Color Image High-Capacity Differential Steganography Algorithm Based on Multiple Adversarial Networks

Bin Ma, *Member, IEEE*, Haocheng Wang, Jian Xu, Xiaoyu Wang, Xiaolong Li, *Member, IEEE*, and Jian Li

***Abstract*— Aiming to mitigate image distortion caused by steganography algorithms at high-capacity information embedding and enhance the steganalysis resistance capability of generated stego images, this paper proposes a high-capacity differential steganography algorithm for color images based on multiple adversarial networks. Instead of directly modifying the pixels of the cover image, the algorithm embeds the secret information into the differential plane generated by the two most similar channels of the cover image. Consequently, the distortion of the stego image is minimized while embedding a secret image of the same size. At the same time, the fidelity of the stego and extracted secret images is continually improved through adversarial training between the generator and discriminator in the proposed steganography network. Furthermore, multiple steganalysis networks are parallelly utilized to enhance the steganalysis resistance capability of stego images. In addition, the Lion optimizer is utilized for the first time to improve the convergence speed of the proposed steganographic network. Experimental results show that the comprehensive performance of the proposed algorithm outperforms other state-of-the-art steganography algorithms significantly.**

***Index Terms*— Differential steganography, high-capacity, Lion optimizer, generative adversarial network.**

Received 9 August 2024; revised 16 November 2024; accepted 25 November 2024. Date of publication 29 November 2024; date of current version 7 April 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62272255, Grant 62302248, and Grant 62302249; in part by the National Key Research and Development Program of China under Grant 2021YFC3340600 and Grant 2021YFC3340602; in part by Taishan Scholar Program of Shandong under Grant tsqn202306251; in part by Shandong Provincial Natural Science Foundation under Grant ZR2020MF054, Grant ZR2023QF018, Grant ZR2023QF032, and Grant ZR2022LZH011; in part by the Ability Improvement Project of Science and Technology SMES in Shandong Province under Grant 2022TSGC2485 and Grant 2023TSGC0217; in part by Jinan "20 Universities"-Project of Jinan Research Leader Studio under Grant 2020GXRC056; in part by Jinan "New 20 Universities"-Project of Introducing Innovation Team under Grant 202228016; in part by the Youth Innovation Team of Colleges and Universities in Shandong Province under Grant 2022KJ124; in part by the "Chunhui Plan" Cooperative Scientific Research Project of Ministry of Education under Grant HZKY20220482; in part by the Achievement Transformation of Science, Education and Production Integration Pilot Project, under Grant 2023CGZH-05; in part by the First Talent Research Project under Grant 2023RCKY131 and Grant 2023RCKY143; and in part by the Integration Pilot Project of Science Education Industry under Grant 2023PX006, Grant 2023PY060, and Grant 2023PX071. This article was recommended by Associate Editor T. Zhang. *(Corresponding author: Jian Li.)*

Bin Ma, Haocheng Wang, and Jian Li are with the Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan 250000, China, and also with Shandong Provincial Key Laboratory of Computer Networks, Shandong Fundamental Research Center for Computer Science, Jinan 250000, China (e-mail: sddxmb@126.com; whckyzy@163.com; ljian_ 20@163.com).

Jian Xu is with the School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China (e-mail: sdfixj@126.com).

Xiaoyu Wang is with the School of Information Science & Technology, Dalian Maritime University, Dalian 116026, China (e-mail: qluwxy@163.com).

Xiaolong Li is with the School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China (e-mail: lixl@bjtu.edu.cn).

Digital Object Identifier 10.1109/TCSVT.2024.3508849

I. Introduction

STEGANOGRAPHY aims to embed confidential information into certain less significant attributes of a cover medium imperceptibly, making it an effective means for secret data transmission. The primary object of image steganography is to embed secret data into an image while minimizing distortion effectively. Generally, image steganography algorithms can be roughly categorized into heuristic-based and deep learning-based approaches.

Among heuristic-based steganography algorithms, the STC (syndrome trellis code) algorithm stands out as one of the most efficient techniques for steganographic encoding [1]. It achieves coding performance that nearly approaches the rate-distortion bound while minimizing distortion efficiently at a given payload [2], [3], [4], [5]. Thereafter, developing a reasonable steganalysis distortion function to reduce the disturbance of the original carrier image content has become mainstream in the field of steganography. Various steganography schemes such as HUGO [6], WOW [7], S-UNIWARD [8], Hill [9], and MiPOD [10] have been proposed from different aspects to optimize the performance of the heuristic-based steganography. These schemes involve adaptively modifying pixel values or frequency coefficients of the cover image according to pre-designed constraint functions for secret data embedding, dynamically adjusting the data embedding cost to reduce the probability of detection by steganalyzers. By decomposing the steganography into secret data encoding and embedding distortion minimization, these approaches achieve high indistinguishability in appearance but are limited in capacity.

On the other hand, with the booming of deep learning technology in recent years, Convolutional Neural Networks (CNNs) have been extensively utilized in the field of steganography to enhance their resistance capability against steganalysis. For instance, Hayes and Danezis [11] devised

an adaptive steganography algorithm through adversarial training, where two networks compete to generate more secure stego images. Baluja [12] constructed multiple deep neural networks for image preprocessing and information hiding, effectively enhancing the quality of stego images. Subsequently, steganographic models based on Generative Adversarial Networks (GAN) have gained more and more attention [13]. Tang et al. [14] developed an adaptive data embedding algorithm to avoid embedding information into smooth regions of an image and introduced XuNet as a steganalyzer for adversarial training, thereby improving the security of stego images. Yang et al. [15] further proposed a simplified distortion calculation function to reduce image steganography distortion. By utilizing a double Tanh activation function, its adversarial training time was significantly minimized compared to Tang's ASDL-GAN-based steganography algorithm. Tan et al. [16] additionally proposed a novel end-to-end generative adversarial network architecture for image steganography, which obtained channel-wise features for deep representation of images by exploiting channel interdependencies. Consequently, a perceptually indistinguishable stego image can be achieved even at high data-embedding capacities. Recently, Liu et al. [17] applied the idea of residual blocks in ResNet to image steganography. In the scheme, image hiding and extraction were obtained by employing encoder and decoder networks. The data hiding capacity and extraction qualities are balanced by the proper setting of parameters during training.

The data-hiding capacity plays an increasingly pivotal role in a steganography scheme. Duan et al. [18] proposed an image steganography scheme based on Xception convolutional neural network and U-Net architecture, where the secret image was compressed and distributed across all available bits of the cover image to achieve large data embedding capacity. Subsequently, certain scholars utilized the Xception architecture to further enhance the steganography capacity and security [19], [20]. Fu et al. [21] concealed a color secret image within another color image of the same size by using residual blocks, they employed adversarial training between the steganography and steganalysis networks to obtain lower distortion and higher visual quality of steganographic images. Ma et al. [22] developed a multilevel adversarial network frame to generate cover images that are more suitable for data embedding. By the utilization of a joint loss function, a secret image can be imperceptibly hidden in the cover image, thereby enhancing its steganalysis resistance capability as well. By introducing a spatial-channel joint attention mechanism, Li et al. [23] devised a deep reversible neural network that can sequentially embed confidential images into a single cover image through a series of flexible cascaded iterative operations, addressing the problem of visual quality degradation of stego images due to high embedding capacity.

The security of stego images is another crucial indicator to be considered. Li et al. [24] construct multiple cross-feedback channels between the contraction path and the expansion path of the generator, which allows the generator can finally learn a sophisticated probability map, thus resulting in a high-security steganography scheme. Zhou et al. [25] proposed a new approach for quickly constructing adversarial images suitable for concealing secret information, ensuring similar discrimination results before and after image steganography. Thereby improving the security of steganographic algorithms. Huang et al. [26] employed multiple steganalyzers rather than a single steganalyzer to enhance the performance of steganographic algorithms. They also designed an adaptive method to update parameters of different steganalyzers during the network training stage, so as to gain high security of the proposed scheme. Recently, Xu et al. [27] introduced the conditional normalizing flow to modify the distribution of the redundant high-frequency component in stego images, improving their robustness while maintaining invisibility. Zheng et al. [28] developed a composition-aware image steganography to achieve both visual security and resistance to deep steganalysis via self-generated supervision. They also improved the naturalness of steganographic images by integrating the rule-based composition method and generative adversarial network to ensure high steganography capability. Cui et al. [29] transferred domain knowledge of steganalysis from the multiple pre-trained source metric networks to the meta metric network, thereby enhancing the generalization ability of the steganographizer, and further improving the security of stego images generated by reversible neural networks.

At the same time, with the boosting of social networks, more and more color images are being disseminated on the internet. However, few scholars have studied the correlations of color image channels to enhance the performance of image steganography [30]. Wang et al. [31] devised a steganographic scheme for spatial color images by exploiting both the similarities and differences among different color channels. The proposed strategy capitalizes on inconsistent modification directions in the R and B channels compared to the G channel, effectively distributing embedding capacity across all three channels and significantly improving resistance against color image steganalysis. Yao et al. [32] integrated RGB channels to generate preliminary disguised images, followed by reversible embedding of collected auxiliary information into these preliminary disguises to produce final disguised images, thereby enabling secret image recovery. Nguyen et al. [33] further introduced a secure channel selection rule for spatial image steganography. The value of the object pixel and the average difference between the object pixel and its neighbors are both considered to determine whether a pixel be modified for data hiding. The results indicate high visual quality and security of stego images under spatial image steganography. As concealing large amounts of information is prone to causing noticeable distortion in an image and making it easier to be detected by a steganalyzer, leveraging inter-channel correlations to enhance image steganography capabilities proves promising as an approach. Although certain steganography algorithms have attempted to utilize image channels for performance enhancement; however, the correlations among different channels have not yet been fully exploited [34] thus limiting the improvement in both embedding capability and imperceptibility of resulting stego images [35], [36].

On the contrary, image steganalysis is considered a technique utilized to identify a stego image whether containing secret messages. Its objective is to evaluate the presence of confidential information within an image by analyzing its

specific image features. Traditional steganalysis algorithms, such as SPAM and SRM (rich model) algorithms [37], identify stego images through analysis of certain high-frequency statistical characteristics to capture the embedding modifications caused by certain steganographic methods and distinguish them with binary classification. In recent years, deep learning technology has witnessed significant advancements. Many researchers have applied various deep-learning networks for steganalysis to improve the accuracy of discrimination. Deep learning-based steganalysis models such as XuNet [38], YeNet [39], SRNet [40], and ZhuNet [41] employed diverse neural networks to explore latent image features with the aim of distinguishing between the cover and stego images effectively. These deep learning-based steganalysis algorithms have demonstrated remarkable improvements compared to traditional methods, posing a substantial threat to the security of steganography.

Considering the fact that the relationship between image channels offers novel insights into information hiding, this paper proposes a new steganography scheme to enhance its capacity. In contrast to conventional steganography algorithms that directly modify pixel values of the original cover image for information-hiding purposes, this algorithm effectively leverages the differences among correlated image channels to achieve high-capacity data hiding. The secret image of identical dimensions is embedded into the differential plane of two closely correlated channels in the original cover image, ensuring that the resulting stego image achieves superior visual quality and robustness against steganalysis. To the best of our knowledge, this is the first work to combine differentials of color image channels and generative adversarial networks together to enhance the capability of an image steganography, while minimizing the image distortion by using the non-addition data hiding scheme on the differential of two close correlated channels. The primary contributions of this paper are outlined as follows:

- A high-capacity differential steganography algorithm for color images based on multiple adversarial networks is proposed for the first time in this paper. By embedding the secret image into the differential plane of a cover image, this approach avoids direct alterations to image pixels while effectively utilizing inter-channel correlation for data hiding, thereby significantly reducing stego image distortions caused by high-capacity data embedding.
- To improve the steganalysis resistance capability of the proposed scheme, an organic combination of certain advanced steganalysis networks is involved to reinforce the security of the resulting stego image. Due to the adversarial of various steganalysis networks, not only the steganalysis resistance capability of the yielding stego image but also the fidelity of the stego and extracted secret images have significantly improved.
- The Lion optimizer is firstly utilized in a steganography algorithm to improve the training effectiveness and efficiency of the generative adversarial networks, enabling the proposed scheme to achieve substantial enhancements in both stability and convergence capability, clearly surpassing those conventional steganographic schemes that use Adam optimizer.

The remaining sections of the paper are organized as follows: Section II provides a brief review of related literature, while Section III elaborates on the proposed high-capacity differential steganography algorithm for color images. Section IV presents qualitative and quantitative analyses of experimental results, followed by a comparison with certain advanced schemes in Section V. Finally, Section VI concludes the study and outlines future research plans.

II. Related Work

A. Generative Adversarial Network

Generative Adversarial Networks (GANs) is a deep learning model that was introduced by Goodfellow in 2014. It has emerged as one of the most promising unsupervised learning methods for modeling complex distributions in recent years. By engaging in a mutual game between two modules, namely the generative network and the discriminative network, this framework generates high-quality image outputs. The primary objective of the generator is to transform Gaussian noise into perceptually similar samples to the training data. Initially, the generator produces a batch of samples from Gaussian noise, which are then evaluated by the discriminator to measure their difference from real samples, known as discriminative error. By iterative minimization of this error during network training, the generator continuously improves its ability to generate high-quality samples. Meanwhile, the discriminator seeks to maximize differentiation between generated and real samples in order to improve its discriminative accuracy. This adversarial process iteratively continues until the generator produces an image of such high quality that it becomes indistinguishable from real ones by the discriminator. GANs have achieved significant success across various domains including image generation, segmentation, and style transfer due to their ability to learn latent features from actual data distribution through adversarial training and generate realistic-looking images. However, GANs face challenges such as mode collapse and instability during their training process; thus careful adjustment of hyper-parameters and network architecture is crucial for achieving satisfactory results. The objective function of GAN is as follows,

$$\min_G \max_D (D, G) = \sum_{x \sim pdata(x)} [\log D(x)] + \sum_{x \sim pz(z)} [\log(1 - D(G(z)))] \quad (1)$$

where $D(x)$ is the discriminative probability of real images, $G(z)$ denotes the generated images produced from input noise z, and $D(G(z))$ signifies the discriminative probability of the generated images.

B. U-Net Architecture

The U-Net architecture is a typical encoder-decoder network that was initially developed for image segmentation tasks. Due to its exceptional image reconstruction capability, it has found

Algorithm 1 Lion Optimizer

given β_1, β_2, λ, η,f
initialize θ_0,$m_0 \leftarrow 0$
while θ_t not converged **do**
  $g_t \leftarrow \nabla_\theta f(\theta_{t-1})$
  update model parameters
  $c_t \leftarrow \beta_1 m_{t-1} + (1-\beta_1) g_t$
  $\theta_t \leftarrow \theta_{t-1} - \eta_t(\text{sign}(c_t) + \lambda\theta_{t-1})$
  update EMA of g_t
  $m_t \leftarrow \beta_2 m_{t-1} + (1-\beta_2) g_t$
end while
return θ_t

extensive applications in various domains of image processing. Its encoding module employs convolutional and pooling layers to progressively downsample the input image and extract latent feature representations. In contrast, the decoding module incrementally restores the original image through upsampling, and enhancing recovery resolution via skip connections to integrate feature maps of corresponding encoder and decoder layers. Thanks to this well-designed framework, the U-Net structure effectively captures both local and global information while preserving high-resolution details, making it particularly suitable for tasks such as image semantic segmentation and reconstruction.

The outstanding performance of U-Net in image reconstruction also makes it an ideal means in the domain of image steganography. Within an image steganography scheme, the encoder part of U-Net combines both the secret and the cover information and transforms them into the feature matrix in latent space, which is then expanded to generate the final stego image through the decoder part.

C. Lion Optimizer

The Lion (Evolved Sign Momentum) optimizer is a gradient-based optimization algorithm developed by Google Brain to enhance the optimization efficiency of a deep learning network [42]. It has been extensively applied in various domains, such as image classification, image-text matching, diffusion models, language model pre-training, and fine-tuning. The Lion optimizer adaptively updates the learning rate according to the gradient of each parameter, thereby accelerating network convergence speed effectively and avoiding local optimal solutions. By accumulating a portion of historical gradients for gradient updates, the Lion optimizer improves parameter alteration stability and resolves network oscillation issues. Furthermore, through analyzing the gradient distribution of network parameters, the Lion optimizer dynamically adjusts gradients to achieve a balanced parameter distribution and ultimately enhances the generalization ability of the network. The gradient update process of the Lion optimizer is as follows:

In the algorithm, g_t represents the gradient of the loss function, β_1 and β_2 are decay rate factors used for updating the first and second-moment estimates, λ, η, f, denote the weight decay coefficient, learning rate, and loss function respectively, θ_t represents the model parameters at iteration t, m and c refer to the mean and exponentially weighted moving average of the gradient g_t. In most tasks, the Lion optimizer has exhibited superior convergence capability compared to certain mainstream optimizers such as Adam, Radam, and AdamW.

III. Proposed Method

A. Overall Structure

In this section, we will elaborately describe the construction of the proposed high-capacity differential steganography algorithm for color images based on multiple adversarial networks. Its objective is to minimize the image distortion caused by large amounts of data embedding and enhance the steganalysis resistance capability of the stego image. The proposed steganographic model primarily consists of three subnetworks: the Differential Steganography Network (DSN), the Steganalysis Adversarial Network (SAN), and the Data Extraction Network (DEN). As illustrated in Fig.1, the DSN comprises a differential image generation module and a generator module based on the U-Net structure. In DSN, the secret image is embedded into the differential plane rather than directly into a single channel of the cover image. Consequently, the spectral power of the secret image would be dispersed across different channels to reduce the disturbance to the cover image, thereby minimizing the distortion caused by data hiding. On the other side, the SAN performs steganalysis on the generated stego images and generates discriminative loss to enhance their steganalysis resistance capability. Ultimately, a stego image not only exhibits high visual fidelity but also possesses excellent steganalysis resilience capability can be yielded with the proposed scheme.

The proposed scheme focuses on the embedding of a color image into another three-channel cover image of the same size. If the data is embedded into one cover channel, it means that a pixel of the cover image needs to contain 24 bits of information, resulting in an extreme distortion in the stego image. The framework of the high-capacity differential steganography algorithm based on multiple adversarial networks for color images is illustrated in Fig.1. It can be observed that the proposed scheme includes two adversarial loops. Initially, the two most similar color channels are chosen to calculate pixel differences and generate the differential plane X_{dp}. Then, the secret image X_{sec} along with the differential plane X_{dp} are jointly fed into the DSN to achieve information embedding, resulting in an embedded differential plane denoted as X_{edp}. The perceptual fidelity of the stego and the cover image is optimized by the steganography adversarial loop. Simultaneously, the stego image is further input into the steganalysis adversarial loop, which involves certain advanced steganalysis algorithms serving as the discriminator, to enhance the steganalysis resistance capability of the stego image. Finally, the steganalysis error, along with the secret image reconstruction error and stego image coherence error are back-propagated into the differential steganography network to optimize the secret data hiding performance iteratively. It should be noted that the image optimization process may lead to pixel overflow, hence a clip operation is employed for double-end cropping of the image pixel value. Specifically, any value below or equal to

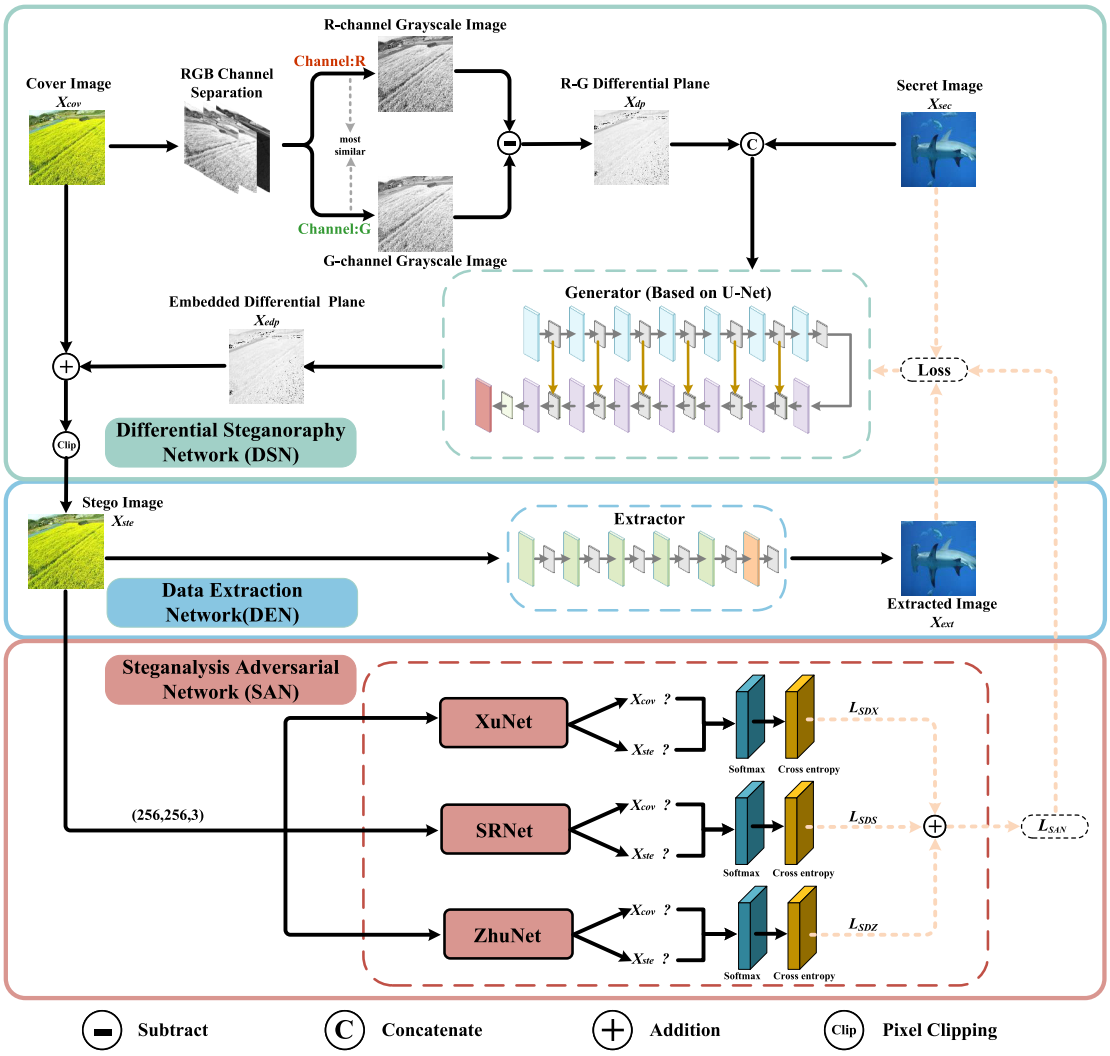


Fig. 1. The structure diagram of the high-capacity differential steganography algorithm for color image based on multiple adversarial networks.

zero is set to zero while any value above or equal to 255 is set to 255. As a result, the stego image not only achieves high visual quality but also obtains strong steganalysis resistance capability.

B. Differential Steganography Network

The Differential Steganography Network comprises two components: the differential plane generation unit and the image embedding unit. In the differential plane generation unit, the cover image is initially separated in RGB color space. Two correlated channels (such as R & G) are subtracted from each other to generate the differential plane. Due to the close correlation between the two similar image channels, the differential planes would be rather sparse, which is advantageous for embedding secret data and effectively minimizing distortion in the stego image. The generated differential plane then served as a new cover for secret data embedding, thereby dispersing the power of the secret image across different color channels. Consequently, it minimizes the interference during secret image hiding and enhances the perceptual fidelity of the stego image. On the other hand, the image embedding unit is mainly composed of an optimized U-Net architecture. This unit is divided into down-sampling and up-sampling modules. The down-sampling module extracts deep latent features from the original cover image by using convolution layers with varying receptive fields. Conversely, the up-sampling module reconstructs the image through deconvolution operations, by incorporating skip connections to fuse shallow and deep features of the original cover image, the quality of the reconstructed image is significantly enhanced in the end.

The differential steganography network is composed of 15 data processing units (As shown in Fig.2). Among them, the initial 7 layers are the encoding network, with each layer comprising a 4×4 convolutional layer, a batch normalization layer (BN), and a Leaky_ReLU activation function. Layers from 8 to 14 belong to the decoding network, where each layer contains a deconvolutional layer with a 4×4 sized kernel, a batch normalization layer (BN), and a ReLU activation function. The 15th layer includes a convolutional layer with a 3×3 sized kernel, a Sigmoid activation function,

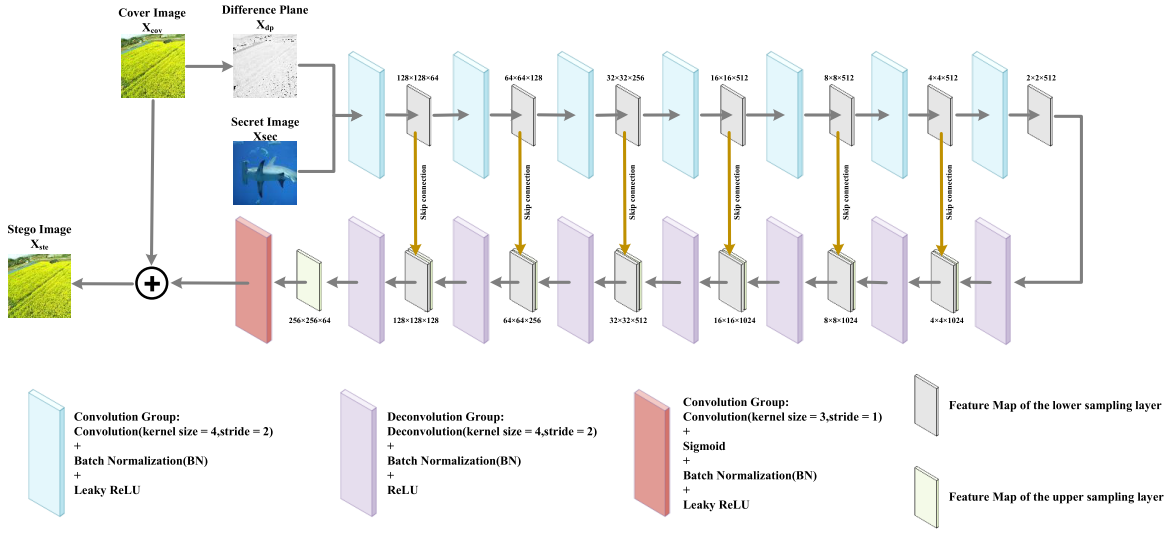


Fig. 2. Architecture of the differential steganography network.

a batch normalization (BN), and a Leaky_ReLU activation function. Additionally, the skip connections are employed to merge features of the i-th layer with those of the (14-i)-th layer in the U-net structure, enabling the decoding network to obtain various dimensional features from the original image during the up-sampling process. This method helps to effectively prevent the loss of trivial content and thus enhances the perceptual quality of generated images. The detailed settings regarding network parameters are provided in Table I.

C. Steganalysis Adversarial Network

The resistance to steganalysis serves as a crucial metric for evaluating the performance of steganography algorithms. The rapid advancement in generative adversarial network technology offers new opportunities for enhancing the steganalysis resistance capability of the stego image. Traditional steganography algorithms typically embed limited information with a heuristic algorithm and assess the anti-detection capability against specific steganalysis methods. It is hard to strike a balance between imperceptibility and high payload capacity. To address this issue, this approach aims to develop a high-capacity differential steganography algorithm based on a generative adversarial network. By competition with certain advanced steganalysis networks, enabling the proposed scheme achieves not only excellent perceptual capability but also high steganography security.

The proposed steganalysis adversarial network incorporates three advanced modules, namely XuNet, SRNet, and ZhuNet, in an additive manner. These three steganalysis modules work in parallel to discriminate the stego image and calculate the corresponding discriminative loss. These three discriminative losses are gathered to provide feedback gradients for the differential steganography network, enabling the stego image to achieve superior steganalysis resistance capability while accommodating a large amount of information. Among them, XuNet is introduced as the first CNN-based steganalysis network. SRNet obtains better steganalysis detection accuracy compared to XuNet by disabling pooling

TABLE I
PARAMETERS OF THE DIFFERENTIAL STEGANOGRAPHY NETWORK

Layer i	Process	Conv/Deconv kernels	Output size
Input	/	/	6×(256×256)
Layer 1	Conv-BN-Leaky ReLU	64×(4×4)	64×(128×128)
Layer 2	Conv-BN-Leaky ReLU	128×(4×4)	128×(64×64)
Layer 3	Conv-BN-Leaky ReLU	256×(4×4)	256×(32×32)
Layer 4	Conv-BN-Leaky ReLU	512×(4×4)	512×(16×16)
Layer 5	Conv-BN-Leaky ReLU	512×(4×4)	512×(8×8)
Layer 6	Conv-BN-Leaky ReLU	512×(4×4)	512×(4×4)
Layer 7	Conv-BN-Leaky ReLU	512×(4×4)	512×(2×2)
Layer 8	DeConv-BN-ReLU	512×(4×4)	512×(4×4)
Skip connection 1	Connect Layer 6 and Layer 8	/	1024×(4×4)
Layer 9	DeConv-BN-ReLU	1024×(4×4)	1024×(8×8)
Skip connection 2	Connect Layer 5 and Layer 9	/	1536×(8×8)
Layer 10	DeConv-BN-ReLU	1024×(4×4)	1024×(16×16)
Skip connection 3	Connect Layer 4 and Layer 10	/	1536×(16×16)
Layer 11	DeConv-BN-ReLU	512×(4×4)	512×(32×32)
Skip connection 4	Connect Layer 3 and Layer 11	/	768×(32×32)
Layer 12	DeConv-BN-ReLU	256×(4×4)	256×(64×64)
Skip connection 5	Connect Layer 2 and Layer 12	/	384×(64×64)
Layer 13	DeConv-BN-ReLU	128×(4×4)	128×(128×128)
Skip connection 6	Connect Layer 1 and Layer 13	/	192×(128×128)
Layer 14	DeConv-BN-ReLU	64×(4×4)	64×(256×256)
Layer 15	Conv-Sigmod-BN-Leaky ReLU	3×(3×3)	3×(256×256)

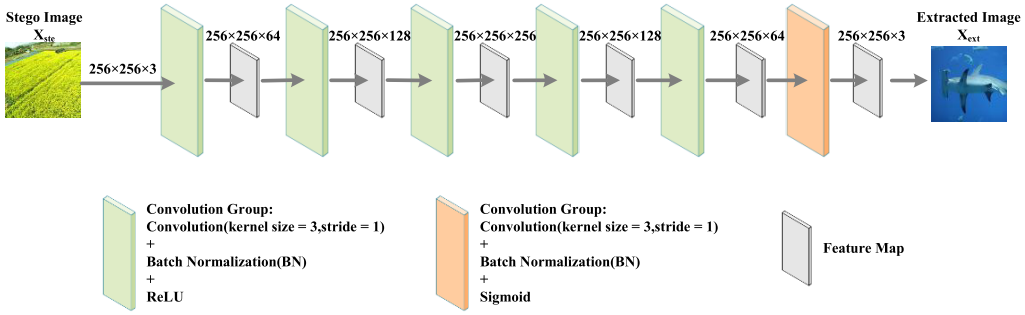


Fig. 3. Main structure of data extraction network.

TABLE II
PARAMETERS OF THE DATA EXTRACTION NETWORK

Layer i	Process	Conv kernels	Padding	Output size
Input	/	/		3×(256×256)
Layer 1	Conv-BN-ReLU	64×(3×3)	Same	64×(256×256)
Layer 2	Conv-BN-ReLU	128×(3×3)	Same	128×(256×256)
Layer 3	Conv-BN-ReLU	256×(3×3)	Same	256×(256×256)
Layer 4	Conv-BN-ReLU	128×(3×3)	Same	128×(256×256)
Layer 5	Conv-BN-ReLU	64×(3×3)	Same	64×(256×256)
Layer 6	Conv-Sigmoid	3×(3×3)	Same	3×(256×256)

operations and preventing convolutional operations from suppressing steganographic signals. Furthermore, ZhuNet adopts smaller convolutional kernels and utilizes separable convolutions to extract residual channel correlations, thereby achieving currently the best performance in steganalysis. Accordingly, this paper leverages these three advanced steganalyzers to construct a high-capability steganalysis adversarial network for enhancing the steganalysis resistance capability of stego images.

D. Data Extraction Network

Data recovery capability is also an important indicator for evaluating the performance of a steganographic scheme. A higher level of similarity between the extracted image and the original one is desired in the data extraction network. In the proposed scheme, a 6-layer convolutional neural network is developed to extract the embedded image comprehensively. As illustrated in Fig.3. The data extraction network consists of encoder and decoder modules. Each convolutional layer employs a 3×3 sized convolution kernel with the stride and padding set as 1. To enhance the non-linear mapping capability of the extractor, batch normalization (BN) and ReLU activation functions are applied after each convolutional layer. Additionally, the sigmoid activation function is utilized in the final layer to facilitate secret image extraction. Notably, the Lion optimizer is employed for the first time in a data extraction network to improve both the network training efficiency and the quality of the extracted image. Specific parameters of the data extraction network are illustrated in Table II.

E. Loss Function

The loss function is an important component of the multi-adversarial differential steganography network for color images. It maps the value of network error to a non-negative real number to represent the "risk" or "loss" of the event, thereby achieving network optimization by minimizing the loss function. In this paper, three kinds of network loss functions, including cross-entropy loss, mean squared error loss, and structural similarity loss, are employed to enhance the performance of the proposed scheme.

1) Cross-Entropy Loss: Cross-entropy is primarily utilized for quantifying the disparity between two probability distributions. In the proposed scheme, Cross-entropies of three steganalyzers, namely XuNet, SRNet, and ZhuNet, are involved to evaluate the differences between the cover and stego images, thereby enhancing the perceptual quality of the generated image. The expressions of three loss functions are as follows,

$$L_{SDX} = -\sum_{SDXi=1}^{2} x'_{SDXi} \log(x_{SDXi}) \tag{2}$$

$$L_{SDS} = -\sum_{SDSi=1}^{2} y'_{SDSi} \log(y_{SDSi}) \tag{3}$$

$$L_{SDZ} = -\sum_{SDZi=1}^{2} z'_{SDZi} \log(z_{SDZi}) \tag{4}$$

where, x_{SDXi} and x'_{SDXi}, y_{SDSi} and y'_{SDSi}, z_{SDZi} and z'_{SDZi} are the probability distributions of the ground truth and the outputs of steganalyzers involve XuNet, SRNet, and ZhuNet, respectively.

2) Mean Squared Error Loss: Mean square error (MSE) is an evaluation index to measure the difference between the ground truth and the estimator. The similarity between the stego image X_{ste} and the cover image X_{cov} serves as an important indicator in assessing the performance of a steganography scheme. Therefore, MSE loss is employed to achieve pixel-level difference detection between the cover and stego images in this approach. The MSE loss between the stego image X_{ste} and the cover image X_{cov} is calculated as follows,

$$L_{S1} = \sum_{S1i=1}^{n} \frac{\left(x_{S1i} - x'_{S1i}\right)^2}{n} \tag{5}$$

where, x_{S1i} and x'_{S1i} are the pixel values of the cover and the stego image, respectively. m denotes the total pixel number of each image. Similarly, the Mean Squared Error (MSE) loss L_{E1} between the secret image X_{sec} and the extracted image X_{ext} is,

$$L_{E1} = \sum_{E1i=1}^{n} \frac{\left(y_{E1i} - y'_{E1i}\right)^2}{n} \tag{6}$$

where, y_{E1i} and y'_{E1i} are the pixel values of the secret image and the extracted image, respectively. n represents the total pixel number of each image.

3) Structure Similarity Loss: Structural Similarity (SSIM) is also an important indicator for evaluating the consistency of content between two images. Considering that natural images always exhibit strong inter-pixel correlations, SSIM measures image distortion by using a combination of three different factors including brightness, contrast, and structure. In this paper, the SSIM Loss involves L_{S2} and L_{E2}, they are employed to calculate the structure differences between the stego image X_{ste} and cover image X_{cov}, as well as between the secret image X_{sec} and extracted image X_{ext}. The formulas are shown as,

$$L_{S2} = \frac{(2\mu_x\mu_{x'} + C_1)(2\sigma_{xx'} + C_2)}{(\mu_x^2 + \mu_{x'}^2 + C_1)(\sigma_x^2 + \sigma_{x'}^2 + C_2)} \tag{7}$$

$$L_{E2} = \frac{(2\mu_y\mu_{y'} + C_3)(2\sigma_{yy'} + C_4)}{(\mu_y^2 + \mu_{y'}^2 + C_3)(\sigma_y^2 + \sigma_{y'}^2 + C_4)} \tag{8}$$

where μ_x, σ_x, $\mu_{x'}$, $\sigma_{x'}$ denote the mean and variance of the cover image and the stego image, respectively. Similarly, μ_y, σ_y, $\mu_{y'}$, $\sigma_{y'}$ denote the mean and variance of the secret image and the extracted image. $\sigma_{xx'}$, $\sigma_{yy'}$ are the covariance between the cover and the stego image, and between the secret image and the extracted image, respectively. Additionally, $C_i = (K_i D)^2 (i = 1, 2, 3, 4)$, where D denotes the dynamic range of image pixels, K_i is a constant. Consequently, SSIM can comprehensively exhibit the perceptual quality of the generated stego image in the semantic view.

4) Total Loss: The total loss L_{Total} of the proposed differential steganography network is composed of three Loss functions including cross-entropy loss, MSE loss, and SSIM loss with different weights. The formula for the total network loss is as follows,

$$L_{Total} = \alpha L_{MSE} + \beta L_{SSIM} + \gamma L_{SAN} \tag{9}$$

where, L_{MSE}, L_{SSIM}, and L_{SAN} are the MSE loss, SSIM loss, and cross-entropy loss, respectively. Specifically, L_{MSE} includes loss of L_{S1} and L_{E1}; L_{SSIM} consists of L_{S2} and L_{E2}; and L_{SAN} is composed of L_{SDX}, L_{SDS}, and L_{SDZ}.

By optimizing the weights of different loss functions, the perceptual fidelity of the stego image, as well as that of the extracted image, is continuously improved. Simultaneously, as three advanced steganalyzers are involved in the proposed scheme to discriminate the cover and the generated stego image, the steganalysis resistance capability of the proposed scheme would be effectively enhanced, thereby achieving a harmonious balance between well steganalysis resistance capability and high stego image visual quality. The detailed optimization process for the proposed scheme is illustrated in Algorithm 2.

Algorithm 2 Training Steps of Color Image High-Capacity Differential Steganography Algorithm Based on Multiple Adversarial Networks

Require:
X_{cov}: cover image
X_{sec}: secret image
SAN: steganalysis adversarial network
G: generator in steganography network
$Concat$: connect by channel dimension
$Clip$: clip the pixel values to the range 0 to 255
Begin:
Step 1: X_{dp} is obtained through two similar channels of X_{cov}
Step 2: $X_{edp} \leftarrow G(Concat(X_{dp}, X_{sec}))$
Step 3: $X_{ste} \leftarrow Clip(X_{cov}+X_{edp})$
Step 4: Input X_{ste} into the SAN to generate L_{SAN}
Step 5: $L_{Total} = \alpha L_{MSE} + \beta L_{SSIM} + \gamma L_{SAN}$, update G with L_{Total}

IV. Experimental Results

A. Experimental Setup

To evaluate the performance of the differential steganography algorithm based on multiple adversarial networks for color images, certain experiments are carried out and the results are discussed in this section. The widely used dataset ImageNet is adopted for network training and performance evaluation. In the experiment, the ImageNet dataset was divided into the training and testing sets. The training set includes two subsets, namely trainA and trainB, each containing 35,000 natural color images with dimensions of 256×256. The testing set comprises 10,000 natural color images with the same dimensions. During the training phase, the Lion optimizer was utilized to balance training efficiency and performance effectively. The learning rate of the Lion optimizer was set to 1×10^{-5}, while the batch size was set to 32. All experiments were conducted on a Dell T630 server with 128GB RAM, and an RTX 3090 GPU card with 24GB VRAM, the software environments are Python 3.7 and TensorFlow 2.4.

B. Qualitative and Quantitative Analysis

In the experiment, images with varying levels of texture complexity were selected from the dataset to evaluate the performance of the proposed network, enabling it to verify its data-hiding capability from different perspectives. The intuitive experimental results are illustrated in Fig.4. Columns 1 and 5 are the cover images and secret images, where, three images with smooth textures (labeled as a, b, c in Fig.4) and three images with complex textures (labeled as d, e, f in Fig.4) are employed to exhibit the performance of the scheme. Columns 2 and 6 displayed their corresponding stego and extracted images. Columns 3, 7, 4, and 8 are histogram distributions of three color channels for cover, secret, stego, and extracted images, respectively.

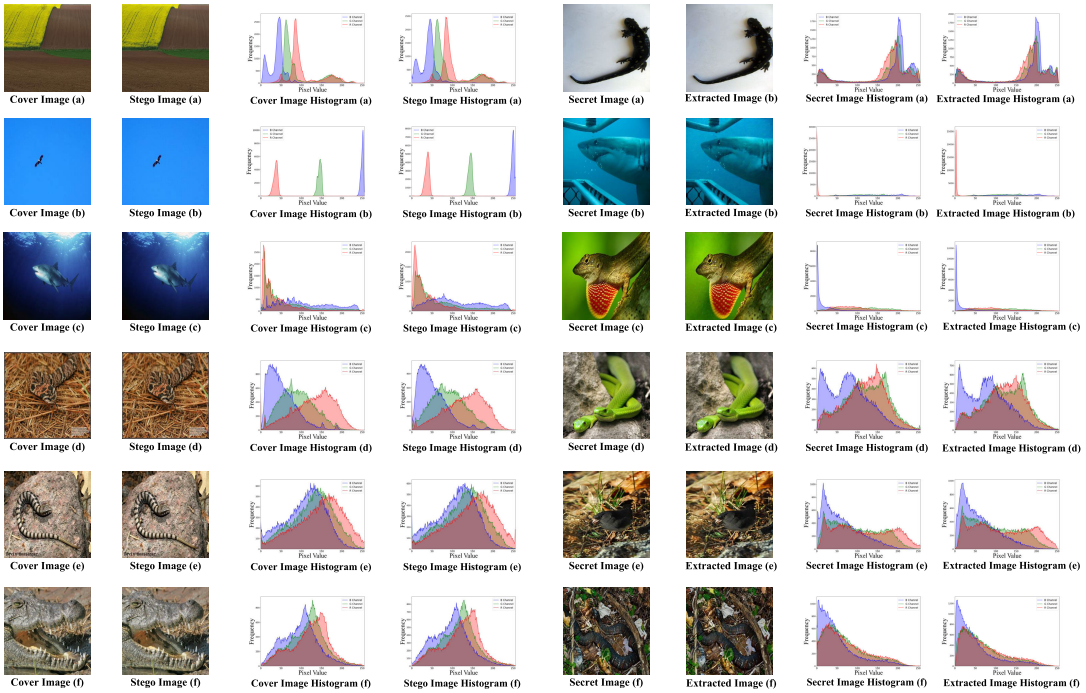


Fig. 4. Samples of the cover and stego images, the secret and extracted images, and their histogram distributions.

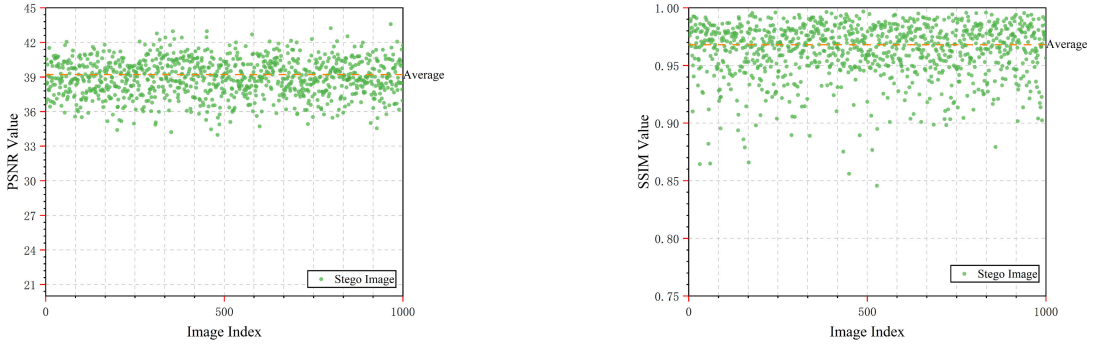


Fig. 5. (a) PSNR scatter-dot diagram of stego images, (b) SSIM scatter-dot diagram of stego images.

From Fig.4, it can be observed that there is a high level of visual similarity between the cover and stego images, as well as between the secret and the extracted images. It is hard to perceptually distinguish the stego image from the cover image, and the extracted image from the embedded secret image perceptually. Simultaneously, The histogram distributions between the stego and the cover images, and those between the extracted and the embedded secret images nearly exhibit the same styles. The results demonstrate that the proposed scheme gains high steganography performance apparently.

The PSNR and SSIM values of stego images from 1000 test images were also presented in Fig.5, where the horizontal axis represents image indices and the vertical axis represents SNR or SSIM values. It can be seen that even after the same-sized image has been embedded into the cover image, the resulting stego image still maintains high visual quality. The average PSNR value of stego images was 39.21 dB. Notably, 27.2% of the images exhibit PSNR values exceeding 40 dB, while 12.1% had PSNR values greater than 41.05 dB. Furthermore, an overwhelming majority (98.9%) possessed PSNR values higher than 36 dB. Additionally, the average SSIM value of resulting stego images surpasses 0.968, with approximately 21.4% of stego images achieving SSIM values greater than 0.98.

Moreover, Fig.6 further shows both PSNR and SSIM distributions of the extracted images. The results indicate that the average PSNR of the extracted secret images was about 34.92 dB. Specifically, 26.7% of the extracted images achieve PSNR values exceeding 35.35 dB. On the other hand, the average SSIM of the extracted images reaches 0.948, and 11.2% of extracted images achieve SSIM values greater than 0.96.

The experimental results demonstrate that the differential steganography algorithm based on multiple adversarial networks for color images can not only generate high-fidelity stego images but also extract the embedded image extremely

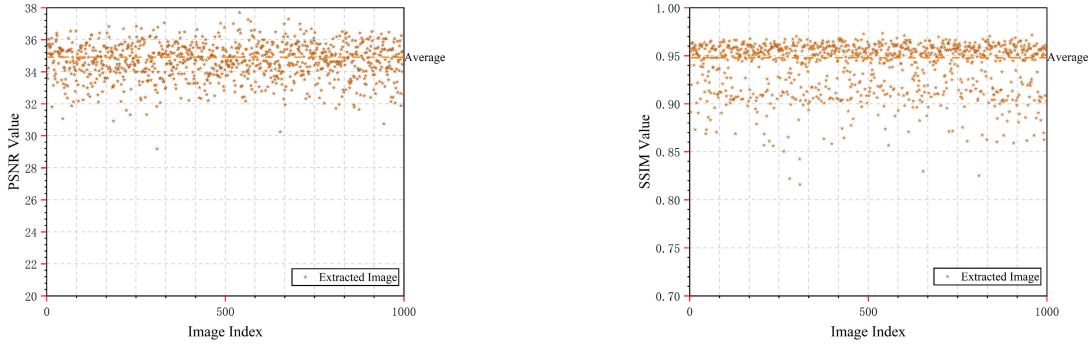


Fig. 6. (a) PSNR scatter-dot diagram of extracted images, (b) SSIM scatter-dot diagram of extracted images.

TABLE III
STEGANOGRAPHY PERFORMANCE OF DIFFERENT DIFFERENTIAL PLANES

	Stego Image (SSIM)	Stego Image (PSNR)	Extracted Image (SSIM)	Extracted Image (PSNR)
B&R	0.956	38.25dB	0.831	27.16dB
G&B	0.911	35.25dB	0.853	27.49dB
R&G	0.968	39.21dB	0.948	34.92dB

consistent with the original one. Consequently, it holds great potential within image steganography applications.

C. Influence of Different Network Coefficients

1) Influence of Difference Plane: The adoption of different color channels for the generation of differential planes also influences the performance of the steganography scheme. In the experiment, three RGB color channels of an image were subtracted from each other to generate three distinct differential planes, thereby analyzing the influence of various differential planes on steganographic performance. As shown in Table III, the experimental results demonstrate that the R and G channels generated differential plane yields PSNR values of 39.21dB for the stego image and 34.92dB for the extracted image. Whereas, the differential planes generated from B and R channels lead to a slight degradation in visual quality for both the stego and the extracted secret images, resulting in the PSNR values of 38.25 and 27.16dB, respectively. Similarly, when the B and G channels are utilized to generate the differential plane, the resulting PSNR values are 35.25dB for the stego image and 27.49dB for the extracted secret image.

The results demonstrate a strong correlation of pixels among the R and G channels, indicating that pixels in these two channels always exhibit similar changing trends. Consequently, the differential plane generated from the R and G channels is particularly suitable for large data hiding.

2) Influence of Loss Function Weights: As previously discussed, different loss function weights were assigned to the loss functions in order to optimize the performance of the proposed scheme. An organic combination of different loss function weights enables the scheme to achieve a superior balance between the visual fidelity and the steganalysis resistance capability of the stego image, as well as to gain high data extraction performance for secret images.

TABLE IV
EXPERIMENTAL RESULTS WITH DIFFERENT VALUES OF α AND β

α	β	SSIM (Stego)	PSNR (Stego)	SSIM (Extracted)	PSNR (Extracted)
0	1	0.971	39.43dB	0.917	30.47dB
0.25	0.75	0.969	39.24dB	0.925	31.18dB
0.5	0.5	0.968	39.21dB	0.948	34.92dB
0.75	0.25	0.965	39.06dB	0.936	33.41dB
1	0	0.952	39.03dB	0.939	33.83dB

As shown in Table IV, the mean square error (MSE) and structural similarity (SSIM) loss function weights are firstly adjusted to achieve optimal visual effects for both stego images and extracted images. Both the PSNR and SSIM are utilized to evaluate the influence of the loss function weights on the performance of the proposed scheme, with the purpose of minimizing the image distortion between the stego image and cover image, as well as between the extracted and embedded secret images.

It can be observed that a smaller α combined with a larger β weight always results in an enhanced visual quality for the stego images but poorer quality for extracted secret images. Conversely, a larger α and a smaller β slightly enhance the visual quality of the extracted secret images but reduce the quality of the stego images correspondingly. The experimental results reveal that the SSIM loss is primarily responsible for the structural similarity of images, and the PSNR loss is highly sensitive to pixel changes. As a result, the optimal weights of α and β are set to 0.5, which enables the proposed scheme to achieve excellent visual quality for both stego and extracted secret images.

Furthermore, the influence of cross-entropy loss γ is also investigated in this paper. As shown in Table V. It is evident that the steganalysis resistance capability of the stego image grows with the weights γ increases. When the value of γ is 1, the steganography resistance capabilities of the stego image are 51.1%, 54.2%, and 58.1% against XuNet, SRNet, and ZhuNet, respectively; however, the resulting image only achieves a PSNR of 36.91dB under this condition. It should be noted that the value 50% indicates equal probabilities between cover and stego images where steganalyzers fail to distinguish them apart. Conversely, when the value of γ is reduced to 0.0025, the resistance capability of stego images against XuNet, SRNet, and ZhuNet steganalyzers declines to 52.4%, 58.2%,

TABLE V
COMPARISON OF EXPERIMENTAL RESULTS WITH DIFFERENT COEFFICIENT γ

γ	SSIM (Stego)	PSNR (Stego)	SSIM (Extracted)	PSNR (Extracted)	XuNet	SRNet	ZhuNet
0	0.981	40.23dB	0.971	36.65dB	70.2%	75.4%	81.2%
0.0025	0.978	39.73dB	0.967	36.02dB	52.4%	58.2%	62.4%
0.01	0.974	39.44dB	0.954	35.21dB	52.2%	57.1%	61.2%
0.05	0.968	39.21dB	0.948	34.92dB	51.8%	54.8%	59.7%
0.25	0.963	37.42dB	0.949	34.61dB	51.7%	54.6%	58.8%
1	0.954	36.91dB	0.944	34.14dB	51.1%	54.2%	58.1%

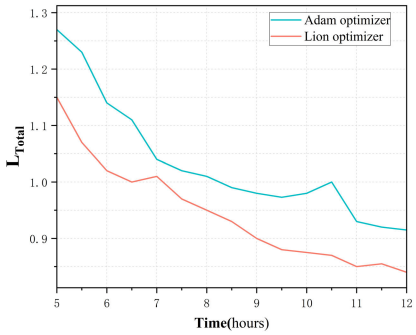


Fig. 7. Comparison of convergence speed between Adam optimizer and Lion optimizer.

and 62.4%, respectively; While the average PSNR of stego images increases up to 39.73 dB. Therefore, the value of γ is set to 0.05 in this experiment, aiming to strike an optimal balance between the visual fidelity and steganalysis resistance capability for stego images.

3) Influence of Optimizer: Deep neural networks approach the global optimal solution by gradient back-propagation. A superior optimizer facilitates a steady and rapid descent of the gradient. Adam optimizer is a common choice for deep learning training, it adjusts the learning rate by calculating the first and second moments of the gradient to achieve a robust and adaptive optimization strategy. However, there is still room for improvement in terms of convergence speed and stability with this optimizer. Therefore, the Lion optimizer is employed for the first time in our proposed image steganography scheme to obtain high network training performance. The Lion optimizer was initially presented by Google in 2023 to address challenges encountered during deep neural network training, such as hard convergence, gradient explosion, or disappearance. As shown in Fig.7, the total loss of the steganography network reduces with the training time increases both under Adam and Lion optimizers. However, it is worth noting that the loss descends significantly faster under the Lion optimizer compared to the Adam optimizer. Specifically, after 5 hours of training, the loss of the differential steganography network under Adam is 1.271, while the loss reduces to 1.149 under the Lion optimizer. Moreover, when the training time reaches 12 hours, the loss of the proposed network under the Lion optimizer declines to 0.835, whereas the loss of the network under the Adam optimizer is 0.927. The convergence speed of the proposed network using the Lion optimizer is approximately 15.5% faster than that using the Adam optimizer within the same training duration.

The reason can be attributed to the following two points, firstly, compared to Adam and other optimizers require both the first and the second order moments, the Lion optimizer only requires momentum, which reduces the additional memory footprint by half. Secondly, the Lion algorithm generates uniform amplitude updates in all dimensions through symbolic operations, which mitigates instability and oscillations caused by the loss difference and promotes model generalization. As a result, the Lion optimizer enables the differential steganography networks to gain superior performance and convergence capability compared to the Adam optimizer.

TABLE VI
PSNR AND SSIM OF STEGO IMAGE AND EXTRACTED IMAGE

	Stego (SSIM)	Stego (PSNR)	Extracted (SSIM)	Extracted (PSNR)
Baluja [12]	0.924	28.41dB	0.922	28.06dB
Duan [18]	0.954	36.67dB	0.963	36.84dB
Fu [21]	0.941	31.05dB	0.942	29.83dB
Han [22]	0.943	34.25dB	0.945	33.87dB
Zheng [28]	0.960	33.09dB	0.947	33.31dB
Cui [29]	0.976	40.23dB	0.951	39.12dB
Ours	0.968	39.21dB	0.948	34.92dB

V. EXPERIMENTAL RESULTS COMPARISON AND DISCUSSIONS

To further demonstrate the superiority of the proposed scheme, this section compared the performance of our differential steganography networks with certain state-of-the-art image steganography schemes. For comparison fairness, each scheme is trained on the same image dataset and testified with its optimal hyperparameters. The visual quality of the stego and the extracted secret images, the steganalysis resistance capability of the stego image, and the network steganography efficiency are all discussed in this section so that the performance of the proposed scheme can be comprehensively evaluated.

A. Visual Quality Comparison With Other Advanced Schemes

To assess the fidelity of stego images, this study compares the proposed scheme with six advanced steganography approaches, namely Baluja's scheme [12], Duan's scheme [18], Fu's scheme [21], Han's scheme [22], Zheng's scheme [28] and Cui's scheme [29]. All of these schemes have achieved excellent image steganography capabilities. Meanwhile, two widely used metrics PSNR and SSIM are employed for evaluating image quality.

As illustrated in Table VI, the experimental results indicate that the stego image generated from the proposed scheme

TABLE VII
DETECTION ACCURACY OF XUNET, SRNET, AND ZHUNET STEGANALYZERS ON SIX STEGANOGRAPHIC SCHEMES

	Baluja [12]	Duan [18]	Fu [21]	Han [22]	Zheng [28]	Cui [29]	Ours
XuNet	98.9%	92.5%	77.8%	59.7%	62.5%	61.2%	51.8%
SRNet	99.5%	94.3%	83.6%	72.4%	68.2%	55.5%	54.8%
ZhuNet	99.8%	96.2%	86.4%	78.6%	70.8%	89.1%	59.7%

TABLE VIII
COMPARISON OF GENERATION TIMES FOR 1000 IMAGES

	Baluja [12]	Duan [18]	Fu [21]	Han [22]	Zheng [28]	Cui [29]	Ours
Time(s)	37.546	38.354	44.099	83.369	57.621	320	42.373

achieves significantly higher PSNR and SSIM values compared to other advanced schemes. Specifically, the average PSNR of our stego images is 10.8dB, 2.54dB, 8.16dB, 4.96dB, and 6.12dB higher than Baluja's scheme, Duan's scheme, Fu's scheme, Han's scheme, and Zheng's scheme after the same sized image hiding, respectively. Additionally, the average SSIM values of our stego images are also increased by 0.044, 0.014, 0.027, 0.025, and 0.008 compared to other corresponding steganography schemes.

Simultaneously, the fidelity of the extracted secret image was also evaluated in experiments. As shown in Table VI, it can be observed that the average PSNR of secret images extracted from our proposed scheme achieves an improvement of 6.86dB, 5.09dB, 1.05dB, and 1.61dB over Baluja's scheme, Fu's scheme, Han's scheme, and Zheng's scheme, respectively. Meanwhile, the average SSIM also has improved by 0.026, 0.006, 0.003, and 0.001 over Baluja's scheme, Fu's scheme, Han's scheme, and Zheng's scheme. Although the average PSNR and SSIM of extracted images are slightly less than that of Duan's scheme by 1.92dB and 0.015, our stego image achieved a higher average PSNR and SSIM than Duan's approach by 2.54 dB and 0.014. It is also worth noting that although the average PSNR and SSIM values of our stego and extracted images are slightly inferior to those of Cui's scheme, the values of the two indicators are very similar to those of Cui's scheme. The quality discrepancy between stego images generated by the two schemes is almost negligible, as well as the extracted secret image. Moreover, by incorporating both the differentiation of image channels and three advanced steganalyzers into the proposed scheme, our stego images achieve superior steganalysis resistance capability than that of Cui's scheme, and the data hiding efficiency is also obviously improved.

In sum, the proposed scheme exhibits superior performance in terms of visual fidelity for both stego and extracted images compared to other advanced image steganographic schemes. This can be attributed to the fact that the proposed scheme embeds information into the differential plane generated from two similar channels of the image. The energy of the embedded image is dispersed across certain closely related channels, thereby avoiding excessive disturbance caused by data embedding to the cover image. As a result, the distortion is effectively mitigated even after high-capacity information embedding, and thus the visual fidelity of the stego image is greatly enhanced.

B. Resistance to Steganalysis Comparison With Other Advanced Schemes

The steganalysis resistance capability of the stego image from the proposed method is also evaluated by comparing it with other excellent steganography schemes. All the schemes are trained on the ImageNet dataset and initialized with identical parameters. The average steganalysis results for stego images generated by different steganographic schemes are illustrated in Table VII. The experimental results indicate that stego images from the proposed scheme gain the strongest steganalysis resistance capabilities compared to other schemes. Specifically, the steganalysis detection rates of stego images from the proposed scheme are 51.8%,54.8%, and 59.7% against XuNet, SRNet, and ZhuNet steganalyzers, surpassing other steganography schemes by at least 7.9%, 0.7%, and 11.1%, respectively. The proposed scheme hides the secret image into the differential plane of closely correlated image channels through multiple adversarial networks and employs certain advanced steganalyzers parallelly to improve the steganalysis resistance capability of the stego image by multilevel network confrontation, thereby enabling the stego image to achieve high anti-steganalysis capability after multiple data hiding adversarial. The experimental results demonstrate superior performance in terms of resisting steganalysis for our proposed steganographic scheme.

C. Efficiency Comparison With Other Advanced Schemes

The computational efficiency is also an important indicator for a deep learning-based steganography scheme. In the experiment, the computational efficiency of the proposed scheme and other steganographic schemes were testified by generating 1000 same-sized stego images, respectively. The average time for each steganography scheme is illustrated in Table VIII. It can be observed that the proposed scheme achieves excellent data-hiding efficiency compared to other advanced steganographic schemes. Although our data hiding time is slightly longer than that of Baluja's scheme and Duan's scheme, the difference is negligible. However, when compared to

other high-performance schemes, our approach demonstrates a substantial improvement in operational efficiency.

The reason is that Baluja's scheme and Duan's scheme solely employ simple convolutional networks to achieve image steganography, the steganalysis resistance capabilities of the stego image have not been considered in the process of data hiding, therefore, their stego image's quality and security is low. On the contrary, the other four high-performance steganography schemes employ the adversarial network to improve the anti-steganalysis capability of stego images, their network structure is complex and the data hiding time is far larger than our proposed scheme for the same-sized image embedding. The results clearly show that the proposed scheme achieves higher data-hiding efficiency than other superior performance schemes that involve modules for steganalysis resistance capability improvement. The slight increase in stego image generation time compared to those low-performance steganography schemes is completely acceptable. In sum, the proposed scheme achieves the best comprehensive performance compared to other state-of-the-art schemes.

VI. Conclusion

This paper proposes a high-capacity differential steganography algorithm based on multiple adversarial networks for color images. By embedding a same-sized secret image into the differential plane of the cover image, the algorithm effectively enhances the steganography capacity, enabling the stego and extracted secret images to gain excellent fidelity compared to other advanced schemes. Furthermore, it incorporates multiple adversarial networks to enhance the steganalysis resistance capability of stego images. As a result, not only the perceptual quantity of the stego and extracted images is enhanced, but also the steganalysis capability of stego images is improved to a great extent. In the experiment, the impacts of different loss functions, image planes, and network optimizers are discussed in detail to evaluate the superiority of the proposed scheme. Moreover, the fidelity of the stego and extracted images, the steganography resistance capability of the stego image, and the data hiding efficiency of the proposed scheme are all investigated by comparing them with other state-of-the-art steganography approaches. Extensive experimental results demonstrate the superiority of the proposed scheme in terms of image steganography.

In future research, we will focus on optimizing the generative adversarial networks to enhance both the visual quality and resistance against steganalysis of stego images.

References

[1] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 920–935, Sep. 2011.

[2] B. Ma, Z. Tao, R. Ma, C. Wang, J. Li, and X. Li, "A high-performance robust reversible data hiding algorithm based on polar harmonic Fourier moments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2763–2774, Apr. 2024.

[3] B. Ma, K. Li, J. Xu, C. Wang, J. Li, and L. Zhang, "Enhancing the security of image steganography via multiple adversarial networks and channel attention modules," *Digit. Signal Process.*, vol. 141, Sep. 2023, Art. no. 104121.

[4] S. Weng, Y. Zhou, T. Zhang, M. Xiao, and Y. Zhao, "Reversible data hiding for JPEG images with adaptive multiple two-dimensional histogram and mapping generation," *IEEE Trans. Multimedia*, vol. 25, pp. 8738–8752, 2023.

[5] J. Mielikainen, "LSB matching revisited," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 285–287, May 2006.

[6] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Proc. 12th Int. Conf. Inf. Hiding*, Jun. 2010, pp. 161–177.

[7] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Costa Adeje, Spain, Dec. 2012, pp. 234–239.

[8] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP J. Inf. Secur.*, vol. 2014, pp. 1–13, Dec. 2014.

[9] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4206–4210.

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

[11] J. Hayes and G. Danezis, "Generating steganographic images via adversarial training," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.

[12] S. Baluja, "Hiding images in plain sight: Deep steganography," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[13] I. J. Goodfellow, "Generative adversarial networks," in *Proc. Neural Inf. Process. Syst.*, vol. 3, Jul. 2014, pp. 2672–2680.

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

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

[16] J. Tan, X. Liao, J. Liu, Y. Cao, and H. Jiang, "Channel attention image steganography with generative adversarial networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 2, pp. 888–903, Mar. 2022.

[17] L. Liu, L. Meng, X. Wang, and Y. Peng, "An image steganography scheme based on ResNet," *Multimedia Tools Appl.*, vol. 81, no. 27, pp. 39803–39820, Nov. 2022.

[18] X. Duan, K. Jia, B. Li, D. Guo, E. Zhang, and C. Qin, "Reversible image steganography scheme based on a U-Net structure," *IEEE Access*, vol. 7, pp. 9314–9323, 2019.

[19] X. Duan, M. Gou, N. Liu, W. Wang, and C. Qin, "High-capacity image steganography based on improved xception," *Sensors*, vol. 20, no. 24, p. 7253, Dec. 2020.

[20] X. Duan, W. Wang, N. Liu, D. Yue, Z. Xie, and C. Qin, "StegoPNet: Image steganography with generalization ability based on pyramid pooling module," *IEEE Access*, vol. 8, pp. 195253–195262, 2020.

[21] Z. Fu, F. Wang, and X. Cheng, "The secure steganography for hiding images via GAN," *EURASIP J. Image Video Process.*, vol. 2020, no. 1, p. 46, Dec. 2020.

[22] B. Ma, Z. Han, J. Li, C. Wang, Y. Wang, and X. Cui, "A high-capacity and high-security generative cover steganography algorithm," in *Proc. Int. Conf. Artif. Intell. Secur.* Cham, Switzerland: Springer, 2022, pp. 411–424.

[23] F. Li, Y. Sheng, X. Zhang, and C. Qin, "ISCMIS: Spatial-channel attention based deep invertible network for multi-image steganography," *IEEE Trans. Multimedia*, vol. 26, pp. 3137–3152, 2023.

[24] F. Li, Z. Yu, and C. Qin, "GAN-based spatial image steganography with cross feedback mechanism," *Signal Process.*, vol. 190, Jan. 2022, Art. no. 108341.

[25] L. Zhou, G. Feng, L. Shen, and X. Zhang, "On security enhancement of steganography via generative adversarial image," *IEEE Signal Process. Lett.*, vol. 27, pp. 166–170, 2020.

[26] D. Huang, W. Luo, M. Liu, W. Tang, and J. Huang, "Steganography embedding cost learning with generative multi-adversarial network," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 15–29, 2024, doi: 10.1109/TIFS.2023.3318939.

[27] Y. Xu, C. Mou, Y. Hu, J. Xie, and J. Zhang, "Robust invertible image steganography," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 7865–7874, doi: 10.1109/CVPR52688.2022.00772.

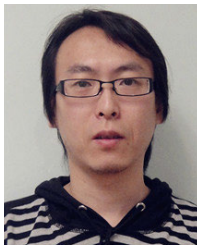
[28] Z. Zheng, Y. Hu, Y. Bin, X. Xu, Y. Yang, and H. T. Shen, "Composition-aware image steganography through adversarial self-generated supervision," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 9451–9465, Nov. 2023, doi: 10.1109/TNNLS.2022.3175627.
[29] Q. Cui, W. Tang, Z. Zhou, R. Meng, G. Nan, and Y.-Q. Shi, "Meta security metric learning for secure deep image hiding," *IEEE Trans. Dependable Secure Comput.*, vol. 21, no. 5, pp. 4907–4920, Sep. 2024, doi: 10.1109/TDSC.2024.3363692.
[30] J. Yu, X. Zhang, Y. Xu, and J. Zhang, "Cross: Diffusion model makes controllable, robust and secure image steganography," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2024, pp. 1–14.
[31] Y. Wang, W. Zhang, W. Li, X. Yu, and N. Yu, "Non-additive cost functions for color image steganography based on inter-channel correlations and differences," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2081–2095, 2020.
[32] H. Yao, X. Liu, Z. Tang, Y.-C. Hu, and C. Qin, "An improved image camouflage technique using color difference channel transformation and optimal prediction-error expansion," *IEEE Access*, vol. 6, pp. 40569–40584, 2018.
[33] T. D. Nguyen, S. Arch-Int, and N. Arch-Int, "A novel secure channel selection rule for spatial image steganography," in *Proc. 12th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, Songkhla, Thailand, Jul. 2015, pp. 230–235.
[34] Y. Kang, F. Liu, C. Yang, X. Luo, and T. Zhang, "Color image steganalysis based on residuals of channel differences," *Comput., Mater. Continua*, vol. 59, no. 1, pp. 315–329, 2019.
[35] I. Shukla, A. Joshi, and S. Girme, "LSB steganography mechanism to hide texts within images backed with layers of encryption," in *Proc. 16th Int. Conf. Secur. Inf. Netw. (SIN)*, Jaipur, India, Nov. 2023, pp. 1–6, doi: 10.1109/SIN60469.2023.10474976.
[36] S. Prasad, A. K. Pal, and S. Mukherjee, "An RGB color image steganography scheme by binary lower triangular matrix," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 6865–6873, Jul. 2023, doi: 10.1109/TITS.2023.3264467.
[37] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 868–882, Jun. 2012.
[38] G. Xu, H.-Z. Wu, and Y.-Q. Shi, "Structural design of convolutional neural networks for steganalysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 708–712, May 2016.
[39] J. Ye, J. Ni, and Y. Yi, "Deep learning hierarchical representations for image steganalysis," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 11, pp. 2545–2557, Nov. 2017.
[40] M. Boroumand, M. Chen, and J. Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 5, pp. 1181–1193, May 2019.
[41] R. Zhang, F. Zhu, J. Liu, and G. Liu, "Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1138–1150, 2020.
[42] X. Chen et al., "Symbolic discovery of optimization algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2024, pp. 1–29.

Bin Ma (Member, IEEE) received the M.S. and Ph.D. degrees from Shandong University, Jinan, China, in 2005 and 2008, respectively. From 2008 to 2013, he was an Associate Professor with the School of Information Science, Shandong University of Political Science and Law, Jinan. From 2013 to 2015, he visited New Jersey Institute of Technology, Newark, NJ, USA, as a Visiting Scholar. He is currently a Professor with the School of Cyber Security, Qilu University of Technology (Shandong Academy of Sciences), Shandong, China. His research interests include reversible data hiding, multimedia security, and image processing. He is a member of ACM. He serves as an Editorial Board Member of a few journals, such as IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, *Journal of Visual Communication and Image Representation*, and *IEEE Signal Processing Magazine*.

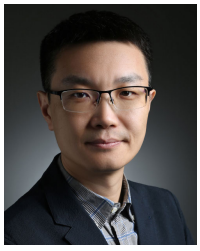
Haocheng Wang was born in Taian, China, in 2000. He is currently pursuing the M.S. degree with Qilu University of Technology (Shandong Academy of Sciences), China. His research interests include multimedia information security, data hiding, and steganography.

Jian Xu was born in Weifang, Shandong, in 1973. She received the master's degree. She is currently an Associate Professor with Shandong University of Finance and Economics. Her research interests include multimedia information security.

Xiaoyu Wang was born in Heze, China, in 1994. She received the B.S. degree in computer science and technology and the M.S. degree in computer application from the Qilu University of Technology (Shandong Academy of Science), Jinan, China, in 2016 and 2019, respectively. She is currently pursuing the Ph.D. degree in computer science and technology from Dalian Maritime University, Dalian, China. Her research interests include reversible data hiding, machine vision, and image processing.



Xiaolong Li (Member, IEEE) received the B.S. degree from Peking University, China, in 1999, the M.S. degree from École Polytechnique, France, in 2002, and the Ph.D. degree in mathematics from the ENS de Cachan, France, in 2006. He was a Post-Doctoral Fellow and then a Researcher with Peking University from 2007 to 2016. He is currently a Professor with the Institute of Information Science, Beijing Jiaotong University, Beijing, China. His research interests include image processing and information hiding.



Jian Li received the bachelor's and master's degrees in computer science from Shandong University and the Ph.D. degree in computer science from Sun Yat-sen University. He is currently with the Qilu University of Technology (Shandong Academy of Sciences), China. With the support of NSFC and NSFC Jiangsu Province, he has authored or co-authored more than 20 research articles most of which are about digital image and video forensics.