IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Adversarial Learning for Invertible Steganography

CHING-CHUN CHANG[1]
[1]Department of Computer Science, University of Warwick, Coventry, UK (e-mail: ching-chun.chang@warwickgrad.net)

ABSTRACT Deep neural networks have revolutionised the research landscape of steganography. However, their potential has not been explored in invertible steganography, a special class of methods that permits the recovery of distorted objects due to steganographic perturbations to their pristine condition. In this paper, we revisit the regular-singular (RS) method and show that this elegant but obsolete invertible steganographic method can be reinvigorated and brought forwards to modern generation via *neuralisation*. Towards developing a renewed RS method, we introduce adversarial learning to capture the regularity of natural images automatically in contrast to handcrafted discrimination functions based on heuristic image prior. Specifically, we train generative adversarial networks (GANs) to predict bit-planes that have been used to carry hidden information. We then form a synthetic image and use it as a reference to provide guidance on data embedding and image recovery. Experimental results showed a significant improvement over the prior implementation of the RS method based on large-scale statistical evaluations.

INDEX TERMS Convolutional neural networks, generative adversarial networks, invertible steganography.

I. INTRODUCTION

STEGANOGRAPHY is the art and science of hiding information within a seemingly innocuous carrier or cover. The word is derived from the Greek *steganós* and *graphein*, literally 'covered writing' [1]. To date, steganography has found a variety of applications, including but not limited to covert communication [2]–[6], copyright protection [7]–[9], tamper detection [10]–[12], broadcast monitoring [13], traitor tracing [14].

Most of the steganographic methods inevitably distort the cover objects with a small amount of noise as the price to pay for carrying hidden data. Although the distortion is often quite imperceptible, this condition might not be acceptable under certain circumstances in which data integrity and high resolution are important, for example, remote sensing and medical imaging. In today's big data era, steganography, or more frequently addressed as watermarking, can be used as a means to help archiving data through inserting digital object identifier, digital signature or metadata, and facilitate verification of the authenticity when distributing the samples. However, recent studies have shown that deep learning models can be susceptible to some deliberately crafted small noise called *adversarial perturbations*, causing the output to change drastically [15]–[19]. While no claim has been made that steganographic noise would to any extent poison and contaminate the dataset like specially engineered perturbations, it is desirable to undo the changes and recover an untainted clean copy of the samples for good measure, as the proverb goes, 'a stitch in time may save nine'.

As far as the author is aware, the concept of *invertible steganography*, (also known as erasable watermarking, lossless embedding, or reversible data hiding), dates back to about two decades ago. One of the earliest schemes was introduced for the purpose of image authentication and was filed in a US patent [20]. It described a method of embedding a digital signature in an image through modulo operations, but it suffers from a drawback of salt-and-pepper artefacts. Another early study suggested to apply lossless compression on bit-planes [21]. However, when the length of an intended message is greater than permissible loading capacity provided by compressing a given bit-plane, it resorts to a higher bit-plane for a higher compression rate, causing the artefacts quickly become visible. By today's standard, these are not considered as a pragmatic approach.

The first practical and elegant methodology was proposed by Goljan, Fridrich and Du [22]. This seminal paper defined a general construction that utilises an invertible noise-adding and a discrimination function to realise invertible steganography. The data embedding and image recovery processes are guided by a map that classifies blocks of pixels into *regular*, *singular* and *unusable* groups, and therefore it is referred to as the regular-singular (RS) method. It led up to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3034936, IEEE Access

vigorous progress and evolution of invertible steganography [23] and was followed by a fair number of methods including recursive code construction [24]–[26], difference expansion [27]–[29], histogram shifting [30]–[33], circular interpretation [34], wavelet transform [35], code division multiplexing [36], to name but a few.

Recent advances in deep learning has driven an unprecedented revolution in academia and industry, and the research area of steganography is no exception [37]. The recent development of various steganographic methods based on deep learning can be characterised as cover modification [38]–[43], cover synthesis [44], and cover selection [45]. However, these methods either make non-erasable alterations to the cover objects or restrict the choice of them, and thus none meets the conventional definition and requirement of invertible steganography.

As we looked back upon the history of invertible steganography, we found that the RS method offers an elegant framework, allowing us to infuse new life into it with deep learning technology, or to *neuralise* it. In this paper, we revisit the RS method and explore *adversarial learning for invertible steganography* (acronym: ALIS). In an attempt to present a new perspective and develop a ground-breaking method based on deep neural networks, we build generative adversarial networks (GANs) to assist synthesis of a realistic reference image, which is subsequently utilised as a hint to guide message embedding and image recovery. Specifically, the proposed GANs are designed to learn the binary-space distribution and reconstruct a bit-plane that has been replaced with the payload from other preserved bit-planes. Experimental results validated the effectiveness of the proposed method and showed a significant performance boost. The major contribution of this paper is to demonstrate the potential of ALIS for bringing an outdated invertible steganographic method forwards into modern generation with a notable improvement and hopefully inspire future research.

The remainder of this paper is organised as follows. Section II revisits the RS method, points out some principal features and recognises a reserve of latent improvement. Section III presents the proposed model for reference-image synthesis. Section IV translates the method into practice and evaluates the performance experimentally. Section V carries out further analyses, discusses the limitations and outlines the directions for future research. The paper is concluded in Section VI.

II. REGULAR-SINGULAR METHOD

The RS method realises lossless data embedding through an invertible noise adding and a discrimination function. Consider an 8-bit greyscale cover image of $H \times W$ pixels with 256 different intensities (i.e. shades of grey), numbered from 0 to 255. To begin with, we divide the image into disjoint blocks of $n \times n$ pixels, written as

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

We define an invertible noise adding operation as an involutory function such that the function itself is its own inverse:

$$f(f(X)) = X \, . \tag{2}$$

This operation can be further parameterised by an amplitude factor α that describes the average change of X by the operation, as denoted by

$$\overline{X} = f_\alpha(X) \, , \tag{3}$$

where

$$\alpha = \frac{\Delta(\overline{X}, X)}{n \cdot n} \, . \tag{4}$$

A straightforward realisation is *bit flipping*. When $\alpha = 1$, we can implement an invertible noise adding operation by LSB flipping. In general, let us denote by β the order of a bit-plane and thus

$$f_\alpha(X) = \mathrm{Flip}_\beta(X) \, , \tag{5}$$

where $\alpha = 2^\beta$. Next, we define a discrimination function that assigns a score to each image block, written as

$$g \colon X \to \mathbb{R} \, . \tag{6}$$

This function is typically designed to capture and reflect the *regularity* of natural images in such a manner that a given block should have a low score if it is in its original condition and high if altered. It can be instantiated to calculate the variance of a block of pixels based on the smoothness prior. On the grounds of the computed score, we discriminate blocks into the following three types:

$$X \in \begin{cases} \text{Regular } (\mathcal{R}) & \text{if } g(X) < g(\overline{X}) \, , \\ \text{Singular } (\mathcal{S}) & \text{if } g(X) > g(\overline{X}) \, , \\ \text{Unusable } (\mathcal{U}) & \text{if } g(X) = g(\overline{X}) \, . \end{cases} \tag{7}$$

We can further derive that

$$\begin{aligned} &\text{if } X \in \mathcal{R} \, , \ \overline{X} \in \mathcal{S} \, , \\ &\text{if } X \in \mathcal{S} \, , \ \overline{X} \in \mathcal{R} \, , \\ &\text{if } X \in \mathcal{U} \, , \ \overline{X} \in \mathcal{U} \, . \end{aligned} \tag{8}$$

By scanning the cover image with the discrimination mechanism, we can form an RS map by assigning 0s to the regular blocks and 1s to the singular blocks while the unusable blocks are simply skipped as they can be unambiguously identified. Based on the premise of a *well-behaved* discrimination function, we heuristically expect a bias between the numbers of regular and singular groups. This allows us to losslessly compress the map. Let the number of regular, singular, and unusable blocks be denoted by $N_\mathcal{R}$, $N_\mathcal{S}$, and $N_\mathcal{U}$ respectively. The relative frequencies (i.e. probabilities) of $N_\mathcal{R}$ and $N_\mathcal{S}$ are

$$p_\mathcal{R} = \frac{N_\mathcal{R}}{N_\mathcal{R} + N_\mathcal{S}} \, , \tag{9}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3034936, IEEE Access

and

$$p_{\mathcal{S}} = \frac{N_{\mathcal{S}}}{N_{\mathcal{R}} + N_{\mathcal{S}}} \,. \tag{10}$$

Following Shannon's source coding theorem [46], the number of bits required to represent the RS map using a context-free entropy coding is given by

$$|\mathrm{Compress}(\text{RS map})| = (N_{\mathcal{R}} + N_{\mathcal{S}}) \cdot \mathrm{H}(\text{RS map}) \,, \tag{11}$$

where $\mathrm{H}(\text{RS map})$ is the unconditional (i.e. zero-order) entropy of the RS map:

$$\mathrm{H}(\text{RS map}) = -(p_{\mathcal{R}} \log_2 p_{\mathcal{R}} + p_{\mathcal{S}} \log_2 p_{\mathcal{S}}) \,. \tag{12}$$

Now we are equipped with all the preliminary concepts of the RS method: the invertible noise adding operation, the discrimination function, and the RS map. Let us denote by m a message bit and by Y a stego block. We can embed a message bit by matching the block type to it and flip the block if not matched, expressed symbolically as

$$Y = \begin{cases} X & \text{if } X = m \,, \\ \overline{X} & \text{if } X \neq m \,. \end{cases} \tag{13}$$

The overall payload is the compressed RS map with the message bits appended after it. Hence, the number of embeddable bits, or capacity, is calculated as

$$\text{Capacity} = N_{\mathcal{R}} + N_{\mathcal{S}} - |\mathrm{Compress}(\text{RS map})| \,. \tag{14}$$

The message extraction is done by simply reading through the β^{th} bit-plane and separating the compressed RS map and the message bits. The image recovery is achieved by flipping back each block in accordance with the uncompressed RS map. Assume that the message is a random bit-stream (e.g. the message is compressed and encrypted). Since it is anticipated that only a half of the regular and singular blocks will be flipped, the overall amount of squared-error distortion can be estimated to be around

$$\text{Distortion} = \frac{\alpha^2 n^2 (N_{\mathcal{R}} + N_{\mathcal{S}})}{2} \,. \tag{15}$$

The prime aim of invertible steganography is to pursue a maximal capacity while keeping the amount of distortion as low as possible. As we can see, a way to achieve high capacity is to minimise $N_{\mathcal{U}}$ so as to increase $N_{\mathcal{R}}$ and $N_{\mathcal{S}}$. In addition, the capacity depends largely on the size of the compressed RS map, which is connected to the bias between $p_{\mathcal{R}}$ and $p_{\mathcal{S}}$. As a consequence, the overarching key to increase the capacity relies on how much the bias can be exacerbated and that is governed by the design of the discrimination function. The problem is therefore narrowed to formulate a good discrimination function that is capable of distinguishing between the distributions of natural and noisy images. Alternatively, if we can predict or generate a *reference image* close enough to the original cover image, then we will be able to tell the real from the fake. And that is where neural networks come in useful.

III. NEURALISATION

For the purpose of aggravating the bias between $p_{\mathcal{R}}$ and $p_{\mathcal{S}}$, we explore adversarial learning and instantiate the idea of discrimination function as to calculate the absolute deviations of X and $\overline{X}$ from a reference block denoted by $\widetilde{X}$, that is,

$$g(X) = \left| X - \widetilde{X} \right| \,, \tag{16}$$

and

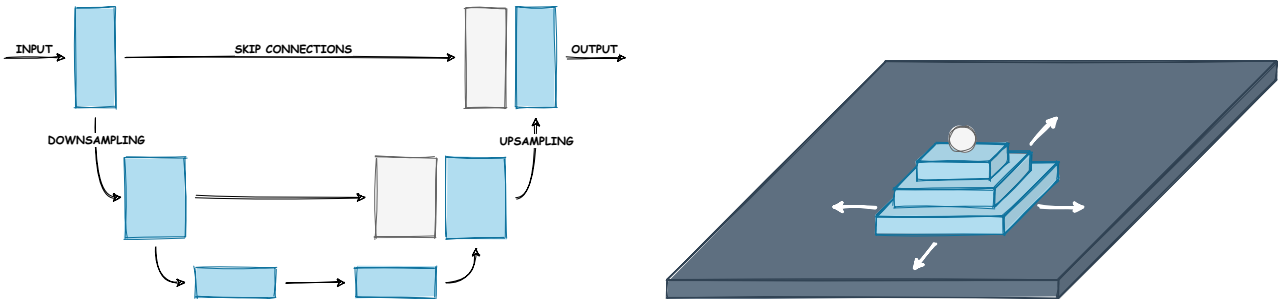
$$g(\overline{X}) = \left| \overline{X} - \widetilde{X} \right| \,. \tag{17}$$

Then we can categorise each block into the regular, singular, or unusable case and obtain the RS map. Note that the reference blocks must be estimated from some distortion-free parts of the image before and after data embedding (i.e. at the sender and receiver ends), so as to ensure the uniqueness of the RS map. Obviously, if we consider the aforementioned bit flipping as our realisation of invertible noise adding, the only part changed is the β^{th} bit-plane, whereas the rest remains untouched. This leads naturally to the idea of exploiting the unmodified bit-planes to predict or synthesise the changed one. Empirically, it is feasible, albeit difficult, to fabricate a bit-plane to a fair degree of accuracy by using a GAN. The task becomes more challenging as we dive into a deeper bit-plane. Hence, we train separate GANs to predict bit-planes of different depths.

A GAN is a class of neural networks invented by Goodfellow *et al.* [47] and it learns to generate synthetic images through a two-player minmax game between a generator and a discrimination. To sum up, the proposed ALIS utilises conditional GANs to synthesise bit-planes. The input is seven pristine distortion-free bit-planes and the target is the β^{th} bit-plane. Apropos of network architectures, we would like to give the credit to the *pix2pix* model [48], a seminal study on various image-to-image translation tasks, by which our GAN is primarily inspired. We configure the generator G as a U-Net and the discriminator D as a Markovian discriminator. These two adversarial neural networks form the backbone of ALIS, as illustrated diagrammatically in Figure 1. The U-Net is a neural network architecture that enables pixel-wise output [49]. It is characterised by its U-shape, which explains the origins of the name, and is comprised of a pair of encoder and decoder with skip connections between mirrored layers. The samples flow from top to bottom through a series of convolutions and downscaling, and then go back through a succession of convolutions and upscaling to the full resolution. The skip connections allow multi-resolution feature maps from previous layers to be retained and concatenated with feature maps at later layers. The Markovian discriminator is a patch-level CNN that learns to classify whether each overlapping patch of an image is synthetic or real. It assumes independence between pixels in different patches and imposes restriction upon model's attention to local image structures, akin to a Markov random field.

A straightforward way to measure the distance between the generated output and the target is to calculate the Manhattan distance (ℓ_1 norm) or the Euclidean distance (ℓ_2 norm). We

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3034936, IEEE Access



(a) U-Net generator (b) Markovian discriminator

FIGURE 1. An illustration of the generator and discriminator of GAN.

opt for the ℓ_1 norm empirically and also on account of a study suggesting that Manhattan distance may be preferable to Euclidean distance for the case of high dimensional data [50]. While ℓ_1 norm is capable of capture *low-frequency* structure, if the model solely rests on the ℓ_1 norm, the outcome would tend to be blurry since this loss function is minimised by averaging all plausible outputs, thus incentivising a blur when uncertainty encountered in complex areas. In view of this problem, we require another measurement of to what degree the model has learnt to represent *high-frequency* structure. Hence, an attention-restricted discriminator (i.e. Markovian discriminator) is introduced to guide the model to learn the minute structure in local image patches. Let x and y denote a pair of input and target. The ℓ_1 norm is written as

$$\mathcal{L}_{\ell_1}(G) = \mathbb{E}[\|y - G(x)\|_1]\,. \tag{18}$$

The adversarial loss is derived from the cross-entropy between the real and synthetic distributions, expressed as

$$\begin{aligned}\mathcal{L}_{\text{GAN}}(G, D) =& \mathbb{E}[\log D(y)] + \\ & \mathbb{E}[\log 1 - D(G(x))]\,.\end{aligned} \tag{19}$$

Therefore, the overall objective is

$$\min_G \max_D \mathcal{L}_{\text{GAN}}(G, D) + \lambda \mathcal{L}_{\ell_1}(G)\,, \tag{20}$$

where λ is a parameter for balancing between two loss terms and is set empirically.

IV. EXPERIMENTS

We validate and evaluate the proposed method experimentally on the USC-SIPI [51] and BOSSbase [52] image datasets and draw a comparison against a variance-based implementation of the RS method. Architectural details and further results are provided in Appendices.

A. EXPERIMENTAL SETUP

1) Datasets

We made use of the BOSSbase for large-scale training and testing and USC-SIPI for the purpose of inference. All the images were converted to 8-bit grayscale and resampled to 256×256 pixels using the Lanczos algorithm [53].

USC-SIPI

The USC-SIPI image database is a collection of scanned and digitalised pictures from a variety of sources and has been used widely to support research in image processing and computer vision. The database is categorised into volumes according to the basic character of the pictures. Images are of either 8-bit greyscale or 24-bit colour format with 256×256, 512×512, or 1024×1024 pixels. Images used in our experiments are from the *Miscellaneous* volume.

BOSSbase

The BOSSbase is considered the most frequently employed image database within the steganography community. The database contains $10,000$ grayscale images captured with several different cameras across a broad spectrum of scenes. We split image samples into training and test sets at the ratio of $8:2$. All the experimental results were obtained from the test set of $2,000$ unseen samples.

2) Implementation

Each model was trained over 100 epochs with the initial learning rate set to 2×10^{-4} and the batch size set to 32. Learning rate decay policy was enforced halfway through the training. The model parameters were handled and updated by the Adam optimiser [54]. The generator was a U-Net with skip connections between downsampling and upsampling layers, whereas the discriminator was a CNN. The applied nonlinear activation functions were ReLU [55] and LeakyReLU [56]. The weight for balancing adversarial loss and ℓ_1 norm was set to $\lambda = 10^3$ for every model. As common strategies against overfitting and to ensure training stability, we performed Xavier initialisation [57], batch normalisation [58], dropout [59], and data augmentation [60]. Blocks for carrying message bits was fixed to the size of 2×2 pixels. Bit flipping was practised on the bottom half of the bit-planes $\beta = 1, 2, 3, 4$, resulting in the distortion amplitudes $\alpha = 1, 2, 4, 8$, respectively. It is worth noting that post-processing was applied to binarise the synthetic bitmaps (i.e. qualify the values to zero and one) since the outputs of GANs are not necessarily binary.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3034936, IEEE Access

3) Metrics

Let X and Y denote the cover and stego images of $H \times W$ pixels. We measure the capacity, distortion, and compressed size as follows.

Capacity

Steganographic capacity rate is represented by the number of message bits embedded in the stego image per pixel (bpp), as given by

$$\text{SCR} = \frac{\text{number of message bits}}{H \times W}. \tag{21}$$

Distortion

We evaluate the distortion and visual quality of stego images by the peak signal-to-noise ratio (in dB), defined as

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{255^2}{\text{MSE}}\right), \tag{22}$$

where MSE denotes mean squared error:

$$\text{MSE} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}[X(i,j) - Y(i,j)]^2. \tag{23}$$

Accuracy

We assess the accuracy of synthetic bitmap by bit error rate:

$$\text{BER} = \frac{\text{number of error bits}}{H \times W}. \tag{24}$$

Compression

Despite the fact that a more sophisticated compression algorithm may be employed, we compress the RS map by context-free arithmetic coding [61] and represent space saving by data compression ratio (%):

$$\text{DCR} = \left(1 - \frac{\text{compressed data size}}{\text{uncompressed data size}}\right) \times 100 \tag{25}$$

4) Baselines

While the research of invertible steganography has undergone rapid development and there are voluminous literatures in this field, the RS method remains largely unexplored. At the time of writing, we are not aware of a related variation of the original implementation. Hence, we use the original implementation of the RS method as the baseline with a slight modification to the discrimination function.

Local Variance (LocVar)

The original RS method instantiated the discrimination function as a forward finite difference:

$$\Delta(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i|. \tag{26}$$

It is a naïve way to detect noise based on the smoothness prior. We modify it slightly by estimating the local variance:

$$\text{Var}(X) = \mathbb{E}\left[(X - \mu)^2\right], \tag{27}$$

where X is a block of pixels and μ is the mean of the block.

B. EXPERIMENTAL RESULTS

We begin by the experiments on the test set of the BOSSbase. We first examine the accuracy of synthetic bitmap. It is followed by a study on how the accuracy affects separate factors related to the capacity. We further investigate the relationship between accuracy, capacity and distortion. A comparison against a variance-based implementation of the RS method is presented. Our experiments end with an evaluation on selected images from the USC-SIPI dataset.

Validation: Accuracy

In order to validate whether ALIS had learnt to generate accurate bitmaps, we measured the BER of synthetic bitmaps. It can be observed from Figure 2 that the error rate diminishes as we move towards a higher bit-plane, verifying the hypothesis that the task of bitmap synthesis becomes less difficult for the bit-plane of a higher order. It is also worth noting that even for the least significant bit-plane, the average performance is much better than random guessing.

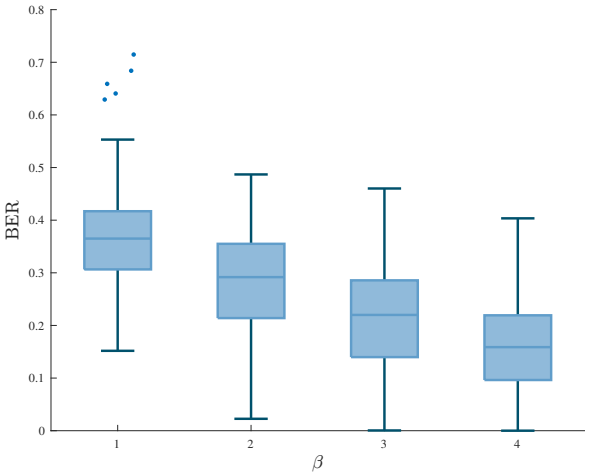


FIGURE 2. BER of synthetic bit-planes of different orders.

Regression Analysis: Accuracy and Capacity

The ability to predict bitmaps has been confirmed. We are, nonetheless, more concerned with whether a well-functioned bitmap predictor can help enhancing the embedding capacity. Recall that the capacity is in connection with the number of unusable blocks and the size of compressed RS map. Thus, we examined the relationship between BER, DCR, and $N_{\mathcal{U}}$ through regression analysis, as shown in Figure 3. There is a clear trend between BER and DCR, confirming that an accurate predictor indeed escalates the bias between $N_{\mathcal{R}}$ and $N_{\mathcal{S}}$. The correlation between BER and $N_{\mathcal{U}}$ is also evident, implying that an effective predictor should display less ambivalent judgement about the regularity of images.

Multivariate Analysis: Accuracy, Capacity, and Distortion

The correlations between accuracy, capacity, and distortion were examined, as shown in Figure 4. It is evident that a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3034936, IEEE Access

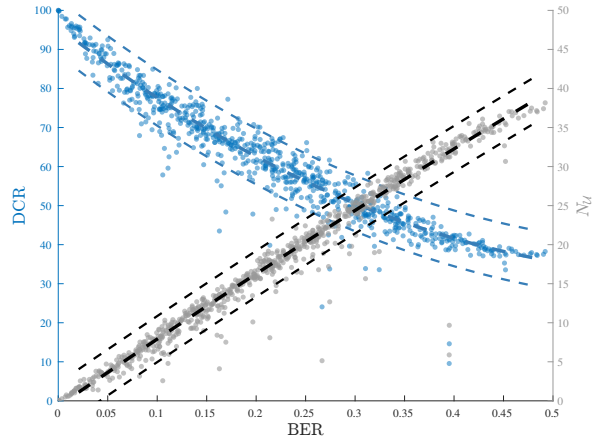


FIGURE 3. Regression Lines that visualise the relationship between BER and DCR (blue line) and the relationship between BER and $N_{\mathcal{U}}$ (black line).

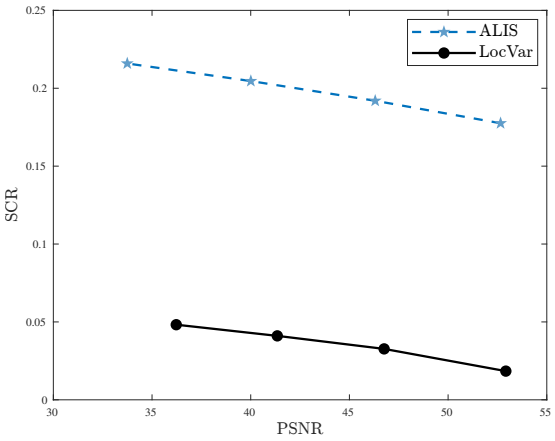


FIGURE 5. Capacity-distortion curves of ALIS and LocVar.

low error rate serves to enhance the capacity. A neat linear association between capacity and distortion can be observed on account of the fact that both measurements are governed by $N_{\mathcal{R}}$ and $N_{\mathcal{S}}$. It can also be inferred that in general hiding information in the bit-plane of a high order results in high accuracy (low BER), high capacity (high SCR), and high distortion (low PSNR).

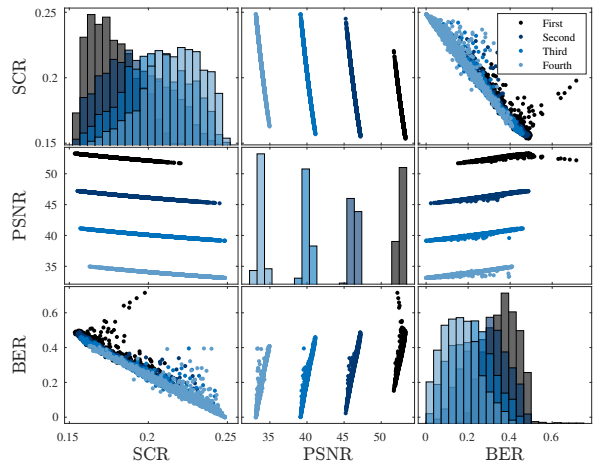


FIGURE 4. Relationships between SCR, PSNR and BER with respect to different target bit-planes.

Comparison: ALIS versus LocVar

In order to evaluate to what extent ALIS can upgrade the RS method, a comparison between ALIS and a variance-based implementation with respect to the capacity-distortion curve is presented in Figure 5. The simulation results demonstrated a significant improvement, validating the potential of ALIS to reinvigorate an obsolete method.

Inference

We close our experiments with a performance evaluation on commonly used test images, as reported in Table 1. The experimental results conformed closely with our findings from the results on the large-scale dataset.

V. DISCUSSION

The experimental results suggest that adversarial learning is able to train a bitmap predictor with the accuracy that is adequate to realise invertible steganography. We have achieved a significant performance boost over the former implementation of the RS method. Nevertheless, we endow ALIS with the potential that has not yet been fully released and unlocked. In view of this, we offer some insights regarding how ALIS may be further improved as follows.

The core of ALIS is the GANs for bitmap synthesis. The present generative model is a *vanilla* GAN, which leaves scope for adaptation (e.g. least squares GAN [62] and Wasserstein GAN [63]). Fine-tuned configurations or *ad hoc* loss functions may further refine the model and result in various outcomes. As a matter of fact, the task of bitmap synthesis may be modelled as that of super-resolution imaging [64]–[66], noise reduction [67], image restoration [68], or bit-depth expansion [69]. These topics have been a major research subject in image processing and their advances may confer great benefit and influence the overall performance of ALIS positively. While we optimise the network as an independent predictive module, we figure that it might limit the top-end effectiveness of ALIS. A joint *end-to-end* training of a GAN with the steganographic method is likely to achieve an even better performance by allowing more information to be integrated and the gradients to flow and backpropagate through all the units.

We would also like to point out that our deployment of the RS method is not optimal. The framework of the RS method is much more flexible and its potential seems to remain largely untapped. In particular, it only prescribes that the noise adding should be invertible, but does not specify that it must be a bit flipping operation. In addition, it should be possible to adjust the block size and flip bits in an adaptive way with a dynamic embedding amplitude, rather than a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3034936, IEEE Access

TABLE 1. Evaluation of accuracy, capacity and distortion on standard test images

β	1^{st}			2^{nd}			3^{rd}			4^{th}		
Images	BER	SCR	PSNR	BER	SCR	PSNR	BER	SCR	PSNR	BER	SCR	PSNR
Airplane	0.3976	0.1681	52.88	0.3276	0.1822	46.51	0.2341	0.2011	40.06	0.1536	0.2171	33.70
Lena	0.4098	0.1635	53.00	0.3244	0.1810	46.54	0.2354	0.2006	40.07	0.1452	0.2198	33.65
Mandrill	0.4796	0.1559	53.21	0.4559	0.1577	47.14	0.4161	0.1648	40.93	0.3467	0.1777	34.58
Peppers	0.4148	0.1632	53.01	0.3337	0.1794	46.58	0.2335	0.2020	40.04	0.1338	0.2229	33.59

TABLE 2. Generator architecture.

Encoder Layer	I/O Channels	Process	Decoder Layer	I/O Channels	Process
1	(7, 64)	Conv-LeakyReLU	16	(128, 1)	ConvTranspose-Tanh
2	(64, 128)	Conv-BatchNorm-LeakyReLU	15	(256, 64)	ConvTranspose-BatchNorm-ReLU
3	(128, 256)	Conv-BatchNorm-LeakyReLU	14	(512, 128)	ConvTranspose-BatchNorm-ReLU
4	(256, 512)	Conv-BatchNorm-LeakyReLU	13	(1024, 256)	ConvTranspose-BatchNorm-ReLU
5	(512, 512)	Conv-BatchNorm-LeakyReLU	12	(1024, 512)	ConvTranspose-BatchNorm-ReLU
6	(512, 512)	Conv-BatchNorm-LeakyReLU	11	(1024, 512)	ConvTranspose-BatchNorm-ReLU
7	(512, 512)	Conv-BatchNorm-LeakyReLU	10	(1024, 512)	ConvTranspose-BatchNorm-ReLU
8	(512, 512)	Conv-ReLU	9	(512, 512)	ConvTranspose-BatchNorm-ReLU

TABLE 3. Discriminator architecture.

Layer	I/O Channels	Process
1	(8, 64)	Conv-LeakyReLU
2	(64, 128)	Conv-BatchNorm-LeakyReLU
3	(128, 256)	Conv-BatchNorm-LeakyReLU
4	(256, 512)	Conv-BatchNorm-LeakyReLU
5	(512, 1)	Conv-Sigmoid-BCELoss

static amplitude, in order to balance between the capacity and the distortion in a more delicate manner. Furthermore, the use of a reference image or a prediction mechanism is in fact not new in invertible steganography [70]–[74]. The ways in which the concept of ALIS can be translated beyond the RS method deserve further investigation.

The worldwide popularisation of cloud computing, accompanied by a growing public awareness of data privacy, has given rise to the research of privacy-preserving invertible steganography [75]–[80]. This task is challenging due to the fact that private data is encrypted and consequently steganography has to be carried out in the encrypted domain. From our perspective, the application of ALIS towards this task is promising. Specifically, Chang *et al.* proposed a privacy-preserving invertible steganographic scheme, which can be viewed as an extension of the RS method in the encrypted domain [81]. The authors described bit flipping methods in the encrypted domain through privacy homomorphisms [82] and utilised Bayesian inference to predict the original state of the changed bits. It seems to be possible that a collaboration between this scheme and ALIS could produce positive research outcomes.

VI. CONCLUSION

In this paper, we present ALIS as a pioneer attempt to neuralise the RS method. Experimental results validated the effectiveness of the proposed method and showed a significant improvement over the prior implementation. Many interesting problems and possible refinements are left open for future work. We envision by further exploring the unfulfilled potential of ALIS, a state-of-the-art performance may be achieved. This paper is intended primarily to shed light on how deep learning can breathe new life into a classic but antiquated steganographic method and improve its performance drastically. We hope this article can serve as a point of departure for future research and herald a new dawn of invertible steganography with deep neural networks.

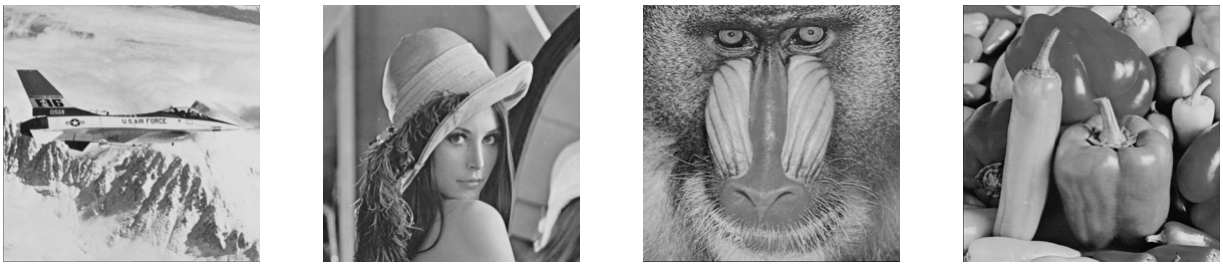
APPENDIX A NETWORK ARCHITECTURES

The architectural details of U-Net generator and Markovian discriminator are listed in Table 2 and Table 3 respectively. We abbreviate convolution as *Conv*, transposed convolution as *ConvTranspose*, batch normalisation as *BatchNorm*, binary cross entropy loss as *BCELoss*. All convolutions are applied with 4×4 kernels and stride 2.

APPENDIX B ADDITIONAL RESULTS

Selected samples from the USC-SIPI and BOSSbase datasets are shown in Figures 6 and 7. RS Maps produced by variance-based and learning-based methods are illustrated in Figure 8. A comparison between real bitmaps and synthetic bitmaps generated by ALIS is presented in Figure 9. Examples of reference and stego images are displayed in Figures 10 and 11. We use subscript to denote the order of target bit-plane. We are aware that it is barely possible to choose a representative sample of the whole dataset. Therefore, the experimental results on *Lena* are for demonstration purposes only.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3034936, IEEE Access



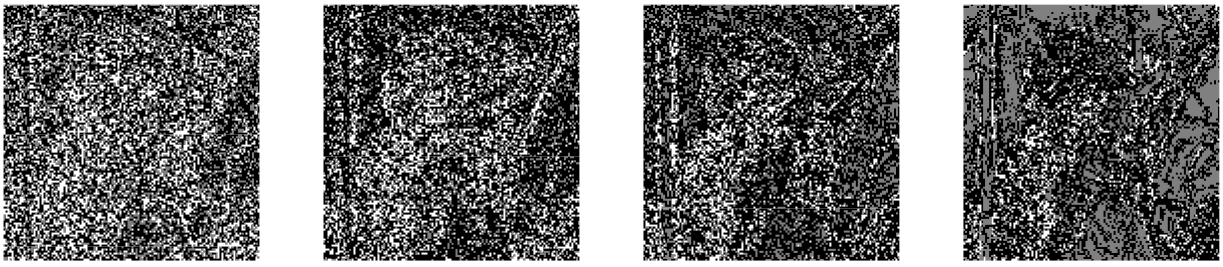
(a) Airplane (b) Lena (c) Mandrill (d) Peppers

FIGURE 6. Samples from the USC-SIPI image database.

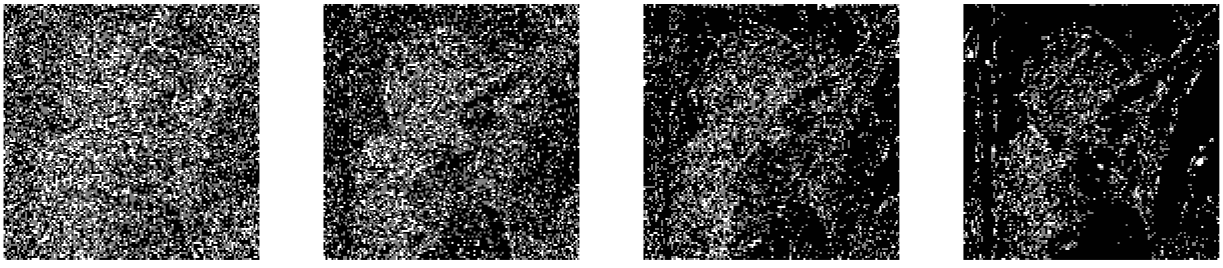
(a) ref. 1 (b) ref. 8 (c) ref. 29 (d) ref. 47

(e) ref. 282 (f) ref. 703 (g) ref. 792 (h) ref. 1678

FIGURE 7. Samples from the BOSSbase.



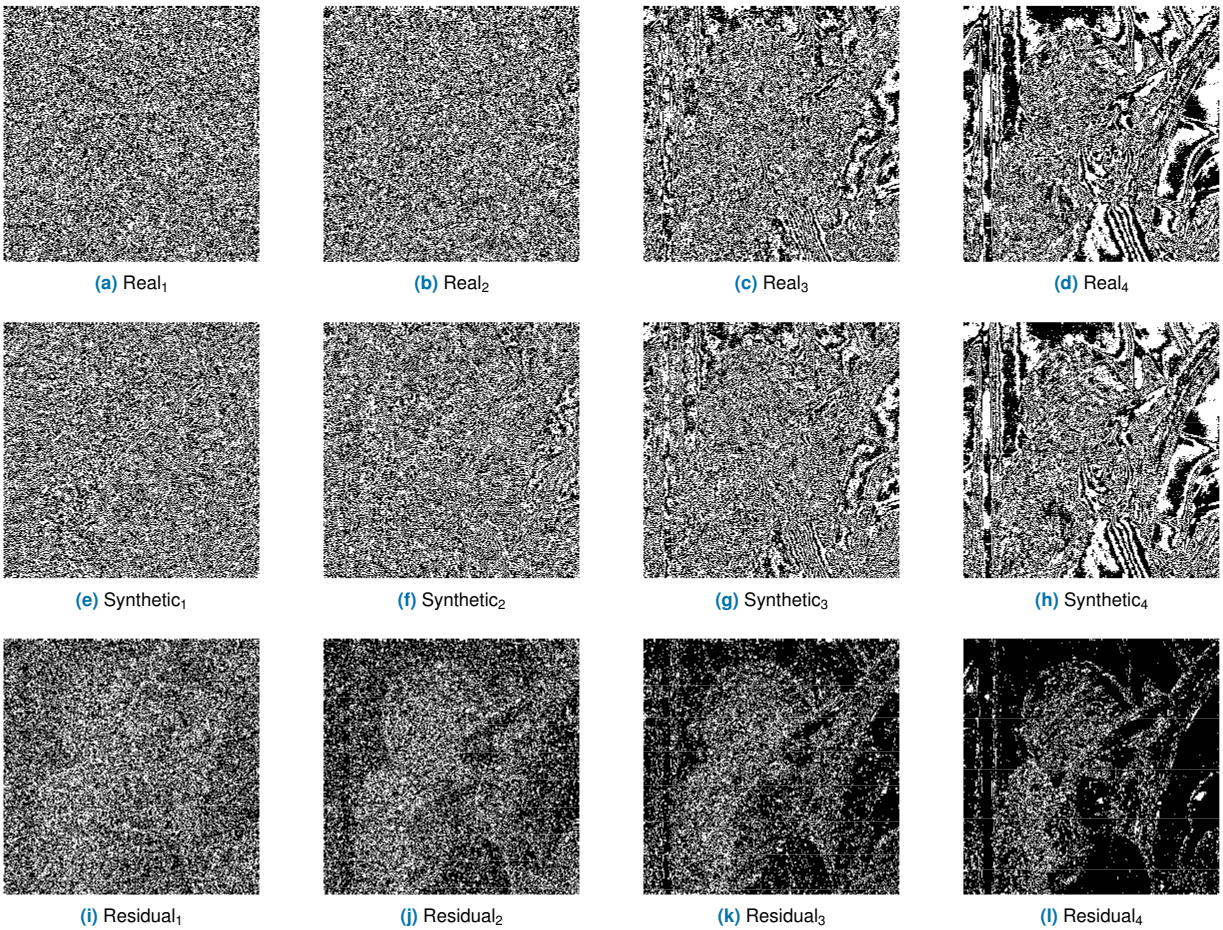
(a) LocVar_1 (b) LocVar_2 (c) LocVar_3 (d) LocVar_4



(e) ALIS_1 (f) ALIS_2 (g) ALIS_3 (h) ALIS_4

FIGURE 8. RS Maps generated by LocVar and ALIS. Regular blocks are coloured in black, singular blocks in white and unusable blocks in grey.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3034936, IEEE Access



(a) Real_1 (b) Real_2 (c) Real_3 (d) Real_4

(e) Synthetic_1 (f) Synthetic_2 (g) Synthetic_3 (h) Synthetic_4

(i) Residual_1 (j) Residual_2 (k) Residual_3 (l) Residual_4

FIGURE 9. Real bitmaps, synthetic bitmaps and their residuals.

(a) Reference_1 (b) Reference_2 (c) Reference_3 (d) Reference_4

FIGURE 10. Reference images.

(a) Stego_1 (b) Stego_2 (c) Stego_3 (d) Stego_4

FIGURE 11. Stego images.

REFERENCES

[1] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, 2009.

[2] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Proceedings of International Workshop on Information Hiding (IH)*, Calgary, AB, Canada, 2010, pp. 161–177.

[3] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *Proceedings of IEEE International Workshop on Information Forensics and Security (WIFS)*, Tenerife, Spain, 2012, pp. 234–239.

[4] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Paris, France, 2014, pp. 4206–4210.

[5] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1–13, 2014.

[6] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, 2016.

[7] 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.

[8] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1064–1087, 1998.

[9] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1127–1141, 1999.

[10] J. Fridrich, "Image watermarking for tamper detection," in *Proceedings of International Conference on Image Processing (ICIP)*, Chicago, IL, USA, 1998, pp. 404–408.

[11] 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.

[12] P. W. Wong and N. Memon, "Secret and public key image watermarking schemes for image authentication and ownership verification," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1593–1601, 2001.

[13] G. Depovere, T. Kalker, J. Haitsma, M. Maes, L. de Strycker, P. Termont, J. Vandewege, A. Langell, C. Alm, P. Norman, G. O'Reilly, B. Howes, H. Vaanholt, R. Hintzen, P. Donnelly, and A. Hudson, "The VIVA project: Digital watermarking for broadcast monitoring," in *Proceedings of International Conference on Image Processing (ICIP)*, Kobe, Japan, 1999, pp. 202–205.

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

[15] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," in *Proceedings of International Conference on Learning Representations (ICLR)*, Banff, Canada, 2014, pp. 1–10.

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

[17] 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)*, Las Vegas, NV, USA, 2016, pp. 2574–2582.

[18] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," in *Proceedings of International Conference on Learning Representations (ICLR)*, Toulon, France, 2017, pp. 1–14.

[19] 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)*, Honolulu, HI, USA, 2017, pp. 86–94.

[20] C. W. Honsinger, P. W. Jones, M. Rabbani, and J. C. Stoffel, "Lossless recovery of an original image containing embedded data," U.S. Patent 6 278 791, 21, 2001.

[21] J. Fridrich, M. Goljan, and R. Du, "Invertible authentication," in *Proceedings of SPIE Security and Watermarking of Multimedia Contents III*, vol. 4314, San Jose, CA, United States, 2001, pp. 197–208.

[22] M. Goljan, J. Fridrich, and R. Du, "Distortion-free data embedding for images," in *Proceedings of International Workshop on Information Hiding (IH)*, Pittsburgh, PA, USA, 2001, pp. 27–41.

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

[24] T. Kalker and F. M. J. Willems, "Capacity bounds and constructions for reversible data-hiding," in *Proceedings of International Conference on Digital Signal Processing (DSP)*, Santorini, Greece, 2002, pp. 71–76.

[25] W. Zhang, B. Chen, and N. Yu, "Improving various reversible data hiding schemes via optimal codes for binary covers," *IEEE Transactions on Image Processing*, vol. 21, no. 6, pp. 2991–3003, 2012.

[26] D. Hou, W. Zhang, Y. Yang, and N. Yu, "Reversible data hiding under inconsistent distortion metrics," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5087–5099, 2018.

[27] 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.

[28] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 1147–1156, 2004.

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

[30] 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.

[31] G. Coatrieux, W. Pan, N. Cuppens-Boulahia, F. Cuppens, and C. Roux, "Reversible watermarking based on invariant image classification and dynamic histogram shifting," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 111–120, 2013.

[32] X. Li, B. Li, B. Yang, and T. Zeng, "General framework to histogram-shifting-based reversible data hiding," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2181–2191, 2013.

[33] J. Wang, X. Chen, J. Ni, N. Mao, and Y.-Q. Shi, "Multiple histograms-based reversible data hiding: Framework and realization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 8, pp. 2313–2328, 2020.

[34] 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.

[35] S. Lee, C. D. Yoo, and T. Kalker, "Reversible image watermarking based on integer-to-integer wavelet transform," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 321–330, 2007.

[36] B. Ma and Y.-Q. Shi, "A reversible data hiding scheme based on code division multiplexing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1914–1927, 2016.

[37] J. Liu, Y. Ke, Z. Zhang, Y. Lei, J. Li, M. Zhang, and X. Yang, "Recent advances of image steganography with generative adversarial networks," *IEEE Access*, vol. 8, pp. 60 575–60 597, 2020.

[38] W. Tang, S. Tan, B. Li, and J. Huang, "Automatic steganographic distortion learning using a generative adversarial network," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1547–1551, 2017.

[39] J. Hayes and G. Danezis, "Generating steganographic images via adversarial training," in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, Long Beach, CA, USA 2017, pp. 1954–1963.

[40] Y. Zhang, W. Zhang, K. Chen, J. Liu, Y. Liu, and N. Yu, "Adversarial examples against deep neural network based steganalysis," in *Proceedings of ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec)*, Innsbruck, Austria, 2018, pp. 67–72.

[41] W. Tang, B. Li, S. Tan, M. Barni, and J. Huang, "CNN-based adversarial embedding for image steganography," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 8, pp. 2074–2087, 2019.

[42] L. Zhou, G. Feng, L. Shen, and X. Zhang, "On security enhancement of steganography via generative adversarial image," *IEEE Signal Processing Letters*, vol. 27, pp. 166–170, Dec. 2019.

[43] J. Yang, D. Ruan, J. Huang, X. Kang, and Y.-Q. Shi, "An embedding cost learning framework using GAN," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 839–851, 2020.

[44] D. Hu, L. Wang, W. Jiang, S. Zheng, and B. Li, "A novel image steganography method via deep convolutional generative adversarial networks," *IEEE Access*, vol. 6, pp. 38 303–38 314, 2018.

[45] Y. Ke, J. Liu, M. Zhang, T. Su, and X. Yang, "Steganography security: Principle and practice," *IEEE Access*, vol. 6, pp. 73 009–73 022, 2018.

[46] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, July 1948.

[47] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, Montreal, QC, Canada, 2014, pp. 2672–2680.

[48] 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)*, Honolulu, HI, USA, 2017, pp. 5967–5976.

[49] 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)*, Munich, Germany, 2015, pp. 234–241.

[50] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *Proceedings of International Conference on Database Theory (ICDT)*, London, UK, 2001, pp. 420–434.

[51] A. G. Weber, "The USC-SIPI image database: Version 5," Signal and Image Processing Institute, USC Viterbi School of Engineering, Tech. Rep., 2006.

[52] 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)*, Prague, Czech Republic, 2011, pp. 59–70.

[53] C. E. Duchon, "Lanczos filtering in one and two dimensions," *Journal of Applied Meteorology*, vol. 18, no. 8, pp. 1016–1022, 1979.

[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, 2015.

[55] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, FL, USA, 2011, pp. 315–323.

[56] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Processings of International Conference on Machine Learning (ICML)*, Atlanta, GA, USA, 2013, pp. 1–6.

[57] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, Sardinia, Italy, 2010, pp. 249–256.

[58] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of International Conference on Machine Learning (ICML)*, Lille, France, 2015, pp. 448–456.

[59] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[60] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 60, pp. 1–48, 2019.

[61] G. G. Langdon, "An introduction to arithmetic coding," *IBM Journal of Research and Development*, vol. 28, no. 2, pp. 135–149, 1984.

[62] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2813–2821.

[63] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of International Conference on Machine Learning (ICML)*, vol. 70, Sydney, Australia, 2017, pp. 214–223.

[64] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[65] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 105–114.

[66] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 5835–5843.

[67] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[68] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 9446–9454.

[69] Y. Zhao, R. Wang, W. Jia, W. Zuo, X. Liu, and W. Gao, "Deep reconstruction of least significant bits for bit-depth expansion," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2847–2859, 2019.

[70] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless generalized-LSB data embedding," *IEEE Transactions on Image Processing*, vol. 14, no. 2, pp. 253–266, 2005.

[71] 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.

[72] V. Sachnev, H. J. Kim, J. Nam, S. Suresh, and Y.-Q. 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.

[73] X. Li, B. Yang, and T. Zeng, "Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3524–3533, 2011.

[74] I. Dragoi and D. Coltuc, "Local-prediction-based difference expansion reversible watermarking," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1779–1790, 2014.

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

[76] 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.

[77] X. Zhang, J. Long, Z. Wang, and H. Cheng, "Lossless and reversible data hiding in encrypted images with public-key cryptography," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1622–1631, 2016.

[78] 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.

[79] 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.

[80] 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. 54 117–54 132, 2019.

[81] C.-C. Chang, C.-T. Li, and Y.-Q. Shi, "Privacy-aware reversible watermarking in cloud computing environments," *IEEE Access*, vol. 6, pp. 70 720–70 733, 2018.

[82] R. L. Rivest, L. Adleman, and M. L. Dertouzos, "On data banks and privacy homomorphisms," *Foundations of Secure Computation*, pp. 169–177, 1978.

CHING-CHUN CHANG received the PhD degree in Computer Science from the University of Warwick, UK in 2019. He received the BBA degree in Information Management from National Central University, Taiwan in 2015. He was granted the Marie-Curie fellowship in 2017. He engaged in a short-term scientific mission supported by European Cooperation in Science and Technology Actions at the Faculty of Computer Science, Otto von Guericke University Magdeburg, Germany in 2016. He participated in a research and innovation staff exchange scheme supported by Marie Skłodowska-Curie Actions at the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, USA during 2017. He was a visiting scholar at the School of Computer and Mathematics, Charles Sturt University, Australia in 2018 and at the School of Information Technology, Deakin University, Australia in 2019. He has been a research fellow at the Department of Electronic Engineering, Tsinghua University, China since 2020. His research interests include steganography, applied cryptography, digital forensics, multimedia security, image processing, computer vision, and machine learning.