RED-Net: Residual and Enhanced Discriminative Network for Image Steganalysis in the Internet of Medical Things and Telemedicine

Kai Chen, Zhengyuan Zhou, Yuchen Li, Xu Ji, Jiasong Wu, Gouenou Coatrieux, Jean-Louis Coatrieux, Yang Chen

RED-Net: Residual and Enhanced Discriminative Network for image steganalysis in the Internet of medical things and telemedicine

Kai Chen, Zhengyuan Zhou, Yuchen Li, Xu Ji, Jiasong Wu, Gouenou Coatrieux, *Senior Member, IEEE*, Jean-Louis Coatrieux,*Life fellow, IEEE*, and Yang Chen, *Senior Member, IEEE*

***Abstract*— Internet of Medical Things(IoMT) and telemedicine technologies utilize computers, communications, and medical devices to facilitate off-site exchanges between specialists and patients, specialists, and medical staff. If the information communicated in IoMT is illegally steganography, tampered or leaked during transmission and storage, it will directly impact patient privacy or the consultation results with possible serious medical incidents. Steganalysis is of great significance for the identification of medical images transmitted illegally in IoMT and telemedicine. In this paper, we propose a Residual and Enhanced Discriminative Network(RED-Net) for image steganalysis in the internet of medical things and telemedicine. RED-Net consists of a steganographic information enhancement module, a deep residual network, and steganographic information discriminative mechanism. Specifically, a steganographic information enhancement module is adopted by the RED-Net to boost the illegal steganographic signal in texturally complex high-dimensional medical image features. A deep residual network is utilized for steganographic feature extraction and compression. A steganographic information discriminative mechanism is employed by the deep residual network to enable it to recalibrate the steganographic features and drop high-frequency features that are mistaken for steganographic information. Experiments conducted on public and private datasets with data hiding payloads ranging from 0.1bpp/bpnzac-0.5bpp/bpnzac in the spatial and JEPG domain led to RED-Net's steganalysis error P_E in the range of 0.0732-0.0010 and 0.231-0.026, respectively. In general, qualitative and quantitative results on public and private datasets demonstrate that the RED-Net outperforms 8 state-of-art steganography detectors.**

***Index Terms*— Telemedicine, the Internet of Medical Things(IoMT), medical image processing, steganalysis, deep learning, and medical information protection.**

This work was supported in part by the National Key Research and Development Program of China (No. 2021ZD0113202), in part by the State Key Project of Research and Development Plan under Grant 2022YFC2401600, in part by the National Natural Science Foundation of China under Grant T2225025, in part by the Key Research and development Programs in Jiangsu Province of China under Grant BE2021703 and BE2022768, in part by China Scholarship Council under Grant 202106090126. Gouenou Coatrieux is the co-corresponding author and Yang Chen is the corresponding author.

Kai Chen is with the School of Cyber Science and Engineering, Southeast University, Nanjing 210096, and also with the Key Laboratory of Computer Network and Information Integration (Ministry of Education), Southeast University, Nanjing 210096, China.

Zhenyuan Zhou, Yuchen Li, Jiasong Wu, and Xu Ji are with the School of Computer Science and Engineering, Southeast University, Nanjing 210096.

Gouenou Coatrieux is with the IMT Atlantique, Inserm, LaTIM UMR1101,Brest 29000, France.

Jean-Louis Coatrieux is with the Laboratoire Traitement du Signal et de l'Image, Universit ´e de Rennes 1, F-35000 Rennes, France, with the Centre de Recherche en Information Biom´edicale Sino-francais,35042 Rennes, France, and also with the National Institute for Health and Medical Research, F-35000 Rennes, France.

Yang Chen is with the Key Laboratory of New Generation Artificial Intelligence Technology and its Interdisciplinary Applications(Southeast University), Ministryof Education, China, Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, School of Computer Science and Engineering, Southeast University, Nanjing, 210096, China, Jiangsu Key Laboratory of Molecular and Functional Imaging, Department of Radiology, Zhongda Hospital, Southeast University, Nanjing 210009, China(e-mail: chenyang.list@seu.edu.cn).

I. Introduction

Internet of Medical Things (IoMT) and telemedicine technologies facilitate the interaction between specialists and medical devices by transferring data, text, voice, and image data to medical information networks [1]–[5]. When a medical imaging device scans a patient, the resulting images are stored in the picture archiving and communication system (PACS) and then transmitted to the workstation of where a physician can analyze them with other information from the hospital information system (HIS). These pieces of information can also be transmitted via IoMT and telemedicine for remote diagnosis, treatment, and consultation of sick and injured patients, as in the context of, for example, remote areas with poor medical conditions, islands, or ships. In such open environments, it is urgent to strengthen the information security and privacy protection of data during transmission, storage, and usage.

Watermarking and steganography have been recently proposed to enhance the protection of medical images. Both are based on data or information hiding which consists of inserting a message into an image by imperceptibly modifying its pixel gray level values. Such modifications are referred to as the watermark or steganographic signal. Watermarking technology can provide powerful technical support for knowing where the image comes from, which patient it belongs to and to detect an image that has been tampered [6]–[10]. In healthcare, steganography interest is more about protecting privacy by embedding sensitive patient data into the image [11]–[15]. As information-hiding technology continues to spread, it has gradually become a double-edged sword. While it provides security for people's communication, it is also used by criminals for personal gain or terrorist attacks. In 2001, CNN, a mainstream media outlet in the United States, published a story about the use of steganography to commit crimes through covert communications [16]. Steganography has been used in cases such as the 2007 Colombian drug cartel and the 2011 Almighty God cult. The illegal and malicious use of medical image steganography is bound to cause medical information leakage. One malicious user can for instance leak sensitive data through externalized images. It can also, result in significant medical errors if images are too heavily modified by the steganographic processes. Steganalysis is a dual technique of steganography, which can differentiate stego-images from original images or, more clearly, detect the steganographic signal in steganographic images. Steganalysis models can be divided into specific and universal methods depending on the scope of application. Specific steganalysis is a specialized model designed when the specific steganographic algorithm is known. In 2005, Andrew developed the feature histogram formula (HCF) for steganalysis stego-signal in grayscale LSB image [17]. Liu et al. utilized the correlation of the lowest two-bit planes of the image as a feature to detect the LSBM steganography scheme [18]. Bohme proposed a dedicated steganalysis algorithm for Jsteg using the LSB algorithm on frequency domain images [19]. Xia et al. designed a special steganalysis model for the LSBM steganography algorithm by analyzing the correlation between adjacent pixels [20]. Gul et al. [21] and Luo et al. [22]proposed a specialized steganalysis model for HUGO, respectively. Tang et al. proposed a steganalysis model for

the WOW embedding algorithm which works both the spatial and frequency domains [23].

In recent years, deep learning has shown significant advantages in the medical field [24]–[29] and medical image steganalysis [30]–[34]. For instance, Lu et al. proposed a framework named DTTSN to support deterministic low-latency communication for large numbers of potential customers with real-time medical demands in smart healthcare application [35]. Tusar et al. used improved preprocessing techniques, an efficient combination of spectral, cepstrum, and periodic features, and the implementation of a gradient enhancer to achieve robust and consistent performance across multiple datasets [36]. Regarding steganalysis, Zeng et al. first proposed a deep learning-based model for frequency domain images to demonstrate that the deep learning-based steganalysis model no longer has detection capability for a single domain [37]. In 2018, Li et al. proposed a structure called ReST-Net, which incorporates width network ideas on top of the Xu-Net model [38]. An alternative method for steganographic analysis of digital images based on convolutional neural networks (CNNs) was proposed by Ye et al [39]. Tan et al. proposed a spatial domain enrichment model, which is a well-trained CNN whose steganalysis performance should be comparable to or even better than the hand-coded SRM [40]. Federated learning is a learning mechanism in which multiple data holders collaborate to train a model without sharing data, and only exchange training parameters in the intermediate stages, avoiding direct exposure of data to third parties and providing natural protection of data privacy [41]. Çukur et al. developed the first federated learning-based personalized MRI synthesis model for transforming source contrast images into target contrast images, which contributes to patient privacy preservation during multi-institutional collaborations [42]. Çukur et al. proposed Federated Learning for Generating Image Priors (FedGIMP) for MRI reconstruction to improve patient privacy, performance, and flexibility in multisite collaboration [43]. Comprehensive experiments on multi-institutional datasets demonstrate the enhanced performance of FedGIMP against both centralized and Federated Learning methods based on conditional models.

Deep learning-based steganalysis methods can not only eliminate the need for professional researchers to manually design feature extraction methods but also take advantage of the end-to-end learning process of deep learning, so that feature extraction and discriminators can be trained simultaneously. Qian et al. enhanced the learning ability of steganalysis models for global statistical information by using traditional steganalysis methods with feature analysis through a transfer learning approach [44], [45]. But migration learning can also lead to a lot of limited effects, not only that, due to the small differences between the carrier image and the densely loaded image. If the embedding rate is lower, the network structure will likely have difficulty converging. To solve this problem, Fran et al. were trained without using the filter kernel of traditional steganalysis as a preprocessing layer and updating the weights in the preprocessing layer during the network training process to take advantage of the powerful fitting ability of deep learning [46]. Notice that, these methods required longer training time and were more likely to overfit.

Besides these weaknesses, previous works also neglect to high-frequency feature recalibrating and discrimination which, as we shall see, are of interest to improve steganographic information detection errors. In this paper, we focus on illegal steganography detection of medical images in IoMT and telemedicine networks and make the following contributions:

- A steganographic information enhancement module(SIEM) has been designed to boost or amplify the steganographic signal. This module combines deep learning and manual approaches to extract horizontal information, vertical information, and diagonal high-frequency information from the image and merge them with the original medical image.
- A steganographic information discriminative mechanism(SIDM) is proposed for the first time. This one automatically obtains the importance of each steganographic signal feature channel to emphasize the high-frequency features of the steganographic signal through the interdependence of the steganographic signal feature channels and suppresses the non-steganographic information such as the natural noise and tissue anatomical structure edges.
- A deep residual network consisting of different blocks for feature enhancement, extraction, compression, and of a classifier combined with the above SIEM and SIDM modules to perform steganalysis within 0.1bpp/bpnzac-0.5bpp/bpnzac payloads in the spatial- and frequency-domain.
- A private dataset of medical images for steganographic analysis is collected. Eight spatial- and frequency-domain steganography methods were considered within 0.1bpp/bpnzac-0.5bpp/bpnzac payloads. Qualitative and quantitative results demonstrate that the dataset is useful for training, validating, and testing deep learning-based steganalysis algorithms for medical images.

The paper is organized as follows. Section II comes back to related works while Section III introduces the principles and advantages of the steganographic information enhancement module and steganographic information discriminative mechanism, presenting the details of our global proposal: RED-Net. Section IV qualitatively and quantitatively evaluate the results of the comparison and ablation experiments. Finally, Section V summarizes the work of this paper.

II. Related works

With the gradual enhancement of steganographic algorithms and the emergence of various steganographic algorithms, the generalized steganalysis model was gradually growing. [47]–[51] were based on steganalysis methods with a manual computational approach to feature extraction. The characteristics of these steganalysis methods were generally calculated by professional researchers relying on their own a priori experience and continuous heuristic attempts. The SRM first trained a classifier by extracting the statistical features of the known images, and then the trained classifier was used to discriminate whether the unknown images contain secrets or not [48]. The maxSRMd2 [49] and PSRM [50] both improved by SRM. Spatial-domain steganalysis detected whether images embedded secret information by analyzing the statistical properties of digital images, while frequency-domain steganalysis discriminated by analyzing the relationship between Discrete Cosine Transform(DCT) coefficients due to different DCT and quantization matrices. Two operations could be optimized at the same time, and it was difficult to reach a heterogeneous equilibrium state. Universal steganalysis detected whether an image contains secret information based on unknown carrier images and steganographic algorithms. With the booming development of deep learning, steganalysis methods based on deep learning methods could not only eliminate the need for professional researchers to manually design the feature extraction method but also took advantage of the end-to-end learning process of deep learning, which made it possible to train the feature extraction and the discriminator at the same time. Deep learning-based models for steganalysis can be mainly categorized into semi-learning models and full-learning models. The semi-learn steganalysis model utilized a fixed filter kernel as a separate preprocessing layer in the steganalysis network, and the internal weight parameters were not involved in backpropagation, while the other network layers were optimized by relying on deep learning methods. Xu et al. employed a 20-layer fully convolutional

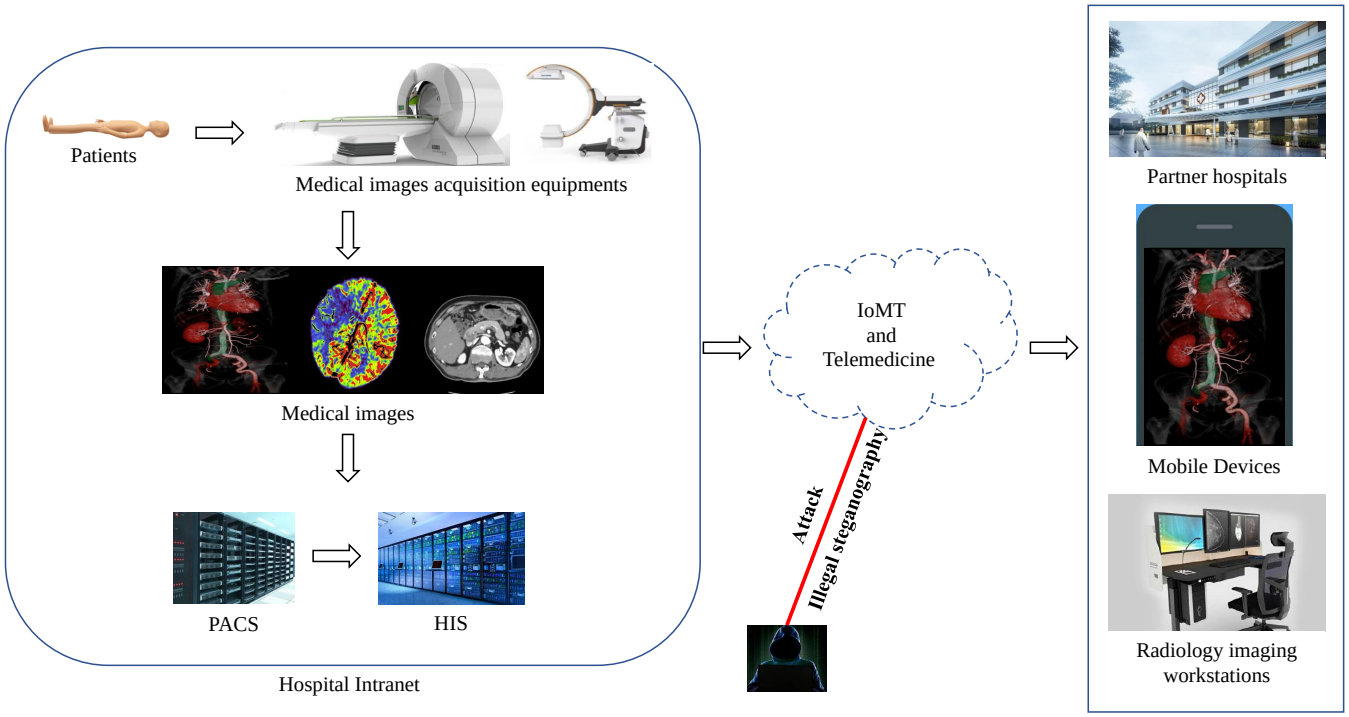

Fig. 1. Medical images transmitted in IoMT and telemedicine suffer from the crisis of illegal steganography.

network to demonstrate that deep learning networks could defeat feature-based steganalysis methods in complex domains, and also demonstrated that deep networks can extract steganographic noise more easily than width networks [52]. Chen et al. [53] also proposed VNet and PNet with JPEG phase sensing based on Xu-Net. PNet and VNet learned the a priori knowledge of DCTR and other frequency domain steganalysis and added a JPEG phase-aware module in the network framework to learn the signal-to-noise ratio information in the frequency domain. The parameters in the preprocessing layer were updated along with the network backpropagation during the training process of the full-learning network [33], [54], [55].The SR-Nert utilized the residual network to simulate the process of traditional SRM in filtering features, which could be applied not only in the spatial domain but also had good results in the JPEG domain. Vit exploited a convolutional visual transformer to capture local and global dependencies between noisy features for spatial domain information steganography. The full-learning models have higher detection accuracy than the traditional steganalysis and half-learning models. However, it takes longer training time and is more prone to overfitting.

In this work, a steganographic information enhancement module based on a deep residual network is exploited to enhance high-frequency information of medical images containing natural noise, organizational boundaries, and steganographic information. A steganographic information discrimination module is employed to recalibrate the amplified information, keeping the steganographic signals and eliminating the mis-boosted fake steganographic information to improve the detection accuracy of RED-Net.

III. Method

As shown in Fig.1, hackers and criminals illegally steganography medical images by hacking into the PACS and HIS in the hospital. When performing remote diagnosis, they can illegally steganography medical images by hacking into IoMT, spreading viruses illegally steganography in medical images to cause crashes in remote diagnosis and IoMT systems. We proposed a RED-Net as shown in detail in Fig.2 consisting of a steganographic information enhancement module, a deep residual network, and a steganographic information discriminative module. The steganographic information enhancement module decomposed the original medical image into horizontal high-frequency information, vertical high-frequency information, and diagonal high-frequency information. The extracted high-frequency information was merged with the convolved original image, and the signal of illegal steganography in medical images is boosted. The images with extracted high-frequency information were fed to a deep residual network with the steganographic information discriminative mechanism for detection.

A. Steganography information enhancement module

A steganographic signal in a medical image containing illegal steganographic information is usually considered to be a high-frequency signal similar to noise. The steganographic information enhancement module (SIEM) is employed to boost high-frequency information in the original illegal steganographic medical images. This high-frequency information is divided into natural noise inherent in the medical image, edges of tissue anatomy, and illegal steganographic information. Inspired by the [56]–[60] , the wavelet transform is adopted in the steganographic information enhancement module to boost the high-frequency signal. For an arbitrary function or signal, the wavelet transform is defined as:

$$W_f(a,b) = \frac{1}{\sqrt{|a|}} \int_R f(x)\bar{\psi}\left(\frac{x-b}{a}\right) dx, \tag{1}$$

where $f(x)$ denotes a illegal steganography medical image, $\bar{\psi}(x)$ is the wavelet mother function. a and b are the scale and translation of the wavelet function. Wavelet series expansion maps a continuously variable function into a sequence of numbers, and if the function to be expanded is a sequence of numbers of sampled values of a continuous function, the resulting coefficients are called the discrete

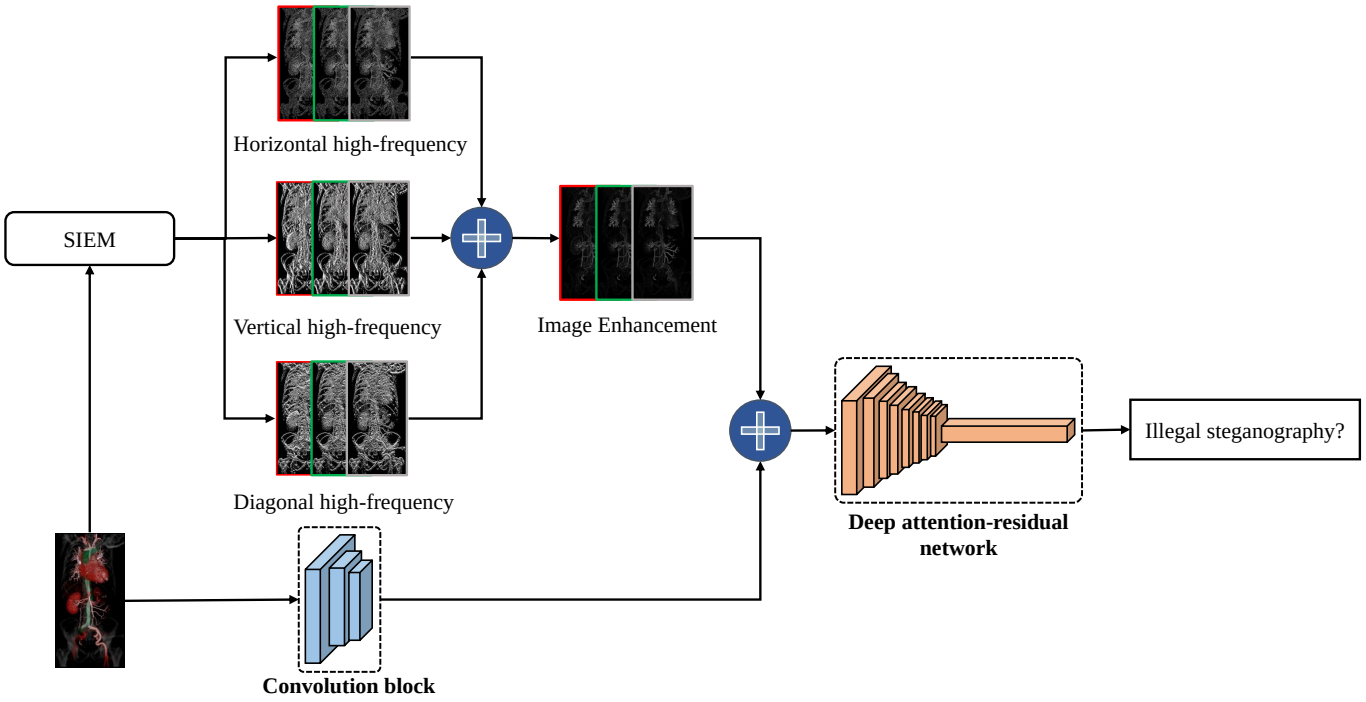

Fig. 2. A scheme for medical image steganography analysis in the IoMT and telemedicine. The extracted horizontal, vertical, and diagonal high-frequency information by the SIEM is merged with the original image and the image with merged high-frequency information is input to a deep residual network with a SIDM.

wavelet transform. The sequence expansion defined by the continuous wavelet transform becomes a DWT transform pair as follows:

$$\begin{array}{l} w_{\varphi}\left(j_{0}, k\right)=\frac{1}{\sqrt{M}} \sum_{x} f(x) \varphi_{j_{0}, k}(x) \\ w_{\psi}(j, k)=\frac{1}{\sqrt{M}} \sum_{x} f(x) \psi_{j, k}(x) \end{array}, \tag{2}$$

for j $\geq j_0$,

$$f(x)=\frac{1}{\sqrt{M}} \sum_{k} w_{\varphi}\left(j_{0}, k\right) \varphi_{j_{0}, k}(x)+\frac{1}{\sqrt{M}} \sum_{j=j_{0}}^{\infty} \sum_{k} w_{\psi}(j, k) \psi_{j, k}(x), \tag{3}$$

where $w_{\varphi}\left(j_{0}, k\right)$ and $w_{\psi}(j, k)$ are the approximation coefficients and wavelet coefficients, respectively. $\varphi_{j_0,k}$ and $\psi_{j,k}(x)$ are the scale function and wavelet function at different scales and locations, respectively. j is the order of the scale, and the larger j is, the smaller the scale, which corresponds to a higher frequency and is closer to the details. k is the offset of the location. The scale function represents the original signal, and as the scale level decreases, the scale becomes larger and larger, and the representation of the original signal becomes more and more inaccurate, we use the wavelet function to represent the difference between the scale function representation part and the original signal. The role of the scale function $\psi(x)$ is to roughly represent the original signal, the scale function by the real numbers, the square can be a product of the function, the scale function for the power of 2 expansion and integer times the translation to get the set of functions $\left\{\varphi_{j,k}(x) \mid j, k \in \mathbb{Z}\right\}$,

$$\varphi_{j,k}(x)=2^{j/2} \varphi\left(2^{j} x-k\right), \tag{4}$$

The wavelet function corrects the difference between the scale function and the original signal. After a given scale function $\varphi(x)$, there must exist a wavelet function $\psi(x)$, similar to the scale function with a power of 2 scalings and an integer multiple of the translation, to obtain the function:

$$\psi_{j,k}(x)=2^{j/2} \psi\left(2^{j} x-k\right), \tag{5}$$

In this paper, the following equation is used for the Haar scale function.

$$\varphi(x)=\begin{cases} 1 & \text{if } 0 \leq x<1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The wavelet function corresponding to the Haar scale function is shown as follows:

$$\psi(x)=\begin{cases} 1 & 0 \leq x<0.5 \\ -1 & 0.5 \leq x<1 \\ 0 & \text{elsewhere} \end{cases} \tag{7}$$

The discrete wavelet transform (DWT) of medical images as shown in Fig.3 is performed with the haar function, and the decomposed horizontal, vertical, and diagonal high-frequency information is merged with the original medical images. The 1D-DWT is first performed on each row of the image to obtain the low-frequency component L and the high-frequency component H of the original image in the horizontal direction, and then 1D-DWT is performed on each column of the transformed data to obtain four different frequency bands, with one approximate component and three detailed components. As shown in Fig.3, the Haar wavelet transform reduces the resolution of the image and does not change the high-frequency information such as the contours of the image. The mean value of the low-frequency information is transformed slowly and the high-frequency difference changes faster, storing detailed information about the noise of the image.

B. Steganography information discriminative mechanism

The steganographic information discriminative mechanism(SIDM) as shown in Fig.4 [61] is combined with the deep residual network to improve the performance of RED-Net. After the deep residual network reduces the dimensionality of the feature map, the SIDM facilitates the modeling of interdependencies between channels and adaptively recorrects the intensity of feature correspondences between channels by the global loss function of the network. The global

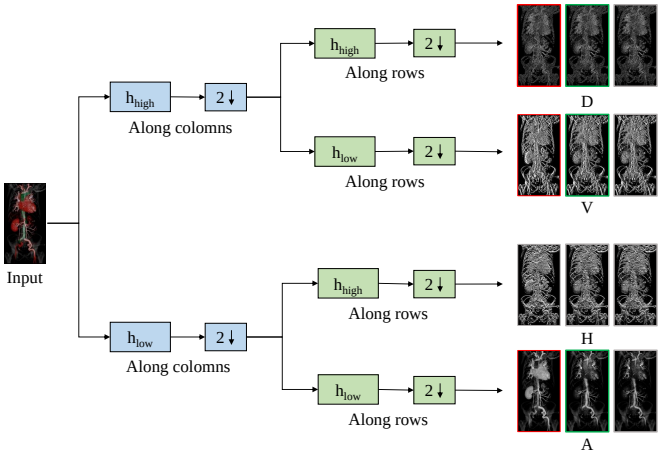

Fig. 3. For two-dimensional medical images Haar wavelet transform performs low-pass and high-pass filtering from both horizontal and vertical directions. After one level of Haar wavelet transform, the low-frequency component A, vertical high-frequency component V, horizontal high-frequency component H, and diagonal high-frequency component D of the original images are obtained.

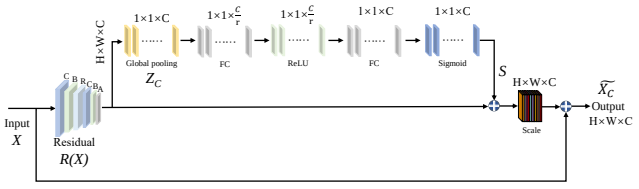

Fig. 4. The architecture of the steganographic information discriminative mechanism introduced in the deep residual network.

average pooling of the feature map after the residual block becomes a $1 \times 1 \times C$ feature vector Z_C, which embeds the global information of each channel. A statistic $\mathbf{Z} \in \mathbb{R}^C$ is generated by $\mathbf{R}_X$ by the spatial dimension $W \times H$, where the c-th element of Z is calculated by:

$$Z_C = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \mathbf{R}_X(i,j), \tag{8}$$

We adopt a sigmoid-activated two-layer fully connected gating mechanism:

$$\mathbf{s} = \sigma(g(\mathbf{z}, \mathbf{W})) = \sigma\left(\mathbf{W}_2 \delta\left(\mathbf{W}_1 \mathbf{z}\right)\right), \tag{9}$$

where δ and σ represent the ReLU function and the Sigmoid function, $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r}} \times C$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$. The final output of the SIDM is obtained by rescaling the transformation output with the activations:

$$\widetilde{X_C} = \mathbf{F}_{\text{scale}}\left(\mathbf{R}_C, s_c\right) = s_c \cdot \mathbf{R}_C, \tag{10}$$

where $\widetilde{X_C}$ is a feature map of a featured channel of $\widetilde{X}$ and s_c is a scalar value in the gating cell vector $\mathbf{s}$. The SIDM relies on the interrelationship between feature channels, and we adopt a new feature-recalibrate strategy. Specifically, the importance of each feature channel is automatically obtained by learning, and then features that are helpful for the current classification task are promoted and features that are not relevant for the current classification task are suppressed based on this level of importance.

C. The deep residual network

The architecture of the deep residual network introducing the SIEM and the SIDM for medical image steganalysis is shown in Fig.5. The deep residual network consists of a feature enhancement block, a feature extraction block, a feature compression block, and a classifier. The detail of the network is listed in Tab.I. The steganographic signal is the high-frequency noise existing in the steganographic medical image. The goal of SIEM and instance-norm layers is to reinforce the noise-like steganographic signals and suppress image content. The proposed RED-Net has taken a deep residual network (DS-Net) as the base model and then has introduced a steganography information enhancement module(SIEM) and a steganography information discriminative mechanism(SIDM) to enhance its performance in steganalysis. The subsequently introduced SIEM and SIDM are refined on the noise extraction and feature compression of DS-Net. The first 7 components of DS-Net mainly consist of a convolutional layer, a BatchNorm layer, and an activation layer in series. The average pooling operation is additionally introduced in the middle 3 components of DS-Net. The difference is because the pooling operation strengthens the content and suppresses noise-like steganographic signals by averaging adjacent embedding variations, which is detrimental to steganography. Average pooling is used by the middle component to reduce the feature dimensions for feature compression. The feature map after information fusion is processed by a 3×3 convolution kernel with step size 1 and padding 1 and then activated by the ReLU function after batchNorm. Two convolutional layers with a 3×3 convolutional kernel of step 1 and padding 1 are employed during the feature extraction phase. The shortcut connections help propagate gradients to the upper layers, which are the hardest to train because of the vanishing gradient phenomenon. The SIDM squeezes the feature map obtained by convolution to obtain the global features at the channel level, then performs the excitation operation on the global features to learn the relationship between the channels, while obtaining the weights of different channels. Essentially, the SIDM works on the channel dimension by doing the gating operation, which allows the model to focus more on the most informative channel features and suppress those unimportant channel features. The feature compression block is dedicated to reducing the dimensionality of the feature map. Pooling in the form of 3 × 3 averaging with stride 2 is applied to feature compression. 512 feature maps of dimension 16×16 are reduced to a 512-dimensional feature vector by computing statistical moments of each 16×16 feature map. This 512-dimensional output enters the classifier part of the network.

The loss function in the proposed network is designed as follows:

$$\text{softmax}\left(x^{(i)}\right) = \begin{bmatrix} p\left(y^{(i)}=0 \mid x^{(i)}\right) \\ p\left(y^{(i)}=1 \mid x^{(i)}\right) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{x_j^{(i)}}} \begin{bmatrix} e^{x_1^{(i)}} \\ e^{x_2^{(i)}} \\ \vdots \\ e^{x_k^{(i)}} \end{bmatrix}, \tag{11}$$

where $x^{(i)}$ is the $i-th$ input and is a k-dimensional vector, $y^{(i)}$ is the result of the $i-th$ time. 0 is not steganography, 1 is steganography. $x_j^{(i)}$ represents the value of the $j-th$ dimension of $y_j^{(i)}$. The proposed network is essentially designed to solve a binary classification problem, so $k = 2$.

The detection performance was measured with the total classification error probability on the testing set under equal prior.

$$P_{\text{E}} = \min_{P_{\text{FA}}} \frac{1}{2}\left(P_{\text{FA}} + P_{\text{MD}}\right), \tag{12}$$

where P_{FA} and P_{MD} are the false-alarm and missed-detection probabilities. The false-alarm probabilities P_{FA} represents the ratio of the number of cover images that are misclassified as stego images to the total number of cover images, and can be given by the following

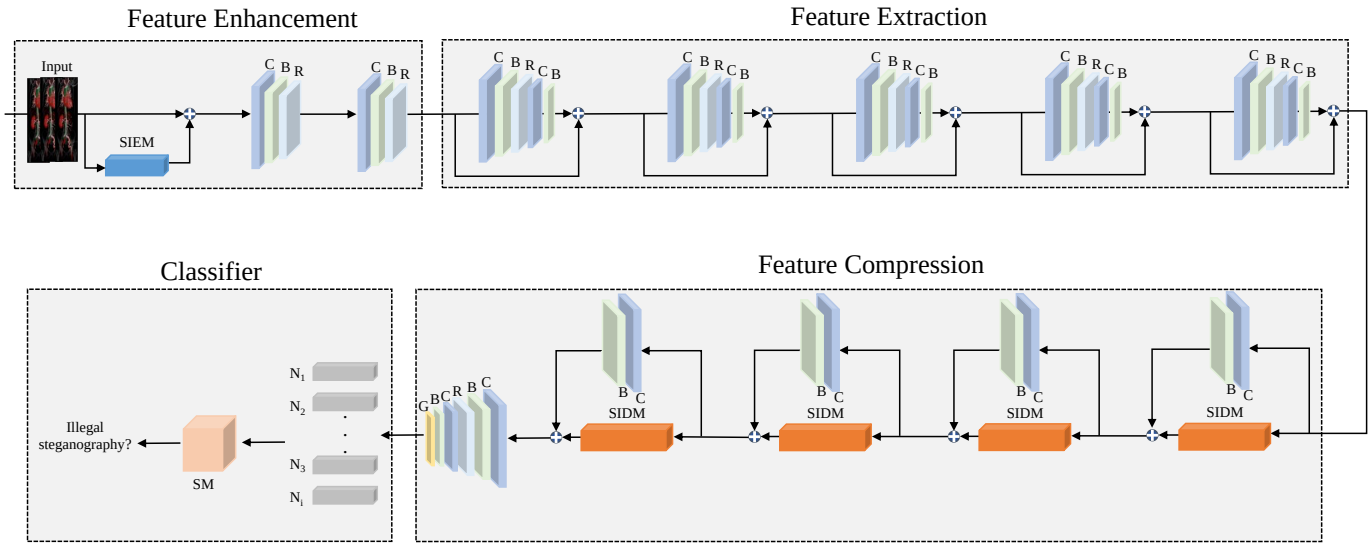

Fig. 5. The architecture of the deep residual network introducing the SIEM and the SIDM for steganalysis.

TABLE I
THE DETAIL OF THE RED-NET ARCHITECTURE

Part	Input → Output shape	Layer Information
Feature Enhancement	(256,256,1) → (256,256,1)	SIEM,IN
	(256,256,1) → (256,256,64)	Conv-(N64,K3,S1,P1),BN,ReLu
	(256,256,64) → (256,256,16)	Conv-(N16,K3,S1,P1),BN,ReLu
Feature Extraction	(256,256,16) → (256,256,16)	Conv-(N16,K3,S1),BN,ReLu,Conv-(N16,K3,S1),BN
	(256,256,16) → (256,256,16)	Conv-(N16,K3,S1),BN,ReLu,Conv-(N16,K3,S1),BN
	(256,256,16) → (256,256,16)	Conv-(N16,K3,S1),BN,ReLu,Conv-(N16,K3,S1),BN
	(256,256,16) → (256,256,16)	Conv-(N16,K3,S1),BN,ReLu,Conv-(N16,K3,S1),BN
	(256,256,16) → (256,256,16)	Conv-(N16,K3,S1),BN,ReLu,Conv-(N16,K3,S1),BN
Feature Compression	(256,256,16) → (128,128,16)	SIDM,Conv-(N16,K1,S2,P1),BN
	(128,128,16) → (64,64,64)	SIDM,Conv-(N64,K1,S2,P1),BN
	(64,64,64) → (32,32,128)	SIDM,Conv-(N128,K1,S2,P1),BN
	(32,32,128) → (16,16,256)	SIDM, Conv-(N256,K1,S2,P1),BN
	(16,16,256) → 512	Conv-(N512,K3,S1,P1),BN,ReLu,Conv(N512,K3,S1,P1),BN,Global Average Pooling
Classifier	512 → 2	Fully Connection-(N512),Softmax

equation:

$$P_{\mathrm{FA}} = \frac{FP}{FP + TN}, \tag{13}$$

where TN(True Negative) indicates the number of images that detected the cover as a cover correctly, and FP(False Positive) indicates the number of images that detected the cover as a stego incorrectly. The missed-detection probabilities P_{MD} represents the ratio of the number of stego images that are misclassified as cover images to the total number of stego images, and can be given by the following equation:

$$P_{\mathrm{MD}} = \frac{FN}{TP + FN}, \tag{14}$$

where FN(False Negative) indicates the number of images that detected the stego as a cover incorrectly, and TP(True Positive) indicates the number of images that detected the stego as a stego correctly.

IV. Analysis and Evaluation of Experimental Results

A. Datasets and experiment environment

A private dataset is collected to train and validate our proposed method. The dataset contains 1639 digital medical images with the size of 256×256 from the Nanjing First Hospital, China, with the approval of the Institutional Review Board and patient consent forms. We increased the number of images to 6556 by using data enhancement with rotation of 90°, 180°, and 270°. Five state-of-the-art embedding methods [62]–[66] were performed on 6556 images of illegal information steganographically. These five embedding methods contained both spatial domain steganography and frequency domain steganography. 2×28280 cover and stego images were used as the training set, 2×2000 as the validation set, and 2×2500 as the test set. The public dataset containing 150,000 medical images is employed to demonstrate the RED-Net's superior performance. The dataset is provided by the US National Institutes of Health and all images are from 30,805 unique patients with an original image size of 1024×1024. The same 5 illegal state-of-the-art embedding methods and the same dataset generation methods are implemented for the private dataset. To sum up, 2×125000 cover and stego images were used as the training set, 2×20000 as the validation set, and 2×5000 as the test set. We first counted the total number of patients in the public dataset and the private dataset, and then divided these two datasets into training set, validation set and test set according to the ratio of 7:2:1 with the number of patients as the basic unit. Then, the data of the trained set was expanded, and the data before and after enhancement were used as the total training data, and the validation and test sets were operated in the same way.

All deep learning-based algorithms were implemented using the

Pytorch framework. The batch normalization parameter is learned by an exponential moving average with a decay rate of 0.9. The weights of the fully connected layer are initialized using an unbiased Gaussian model with a mean of 0 and a standard deviation of 0.01. The networks were trained and tested on a computer with configurations: CPU is Intel Core i9-9900KF @ 3.60GHz; GPU is NVIDIA RTX 3090 with 24 GB memory. The learning rate was initially set to 0.001 and then decayed to 0.0001 as the network was trained, with a batch size of 16. The Adamax optimizer with default settings was used. The proposed network was trained with a total of 457 epochs and 400,000 iterations.

B. Comparison experiment

1) Analysis of spatial domain comparison experiment results: On both public and private datasets, we compare the performance of the proposed the RED-Net with the three advanced steganography detectors, which are maxSRMd2 [49], PSRM [50], SRNet [33], and Vit [54] in the spatial domain under the payload range of 0.1bpp to 0.5bpp. Tab.II and Tab.III show the performance of comparison methods in the spatial domain in terms of P_{FA}, P_{MD}, and P_{E}. From Tab.II, we can observe that the proposed RED-Net achieves the best P_{E} tested on the public dataset compared to other spatial domain steganalysis algorithms. The PSRM algorithm has the worst detection error on the public and private datasets with 0.1383 and 0.4647, respectively, at the lowest embedding rate of 0.1bpp. The difference in detection error between the RED-Net and PSRM algorithm at 0.1bpp is the most prominent between the RED-Net and other spatial domain steganography detectors in terms of other embedding rates. Although Vit achieved the best P_{FA} at 0.1bpp, 0.2bpp, and 0.5bpp, the P_{MD} at this moment is hundreds of times more accurate compared to the maximum difference of RED-Net. According to Eq.12 for P_{E}, RED-Net's excellent performance in P_{MD} made up the gap with Vit in P_{FA}. The same results are listed in TableIII. In Tab.III, the PSRM algorithm achieves the lowest P_{MD} at 0.01 bpp, but P_{E} is the highest result of all the spatial comparison methods tested on the private dataset. The reason for this phenomenon is that the PSRM steganography detector has the highest P_{FA} at 0.1 bpp, and the proportion of negative samples predicted to be positive to the total negative samples is the largest of all algorithms. As far as P_{MD} was concerned, PSRM got the best P_{MD} at 0.1bpp and 0.4bpp, but at this time the P_{FA} was as high as 90.53% and 18.80%. The P_{MD} of RED-Net at 0.1bpp and 0.4bpp were 18% and 3.76%, which were much smaller than the values of PSRM. The detection error P_{E} we designed takes into account the combined effect of P_{FA} and P_{MD} on the steganography detector, not one of them alone. The same phenomenon happened with Vit, where RED-Net's P_{FA} at 0.1bpp was nearly 1.8 times that of Vit, but the P_{MD} was a quarter of that of Vit. Vit and RED-Net achieved about the same P_{FA} at 0.3bpp, but Vit had more than 3 times the P_{MD} of RED-Net. In general, the test results on the public dataset appear to be significantly better than those on the private dataset. This is because the amount of data in the public dataset greatly exceeds that in the private dataset, and the performance of the RED-Net is fully developed as it is fully trained.

In Fig.6, we further illustrate the detection error of the steganography detector in the spatial domain in the comparison experiment. As can be seen in Fig.6, the deep learning-based methods, SRNet and the RED-Net, outperform the traditional maxSRM and PSRM algorithms, which are based on extracting manual features, in terms of detection error at each payload rate. The reason is that the SRM algorithm and maxSRM algorithm only consider the correlation of pixel points and their neighboring position pixels, without considering the texture features of the image and without predicting the unknown at the time of steganography, so there is a certain decrease in the accuracy rate when analyzing the adaptive steganography method. The detection errors of SRNet and the RED-Net are very close at 0.2bpp to 0.4bpp payload on both public and private datasets. The proposed RED-Net and SRNet perform feature extraction and feature compression for steganographic noise by employing a residual connection. Since image steganography is the embedding of a binary bit stream into the spatial or JPEG domain of a carrier image, it is the same as adding a weak noise to the carrier image to generate a dense image. Thus superimposing noise into it changes the correlation of neighboring pixels in the original image, as well as the residual image, so using the residual image as the substrate for feature extraction reduces the impact of image content on feature extraction for steganalysis.

2) Analysis of JEPG domain comparison experiment results: For the JPEG domain, PNet [53], VNet [53], DCTR [51], and UCNet [55] for payloads 0.1-0.5bpnzac(bit per non-zero AC DCT coefficient) are tested for quality factors 75. The results of the JEPG domain comparison experiment are listed in Tab.IV and Tab.V.

From Tab.V, it can be observed that the DCTR steganalysis algorithm achieves the best P_{MD} at 0.1bpnzac for the private dataset. But the P_{FA} is the worst with 94.60%, nearly four times the 28% achieved by RED-Net at this time. The P_{FA} of UCNet at 0.2bpnzac and 0.4bpnzac was 11.68% and 2.80%, and the P_{MD} was 28.44% and 17.12%. The P_{MD} of RED-Net at 0.2bpnzac and 0.4bpnzac was 14.32% and 3.76%, and the P_{MD} was 13.32% and 5.12%. Although UCNet performs better than RED-Net in terms of P_{FA}, it can be seen from Eq.12 that the performance of the model is not determined by P_{FA} alone but also depends on P_{MD}, and UCNet's performance in P_{MD} is much worse than that of RED-Net. The RED-Net has the largest improvement in detection error of 80% at 0.2bpnzac. The optimal P_{E} is achieved by the RED-Net on both public and private datasets. In Fig.7, we graphically illustrate the detection error of the steganalysis algorithm in the frequency domain comparison test at 0.1 bpnzac to 0.5 bpnzac. We can observe from Fig.7 that the deep learning-based steganalysis methods outperform the traditional steganalysis algorithm DCTR on both public and private datasets. PNet and VNet learn the a priori knowledge of frequency domain steganalysis such as DCTR and incorporate the JPEG phase sensing module in the network framework to improve the steganography detection accuracy. The JPEG phase sensing module is incorporated into the network framework to learn the signal-to-noise ratio information in the frequency domain to improve steganography detection accuracy. Compared to PNet, the architecture of VNet is smaller and the detection error on the public dataset is worse than that of PNet. This is because VNet has introduced three additional filter kernels as a fixed preprocessing layer to learn some steganographic noise with directional characteristics, and the filter kernels act as a catalyst in the preprocessing layer. The PNet and VNet depend on a fixed DCT kernel and a threshold setting for the feature set. The RED-Net adopts a skip connection structure in the network to prevent too-deep convolution layers from causing gradient dispersion or gradient explosion in the network during training. The RED-Net improves the quality of representation of high-frequency steganographic signal features by modeling the interdependencies between the network evolution feature channels through the SE attention mechanism, based on the wavelet transform preprocessing layer. For this purpose, the network is allowed to perform feature recalibration of steganographic signals. Through this mechanism, it learns to use global information to selectively emphasize valid high-frequency true steganographic informative features and suppress less reliable false steganographic features. To further evaluate the performance, the receiver operating characteristic (ROC) curves and the corresponding area under the

TABLE II
RESULTS OF COMPARISON EXPERIMENTS IN THE SPATIAL DOMAIN ON THE PUBLIC DATASET

Model	Evaluation index	0.1bpp	0.2bpp	0.3bpp	0.4bpp	0.5bpp
maxSRMd2	P_{FA}	18.67%	5.26%	2.40%	1.13%	0.83%
	P_{MD}	7.53%	3.74%	0.60%	0.2%	0.04%
	P_{E}	0.1160	0.0163	0.0048	0.0059	0.0042
PSRM	P_{FA}	14.86%	6.53%	5.93%	2.40%	2.27%
	P_{MD}	12.80%	4.80%	0.13%	0.03%	0.01%
	P_{E}	0.1383	0.0567	0.0303	0.0032	0.0114
SRNet	P_{FA}	10.63%	5.37%	0.43%	0.49%	0.38%
	P_{MD}	8.36%	2.68%	0.78%	0.34%	0.11%
	P_{E}	0.0961	0.0412	0.0068	0.0041	0.023
Vit	P_{FA}	**7.66%**	**1.06%**	6.50%	4.84%	**0.08%**
	P_{MD}	18.62%	12.88%	**0.06%**	**0.04%**	4.68%
	P_{E}	0.1314	0.0697	0.0328	0.0244	0.0238
RED-Net	P_{FA}	8.88%	2.12%	**0.3%**	**0.32%**	0.16%
	P_{MD}	**5.76%**	**0.96%**	0.6%	0.2%	**0.04%**
	P_{E}	**0.0732**	**0.0154**	**0.0045**	**0.0025**	**0.0010**

TABLE III
RESULTS OF COMPARISON EXPERIMENTS IN THE SPATIAL DOMAIN ON THE PRIVATE DATASET

Model	Evaluation index	0.1bpp	0.2bpp	0.3bpp	0.4bpp	0.5bpp
maxSRMd2	P_{FA}	84.93%	23.20%	17.06%	8.67%	4.13%
	P_{MD}	5.73%	25.33%	12.40%	11.07%	3.73%
	P_{E}	0.4533	0.2424	0.1473	0.0987	0.0393
PSRM	P_{FA}	90.53%	29.06%	17.06%	18.80%	5.20%
	P_{MD}	**4.80%**	20.00%	16.01%	**3.87%**	3.94%
	P_{E}	0.4767	0.2453	0.1653	0.1134	0.0457
SRNet	P_{FA}	20.32%	16.68%	11.61%	6.34%	5.39%
	P_{MD}	31.67%	15.20%	11.92%	8.16%	4.21%
	P_{E}	0.459	0.1461	0.0094	0.053	0.032
Vit	P_{FA}	**10.48%**	15.60%	26.80%	4.08%	3.64%
	P_{MD}	82.76%	28.60%	**7.24%**	18.68%	5.52%
	P_{E}	0.4662	0.221	0.1702	0.1138	0.0458
RED-Net	P_{FA}	18.2%	**14.32%**	**8.32%**	**3.76%**	**3.40%**
	P_{MD}	28%	**13.32%**	8.76%	5.12%	**1.92%**
	P_{E}	**0.231**	**0.138**	**0.085**	**0.044**	**0.026**

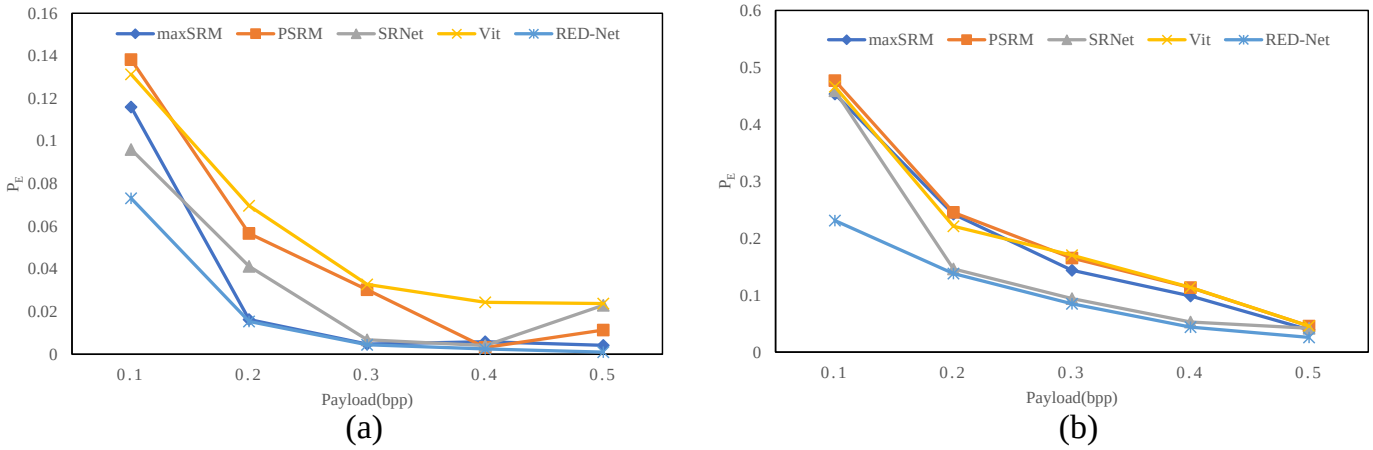

Fig. 6. Detection error of spatial domain steganography detector at the payload of 0.1bpp to 0.5bpp.(a)Detection errors of the maxSRM, PSRM, SR-Net, Vit, and RED-Net on the public dataset. (b)Detection errors of the maxSRM, PSRM, SR-Net, Vit, and the RED-Net on the private dataset.

curve shown in Fig.8 and Fig.9 were used. Overall, RED-Net has a competitive performance compared to other deep learning-based steganographic detectors.

C. Ablation study

In this section, an ablation study is conducted to investigate the proposed SIEM and SIDM of the RED-Net. First, a deep residual network(DS-Net) without a SIEM and a SIDM is considered a baseline model. Then, the baseline model employs a SIEM to enhance steganographic signal extraction. Further, the RED-Net, which introduces both a SIEM and a SIDM, is employed for steganalysis.

The qualitative results tested on the public dataset and private dataset are listed in Tab.VI and Tab.VII. As shown in Tab.VI, the RED-Net, which introduces both the SIEM and the SIDM, achieves the best detection error P_{E} on the public dataset. The baseline model DS-Net achieved the lowest Pmd at 0.2 bpp, but at the same time, it resulted in the highest P_{FA}. As shown in Tab.VI, the RED-Net achieves satisfactory P_{MD} at relatively large embedding rates. As the

TABLE IV
RESULTS OF COMPARISON EXPERIMENTS IN JEPG DOMAIN ON THE PUBLIC DATASET

Model	Evaluation index	0.1bpnzac	0.2bpnzac	0.3bpnzac	0.4bpnzac	0.5bpnzac
DCTR	P_{FA}	35.32%	20.30%	15.64%	15.33%	15.94%
	P_{MD}	73.44%	12.46%	4.36%	8.36%	17.25%
	P_{E}	0.2133	0.1638	0.0995	0.0763	0.0812
PNet	P_{FA}	13.30%	13.04%	6.22%	1.86%	4.26%
	P_{MD}	22.48%	12.62%	7.66%	9.24%	4.76%
	P_{E}	0.1789	0.1282	0.0694	0.0537	0.0451
VNet	P_{FA}	20.56%	12.28%	12.18%	6.84%	6.21%
	P_{MD}	18.60%	17.14%	5.68%	4.66%	4.58%
	P_{E}	0.1958	0.1471	0.0893	0.0575	0.0529
UCNet	P_{FA}	11.64%	3.36%	0.60%	1.12%	0.36%
	P_{MD}	8.86%	3.40%	2.48%	0.92%	0.90%
	P_{E}	0.1025	0.0338	0.0154	0.0102	0.0063
RED-Net	P_{FA}	**8.88%**	**2.12%**	**0.3%**	**0.32%**	**0.16%**
	P_{MD}	**5.76%**	**0.96%**	**0.6%**	**0.2%**	**0.04%**
	P_{E}	**0.0732**	**0.0154**	**0.0045**	**0.0025**	**0.0010**

TABLE V
RESULTS OF COMPARISON EXPERIMENTS IN JEPG DOMAIN ON THE PRIVATE DATASET

Model	Evaluation index	0.1bpnzac	0.2bpnzac	0.3bpnzac	0.4bpnzac	0.5bpnzac
DCTR	P_{FA}	94.60%	54.32%	59.31%	43.73%	40.34%
	P_{MD}	**8.00%**	37.08%	45.39%	35.68%	39.14%
	P_{E}	0.4996	0.4583	0.3927	0.3692	0.391
PNet	P_{FA}	23.92%	47.36%	10.88%	22.96%	25.04%
	P_{MD}	69.88%	29.08%	55.84%	28.12%	15.12%
	P_{E}	0.4690	0.3822	0.3336	0.2554	0.2008
VNet	P_{FA}	50.8%	30.84%	19.52%	16.48%	13.32%
	P_{MD}	45.28%	32.20%	25.72%	12.28%	11.64%
	P_{E}	0.4804	0.3152	0.2262	0.1438	0.1248
UCNet	P_{FA}	78.64%	**11.68%**	14.40%	**2.80%**	3.64%
	P_{MD}	17.24	28.44%	13.20%	17.12%	5.52%
	P_{E}	0.4794	0.2006	0.138	0.0996	0.0458
RED-Net	P_{FA}	**18.2%**	14.32%	**8.32%**	3.76%	**3.40%**
	P_{MD}	28%	**13.32%**	**8.76%**	**5.12%**	**1.92%**
	P_{E}	**0.231**	**0.138**	**0.085**	**0.044**	**0.026**

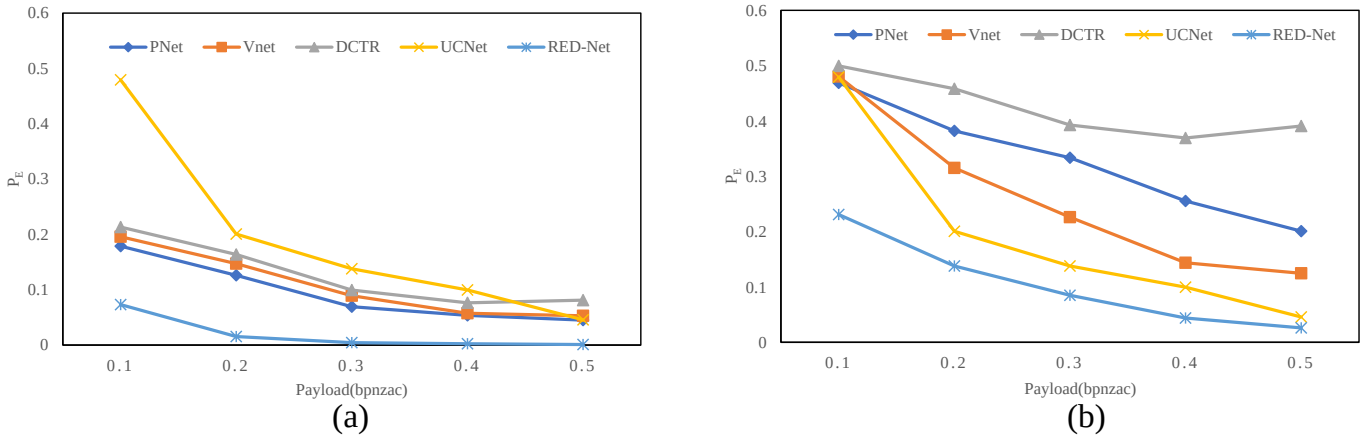

Fig. 7. Detection error of JEPG domain steganography detector at the payload of 0.1bpnzac to 0.5bpnzac.(a)Detection errors of the DCTR, PNet, VNet, UCNet, and the RED-Net on the public dataset. (b)Detection errors of the DCTR, PNet, VNet, UCNet, and the RED-Net on the private dataset.

embedding rate decreases, the P_{MD} of the RED-Net gets worse, and consequently, the P_{FD} slowly gets better. The increasingly improved P_{FA} metrics can compensate for the continuously deteriorating P_{MD}, so DS-Net+SIEM+SIDM can achieve the best detection error P_{E}. As shown in Tab.VII, DS-Net+SIEM+SIDM achieved the best detection error P_{E} in the range of 0.1bpp-0.5bpp on the private dataset. Different from the public dataset, the P_{MD} of DS-Net+SIEM+SIDM becomes worse as the embedding rate decreases, but the P_{FA} becomes better as the embedding rate decreases. At the worst P_{MD} of 0.1 bpp, the P_{FA} of DS-Net+SIEM+SIDM is the lowest. The trend of getting better P_{FA} counteracts the effect of increasingly poor P_{MD} on detection error. As can be observed in Fig.10, the P_{E} of each method in the ablation experiment decreases as the embedding rate increases. This is because the larger the capacity of the illegal information embedded in the carrier, the easier it is to be detected by the steganography detector. The baseline network DS-Net has the worst detection error P_{E} both on the public and private datasets. With the introduction of SIEM alone, the original medical image and its high-frequency information are superimposed and enhanced, reinforcing all the high-frequency information in the

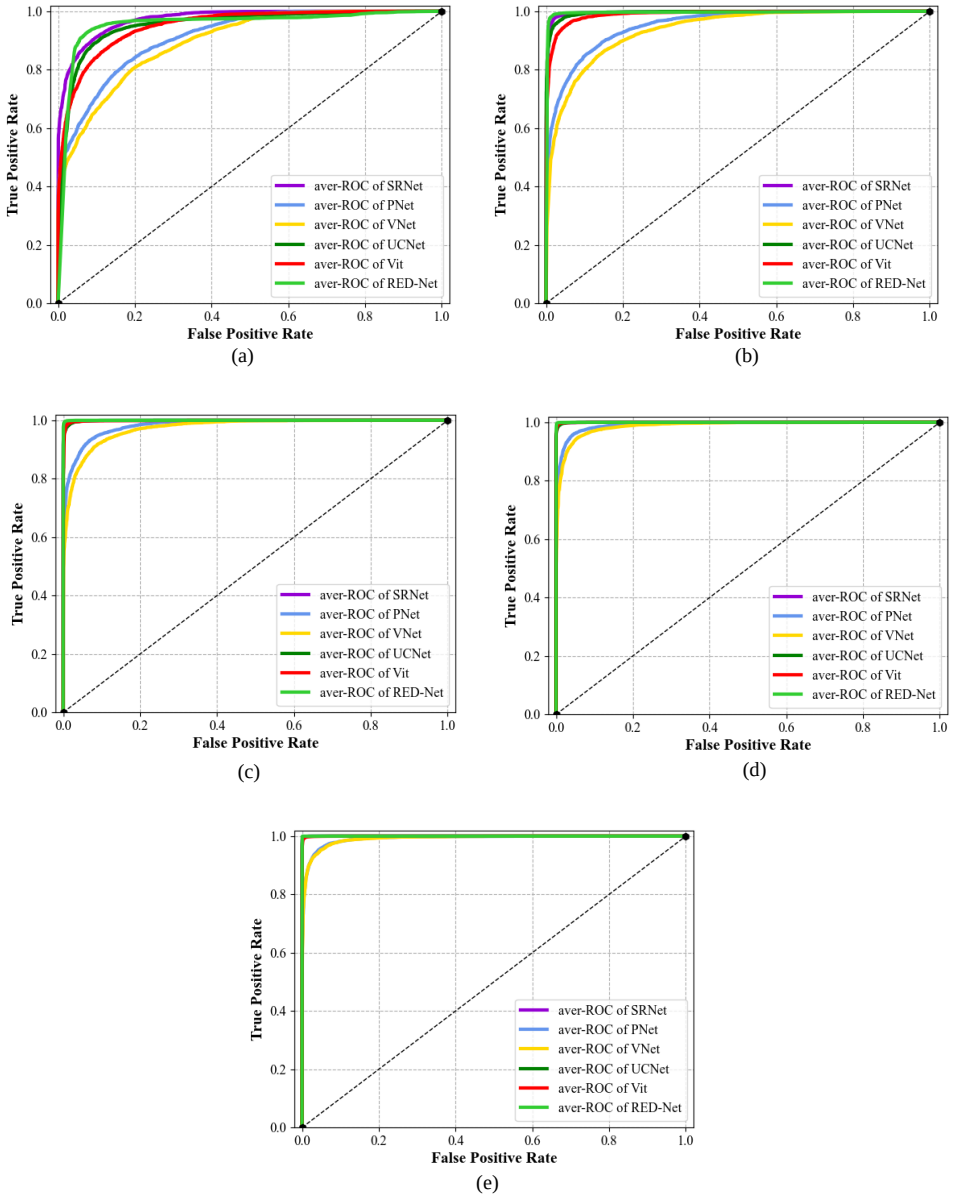

Fig. 8. Comparison of test accuracies of RED-Net with other deep learning-based methods on the public datasets. (a) Test accuracy of RED-Net with SRNet, PNet, VNet, UCNet, and Vit at 0.1bpp/bpnzac payload. (b) Test accuracy of RED-Net versus SRNet, PNet, VNet, UCNet, and Vit at 0.2bpp/bpnzac payload. (c) Test accuracy of RED-Net with SRNet, PNet, VNet, UCNet, and Vit at 0.3bpp/bpnzac payload. (d) Test accuracy of RED-Net with SRNet, PNet, VNet, UCNet, and Vit at 0.4bpp/bpnzac payload. (e) Test accuracy of RED-Net with SRNet, PNet, VNet, UCNet, and Vit at 0.5bpp/bpnzac payload.

TABLE VI

RESULTS OF ABLATION EXPERIMENTS ON THE PUBLIC DATASET

Model	Evaluation index	0.1bpp	0.2bpp	0.3bpp	0.4bpp	0.5bpp
DS-Net	P_{FA}	12.06%	7.14%	1.32%	1.2%	0.84%
	P_{MD}	7.68%	**0.4%**	1.02%	0.24%	0.12%
	P_{E}	0.0987	0.0377	0.0117	0.0072	0.0048
DS-Net+SIEM	P_{FA}	10.98%	4.14%	1.4%	**0.06%**	**0.14%**
	P_{MD}	7.94%	0.78%	**0.26%**	1.02%	0.18%
	P_{E}	0.0946	0.0246	0.0083	0.0054	0.0016
DS-Net+SIDM	P_{FA}	12.66%	5.02%	1.08%	0.46%	0.32%
	P_{MD}	**4.72%**	0.92%	0.3%	0.44%	0.06%
	P_{E}	0.0869	0.0297	0.0069	0.0045	0.0019
DS-Net+SIEM+SIDM(RED-Net)	P_{FA}	**8.88%**	**2.12%**	**0.3%**	0.32%	0.16%
	P_{MD}	5.76%	0.96%	0.6%	**0.2%**	**0.04%**
	P_{E}	**0.0732**	**0.0154**	**0.0045**	**0.0025**	**0.0010**

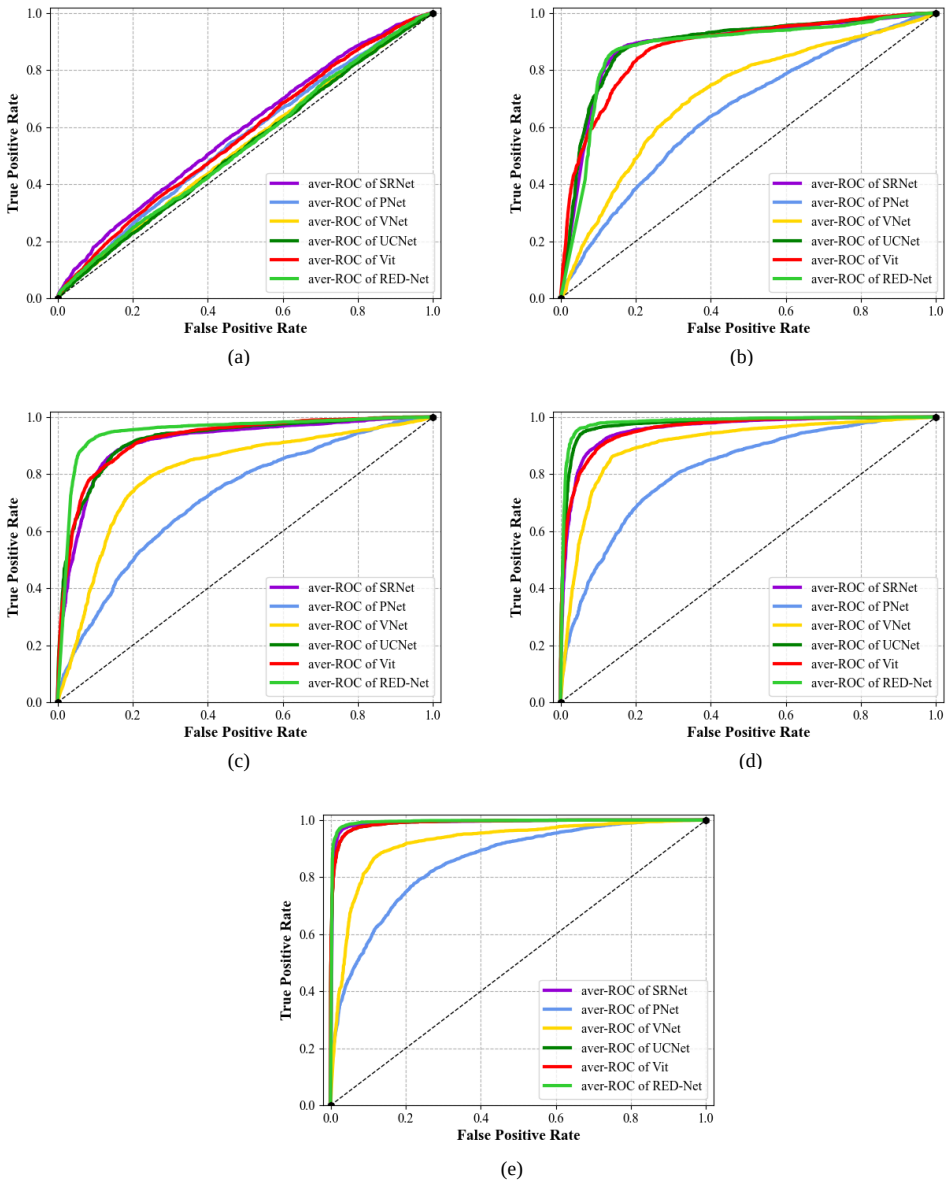

Fig. 9. Comparison of test accuracies of RED-Net with other deep learning-based methods on the private dataset. (a) Test accuracy of RED-Net with SRNet, PNet, VNet, UCNet, and Vit at 0.1bpp/bpnzac payload. (b) Test accuracy of RED-Net versus SRNet, PNet, VNet, UCNet, and Vit at 0.2bpp/bpnzac payload. (c) Test accuracy of RED-Net with SRNet, PNet, VNet, UCNet, and Vit at 0.3bpp/bpnzac payload. (d) Test accuracy of RED-Net with SRNet, PNet, VNet, UCNet, and Vit at 0.4bpp/bpnzac payload. (e) Test accuracy of RED-Net with SRNet, PNet, VNet, UCNet, and Vit at 0.5bpp/bpnzac payload.

TABLE VII
RESULTS OF ABLATION EXPERIMENTS ON PRIVATE DATASET

Model	Evaluation index	0.1bpp	0.2bpp	0.3bpp	0.4bpp	0.5bpp
DS-Net	P_{FA}	35.76%	**10.8%**	7.24%	12.4%	**2.80%**
	P_{MD}	54.24%	27.6%	20.3%	6.92%	3.68%
	P_{E}	0.450	0.193	0.138	0.096	0.033
DS-Net+SIEM	P_{FA}	51.84%	12.88%	8.68%	4.48%	2.96%
	P_{MD}	37.96%	14.76%	9.06%	7.32%	3.28%
	P_{E}	0.449	0.139	0.088	0.059	0.032
DS-Net+SIDM	P_{FA}	32.88%	14.24%	**6.4%**	5.8%	3.04%
	P_{MD}	**27.48%**	15.44%	9.76%	6.44%	3.20%
	P_{E}	0.302	0.149	0.083	0.060	0.032
DS-Net+SIEM+SIDM(RED-Net)	P_{FA}	**18.2%**	14.32%	8.32%	**3.76%**	3.40%
	P_{MD}	28%	**13.32%**	**8.76%**	**5.12%**	**1.92%**
	P_{E}	**0.231**	**0.138**	**0.081**	**0.044**	**0.026**

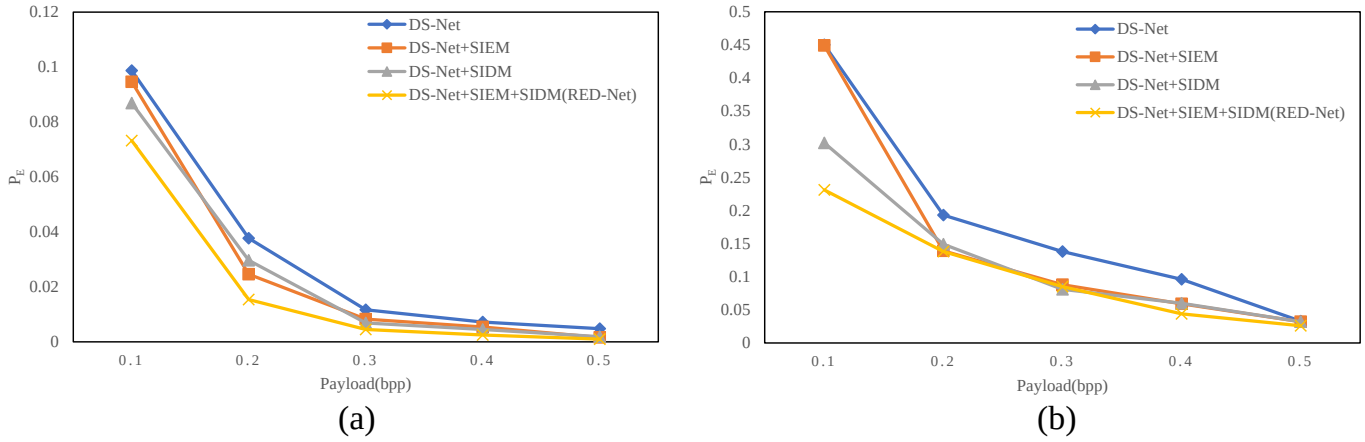

Fig. 10. Detection error of steganography detector in ablation experiment method at the payload of 0.1 to 0.5 bpp. (a)Detection errors of the DS-Net, DS-Net+SIEM, DS-Net+SIDM, and DS-Net+SIEM+SIDM(RED-Net) on the public dataset. (b)Detection errors of the DS-Net, DS-Net+SIEM, DS-Net+SIDM, and DS-Net+SIEM+SIDM(RED-Net) on the private dataset.

medical image. As can be seen in Fig.10, the steganographic signal enhancement capability of SIEM is more effective than SIDM in improving the RED-Net steganalysis performance for relatively small payloads. As the payload is increasing, the high-frequency information of medical images containing illegal steganographic information becomes more and more prominent. The high-frequency signals that include steganographic information, natural noise in medical images, and edges in tissue anatomy are increasingly detected. In the load 0.2bpp-0.5bpp, the SIDM can boost the high-frequency features of the steganographic signal by the interdependence between the channel steganographic signal channels, suppress and remove the edges of the natural medical image noise and tissue anatomical structures that RED-Net misidentified as steganographic signals, and perform the steganographic high-frequency feature recalibration.

TABLE VIII
MODEL COMPLEXITY AND COMPUTATIONAL TIME OF DIFFERENT DEEP LEARNING-BASED METHODS(G: GIGA, M: MILLION, S: SECOND)

Method	FLOPs(G)	Parameter(M)	Time(S)
SRNet	193.3596	4.779618	20.8290
PNet	12.3069	0.031882	41.4546
VNet	13.4528	0.302698	40.0894
Vit	230.1247	1.117296	21.8836
UCNet	106.9785	51.434498	16.6707
RED-Net	243.0435	4.790895	16.8874

D. Algorithmic Complexity

Three universally used metrics, floating-point operations(FLOPs), number of parameters, and throughput, were adopted to compare the complexity of different DL-based steganalysis algorithms. The results listed in Tab.VIII were calculated on the public dataset. PNet and VNet are the two methods with lower model complexity among all deep learning-based methods. Because PNet and VNet are semi-learned steganalysis models based on wide networks, they utilize a fixed filter kernel as an independent preprocessing layer, and the internal weight parameters do not participate in backpropagation, while the other network layers rely on deep learning methods for optimization. They also take the longest time to process the data due to the presence of fixed complex filter kernels. Among the remaining four fully-learned steganalysis models, the networks of UCNet and RED-Net are more complex than those of SRNet and Vit. The main reason for this is that RED-Net introduces SIEM and SIDM on a basic residual network similar to SRNet. Compared to UCNet, RED-Net has a much smaller number of network parameters, and the test time is essentially the same as its. In terms of steganography performance, the detection performance of UCNet is inferior to that of RED-Net, which shows that blindly adding the number of network layers and parameters does not improve the fitting effect and detection accuracy of the network. Taking into account the model complexity and steganography detection accuracy of RED-Net, RED-Net is still the most cost-effective steganography detector once trained compared to the other methods in this paper.

V. CONCLUSION AND DISCUSSION

Recently, numerous methods [33], [49]–[51], [53] have been developed to deal with the problem that illegal steganography and they have achieved remarkable achievements in steganography analysis. However, none of these methods are explored and exploited for specific steganalysis of medical image steganography. The proposed RED-Net incorporates a steganographic information enhancement module and a steganographic information discriminative mechanism based on a deep residual network to perform steganographic analysis of medical images. A joint hand-designed and deep-learning steganographic information enhancement module(SIEM) merges the horizontal high-frequency information, vertical high-frequency information, and diagonal high-frequency information of the original medical images with the medical images containing illegal steganographic information to boost the high-frequency information. A steganographic information discriminative mechanism(SIDM) that can rely on the relationship of the steganographic signal feature channels to distinguish between the steganographic signal in the high-frequency information extracted from medical images and the noise of natural medical images and the edges of tissue anatomical structures is introduced in the deep residual network. This deep residual network can perform steganographic signal feature extraction, feature compression, and steganographic analysis on illegal steganographic medical images. In the comparison experiments, we tested RED-Net in the spatial domain and JEPG on public and private datasets, respectively. For the public dataset, RED-Net's steganalysis error P_{E} are 0.0732, 0.0154, 0.0045, 0.0025. For the private dataset, RED-Net's steganalysis error P_{E} are 0.231, 0.138, 0.085, 0.044, and 0.026. The proposed RED-Net is compared with 8 state-of-art detections in the spatial and JEPG domain in the payload range of 0.1bpp/bpnzac to 0.5 bpp/bpnzac, and the test results of both private and public datasets show that RED-Net

outperforms steganalysis. The results of the ablation experiment verified the effectiveness of SIEM and SIDM. The SIEM can effectively boost the high-frequency information of medical images containing steganographic signals within the payload range of 0.1bpp-0.2bpp, and the SIDM can discriminate the steganographic signals from the natural noise and the edges of tissue anatomy in the medical images within the payload range of 0.2bpp-0.5bpp.

Although the RED-Net demonstrates encouraging improvement in the steganalysis of medical images, some limitations are still to be noticed. (1) Although a private medical image steganalysis dataset was designed, the quality and quantity of this dataset still have significant space for improvement due to practical reasons. (2) We only perform medical image illegal steganography detection for the proposed RED-Net in the payload range of 0.1bpp/bpnzac-0.5bpp/bpnzac, and further research should be devoted to improving the performance of the algorithm to cope with lower load steganography algorithms. (3) The speed of fitting needs to be improved. Deep learning-based steganalysis network has uncertainty in the process of training the network due to a large number of network parameters and is very dependent on the training of the network parameters, and the number of rounds of training of the network itself is relatively long. (4) At this stage, RED-Net can only be applied to the steganography detection of png format medical images containing 2D medical information.

Future research will be devoted to as follows: (1) It is necessary to analyze not only whether the image content is steganographic or not, but also the possible steganographic methods, the region modified by steganography, and finally, the secret information is intercepted by speculating the steganographic methods and the location of steganography. (2) Complementing the effects of SIEM and SIDM in the existing loss function and designing a new multi-level loss function to enhance the medical image detection accuracy. (3) Guaranteeing excellent steganography detection and avoiding overfitting based on small-scale dataset training. (4) Considering RED-Net as a discriminator for CycleGan combined with federated learning to develop a medical image steganography method. (5) We will work on extending the application of RED-Net to steganography containing 3D medical image information in the future.

VI. Acknowledgment

The authors would like to thank the Nanjing First Hospital, China, for providing clinical data, and data are adherent to the tenets of the Declaration of Helsinki.

References

[1] C. F. Pasluosta, H. Gassner, J. Winkler, J. Klucken, and B. M. Eskofier, "An emerging era in the management of parkinson's disease: wearable technologies and the internet of things," *IEEE journal of biomedical and health informatics*, vol. 19, no. 6, pp. 1873–1881, 2015.

[2] L. Liu, J. Xu, Y. Huan, Z. Zou, S.-C. Yeh, and L.-R. Zheng, "A smart dental health-iot platform based on intelligent hardware, deep learning, and mobile terminal," *IEEE journal of biomedical and health informatics*, vol. 24, no. 3, pp. 898–906, 2019.

[3] A. I. Siam *et al.*, "Portable and real-time iot-based healthcare monitoring system for daily medical applications," *IEEE Transactions on Computational Social Systems*, 2022.

[4] N. K. Dewangan and P. Chandrakar, "Patient-centric token-based healthcare blockchain implementation using secure internet of medical things," *IEEE Transactions on Computational Social Systems*, 2022.

[5] V. Ravi, T. D. Pham, and M. Alazab, "Attention-based multidimensional deep learning approach for cross-architecture iomt malware detection and classification in healthcare cyber-physical systems," *IEEE Transactions on Computational Social Systems*, 2022.

[6] S. Thakur, A. K. Singh, S. P. Ghrera, and M. Elhoseny, "Multi-layer security of medical data through watermarking and chaotic encryption for tele-health applications," *Multimedia tools and Applications*, vol. 78, no. 3, pp. 3457–3470, 2019.

[7] Z. Xia, X. Wang, W. Zhou, R. Li, C. Wang, and C. Zhang, "Color medical image lossless watermarking using chaotic system and accurate quaternion polar harmonic transforms," *Signal Processing*, vol. 157, pp. 108–118, 2019.

[8] A. Anand and A. K. Singh, "An improved dwt-svd domain watermarking for medical information security," *Computer Communications*, vol. 152, pp. 72–80, 2020.

[9] A. K. Singh, "Robust and distortion control dual watermarking in lwt domain using dct and error correction code for color medical image," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 30 523–30 533, 2019.

[10] K. Swaraja, K. Meenakshi, and P. Kora, "An optimized blind dual medical image watermarking framework for tamper localization and content authentication in secured telemedicine," *Biomedical Signal Processing and Control*, vol. 55, p. 101665, 2020.

[11] X. Liao, J. Yin, S. Guo, X. Li, and A. K. Sangaiah, "Medical jpeg image steganography based on preserving inter-block dependencies," *Computers & Electrical Engineering*, vol. 67, pp. 320–329, 2018.

[12] S. Edward Jero, P. Ramu, and S. Ramakrishnan, "Discrete wavelet transform and singular value decomposition based ecg steganography for secured patient information transmission," *Journal of medical systems*, vol. 38, no. 10, pp. 1–11, 2014.

[13] M. Sajjad *et al.*, "Mobile-cloud assisted framework for selective encryption of medical images with steganography for resource-constrained devices," *Multimedia Tools and Applications*, vol. 76, no. 3, pp. 3519–3536, 2017.

[14] A. Mohsin *et al.*, "Real-time medical systems based on human biometric steganography: A systematic review," *Journal of medical systems*, vol. 42, no. 12, pp. 1–20, 2018.

[15] B. Abd-El-Atty, A. M. Iliyasu, H. Alaskar, A. El-Latif, and A. Ahmed, "A robust quasi-quantum walks-based steganography protocol for secure transmission of images on cloud-based e-healthcare platforms," *Sensors*, vol. 20, no. 11, p. 3108, 2020.

[16] J. D. Ballard, J. G. Hornik, and D. McKenzie, "Technological facilitation of terrorism: Definitional, legal, and policy issues," *American Behavioral Scientist*, vol. 45, no. 6, pp. 989–1016, 2002.

[17] A. D. Ker, "Steganalysis of lsb matching in grayscale images," *IEEE signal processing letters*, vol. 12, no. 6, pp. 441–444, 2005.

[18] Q. Liu, A. H. Sung, Z. Chen, and J. Xu, "Feature mining and pattern classification for steganalysis of lsb matching steganography in grayscale images," *Pattern Recognition*, vol. 41, no. 1, pp. 56–66, 2008.

[19] R. Böhme, "Weighted stego-image steganalysis for jpeg covers," in *International Workshop on Information Hiding*. Springer, 2008, pp. 178–194.

[20] Z. Xia, X. Wang, X. Sun, Q. Liu, and N. Xiong, "Steganalysis of lsb matching using differences between nonadjacent pixels," *Multimedia Tools and Applications*, vol. 75, no. 4, pp. 1947–1962, 2016.

[21] G. Gul and F. Kurugollu, "A new methodology in steganalysis: breaking highly undetectable steganograpy (hugo)," in *International Workshop on Information Hiding*. Springer, 2011, pp. 71–84.

[22] X. Luo *et al.*, "Steganalysis of hugo steganography based on parameter recognition of syndrome-trellis-codes," *Multimedia Tools and Applications*, vol. 75, no. 21, pp. 13 557–13 583, 2016.

[23] W. Tang, H. Li, W. Luo, and J. Huang, "Adaptive steganalysis against wow embedding algorithm," in *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security*, 2014, pp. 91–96.

[24] K. Chen *et al.*, "Robust restoration of low-dose cerebral perfusion ct images using ncs-unet," *Nuclear Science and Techniques*, vol. 33, no. 3, pp. 1–15, 2022.

[25] D. Hu *et al.*, "Hybrid-domain neural network processing for sparse-view ct reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 88–98, 2020.

[26] T. Lyu *et al.*, "Estimating dual-energy ct imaging from single-energy ct data with material decomposition convolutional neural network," *Medical image analysis*, vol. 70, p. 102001, 2021.

[27] Y. Zhang *et al.*, "Cd-net: Comprehensive domain network with spectral complementary for dect sparse-view reconstruction," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 436–447, 2021.

[28] ——, "Clear: comprehensive learning enabled adversarial reconstruction for subtle structure enhanced low-dose ct imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3089–3101, 2021.

[29] D. Hu *et al.*, "Special: single-shot projection error correction integrated adversarial learning for limited-angle ct," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 734–746, 2021.

[30] D. Hu, Y. Zhang, J. Liu, S. Luo, and Y. Chen, "Dior: Deep iterative optimization-based residual-learning for limited-angle ct reconstruction," *IEEE Transactions on Medical Imaging*, 2022.

[31] Y. Qian, J. Dong, W. Wang, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," in *Media Watermarking, Security, and Forensics 2015*, vol. 9409. SPIE, 2015, pp. 171–180.
[32] G. Xu, H.-Z. Wu, and Y.-Q. Shi, "Structural design of convolutional neural networks for steganalysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 708–712, 2016.
[33] M. Boroumand, M. Chen, and J. Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, 2018.
[34] M. Yedroudj, F. Comby, and M. Chaumont, "Yedroudj-net: An efficient cnn for spatial steganalysis," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2092–2096.
[35] Y. Lu, G. Zhao, C. Chakraborty, C. Xu, L. Yang, and K. Yu, "Time sensitive networking-driven deterministic low-latency communication for real-time telemedicine and e-health services," *IEEE Transactions on Consumer Electronics*, 2023.
[36] T. K. Dash, C. Chakraborty, S. Mahapatra, and G. Panda, "Gradient boosting machine and efficient combination of features for speech-based detection of covid-19," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 11, pp. 5364–5371, 2022.
[37] J. Zeng, S. Tan, B. Li, and J. Huang, "Pre-training via fitting deep neural network to rich-model features extraction procedure and its effect on deep learning for steganalysis," *Electronic Imaging*, vol. 2017, no. 7, pp. 44–49, 2017.
[38] B. Li, W. Wei, A. Ferreira, and S. Tan, "Rest-net: Diverse activation modules and parallel subnets-based cnn for spatial image steganalysis," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 650–654, 2018.
[39] J. Ye, J. Ni, and Y. Yi, "Deep learning hierarchical representations for image steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2545–2557, 2017.
[40] S. Tan and B. Li, "Stacked convolutional auto-encoders for steganalysis of digital images," in *Signal and information processing association annual summit and conference (APSIPA), 2014 Asia-Pacific*. IEEE, 2014, pp. 1–4.
[41] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.
[42] O. Dalmaz *et al.*, "One model to unite them all: Personalized federated learning of multi-contrast mri synthesis," *arXiv preprint arXiv:2207.06509*, 2022.
[43] G. Elmas *et al.*, "Federated learning of generative image priors for mri reconstruction," *IEEE Transactions on Medical Imaging*, 2022.
[44] Y. Qian, J. Dong, W. Wang, and T. Tan, "Learning and transferring representations for image steganalysis using convolutional neural network," in *2016 IEEE international conference on image processing (ICIP)*. Ieee, 2016, pp. 2752–2756.
[45] ——, "Learning representations for steganalysis from regularized cnn model with auxiliary tasks," in *Proceedings of the 2015 International Conference on Communications, Signal Processing, and Systems*. Springer, 2016, pp. 629–637.
[46] C. Fran *et al.*, "Deep learning with depth wise separable convolutions," in *IEEE conference on computer vision and pattern recognition (CVPR)*, 2017.
[47] N. Jindal and B. Liu, "Review spam detection," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 1189–1190.
[48] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Transactions on information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012.
[49] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," in *2014 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2014, pp. 48–53.
[50] V. Holub and J. Fridrich, "Random projections of residuals for digital image steganalysis," *IEEE Transactions on information forensics and security*, vol. 8, no. 12, pp. 1996–2006, 2013.
[51] ——, "Low-complexity features for jpeg steganalysis using undecimated dct," *IEEE Transactions on Information forensics and security*, vol. 10, no. 2, pp. 219–228, 2014.
[52] G. K. Wallace, "The jpeg still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 30–44, 1991.
[53] M. Chen, V. Sedighi, M. Boroumand, and J. Fridrich, "Jpeg-phase-aware convolutional neural network for steganalysis of jpeg images," in *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, 2017, pp. 75–84.
[54] G. Luo, P. Wei, S. Zhu, X. Zhang, Z. Qian, and S. Li, "Image steganalysis with convolutional vision transformer," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3089–3093.
[55] K. Wei, W. Luo, S. Tan, and J. Huang, "Universal deep network for steganalysis of color image based on channel representation," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 3022–3036, 2022.
[56] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE transactions on information theory*, vol. 36, no. 5, pp. 961–1005, 1990.
[57] L. B. Almeida, "The fractional fourier transform and time-frequency representations," *IEEE Transactions on signal processing*, vol. 42, no. 11, pp. 3084–3091, 1994.
[58] R. R. Ernst and W. A. Anderson, "Application of fourier transform spectroscopy to magnetic resonance," *Review of Scientific Instruments*, vol. 37, no. 1, pp. 93–102, 1966.
[59] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
[60] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.
[61] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
[62] T. Pevnỳ, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *International Workshop on Information Hiding*. Springer, 2010, pp. 161–177.
[63] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *2012 IEEE International workshop on information forensics and security (WIFS)*. IEEE, 2012, pp. 234–239.
[64] 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, 2015.
[65] J. Fridrich, T. Pevnỳ, and J. Kodovskỳ, "Statistically undetectable jpeg steganography: dead ends challenges, and opportunities," in *Proceedings of the 9th workshop on Multimedia & security*, 2007, pp. 3–14.
[66] 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.