibility of assembling handcrafted and learnt features. We hope this article can prove instructive for future research on cryptospace invertible steganography with deep learning.

Data Availability

The image data and the neural network model used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] I. Agrafiotis, J. R. C. Nurse, M. Goldsmith, S. Creese, and D. Upton, "A taxonomy of cyber-harms: defining the impacts of cyber-attacks and understanding how they propagate," *Journal of Cybersecurity*, vol. 4, no. 1, p. 10, 2018.

[2] Z. Erkin, A. Piva, S. Katzenbeisser et al., "Protection and retrieval of encrypted multimedia content: when

cryptography meets signal processing," *EURASIP Journal on Information Security*, vol. 20, p. 17, 2007.

[3] R. L. Lagendijk, M. Barni, and M. Barni, "Encrypted signal processing for privacy protection: conveying the utility of homomorphic encryption and multiparty computation," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 82–105, 2013.

[4] C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat, and R. Sirdey, "Recent advances in homomorphic encryption: a possible future for signal processing in the encrypted domain," *IEEE Signal Processing Magazine*, vol. 30, no. 2, pp. 108–117, 2013.

[5] J. R. Troncoso-Pastoriza and F. Perez-Gonzalez, "Secure signal processing in the cloud: enabling technologies for privacy-preserving multimedia cloud processing," *IEEE Signal Processing Magazine*, vol. 30, no. 2, pp. 29–41, 2013.

[6] M. Barni, G. Droandi, and R. Lazzeretti, "Privacy protection in biometric-based recognition systems: a marriage between cryptography and signal processing," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 66–76, 2015.

[7] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding-a survey," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1062–1078, 1999.

[8] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, "Digital image steganography: survey and analysis of current methods," *Signal Processing*, vol. 90, no. 3, pp. 727–752, 2010.

[9] M. T. Ahvanooey, Q. Li, H. J. Shim, and Y. Huang, "A comparative analysis of information hiding techniques for copyright protection of text documents," *Security and Communication Networks*, vol. 2018, 2018.

[10] J. Fridrich, M. Goljan, P. Lisonek, and D. Soukal, "Writing on wet paper," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3923–3935, 2005.

[11] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

[12] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1167–1180, 1999.

[13] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Transactions on Information Theory*, vol. 44, no. 5, pp. 1897–1905, 1998.

[14] Y.-Q. Shi, X. Li, X. Zhang, H.-T. Wu, and B. Ma, "Reversible data hiding: advances in the past two decades," *IEEE Access*, vol. 4, pp. 3210–3237, 2016.

[15] J. A. Stankovic, "Research directions for the Internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 3–9, 2014.

[16] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.

[17] C.-C. Chang, C.-T. Li, and C.-T. Li, "Algebraic secret sharing using privacy homomorphisms for IoT-basedhealthcare systems," *Mathematical Biosciences and Engineering*, vol. 16, no. 5, pp. 3367–3381, 2019.

[18] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proceedings of International Conference on Learning Representations (ICLR)*, pp. 1–11, San Diego, CA, USA, 2015.

[19] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "DeepFool: a simple and accurate method to fool deep neural networks," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2574–2582, Las Vegas, NV, USA, 2016.

[20] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, "Universal adversarial perturbations," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 86–94, Honolulu, HI, USA, 2017.

[21] C. De Vleeschouwer, J.-F. Delaigle, and B. Macq, "Circular interpretation of bijective transformations in lossless watermarking for media asset management," *IEEE Transactions on Multimedia*, vol. 5, no. 1, pp. 97–105, 2003.

[22] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 890–896, 2003.

[23] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 354–362, 2006.

[24] D. M. Thodi and J. J. Rodriguez, "Expansion embedding techniques for reversible watermarking," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 721–730, 2007.

[25] V. Sachnev, H. J. Hyoung Joong Kim, J. Jeho Nam, S. Suresh, and Y.-Q. Yun Qing Shi, "Reversible watermarking algorithm using sorting and prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 7, pp. 989–999, 2009.

[26] D. Coltuc, "Low distortion transform for reversible watermarking," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 412–417, 2012.

[27] C.-C. Chang, C.-T. Li, and Y.-Q. Shi, "Privacy-aware reversible watermarking in cloud computing environments," *IEEE Access*, vol. 6, 2018.

[28] F. Huang, J. Huang, and Y.-Q. Shi, "New framework for reversible data hiding in encrypted domain," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2777–2789, 2016.

[29] W. Zhang, H. Wang, D. Hou, and N. Yu, "Reversible data hiding in encrypted images by reversible image transformation," *IEEE Transactions on Multimedia*, vol. 18, no. 8, pp. 1469–1479, 2016.

[30] C. Yu, X. Zhang, Z. Tang, and X. Xie, "Separable and error-free reversible data hiding in encrypted image based on two-layer pixel errors," *IEEE Access*, vol. 6, 2018.

[31] S. Yi and Y. Zhou, "Separable and reversible data hiding in encrypted images using parametric binary tree labeling," *IEEE Transactions on Multimedia*, vol. 21, no. 1, pp. 51–64, 2019.

[32] Y. Wang, Z. Cai, and W. He, "A new high capacity separable reversible data hiding in encrypted images based on block selection and block-level encryption," *IEEE Access*, vol. 7, 2019.

[33] K. Ma, W. Zhang, X. Zhao, N. Yu, and F. Li, "Reversible data hiding in encrypted images by reserving room before encryption," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 3, pp. 553–562, 2013.

[34] X. Cao, L. Du, X. Wei, D. Meng, and X. Guo, "High capacity reversible data hiding in encrypted images by patch-level sparse representation," *IEEE Transactions on Cybernetics*, vol. 46, no. 5, pp. 1132–1143, 2016.

[35] P. Puteaux and W. Puech, "An efficient MSB prediction-based method for high-capacity reversible data hiding in encrypted images," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1670–1681, 2018.

[36] Z. Yin, Y. Xiang, and X. Zhang, "Reversible data hiding in encrypted images based on multi-MSB prediction and Huffman coding," *IEEE Transactions on Multimedia*, vol. 22, no. 4, pp. 874–884, 2020.

[37] D. Xiao, F. Li, M. Wang, and H. Zheng, "A novel high-capacity data hiding in encrypted images based on compressive sensing progressive recovery," *IEEE Signal Processing Letters*, vol. 27, pp. 296–300, 2020.

[38] L. Liu, A. Wang, and C.-C. Chang, "Separable reversible data hiding in encrypted images with high capacity based on median-edge detector prediction," *IEEE Access*, vol. 8, 2020.

[39] I. C. Dragoi and D. Coltuc, "On the security of reversible data hiding in encrypted images by MSB prediction," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 187–189, 2021.

[40] X. Zhang, "Separable reversible data hiding in encrypted image," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 826–832, 2012.

[41] W. Hong, T.-S. Chen, and H.-Y. Wu, "An improved reversible data hiding in encrypted images using side match," *IEEE Signal Processing Letters*, vol. 19, no. 4, pp. 199–202, 2012.

[42] J. Zhou, W. Sun, L. Dong, X. Liu, O. C. Au, and Y. Y. Tang, "Secure reversible image data hiding over encrypted domain via key modulation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 441–452, 2016.

[43] Z. Qian, X. Zhang, and G. Feng, "Reversible data hiding in encrypted images based on progressive recovery," *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1672–1676, 2016.

[44] C.-C. Chang, C.-T. Li, and K. Chen, "Privacy-preserving reversible information hiding based on arithmetic of quadratic residues," *IEEE Access*, vol. 7, pp. 54 117–154 132, 2019.

[45] X. Zhang, "Reversible data hiding in encrypted image," *IEEE Signal Processing Letters*, vol. 18, no. 4, pp. 255–258, 2011.

[46] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, pp. 1097–1105, 2012.

[48] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 27, pp. 3104–3112, 2014.

[49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of International Conference on Learning Representations (ICLR)*, pp. 1–14, New York, NY, USA, 2015.

[50] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: a neural image caption generator," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3156–3164, New York, NY, USA, 2015.

[51] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, New York, NY, USA, 2015.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, New York, NY, USA, 2016.

[53] C.-C. Chang, "Adversarial learning for invertible steganography," *IEEE Access*, vol. 8, 2020.

[54] J. Fridrich, M. Goljan, and R. Du, "Lossless data embedding-new paradigm in digital watermarking," *EURASIP Journal on Advances in Signal Processing*, vol. 986842, 2002.

[55] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

[56] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[57] K. A. Zhang, A. Cuesta-Infante, L. Xu, and K. Veeramachaneni, "SteganoGAN: High Capacity Image Steganography with GANs," 2019.

[58] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, Honolulu, HI, USA, 2017.

[59] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, Munich, Germany, 2015.

[60] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system: the ins and outs of organizing BOSS," in *Proceedings of International Workshop on Information Hiding (IH)*, pp. 59–70, Prague, Czech Republic, 2011.

[61] A. G. Weber, "The USC-SIPI image database: version 5, USC viterbi School of engineering, signal and image processing Institute," 2006.