Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

SRIS-Net: robust image steganography based on feature score maps*

Ai XIAO†, Zhi LI†‡, Guomei WANG†, Long ZHENG†, Haoyuan SUN

School of Computer Science and Technology, Guizhou University, Guiyang, 550025, China

†E-mail: gs.axiao22@gzu.edu.cn; zhili@gzu.edu.cn; 306252084@qq.com; zhenglong178@163.com

Received Jan. 29, 2024; Revision accepted June 24, 2024; Crosschecked

Abstract: Image steganography algorithms based on deep learning are often trained using either spatial or frequency domain features. It is difficult for features from a single domain to comprehensively express the content of an entire image, which usually leads to poor performance because steganography is commonly a multi-task. To solve this problem, this paper proposes a robust image steganography algorithm based on feature score maps, called the secure and robust image steganography network (SRIS-Net). First, instead of spatial domain steganography, our proposed algorithm utilizes a convolutional neural network to obtain shallow spatial domain features. These features are decomposed by laplace pyramid frequency domain decomposition (LPFDD) to hide secret information in the different frequency sub-bands with a progressive assisted hiding strategy that significantly reduces the influence of the secret information on the cover image, effectively achieving significant invisibility and robust performance. In addition, we propose a global-local embedding module (GLEM) to achieve embedding by considering the overall structure of the image and the local details, and a dual multi-scale aggregation sub-network (DMSubNet) to preform multi-scale reconstruction to improve the quality of the carrier image. For security, we propose a dual-task discriminator structure, while giving a real/fake judgment of the image, which can generate a feature score map of the cover image region of interest (ROI) to guide the embedding module to generate a carrier image with higher imperceptibility and undetectability. Experimental results on the BOSSBase show that our SRIS-Net outperforms other mainstream methods in terms of undetectability and robustness, with more than 9.2 dB and 3.4 dB improvement in visual quality, respectively, and the capacity can be increased up to approximately 72 to 96 bpp.

Key words: Image steganography; Robustness; Un-detectability
https://doi.org/10.1631/FITEE.2400069 **CLC number:**

1 Introduction

Image steganography (Cheddad et al., 2010; Wengrowski and Dana, 2019; Tancik et al., 2020) hides information by exploiting image redundancy. A secret image is embedded in a cover image to generate a carrier image, while maintaining better visual quality of the carrier image and higher accuracy of the secret image extraction. Based on security considerations, the carrier image is usually required to be visually indistinguishable from the cover image. Traditional methods often hide the secret information using the least significant bit (LSB) (Barni et al., 2001; Li XL et al., 2009). These methods can hide only a limited amount of information and are not robust enough against various attacks.

In recent years, deep learning (DL)-based image steganography has shown promising results (Hu et al., 2018; Rehman et al., 2018). Baluja (2017) proposed the first convolutional neural network (CNN) in this domain. Researchers further improved image steganography performance by improving the net-

‡ Corresponding author
* Project supported by the National Natural Science Foundation of China (No. 62062023) and Guizhou Science and Technology Plan Project, China (No. ZK2021YB314)
ORCID: Ai XIAO, https://orcid.org/0009-0002-9839-1718

work structure (Baluja, 2020) and introducing new loss functions (Singh et al., 2022). Invertible neural networks (INNs) demonstrate superior performance in image steganography (Lu et al., 2021; Yang et al., 2024) due to their precise fulfillment of the inverse relationship between embedding and extraction processes. Some studies (Chen et al., 2020) focus on designing frameworks with good robustness and other works improve robustness by introducing noise in training (Ying et al., 2022). ADH-GAN (Yu, 2020) guarantees both robustness and capacity using a generative adversarial network (GAN). However, extensive training of an attention module is required to obtain attention weight maps for steganography assistance. Inspired by PatchGan (Isola et al., 2017), a discriminator can evaluate the entire image generated by the generator and feed the attention back to each patch. As for the image steganography task, if the region of interest (ROI) of the steganalysis can be known before embedding, higher security can be achieved. The above methods directly concatenate or add the cover image and the secret image, or extract their features and then concatenate or add them, without considering the correlation and adaptability between features, and thus it is not optimal for the network to excel in only some aspects of performance but perform poorly in others. Moreover, image steganography based on DL is usually trained using features in a single domain, such as the spatial (Lu et al., 2021) or frequency domain (Jing et al., 2021), which makes comprehensively expressing the content of the image difficult. Compared with the pixel domain, the embedding capacity of the feature domain is larger (Xu et al., 2022) and more robust. In addition, embedding the features of a secret image instead of the original itself can assist in encrypting the secret image (Chen et al., 2020).

Based on the above, we designed a robust image steganography algorithm named the secure and robust image steganography network (SRIS-Net). SRIS-Net uses laplace pyramid frequency domain decomposition (LPFDD) (Lai et al., 2018) to perform multi-scale decomposition of the spatial-domain features of the cover image, and uses these features to achieve progressive embedding and reconstruction in the sub-bands of different frequency domains. Based on a progressive assisted hiding strategy, SRIS-Net uses these frequency-domain features to enhance robustness. Additionally, SRIS-Net employs a dual-task discriminator that not only identifies the authenticity of the image but also provides the feature score map of the discriminator's ROI to guide the embedding process. Consequently, SRIS-Net achieves good performance on steganographic quality, security, robustness, and capacity. The main contributions of this paper include the following:

1. We propose a robust image steganography algorithm, SRIS-Net, based on feature score maps, which performs LPFDD on the image spatial domain and implements incremental embedding and reconstruction across various frequency sub-bands. This strategy minimizes the impact of hidden information on the cover image, and ensures significant invisibility and robust performance.

2. An embedding block (the global-local embedding module, or GLEM) and a dual multi-scale aggregation sub-network (DMSubNet) are introduced, which enhance embedding by considering the correlation and adaptation between the cover and secret images. These components reconstruct image features to improve the quality of image steganography.

3. A dual-task discriminator structure is proposed for security, which can not only determine a real (cover) or fake (carrier), but generate a feature score map of the ROI for the cover image to guide the embedding module to generate a carrier image with higher imperceptibility and undetectability.

2 Related works

Steganography is a technique for hiding a message, audio, image, or video in other media to avoid arousing suspicion. Least significant bit (LSB) (Tamimi et al., 2013) is the most traditional spatial-domain-based method in steganography, but it suffers from texture replication artifacts, particularly in smooth regions, making it vulnerable to steganalysis (Xu et al., 2016; Ye et al., 2017). Frequency-domain methods, such as the discrete cosine transform (DCT) (Ruanaidh et al., 1996) and the discrete wavelet transform (DWT) (Barni et al., 2001), offer more robustness and are harder to detect, although they only conceal bit-level information.

The combination of DL and image steganography has significantly improved performance. Shi et al. (2018) used a GAN-based encoder-decoder architecture to enhance resistance to steganalysis. Many image steganography methods have been pro-

posed based on different network architectures, including the method by Liu et al. (2022) based on ResNet, residual structures (Wu et al., 2018), U-Net structure (Duan et al., 2019), and the Xception block-based method (Duan et al., 2020b), which are mainly designed to optimize capacity and visual quality. StegGAN (Singh et al., 2022) proposed a GAN-based architecture for image steganography, which combined with dual discriminators, significantly improves the quality of both the carrier and secret images but only ensures noiseless extraction. GAN-based steganography algorithms have proved more securely than normal. However, the previous GAN-based steganography primarily used the discriminator for real/false judgments during training, discarding it afterward. There is potential for the discriminator to provide more valuable feedback and aid the generator in creating more realistic and secure carrier images.

LPFDD. The Laplacian pyramid (Lai et al., 2018) is a multi-scale image representation method widely used in the fields of image super-segmentation and image reconstruction. It is created by convolving the original image through successive Gaussian filters to generate blurred images at various scales and then subtracting each Gaussian pyramid layer from the up-sampled image of its predecessor, as defined in Eq. (1).

$$L_i = G_i - Expand(G_{i+1}), \tag{1}$$

where L_i denotes the i-th layer of the Laplacian pyramid image, G_i represents the i-th layer of the Gaussian pyramid image, and $Expand(G_{i+1})$ is the up-sampled previous layer Gaussian pyramid image. This construction allows for multi-scale analysis and processing of the image. By using LPFDD, features are embedded and reconstructed in different subbands, with lower-frequency features assisting the higher-frequency to complete the image steganography task. This approach minimizes the impact of secret information on the cover image, effectively improving visual quality and capacity.

3 Method

SRIS-Net consists of three main components: an LPFDD embedding network, an LPFDD extraction network, and a dual-task discriminator network. Fig. 1 shows the overall architecture of SRIS-Net.

Table 1 Explanation of symbols in this paper

Notation	Description
I_{co}	The image to hide secret information
I_{ca}	The image with secret information inside
I_{se}	The image to be hidden
I_{re}	The recovered secret image from carrier image

Table 1 shows the notation used in this paper.

3.1 LPFDD embedding network

Information embedded in the high-frequency regions is hard to detect but susceptible to attacks (Jing et al., 2021), while embedding information in the low-frequency regions improves robustness but often arouses visual suspicion (Li ZZ et al., 2022). Both domains have suitable embedding regions, and using only one impacts capacity, image quality, and security. Therefore, our method embeds information across multiple frequency domains, finding suitable locations through the global-local embedding module (GLEM) and reconstructing features with the Dual Multi-Scale Aggregation Sub-Network (DM-SubNet) to enhance embedding capacity and image quality.

As shown in Fig. 1, the LPFDD embedding network we proposed divides image steganography into high-frequency (h_0), medium-frequency (h_1), and low-frequency (h_2) stages for progressive embedding and reconstruction. Given a cover image $I_{\mathrm{co}} \in R^{h\times w\times 1}$, we first extract the shallow features $I_0 \in R^{h\times w\times c}$, where h, w is the spatial dimension and c is the channel number. The LPFDD is then used to obtain the frequency features H = [h_0, h_1, h_2], with resolutions $\frac{h}{2^i}\times\frac{w}{2^i}\times c, i=0,1,2$. For a secret image $I_{\mathrm{se}} \in R^{h\times w\times 1}$, features S= [$s_0$, s_1, s_2] are extracted progressively by the Secret Image Feature Extraction Module with resolutions of $\frac{h}{2^i}\times\frac{w}{2^i}\times c, i=0,1,2$ (Fig. 1). Then, guided by the dual-task discriminator feature maps, the information of H and S are fed into three network branches for hierarchical embedding and reconstruction, respectively.

The low-frequency embedding and reconstruction process is shown in Eq. (2):

$$\hat{h_2} = DM_2\left(EM_2\left(h_2, s_2, map_2\right)\right), \tag{2}$$

where $h_2 \in R^{\frac{h}{4}\times\frac{w}{4}\times c}$, $s_2 \in R^{\frac{h}{4}\times\frac{w}{4}\times c}$, and $map_2 \in R^{\frac{h}{4}\times\frac{w}{4}\times c}$ are the low-frequency features of the cover image, secret image, and feature score maps from the discriminator, respectively. EM is the mod-

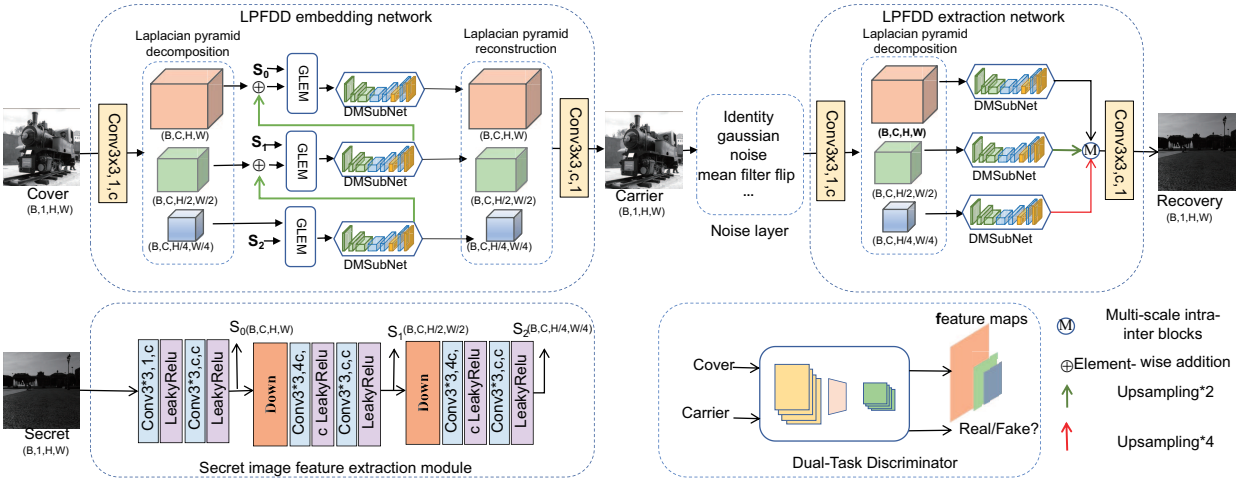

Fig. 1 SRIS-Net framework. First, through the secret image feature extraction module, the progressive resolution features $\mathbf{S} = [s_0, s_1, s_2]$ of the secret image are extracted. Then, $\mathbf{S} = [s_0, s_1, s_2]$ and the cover image features are separately embedded through the GLEM (Fig. 2) guided by the dual-task discriminator feature maps, and reconstructed through DMSubNet (Fig. 3) of the three branches in the LPFDD embedding network. At the receiver end, the carrier image is processed by the LPFDD extraction network to extract the recovered secret image. In the bottom right corner is the dual-task discriminator, which can not only distinguish real versus fake, but also provide a feature scores map for the ROI to guide the embedding of secret information

ule for GLEM, and DM is the DMSubNet. The $\hat{h}_2 \in R^{\frac{h}{4}\times\frac{w}{4}\times c}$ is the reconstructed feature of the low-frequency region.

Given the relatively small amount of information in h_0 and h_1, direct embedding may cause serious information loss, so we adopted a progressive embedding strategy. We up-sample $\hat{h}_2$ to $\hat{h}_2\uparrow$ with the size of h_1 using bilinear interpolation. Then we add it to h_1 to assist in embedding. This process is repeated for h_0, with each step similar to the h_2 embedding process described in Eq. (3). The progressive embedding strategy ensures better preservation of information and improved embedding quality and can effectively break the limitation of single-domain hiding on capacity.

$$\hat{h_i} = DM_i\left(EM_i\left(h_i + \hat{h_{i+1}}\uparrow, s_i, map_i\right)\right). \quad (3)$$

GLEM. The global-local embedding module embeds secret information adaptively by considering both the overall image structure and local details (Fig. 2). Guided by the $map_i \in R^{\frac{h}{2^i}\times\frac{w}{2^i}\times c}$, GLEM embeds $s_i \in R^{\frac{h}{2^i}\times\frac{w}{2^i}\times c}$ into $h_i \in R^{\frac{h}{2^i}\times\frac{w}{2^i}\times c}$ through global and local branches, so the secret information embedding is more adaptive to the cover image, thus improving robustness and security. The global branch uses a self-attention mechanism (Zamir et al., 2022), and the local branch uses Conv1*1 for detail embedding. Q, K, and V are generated with Conv1*1 and DWConv3*3 and T denotes the matrix transposition in Fig. 2.

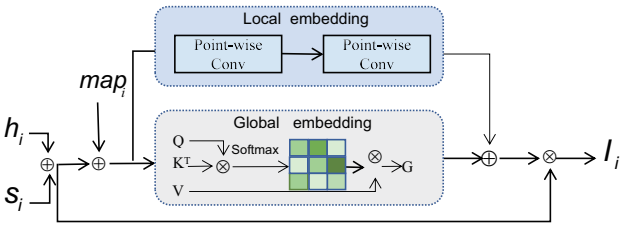

Fig. 2 Global-Local Embedding Module (GLEM) embeds the secret image in the cover image under the guidance of the feature map, and includes a global embedding branch and a local embedding branch

DMSubNet. To handle multi-frequency domain embedding and extraction, DMSubNet uses Multi-Scale Intra-blocks (MSIB) and a Selective Fusion Module (SFM) (Fig. 3). DMSubNet down-samples and up-samples (Shi et al., 2016) features hierarchically, leveraging multi-scale representations to accurately process image details. Meanwhile, for the image steganography reconstruction and extraction process, the manner in which shallow features are directly added with deep features is not applicable. Especially in reconstruction of the secret image, shallow features often contain more carrier image information. Therefore, secret image features are

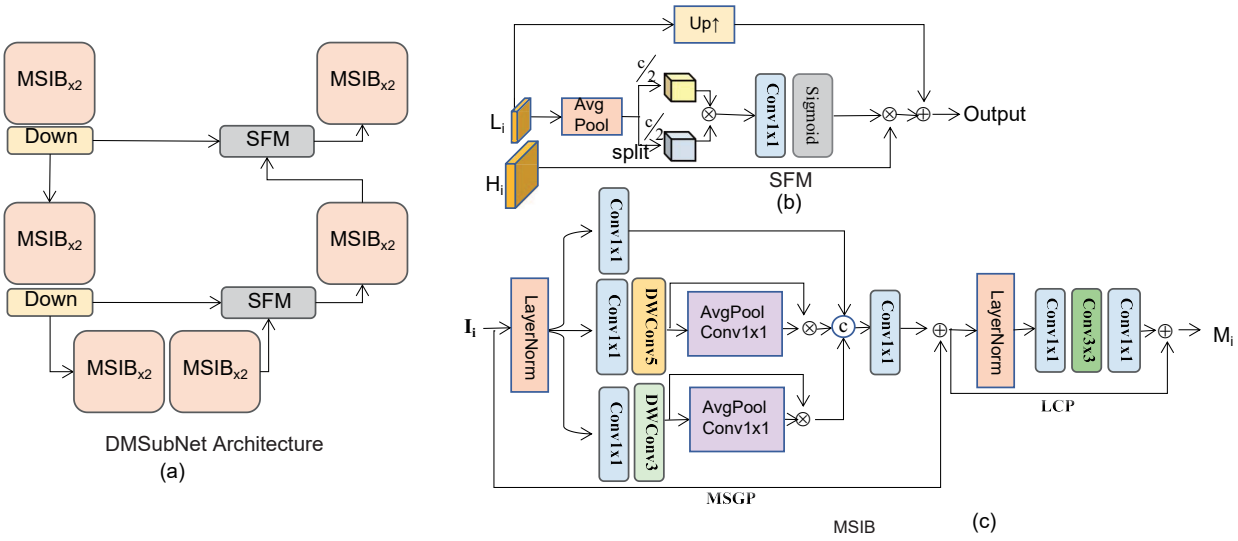

Fig. 3 DMSubNet and related modules. Bottom right, MSIB uses different convolutional kernels for multi-scale feature representation. Top right is our SFM module

extracted selectively, so the addition operation will directly lead to recover the secret image unsuccessfully. Therefore, the SFM enables aggregation of the low-level image features and high-level features selectively, preserving fine structural and textural details for better reconstruction and extraction.

SFM (Fig. 3) achieves fusion of the shallow features and the deeper features selectively. Specifically, we generate the attention map by global average pooling of the low-resolution deep feature map L_i , and then split it along the channels. Then we apply element-by-element multiplication, followed by Conv1 $\times$ 1 to transform the channels. The high-resolution shallow feature H_i is then multiplied with this attention weight map to selectively extract features that are useful for low-resolution deep feature reconstruction. The low-resolution deep features are then up-sampled (Shi et al., 2016) and added to the extracted high-resolution feature to obtain the final feature.

MSIB (Fig. 3) enhances multi-level feature understanding with a multi-scale global perceptual branch (MSGP) and a local contextual refinement branch (LCP). It uses different DWConv kernels for global multi-scale information and Conv3 $\times$ 3 for local details. LayerNorm is applied for better performance. The LCP uses a bottleneck design for efficient local spatial modeling (Zhou et al., 2020).

DMSubNet allows adaptive and robust information embedding across multiple frequency domains, improving both the capacity and quality of the steganographic images.

3.2 LPFDD extraction network

The LPFDD extraction network adopts a similar structure to the LPFDD embedding network. It begins by taking the carrier image as input and extracting shallow features using Conv3 $\times$ 3. The high-, medium-, and low-frequency features are then obtained via LPFDD. These features are processed by DMSubNet to extract the secret information. The low-frequency features are up-sampled (Shi et al., 2016) by a factor of 4, and the medium-frequency features by a factor of 2, and then concatenated with the high-frequency features. The combined features are passed through two MSIB modules to produce the recovered secret image I_{re}, as shown in Fig. 1. To enhance robustness, a noise layer is introduced before extraction, including a mean filter, Gaussian filter, sharpening, Gaussian noise, flipping, and identity processing.

3.3 Dual-task discriminator

Traditional discriminators focus on classifying images as real (cover) or fake (carrier) (Fu et al., 2020) and aid in generating a carrier image that closely resembles the cover. These discriminators are usually discarded after training. However, we argue that the role of the discriminator depends on

not only classifying the inputs as fake or real, but also having the ability to evaluate the whole image, generating a map of evaluation scores concerning the ROI. This evaluation score map can be regarded as a kind of a priori knowledge of the embedding. This knowledge is fed back to the generator to guide it to pay more attention to the regions that are more suitable for embedding during embedding and reconstruction, and to generate a more realistic image of the carrier image. Meanwhile, the classification of the discriminator is based on the difference between the distinguishing regions in the original and the generated image, and the feature map weights of each layer in the network are of great significance in the final results of the classifier. Based on this, our study improves the classical steganalysis ZhuNet (Zhang et al., 2019b), as shown in Fig. 4, which includes a preprocessing layer, two separable convolution (sepconv) blocks, four basic blocks for feature extraction, a spatial pyramid pooling (SPP) module, a fully connected layer, and softmax. We up-sample the features from each basic block to match the original feature size by bilinear interpolation. This is followed by channel concatenation, average pooling, and normalization to obtain the feature score map. The score map is then decomposed using LPFDD into multi-scale feature sub-bands, which are fed back to the generator to guide the embedding process. This dual-task discriminator not only classifies but also provides detailed guidance, improving the security and visual quality of the generated images, as demonstrated in Table 7.

3.4 Loss function

The SRIS-Net loss function primarily comprises the generator's reconstruction loss, extraction loss, adversarial loss, and the discriminator's loss.

Reconstruction loss: to ensure that the carrier image I_{ca} generated by hiding I_{se} in I_{co} is indistinguishable from I_{co}, we also propose a specialized Laplace frequency domain loss. This enhances the similarity between the carrier and cover images in both pixel and frequency domains, improving security and visual quality. The reconstruction loss L_s as Eq. (4):

$$\begin{aligned} L_s = {} & L1loss\left(I_{\mathrm{co}}, I_{\mathrm{ca}}\right) + L_{\mathrm{lp}}\left(I_{\mathrm{co}}, I_{\mathrm{ca}}\right) \\ & + \alpha L_{\mathrm{per}}\left(I_{\mathrm{co}}, I_{\mathrm{ca}}\right). \end{aligned} \tag{4}$$

where $L1loss$ is the L1 loss of I_{co} and I_{ca} in the pixel domain (Eq. (5)):

$$L1loss(X, X^{'}) = \frac{1}{HWC}\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{C}\left\|X_{\mathrm{ijk}} - X_{\mathrm{ijk}}^{'}\right\|. \tag{5}$$

L_{lp} is the frequency domain L1loss of I_{co} and I_{ca} using LPFDD (Eq. (6)):

$$L_{lp}(X, X^{'}) = \sum_{i=0,1,2} L1loss\left(LP\left(X\right)_i, LP\left(X^{'}\right)_i\right). \tag{6}$$

I_{per} is the perceptual loss (Johnson et al., 2016), which can help enhance image details and texture information. α is the the trade-off coefficient adjusts the trade-off coefficient which adjusts the importance of different losses to the total loss function; we set α=0.1.

Extraction loss: the extraction network aims to recover the secret image I_{se} from the carrier image

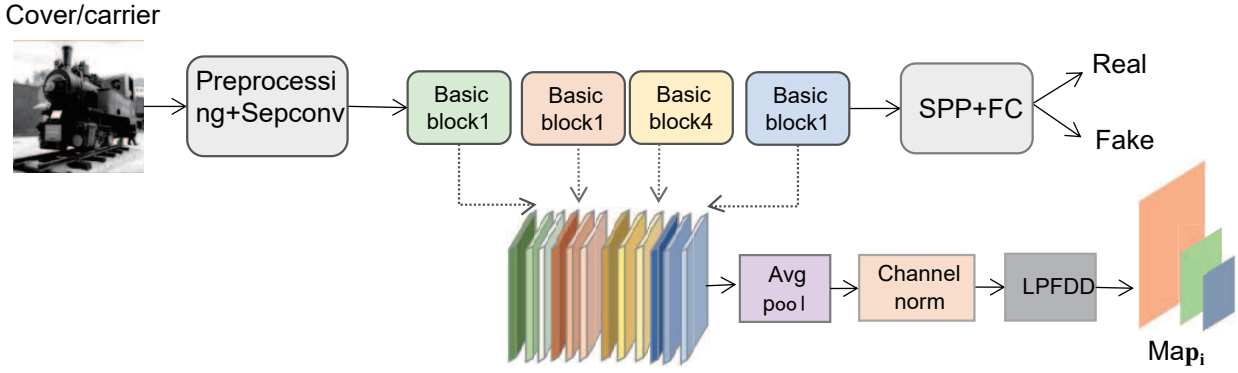

Fig. 4 Dual-task discriminator structure; we up-sample the features of each of the four basic blocks to obtain the feature score map

I_{ca}. The extraction loss I_{re} is defined as in Eq. (7):

$$L_{\mathrm{recon}} = L1loss\left(I_{\mathrm{se}}, I_{\mathrm{re}}\right) + L_{\mathrm{per}}\left(I_{\mathrm{se}}, I_{\mathrm{re}}\right). \quad (7)$$

Adversarial Loss: we use LSGAN (Mao et al., 2017) for the adversarial loss L_{adv} to ensure stable training, as in Eq. (8).

$$L_{\mathrm{adv}} = E_{z\sim p(z)}\left[(D(G(X))) - 1\right]. \quad (8)$$

The total loss of the SRIS-Net generator is then given by Eq. (9), where β is the trade-off factor, set to 0.0004.

$$L_{\mathrm{total}} = L_s + L_{\mathrm{recon}} + \beta L_{\mathrm{adv}}. \quad (9)$$

The dual-task discriminator loss also uses LSGAN (Mao et al., 2017), defined as in Eq. (10).

$$L_D = \frac{E_{z\sim p(z)}\left[(D(G(X))) - 0\right] + E_{x\sim p(x)}\left[(D(X)) - 1\right]}{2}. \quad (10)$$

This structured approach ensures that each component of SRIS-Net is optimized for its specific task, resulting in improved steganography and extraction performance.

4 Experiments

4.1 Setup

The proposed model is trained for 160 epochs on an NVIDIA A100 GPU. The number of channels c=48. The initial learning rate is 0.0004, adjusted using StepLR with step_size=25 and gamma=0.5. The batch size is 8. The generator is optimized by Adam with standard parameters, and the dual-task discriminator is optimized by SGD with momentum=0.9 and weight_decay=0.0005. The generator and discriminator are trained jointly with alternating updates.

We evaluated SRIS-Net on the BOSSBase (Bas et al., 2011), which contains 10,000 grayscale images at resolution 512×512. We set the input size to 160×160 using center-cropping. The dataset is split into 9,000 training images and 1,000 test images. Both the cover and secret image are randomly paired from the training set.

There are five metrics adopted to measure the quality of cover/carrier image pairs and secret/recovery image pairs: MSE, PSNR, SSIM, MS-SSIM, and SCC (Otazu et al., 2005).

4.2 Comparison

To verify the effectiveness of our method, we compared it with SimultaneousCNN (Van et al., 2019), ISGAN (Zhang et al., 2019a), StegNet (Wu et al., 2018), U-Net Structure (Duan et al., 2019), Huang (Huang et al., 2019), Encoder-Decoder (Rehman et al., 2018), Baluja (Baluja, 2017), SteganoCNN (Duan et al., 2020b), Liu (Shi et al., 2018), Improved Xception (Duan et al., 2020a), StegGAN (Singh et al., 2022), DBPSNet (Li ZZ et al., 2022), DAH-Net (Zhang et al., 2023), and PRIS (Yang et al., 2024). All of methods were trained and tested using gray images (Bas et al., 2011).

4.2.1 Steganographic quality

Quantitative results. As shown in Table 2, SRIS-Net significantly outperforms all other methods in terms of five metrics for the cover/carrier, where MSE (0.00006), PSNR (43.55), SSIM (0.9990), MS-SSIM (0.9990), and SCC (0.9979). For the secret/recovery, PRIS (Yang et al., 2024) achieves optimal performance due to the use of INNs with MSE (0.00011), PSNR (40.58), SSIM (0.9981), MS-SSIM (0.9977), and SCC (0.9964). But its strict reversibility often exhibits vulnerability when subjected to noise and other attacks. From Table 3 and Table 4, it is evident that PRIS is inferior to our method in terms of security and robustness.

Qualitative results. We randomly chose a pair of cover and secret images and compared the carrier and recovery images generated by different methods. Considering the length of this paper, we only show the comparison of some of the methods. As shown Fig. 5, HCRGAN (Chen et al., 2020), DAH-Net (Zhang et al., 2023), and StegGAN (Singh et al., 2022) generate blurred images, while our model and DBPSNet (Li ZZ et al., 2022) can generate clear carrier images. However, our method produces even clearer carrier images. We further show the residual frequency histogram of cover/carrier pairs and secret/recovery pairs. As shown in Fig. 5, our method achieves a frequency of 0 pixel value greater than 0.8 and the average pixel value of the residual is 0.2948, which achieves the best visual results compared with other methods. Although our method provide sub-optimal average pixel value compared to DBPSNet (Li ZZ et al., 2022), our method also achieves no pixel

Table 2 Image hiding quality results, with the best results in bold and second bests underlined

Method	Cover/Stego					Secret/Recovery				
	MSE↓	PSNR↑	SSIM↑	MS-SSIM↑	SCC↑	MSE↓	PSNR↑	SSIM↑	MS-SSIM↑	SCC↑
SimultaneousCNN	0.00215	25.63	0.9420	0.9774	0.9790	0.00200	25.33	0.9034	0.9492	0.9632
ISGAN	0.01120	18.95	0.8863	0.9036	0.9637	0.00621	21.09	0.8887	0.9353	0.9316
StegNet	0.00412	22.38	0.9069	0.9554	0.9082	0.00293	24.21	0.9134	0.9599	0.9756
U-Net Structure	0.00121	27.76	0.9430	0.9805	0.9795	0.00139	26.69	0.9276	0.9707	0.9699
Huang	0.00974	19.97	0.9023	0.9142	0.9528	0.00526	21.86	0.9052	0.9391	0.9533
HCRGAN	0.00995	18.29	0.7754	0.8961	0.7470	0.01033	16.14	0.6933	0.8448	0.7097
Encoder-Decoder	0.00051	31.38	0.9099	0.9891	0.8565	0.00092	28.7	0.8972	0.9747	0.9207
Baluja	0.00035	33.26	0.9338	0.9923	0.9024	0.00037	31.38	0.9310	0.9798	0.9577
SteganoCNN	0.00032	33.47	0.9568	0.9922	0.9126	0.00067	30.88	0.9651	0.9878	0.9804
Liu	0.00030	33.89	0.9721	0.9916	0.9534	0.00032	32.64	0.9398	0.9847	0.9726
Improved Xception	0.00020	35.86	0.9651	0.9959	0.9482	0.00158	26.42	0.9502	0.9788	0.9828
StegGAN	0.00041	33.19	0.9574	0.9834	0.9154	0.00043	31.54	0.9293	0.9485	0.9667
DBPSNet	0.00025	34.26	0.9922	0.9921	0.9811	0.00023	36.31	0.9902	0.9966	0.9957
DAH-Net	0.00062	32.35	0.9920	0.9900	0.9797	0.00068	32.21	0.9892	0.9817	0.9805
PRIS	0.00010	40.88	0.9982	0.9978	0.9964	**0.00011**	**40.58**	**0.9981**	**0.9977**	**0.9964**
Ours	**0.00006**	**43.55**	**0.9990**	**0.9990**	**0.9979**	0.00013	39.78	0.9979	0.9956	0.9952

values of 0.55 and average pixel value of 0.5551, indicating that the performance quality of the secret image reconstructed by our method remains impressive.

Through both quantitative and qualitative experimental results, our method demonstrates significant hiding and restoration capabilities. It can generate a visually undetectable carrier and accurately extract the secret image from the carrier image.

4.2.2 Security

Security is crucial for image steganography algorithms. We evaluated the ability of SRIS-Net to resist steganalysis detection. We chose the classical DL-based algorithm XuNet (Xu et al., 2016) to assess steganalysis detection performance. We pretrained XuNet with carrier images generated by the S-UNIWARD algorithm (Holub et al., 2014) from 10,000 pairs of images in BOSSBase at a 0.4bpp payload. Then we randomly paired 1,000 images in the test set with 500 carrier-secret pairs and used various models to generate 500 carrier images. From these, 450 carrier images were used to fine-tune XuNet (Xu et al., 2016), which then detected the remaining carriers and calculated the error detection rate. As shown in Table 3, SRIS-Net achieves the best error detection rate of 36%, 2% higher than DBPSNet (Li ZZ et al., 2022), whereas all other methods score below 30%.

Then, we used traditional SRM (Fridrich and Kodovsky, 2012) and CSR (Denemark et al., 2014) for steganalysis evaluation, employing Ensemble (Kodovsky et al., 2012) as the classifier. We split the carrier and cover images for training and testing, averaging results over 10 cross-tests. SRIS-Net achieved the best results with 24.38% for SRM and 40.60% for CSR.

Table 3 Assessment of steganography security, with the best results in bold and second best underlined

Method	SRM	CSR	XuNet
SimultaneousCNN	10.52%	14.42%	23%
ISGAN	2.08%	3.97%	11%
StegNet	12.10%	14.39%	19%
U-Net Structure	12.90%	15.03%	23%
Huang	4.08%	6.49%	12%
HCRGAN	2.18%	3.69%	4%
Encoder-Decoder	5.26%	8.39%	14%
Baluja	9.06%	11.27%	17%
SteganoCNN	15.22%	16.24%	14%
Liu	17.36%	18.73%	20%
Improved Xception	16.40%	17.85%	16%
StegGAN	15.97%	18.60%	16%
DBPSNet	22.56%	27.51%	34%
DAH-Net	15.80%	18.98%	16%
PRIS	18.80%	20.00%	17%
Ours	**24.38%**	**40.60%**	**36%**

These evaluations demonstrate that SRIS-Net, combined with discriminator feature score maps, can generate highly secure carrier images. The discriminator guides the generator to distribute secret signals to more undetectable regions, enhancing the security of the steganography process.

4.2.3 Anti distortion capability

To increase the model's robustness, we introduced a noise layer that trained the model to reveal the secret image in noisy conditions, as shown in Fig. 1. We chose MSE, PSNR, and SSIM as ro-

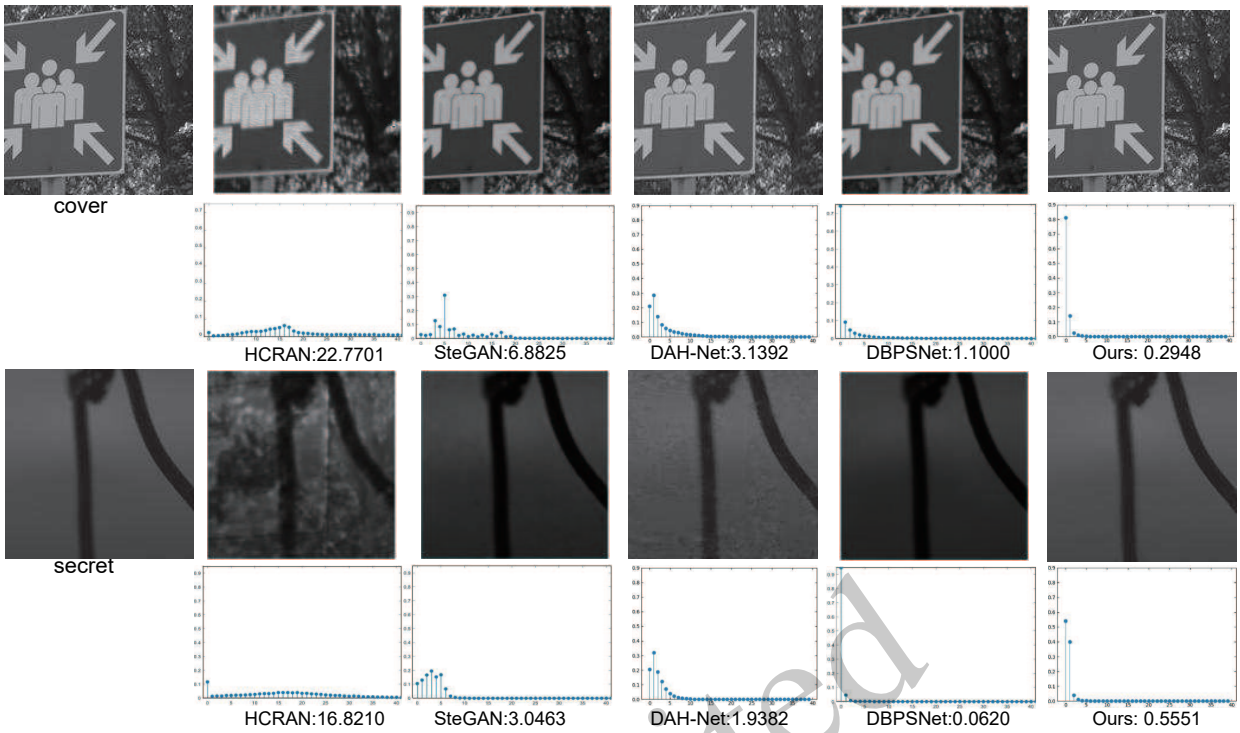

Fig. 5 Random visualization of the same cover/carrier and secret/recovery pairs generated through different methods

bustness evaluation metrics. As shown in Table 4, SRIS-Net performs well under various noise conditions, including mean filtering, Gaussian filtering, sharpening, Gaussian noise, and flipping. Specifically, with mean and Gaussian filtering using 5 × 5 filters, Laplacian sharpening, and Gaussian noise with a variance of 0.01, our extraction network effectively reveals the secret image. SRIS-Net outperforms other methods in MSE, PSNR, and SSIM. Notably, when the carrier image is attacked by flipping, SRIS-Net still reconstructs the secret image with a PSNR of 26.95.

To further demonstrate the model's performance under multiple distortions, we conducted experiments using various combinations of Gaussian filter, Gaussian noise, mean filter, and sharpening, as well as constructing a noise pool, as shown in Table 5. The SRIS-Net performance under combined noise conditions is indeed worse than under single noise conditions, but it still achieves relatively good results. Specifically, under Gaussian filtering and sharpening, it attained MSE, PSNR, and SSIM values of 0.00017, 38.46 dB, and 0.9972, respectively. The worst performance was with the combination of mean filtering and Gaussian noise, yet it still achieved an MSE of 0.00063, PSNR of 32.51 dB, and SSIM of 0.9899. From row 7 of Table 5, SRIS-Net still achieved MSE of 0.00030, PSNR of 36.08 dB, and SSIM of 0.9951 under random noise. These results demonstrate that SRIS-Net possesses robust steganographic capabilities.

4.2.4 Capacity

In some scenarios, capacity is extremely important. We compared SRIS-Net with StegGAN (Singh et al., 2022) by conducting experiments embedding 2, 3, and 4 images in one cover image. Table 6, shows the average PSNR results for the cover/carrier and secret/recovery image pairs. When embedding 2 images, our method achieves an average PSNR of 41.96 dB for cover/carrier images and 34.65 dB for secret/recovery images. Even when embedding 3 images, the performance remains good. Remarkably, with 4 images, the average PSNR is 41.46 dB for cover/carrier images and 30.85 dB for secret/recovery images. However, the StegGAN (Singh et al., 2022) performance significantly deteriorates with 2 images, showing an average PSNR of 33.21 dB for cover/carrier pairs and 21.41 dB for se-

Table 4 Image steganography robustness evaluation. Under noise or attack scenarios, we calculated the MSE, PSNR, and SSIM of the secret/recovery pairs extracted by different methods

Method	Mean			Gaussian			Sharpening			Gaussian Noise			Flip		
	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑
SimultaneousCNN	0.01510	16.74	0.4503	0.04566	13.71	0.3073	0.01904	16.89	0.6210	0.00772	20.44	0.5404	0.04595	12.65	0.3541
ISGAN	0.02876	13.65	0.2099	0.02969	12.25	0.3391	0.00857	19.26	0.8437	0.05360	11.81	0.1260	0.05660	11.38	0.3282
StegNet	0.02109	17.20	0.2474	0.02308	17.68	0.4810	0.01700	18.35	0.6220	0.03161	13.66	0.1994	0.05003	12.75	0.2778
U-Net Structure	0.01494	16.56	0.4490	0.04898	13.10	0.3267	0.01602	17.80	0.6714	0.03453	14.29	0.1944	0.04417	12.68	0.3418
Huang	0.02462	16.03	0.2432	0.03928	14.76	0.3605	0.00711	20.60	0.8529	0.04625	12.94	0.1374	0.05400	11.96	0.3096
HCRGAN	0.02707	12.07	0.3507	0.02195	13.01	0.4799	0.01295	15.03	0.6092	0.03230	10.98	0.1866	0.04789	9.84	0.3068
Encoder-Decoder	0.07654	6.67	0.1190	0.06529	5.11	0.2060	0.79431	14.07	0.1600	0.01756	17.15	0.3316	0.05079	14.92	0.2109
Baluja	0.06280	7.47	0.2420	0.05799	11.25	0.3019	0.60221	2.32	0.2266	0.04124	13.44	0.2256	0.04758	12.48	0.2006
SteganoCNN	0.01667	18.22	0.4638	0.03976	14.19	0.3861	0.00214	26.29	0.8764	0.04870	12.41	0.1522	0.04467	12.72	0.3264
Liu	0.01621	16.29	0.4466	0.03897	14.69	0.3469	0.00945	17.55	0.7869	0.00813	18.69	0.4536	0.04683	12.65	0.3217
Improved Xception	0.01721	18.13	0.4153	0.01963	18.75	0.5387	0.03998	15.02	0.5305	0.03532	14.45	0.1836	0.06891	11.82	0.1982
StegGAN	0.14370	18.38	0.4874	0.03415	15.46	0.4956	0.00792	19.27	0.8004	0.00778	19.27	0.4818	0.04288	13.12	0.3787
DBPSNet	0.00068	31.30	0.9611	0.00023	36.09	0.9916	0.00026	35.86	0.9905	0.00027	34.76	0.9870	0.00391	13.80	0.5623
DAH-Net	0.00122	29.72	0.9805	0.00088	31.13	0.9864	0.00045	34.05	0.9931	0.00114	29.80	0.9817	0.00359	25.01	0.9420
PRIS	0.00061	32.44	0.9887	0.00091	31.27	0.9855	0.00016	39.07	0.9975	0.00048	33.49	0.9921	0.04956	13.29	0.1119
Ours	**0.00013**	**36.62**	**0.9957**	**0.00015**	**39.23**	**0.9976**	**0.00015**	**39.29**	**0.9976**	**0.00023**	**36.77**	**0.9964**	**0.00245**	**26.95**	**0.9635**

cret/recovery pairs. As the number of embedded images increases, the StegGAN image quality declines sharply.

Table 5 Quantitative evaluation of combined noise and random noise; the random noise is applied by randomly using one of the noise types

Combination	MSE↓	PSNR↑	SSIM↑
Gaussian+Gaussian noise	0.00047	33.88	0.9926
Gaussian+Mean	0.00039	34.95	0.9941
Gaussian noise+Sharpening	0.00019	37.86	0.9971
Gaussian+Mean	0.00026	36.90	0.9959
Mean+Gaussian noise	0.00063	32.51	0.9899
Gaussian+Sharpening	0.00017	38.46	0.9972
Random	0.00030	36.08	0.9951

Table 6 Our results and SteGAN under n=2,3,4

n	StegGAN		**Ours**	
	Cover/Stego	Secret/Recovery	Cover/Stego	Secret/Recovery
2	33.21	21.41	41.96	34.65
3	31.35	18.56	41.56	32.33
4	29.87	16.18	41.46	30.85

Fig. 6 shows examples of cover/carrier and secret/recovery pairs constructed by SRIS-Net and StegGAN (Singh et al., 2022) when embedding 2 images. Despite some distortion, the secret image remains relatively clear with SRIS-Net. Fig. 7 visualizes pairs generated by our method with 4 embedded images, demonstrating that the images are still relatively clear. These results indicate that SRIS-Net can achieve a maximum payload of approximately 72−96 bpp, demonstrating superior capacity both visually and quantitatively.

4.3 Ablation study

Effectiveness of the dMSubNet. To verify the effectiveness of the DMSubNet, we constructed one variation, SSubNet: the single-scale reconstruction subnetwork maintains a fixed inter-block scale and includes only one branch of 3×3 as the basic block, removing the SFM inter-scale correlation fusion module. For shallow and deep feature fusion, residual addition is used. As shown in the first and second rows of Table 7, the PSNR value increases by 2.31 dB and 4.31 dB for cover/carrier and secret/recovery pairs, and the error rate of CSR increases by 8.84%, respectively, indicating that the DMSubNet with the dual multi-scale network design is more effective in hiding information.

Effectiveness of the GLEM. The DMSubNet fuses the cover image and secret image by adding directly. However, DMSubNet+GLEM uses the GLEM module for fusion. As shown in the second and third rows of Table 7, the PSNR values increase by 0.34 dB and 0.15 dB for cover/carrier and secret/recovery pairs, respectively. The ability to resist CSR and SRM detection has been improved by 4.14% and 2.66%, respectively.This improvement is due to the global-local adaptive embedding, which helps the secret signals better adapt to the structure and details of the cover image. Thus, the GLEM significantly enhances embedding, extraction performance, and security.

Effectiveness of the dual-task discriminator. Although our network achieves high image quality without the discriminator, it is vulnerable to steganalysis detection. The misdetection rate by SRM (Fridrich and Kodovsky, 2012) is only 6.30%. To demonstrate the guidance provided by our dual-task discriminator, we compared DMSubNet+GLEM+GAN, which uses only a plain discriminator, with our approach. As shown in the third and fourth rows of Table 7, there is a reduction in visual quality due to conflicting metrics between vi-

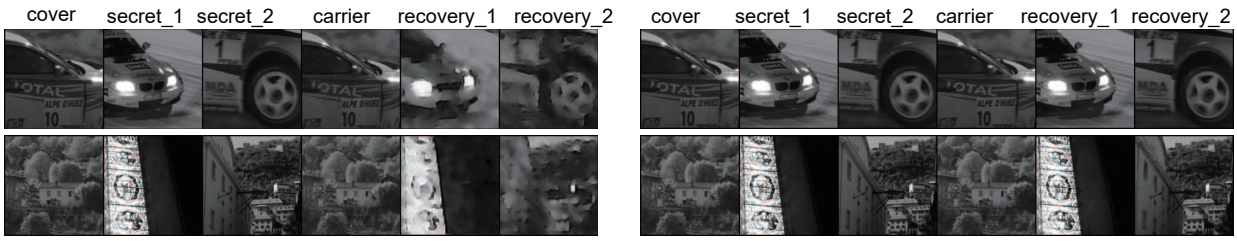

Fig. 6 SteGAN vs. SRIS-Net, when the capacity n=2. On the left is SteGAN, the right is SRIS-Net

Fig. 7 SRIS-Net method of embedding 4 secret images on 1 cover image. Obviously, the recovery images are all still clearly visible

sual quality and resistance to steganalysis. However, resistance to SRM steganalysis detection improves by 13.5%. Our dual-task discriminator iteratively guides embedding using the discriminator feature map. Comparing the fourth and fifth rows of Table 7, visual quality improves by 0.24 dB and 0.54 dB, respectively, relative to DMSubNet+GLEM+GAN. Additionally, in terms of undetectability, the false detection rates for CSR (Denemark et al., 2014) and SRM (Fridrich and Kodovsky, 2012) are 40.60% and 24.38%, respectively, improved by 4.36% and 4.58%. These experiments illustrate that our dual-task discriminator effectively improves the security and quality of the carrier image by providing valuable embedding guidance to the generator. We randomly show the visual effect of 4 sets of image pairs constructed by our method in Fig. 8. It is difficult to distinguish them from the appearance of the images. For better observation, we magnified the residuals by 5, and still, hardly any differences are visible. The frequency histograms of these images are shown in Figs. 8e–h. The horizontal axis ranges from 0 to 40, covering almost all pixel values in the residual image, and the vertical axis represents the frequency of each pixel value; the higher frequency of low pixel values indicates higher similarity between the carrier and cover images. The average pixel values of the residual maps are given, with the lowest being 0.0968 and the highest 1.0457. These quantitative and qualitative results demonstrate the effectiveness of our proposed algorithm.

5 Conclusions

This paper proposes a robust image steganography algorithm, SRIS-Net, based on feature score maps. SRIS-Net integrates spatial and frequency-domain features, employing a progressive assisted hiding strategy. It utilizes a global-local embedding module (GLEM) and the dual multi-scale aggrega-

Table 7 Results of ablation experiments. The sixth row represents SRIS-Net

Method	Cover/Stego					Secret/Recovery					Steganalysis	
	MSE↓	PSNR↑	SSIM↑	MS-SSIM↑	SCC↑	MSE↓	PSNR↑	SSIM↑	MS-SSIM↑	SCC↑	SRM	CSR
SSubNet	0.00009	41.18	0.9986	0.9983	0.9969	0.00025	36.78	0.9964	0.9932	0.9931	1.32%	18.90%
DMSubNet	0.00006	43.49	0.9989	0.9989	0.9979	0.00010	41.09	0.9983	0.9973	0.9963	3.64%	27.74%
DMSubNet+GLEM	0.00006	43.83	0.9990	0.9990	0.9980	0.00010	41.24	0.9984	0.9972	0.9963	6.30%	31.88%
DMSubNet+GLEM+GAN	0.00006	43.31	0.9989	0.9989	0.9977	0.00014	39.24	0.9976	0.9949	0.9947	19.80%	36.24%
Ours	0.00006	43.55	0.9990	0.9990	0.9979	0.00013	39.78	0.9979	0.9956	0.9952	24.38%	40.60%

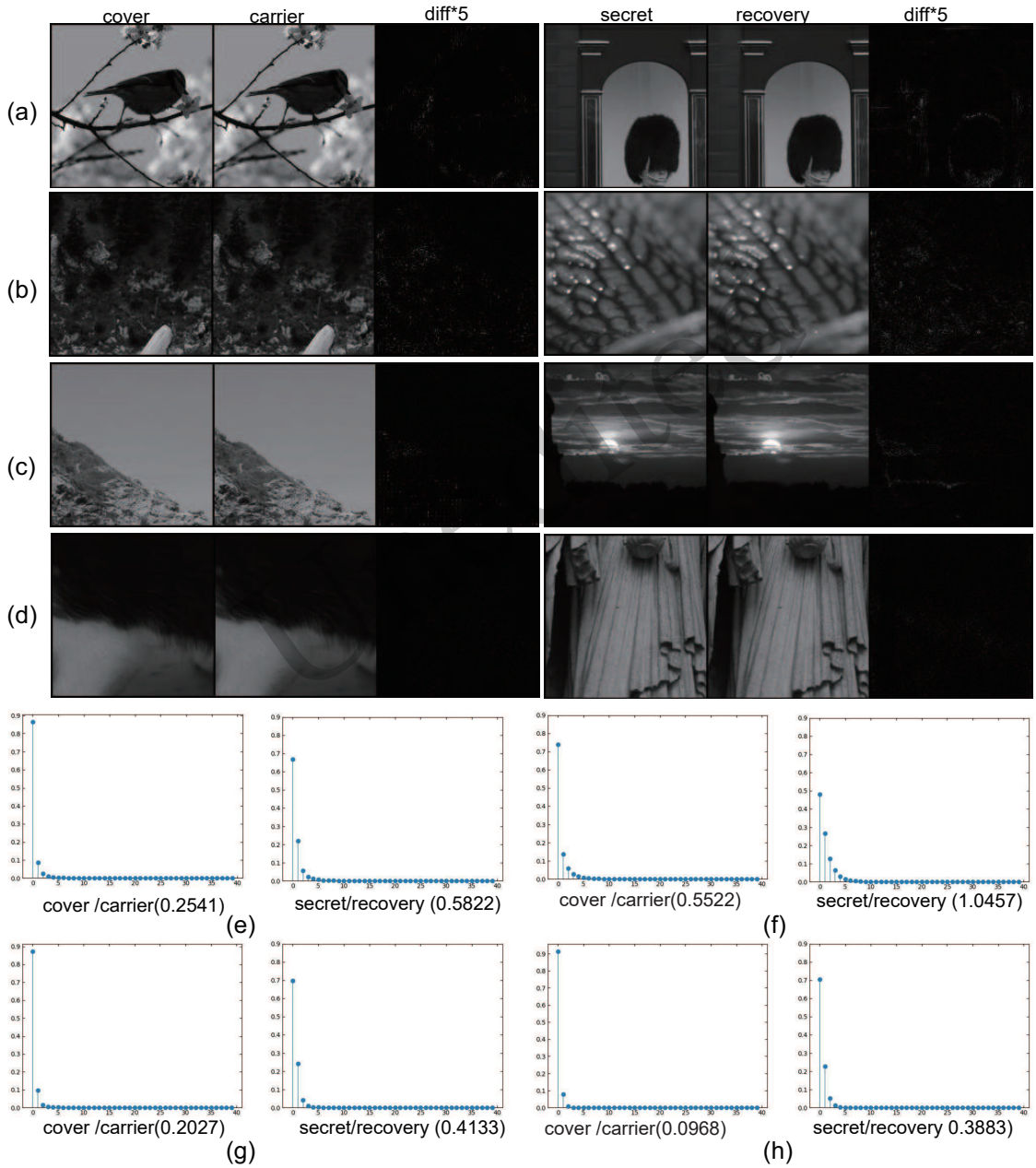

Fig. 8 Four sets of cover/carrier pairs and the secret/recovery pairs generated by our method are shown randomly; we magnify the residuals of each set of image pairs by a factor of 5 to make the difference almost invisible. We show the residual frequency histograms of the image pairs, where (e), (f), (g), and (h) correspond to (a), (b), (c), and (d)

tion subnetwork (DMSubNet) for progressive embedding multi-scale feature reconstruction. These techniques minimize the impact of hidden information on the cover image,enhancing the robustness, visual quality, and security of the carrier image. The proposed dual-task discriminator structure assesses real/fake images and generates feature score maps of the cover image's region of interest (ROI), guiding the embedding module to achieve higher imperceptibility and undetectability. Extensive experiments demonstrate the SRIS-Net superior performance in terms of capacity, visual quality, security, and robustness, validating the efficacy of the algorithm.

Contributors

Ai XIAO designed the research and drafted the manuscript. Zhi LI, Guomei WANG, Long ZHENG and Haoyuan SUN helped organize the manuscript. Zhi LI revised and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Baluja S, 2017. Hiding images in plain sight: deep steganography. Proc 31st Int Conf on Neural Information Processing Systems, p.2066-2076.

Baluja S, 2020. Hiding images within images. *IEEE Trans Pattern Anal Mach Intell*, 42(7):1685-1697. https://doi.org/10.1109/TPAMI.2019.2901877

Barni M, Bartolini F, Piva A, 2001. Improved wavelet-based watermarking through pixel-wise masking. *IEEE Trans Image Process*, 10(5):783-791. https://doi.org/10.1109/83.918570

Bas P, Filler T, Pevný T, 2011. "Break our steganographic system": the ins and outs of organizing boss. Proc 13th Int Conf on Information Hiding, p.59-70. https://doi.org/10.1007/978-3-642-24178-9_5

Cheddad A, Condell J, Curran K, et al. 2010. Digital image steganography: survey and analysis of current methods. *Signal Process*, 90(3):727-752. https://doi.org/10.1016/j.sigpro.2009.08.010

Chen BJ, Wang JX, Chen YY, et al., 2020. High-capacity robust image steganography via adversarial network. *KSII Trans Int Inf Syst*, 14(1):366-381. https://doi.org/10.3837/tiis.2020.01.020

Denemark T, Fridrich J, Holub V, 2014. Further study on the security of s-UNIWARD. *Med Waterm, Secur, Forensics,* 9028:902805. https://doi.org/10.1117/12.2044803

Duan XT, Jia K, Li BX, et al., 2019. Reversible image steganography scheme based on a U-Net structure. *IEEE Access*, 7:9314-9323. https://doi.org/10.1109/ACCESS.2019.2891247

Duan XT, Gou MX, Liu N, et al., 2020a. High-capacity image steganography based on improved Xception. *Sensors*, 20(24):7253. https://doi.org/10.3390/s20247253

Duan XT, Liu N, Gou MX, et al., 2020b. SteganoCNN: image steganography with generalization ability based on convolutional neural network. *Entropy*, 22(10):1140. https://doi.org/10.3390/e22101140

Fridrich J, Kodovsky J, 2012. Rich models for steganalysis of digital images. *IEEE Trans Inf Forensics Secur*, 7(3):868-882. https://doi.org/10.1109/TIFS.2012.2190402

Fu ZJ, Wang F, Cheng X, 2020. The secure steganography for hiding images via GAN. *EURASIP J Image Video Process*, 2020(1):46. https://doi.org/10.1186/s13640-020-00534-2

Holub V, Fridrich J, Denemark T, 2014. Universal distortion function for steganography in an arbitrary domain. *EURASIP J Inf Secur*, 2014(1):1-13. https://doi.org/10.1186/1687-417X-2014-1

Hu DH, Wang L, Jiang WJ, et al., 2018. A novel image steganography method via deep convolutional generative adversarial networks. *IEEE Access*, 6:38303-38314. https://doi.org/10.1109/ACCESS.2018.2852771

Huang JJ, Cheng SY, Lou SH, et al., 2019. Image steganography using texture features and GANS. Proc Int Joint Conf on Neural Networks, p.1-8. https://doi.org/10.1109/IJCNN.2019.8852252

Isola P, Zhu JY, Zhou TH, et al., 2017. Image-to-image translation with conditional adversarial networks. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.1125-1134. https://doi.org/10.1109/CVPR.2017.632

Jing JP, Deng X, Xu M, et al., 2021. HiNet: deep image hiding by invertible network. Proc IEEE/CVF Int Conf on Computer Vision, p.4733-4742. https://doi.org/10.1109/ICCV48922.2021.00469

Johnson J, Alahi A, Li FF, 2016. Perceptual losses for real-time style transfer and super-resolution. Proc 14th European Conf on Computer Vision, p.694-711. https://doi.org/10.1007/978-3-319-46475-6_43

Kodovsky J, Fridrich J, Holub V, 2012. Ensemble classifiers for steganalysis of digital media. *IEEE Trans Inf Forensics Secur*, 7(2):432-444. https://doi.org/10.1109/TIFS.2011.2175919

Lai WS, Huang JB, Ahuja N, et al., 2019. Fast and accurate image super-resolution with deep Laplacian pyramid networks. *IEEE Trans Pattern Anal Mach Intell*, 41(11):2599-2613. https://doi.org/10.1109/TPAMI.2018.2865304

Li XL, Yang B, Cheng DF, et al., 2009. A generalization of LSB matching. *IEEE Signal Process Lett*, 16(2):69-72. https://doi.org/10.1109/LSP.2008.2008947

Li ZZ, Yang XY, Shen KQ, et al., 2022. Dual branch parallel steganographic framework based on multi-scale distillation in framelet domain. *Neurocomputing*, 514:182-194. https://doi.org/10.1016/j.neucom.2022.09.146

Liu LS, Meng LZ, Wang XL, et al., 2022. An image steganography scheme based on ResNet. *Multimed Tools Appl*, 81(27):39803-39820. https://doi.org/10.1007/s11042-022-13206-2

Lu SP, Wang R, Zhong T, et al., 2021. Large-capacity image steganography based on invertible neural networks. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.10816-10825. https://doi.org/10.1109/CVPR46437.2021.01067

Mao XD, Li Q, Xie HR, et al., 2017. Least squares generative adversarial networks. Proc IEEE Int Conf on Computer Vision, p.2794-2802. https://doi.org/10.1109/ICCV.2017.304

Otazu X, Gonzalez-Audicana M, Fors O, et al., 2005. Introduction of sensor spectral response into image fusion methods. *IEEE Trans Geosci Remote Sens*, 43(10):2376-2385. https://doi.org/10.1109/TGRS.2005.856106

Ruanaidh JJKO, Dowling WJ, Boland FM, 1996. Phase watermarking of digital images. Proc 3rd IEEE Int Conf on Image Processing, p.239-242. https://doi.org/10.1109/ICIP.1996.560428

Shi HC, Dong J, Wang W, et al., 2018. SSGAN: secure steganography based on generative adversarial networks. Proc 18th Pacific-Rim Conf on Multimedia on Advances in Multimedia Information Processing, p.534-544. https://doi.org/10.1007/978-3-319-77380-3_51

Shi WZ, Caballero J, Huszár F, et al., 2016. Real-time single image and video super-resolution using an efficient subpixel convolutional neural network. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.1874-1883. https://doi.org/10.1109/CVPR.2016.207

Singh B, Sharma PK, Huddedar SA, et al., 2022. StegGAN: hiding image within image using conditional generative adversarial networks. *Multimed Tools Appl*, 81(28):40511-40533. https://doi.org/10.1007/s11042-022-13172-9

Tamimi AA, Abdalla AM, Al-Allaf O, 2013. Hiding an image inside another image using variable-rate steganography. *Int J Adv Comput Sci Appl*, 4(10):18-21. https://doi.org/10.14569/IJACSA.2013.041004

Tancik M, Mildenhall B, Ng R, 2020. StegaStamp: invisible hyperlinks in physical photographs. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.2117-2126. https://doi.org/10.1109/CVPR42600.2020.00219

ur Rehman A, Rahim R, Nadeem S, et al., 2018. End-to-end trained CNN encoder-decoder networks for image steganography. Proc Computer Vision-ECCV 2018 Workshops, p.723-729. https://doi.org/10.1007/978-3-030-11018-5_64

Van TP, Dinh TH, Thanh TM, 2019. Simultaneous convolutional neural network for highly efficient image steganography. Proc 19th Int Symposium on Communications and Information Technologies, p.410-415. https://doi.org/10.1109/ISCIT.2019.8905216

Wengrowski E, Dana K, 2019. Light field messaging with deep photographic steganography. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.1515-1524. https://doi.org/10.1109/CVPR.2019.00161

Wu P, Yang Y, Li XQ, 2018. StegNet: mega image steganography capacity with deep convolutional network. *Future Int*, 10(6):54. https://doi.org/10.3390/fi10060054

Xu GS, Wu HZ, Shi YQ, 2016. Structural design of convolutional neural networks for steganalysis. *IEEE Signal Process Lett*, 23(5):708-712. https://doi.org/10.1109/LSP.2016.2548421

Xu YM, Mou C, Hu YJ, et al., 2022. Robust invertible image steganography. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.7875-7884. https://doi.org/10.1109/CVPR52688.2022.00772

Yang H, Xu YT, Liu XH, et al., 2024. PRIS: practical robust invertible network for image steganography. *Eng Appl Artif Intell*, 133:108419. https://doi.org/10.1016/j.engappai.2024.108419

Ye J, Ni JQ, Yi Y, 2017. Deep learning hierarchical representations for image steganalysis. *IEEE Trans Inf Forensics Secur*, 12(11):2545-2557. https://doi.org/10.1109/TIFS.2017.2710946

Ying QC, Zhou H, Zeng XH, et al., 2022. Hiding images into images with real-world robustness. Proc IEEE Int Conf on Image Processing, p.111-115. https://doi.org/10.1109/ICIP46576.2022.9897931

Yu C, 2020. Attention based data hiding with generative adversarial networks. *Proc AAAI Conf Artif Intell*, 34(01):1120-1128. https://doi.org/10.1609/aaai.v34i01.5463

Zamir SW, Arora A, Khan S, et al., 2022. Restormer: efficient transformer for high-resolution image restoration. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.5728-5739. https://doi.org/10.1109/CVPR52688.2022.00564

Zhang L, Lu Y, Li J, et al., 2023. Deep adaptive hiding network for image hiding using attentive frequency extraction and gradual depth extraction. *Neural Comput Appl*, 35(15):10909-10927. https://doi.org/10.1007/s00521-023-08274-w

Zhang R, Dong SQ, Liu JY, 2019. Invisible steganography via generative adversarial networks. *Multimed Tools Appl*, 78(7):8559-8575. https://doi.org/10.1007/s11042-018-6951-z

Zhang R, Zhu F, Liu JY, et al., 2020. Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis. *IEEE Trans Inf Forensics Secur*, 15:1138-1150. https://doi.org/10.1109/TIFS.2019.2936913

Zhou DQ, Hou QB, Chen YP, et al., 2020. Rethinking bottleneck structure for efficient mobile network design. Proc 16th European Conf on Computer Vision, p.680-697. https://doi.org/10.1007/978-3-030-58580-8_40