remote sensing

MDPI

Article

Remote Sensing Images Secure Distribution Scheme Based on Deep Information Hiding

Peng Luo [1,2], Jia Liu [2,*], Jingting Xu [1], Qian Dang [2] and Dejun Mu [1]

[1] School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710072, China; lp_nwpu@mail.nwpu.edu.cn (P.L.); xujingting@mail.nwpu.edu.cn (J.X.); mudejun@mail.nwpu.edu.cn (D.M.)
[2] School of Cryptography Engineering, Engineering University of PAP, Xi'an 710086, China; dangqian@mail.nwpu.edu.cn
* Correspondence: liujia1022@gmail.com

Abstract: To ensure the security of highly sensitive remote sensing images (RSIs) during their distribution, it is essential to implement effective content security protection methods. Generally, secure distribution schemes for remote sensing images often employ cryptographic techniques. However, sending encrypted data exposes communication behavior, which poses significant security risks to the distribution of remote sensing images. Therefore, this paper introduces deep information hiding to achieve the secure distribution of remote sensing images, which can serve as an effective alternative in certain specific scenarios. Specifically, the Deep Information Hiding for RSI Distribution (hereinafter referred to as DIH4RSID) based on an encoder–decoder network architecture with Parallel Attention Mechanism (PAM) by adversarial training is proposed. Our model is constructed with four main components: a preprocessing network (PN), an embedding network (EN), a revealing network (RN), and a discriminating network (DN). The PN module is primarily based on Inception to capture more details of RSIs and targets of different scales. The PAM module obtains features in two spatial directions to realize feature enhancement and context information integration. The experimental results indicate that our proposed algorithm achieves relatively higher visual quality and secure level compared to related methods. Additionally, after extracting the concealed content from hidden images, the average classification accuracy is unaffected.

Keywords: remote sensing image; distribution; deep information hiding; attention mechanism

Citation: Luo, P.; Liu, J.; Xu, J.; Dang, Q.; Mu, D. Remote Sensing Images Secure Distribution Scheme Based on Deep Information Hiding. *Remote Sens.* **2024**, *16*, 1331. https://doi.org/10.3390/rs16081331

Academic Editors: Hossein M. Rizeei, Qi Zhao, Guangliang Cheng and Paolo Tripicchio

Received: 8 January 2024
Revised: 25 March 2024
Accepted: 2 April 2024
Published: 10 April 2024

1. Introduction

As advancements in spatial, informational, and communication technologies persist, the primary method for applying spatial information has shifted towards the utilization of digital products and network integration [1], which has greatly facilitated the communication and sharing of remote sensing information, allowing for further utilization of remote sensing images. However, it has also introduced security concerns during distributing such information through networks [2]. RSIs obtain data from vulnerable sites like military installations, oil fields, and airports, posing potential misuse and theft risks. Moreover, processing these images without implementing proper security measures can simplify unauthorized data retrieval and offer uninterrupted access to confidential information, which could subsequently be used for illicit purposes [3]. Adopting a robust content security protection strategy is essential to safeguard the confidentiality of highly sensitive remote sensing data. Encryption is commonly acknowledged as a prevalent technique for protecting the confidentiality of remote sensing images during their transmission and storage [4]. That is, using symmetric or public key cryptographic algorithms to transform plaintext into ciphertext, then transmitting it through a public channel to the recipient. Because ciphertext presents itself as unintelligible or scrambled text, it readily draws the focus of attackers monitoring the communication channel, which can lead to two consequences:

firstly, if the attacker intercepts the ciphertext, they can attempt to decrypt it using various attack methods against different cryptographic systems [5].

Secondly, if the attacker cannot decrypt the ciphertext, they may disrupt the channel to prevent distribution. Indeed, exploring alternative methods for the secure distribution of RSIs in certain cases is necessary. The information-hiding technology in information security has begun to attract increasingly more attention [6,7]. Unlike encryption, in an information-hiding scheme, secret information is embedded in seemingly harmless host information, and attackers cannot intuitively determine whether the information they are monitoring contains secret information. In other words, hosts containing hidden information will not attract the attacker's attention and suspicion. The purpose of information hiding is to make enemies unaware of where there are secrets, as it hides the form of information that exists. The comparison between the idea of our scheme and the traditional encryption distribution method is shown in Figure 1. Although there is relatively little publicly published literature on the use of information-hiding technology for remote sensing image distribution, some scholars' research results [8–12] provide us with good insights. These methods mainly achieve the hiding of one image within another of the same size. However, they do not address discussions and research regarding RSIs. Due to the characteristics of RSIs, such as being taken from a long distance and at a steep angle, they do, indeed, contain complex backgrounds and scattered small objects, which are significantly different from natural images captured in everyday life [13], as shown in Figure 2. Therefore, this paper draws on their ideas to design a new algorithm suitable for the secure distribution task of RSIs. Although there are many categories of RSIs, such as hyperspectral RSIs, multispectral RSIs, and panchromatic RSIs, this study only investigated the multispectral images that have been processed into RGB format after compression. The primary contributions of our proposed scheme are as follows:

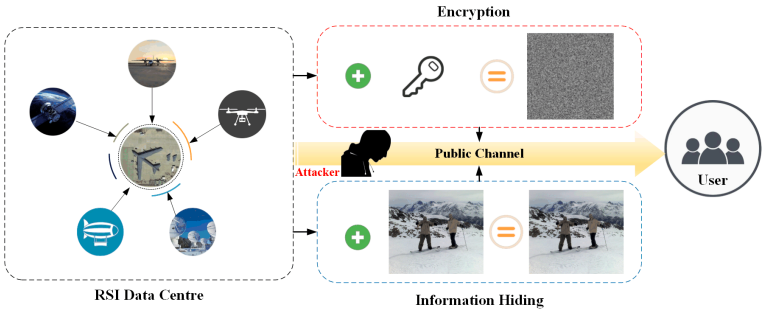

Figure 1. The comparison between steganography and the traditional encryption distribution methods. The encrypted RSI is presented in garbled code, which is easy for attackers to pay attention to. In steganography, the remote sensing image is hidden in the ordinary image in an invisible form, which can be more effective and secure for distribution.

Figure 2. By comparing RSIs (from NWPU-RESISC45 dataset [14]) with regular natural images (from ImageNet dataset [15]), it is evident that RSIs contain complex backgrounds and scattered small objects, whereas regular images typically have simpler backgrounds and more distinct foreground elements.

- To our knowledge, DIH4RSID is the first to explicitly propose using deep information-hiding technology to ensure the secure distribution of RSIs. Therefore, our method opens a new way of thinking about RSI security distribution and expands the application fields of information-hiding technology.
- Unlike the existing HIWI (Hiding images within images) framework, our study proposes a novel preprocessing network architecture, which is designed based on Inception networks and crafted to conform to the unique properties of remote sensing images, which can capture detailed information of objects at different scales.
- According to the characteristics of our tasks and RSIs, a new attention mechanism, PAM, is designed in this paper, which carries out two kinds of pooling from two dimensions, respectively. Convolution operations can then capture cross-channel relationships and spatial remote dependencies.
- A discriminator is added to the scheme, and iterative training is carried out by WGAN-GP (Wasserstein Generative Adversarial Network with Gradient Penalty), which improves the model's stability and correct convergence speed.
- In this study, a functional integrity retention test was performed on the extracted RSI, specifically an accuracy test of scene classification for the RSI after extraction. This offers a new perspective for assessing the performance of high-capacity information-hiding technologies.

2. Related Work

The approach in this article draws on many concepts and methods from the field of image information hiding; therefore, our analysis of related work is confined to the domain of information hiding. The origins of information hiding technology can be traced back to ancient covert means of communication, such as invisible ink and miniaturized fonts. With time, information hiding has gradually changed from traditional physical means to digital technology. Based on the context of the application, information hiding can be divided into digital steganography and digital watermarking. Generally speaking, the former is mainly used for covert communication, while the latter is used for copyright protection. According to such categorization, our DIH4RSID is essentially an application of steganography in RSI distribution. The research on steganography can be roughly divided into three stages. Early information hiding technologies were mainly based on non-adaptive hiding strategies, among which the most typical representative is LSB (Least Significant Bit) [16]. LSB is a steganography method that modifies and stores information based on the least significant bit of an image. Using the insensitivity of human eyes to color differences, the secret information is put into the least significant bit of the picture by a certain embedding method so that the information we need to hide is put into the least significant bit of the picture by a certain method. Because non-adaptive steganography does not consider the characteristics of the cover image itself, it is not safe and is easy to detect and analyze. Based on this, adaptive steganography came into being, representing the second stage of steganography. Adaptive steganography considers the properties of the cover image itself, such as the texture information and edge information of the image content. According to the characteristic of difficulty in detecting complex areas of image texture, secret information is selectively embedded into areas with complex textures or rich edges of the cover, which improves the anti-steganographic detection ability of loaded images. At the same time, all kinds of adaptive steganography algorithms are combined with STC (Syndrome trellis codes) [17] encoding methods. The difference is that the distortion function is different. Such algorithms are represented by HUGO [18], WOW [19], UNIWARD [20], and HILL [21]. Although adaptive steganography methods have achieved high performance, they are confronted with several challenges in terms of both content-adaptive and statistics-based approaches. Firstly, such algorithms can only embed a small number of bits or text information and cannot embed multimedia information such as images [12]. At the same time, these methods often require specialized knowledge to design elaborate distortion cost functions. With the continuous development of analysis algorithms based on deep learning,

the security of these traditional human-designed information hiding algorithms faces great challenges, making researchers turn their attention to deep learning and attempt to use deep learning's powerful feature fusion ability to realize information hiding. Frameworks for information hiding based on deep learning, such as HiDDeN [22] and SteganoGAN [9], have been developed to accomplish the tasks of hiding and extracting information. This development signifies the progression of information hiding into its third phase, known as deep information hiding. These frameworks eliminate the need for the manual design of embedding strategies and achieve higher payloads. However, they still only enable the covert transmission of small data. To address the challenge of hiding large image data, Baluja [23] presented a system to embed a full-color image into another of identical size while minimizing the quality degradation of both images, which is achieved by concurrently training deep neural networks to carry out both the embedding and extraction processes, which are specifically tailored to function in tandem. While this approach represents a significant innovation and yields impressive visual results, its robustness against analytical attacks leaves something to be desired. Rehman et al. [8] endeavored to develop an encoder–decoder architecture rooted in convolutional neural networks, accomplishing complete network training by adopting a novel loss function. While this approach proficiently conserved the fidelity of the concealed image, the visual quality of the crafted stego was subpar. In addition, Duan et al. [24] introduced a reversible information-hiding network that utilizes a U-Net architecture to improve the hiding performance. The approach yielded pleasing outcomes in synthesizing concealed images as well as in the accurate retrieval of secret images. Nonetheless, their research did not delve into an in-depth examination of security concerns. Chen et al. [12] posited that certain secret images might possess intricate spatial characteristics. To address this, he proposed a multi-tiered robust auxiliary module to augment the feature representation, subsequently elevating the restoration quality of secret images. However, due to the absence of a discriminator within the framework, the enhancement in performance was not markedly evident. Some researchers introduced the attention mechanism into the field of deep information hiding [10,25–27], and promising results have been achieved. Tan et al. [26] proposed a new end-to-end image network architecture based on a channel–attention mechanism that generates adversarial networks, which can produce perceptively indistinguishable stego of different capacities. However, their programs suitable for embedding and transmitting binary text are not directly applicable to RSIs due to their inherent nature. According to the above analysis, the existing method cannot be directly applied to the safe distribution of remote sensing images, so it needs to be further modified to adapt to this task.

3. Proposed Scheme

In this section, the general structure of DIH4RSID is introduced first, and then, all parts of DIH4RSID are described in detail. Finally, we delineate the various loss functions and outline the training methods used.

3.1. Overview

As show in Figure 3, DIH4RSID's workflow can be described as follows. Firstly, the sender (RSI's owner) extracts feature maps through a preprocessing network and then hides that into a nature image (cover) by an embedding network, which outputs a stego (also called the hidden image). The stego could be transmitted through public channels to the receiver, who decodes the information by an extracting network to reconstruct the RSI. During the embedding process, a discriminating network acts as an attacker to improve the indistinguishability between the cover and stego.

The secure distribution schema for RSI based on information hiding must satisfy three properties: (1) to avoid attracting the attention of attackers, an embedding process should have a minimal visual impact on the cover; (2) in order to enhance the practicality of the algorithm, the semantics of the extracted RSI should be preserved, and (3) to improve the security of the model, it is important to restrict the success rate of detecting algorithms to a

minimal level. Considering these three requirements, we adopted three unique designs based on the traditional encoder–decoder network, namely, the preprocessing network based on an Inception structure, a PAM, and a discriminator based on XuNet [28].

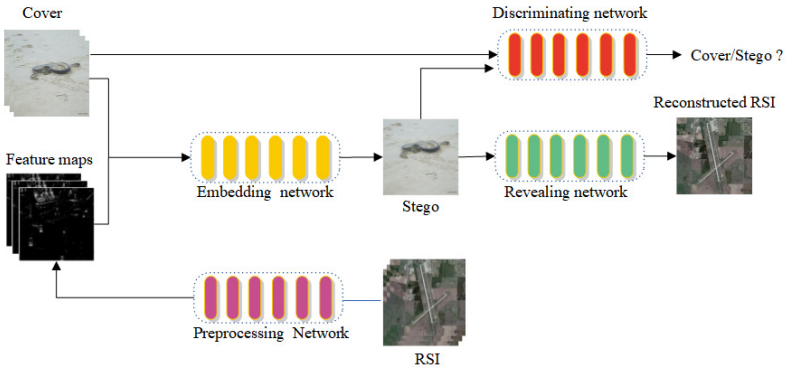

Figure 3. DIH4RSID flowchart: Sender (RSI's owner) transmits RSI to receiver.

3.2. Preprocessing Network

RSIs feature large-scale variations in objects, rich details, complex structures, and ambiguous distinctions between subjects and backgrounds. Baluja's [23] preprocessing network is not very suitable for RSIs, as these images require multiple convolutional kernels of different sizes to capture the features of targets at different scales. To address this issue, we integrate the Inception [29] structure into the design of the preprocessing network, as depicted in Figure 4. The Inception structure is a concept popularized by the inception architecture of convolutional neural networks, which employs three distinct sizes of convolutional kernels and one max pooling in parallel to obtain spatial features at different scale-receptive fields. In the parallel branches, dimensions are first adjusted through a 1×1 convolution and then subjected to convolution kernels of varying scales and pooling operations. Finally, the extracted multi-scale spatial features are fused through Concat operations, enhancing the network's ability to capture a more complete description of target spatial structural features at different depth levels.

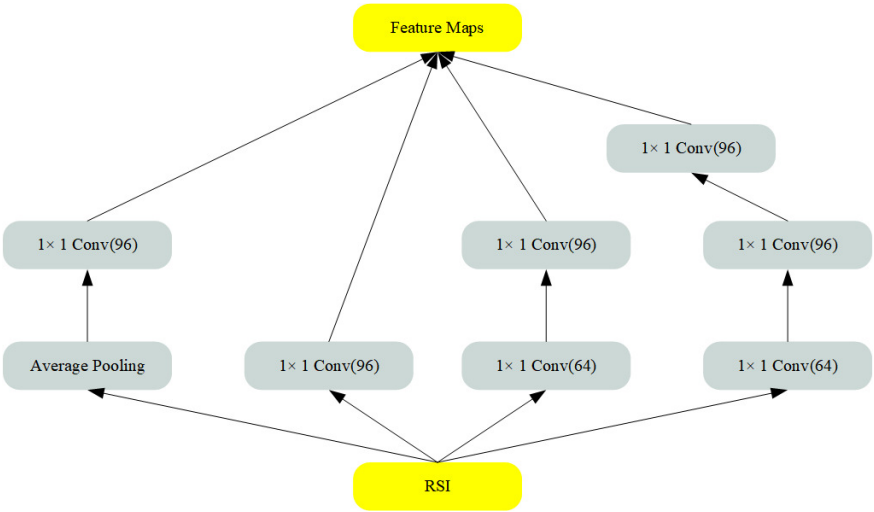

Figure 4. Preparing network architecture to extract the feature of RSI.

3.3. Parallel Attention Mechanism

In the realm of deep learning, the attention mechanism trains networks to concentrate on salient features while disregarding those that are irrelevant. For CNN (Convolutional Neural Network)-based information-hiding schemes that directly produce a stego, the secret information is ingrained through integration with the cover features. Given that the

value of these features varies with respect to information hiding, applying an attention mechanism could enhance the system's performance. Our task requires an accurate classification of the extracted remote sensing images, with a particular focus on preserving small targets in RSIs. Therefore, the existing Convolutional Block Attention Module (CBAM) [30] attention mechanism is not suitable for our task. Drawing inspiration from the Coordinate Attention [31] (CA) and Efficient Channel Attention [32] (ECA) mechanisms, we design the PAM to effectively grasp the inter-channel relationships and extensive spatial dependencies, incorporating precise positional details to enhance the embedding effect of remote sensing images and improve the accuracy of feature extraction. The detailed structure of PAM is shown in Figure 5.

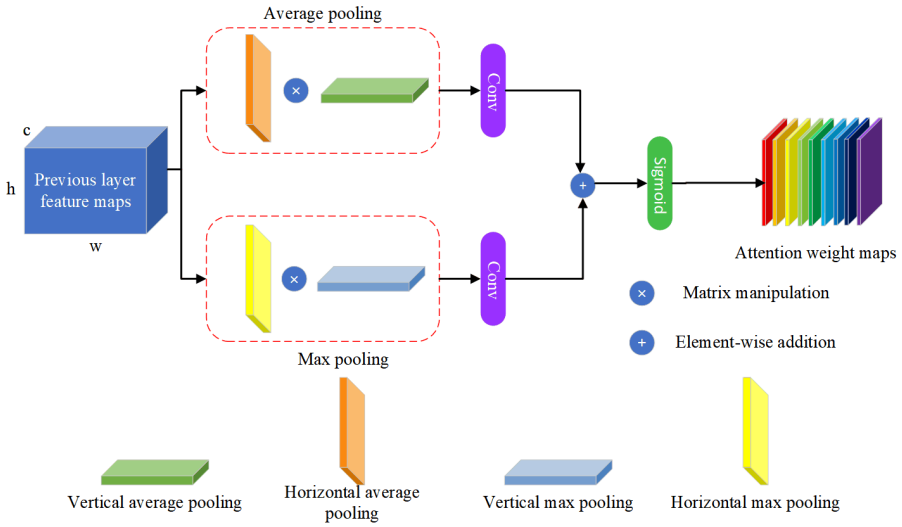

Figure 5. Parallel attention mechanism contains four types of pooling strategies.

Firstly, the horizontal average pooling, vertical average pooling, horizontal max pooling, and vertical max pooling are performed for each channel of the input feature map denoted as $F(h, w, c)$ (h denotes the height of the feature map, w denotes the width of the feature map, and c denotes the channel of the feature map), respectively, as are formulated by Equations (1)–(4).

$$F_{hap}(h, c) = \frac{1}{W} \sum_{i=1}^{w} F(h, i, c) \tag{1}$$

$$F_{vap}(w, c) = \frac{1}{h} \sum_{i=1}^{h} F(i, w, c) \tag{2}$$

$$F_{hmp}(h, c) = \max_{0 \le i < w} F(h, i, c) \tag{3}$$

$$F_{vmp}(w, c) = \max_{0 \le i < h} F(i, w, c) \tag{4}$$

Subsequently, to prevent the reduction in channel dimension during cross-channel interactions, we employ Conv2D with a flexible kernel size. This is designed to produce attention weights across both spatial dimensions. Additionally, we incorporate two distinct types of pooling strategies. These processes, collectively, can be expressed as Equations (5) and (6).

$$\hat{F}_{AP}(h, w, c) = \text{Conv2D}\left(F_{vap} \times F_{hap}, kernal\right) \tag{5}$$

$$\hat{F}_{MP}(h, w, c) = \text{Conv2D}\left(F_{hmp} \times F_{vmp}, kernal\right) \tag{6}$$

$$kernal = \left| \frac{\log_2(c)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{7}$$

According to ECA, the kernel of Conv2D is computed by (7), where $||_{odd}$ denotes the nearest odd number, and b and γ are set to 1 and 2.

Finally, to determine the weight of the coordinate attention, we combine the adaptive pooling features; $\hat{F}_{AP}$ and $\hat{F}_{MP}$ are combined using the operation of matrix multiplication. This procedure can be mathematically represented as Equation (8).

$$Weight_{pam}(h, w, c) = \text{sigmoid}\left(\hat{F}_{AP} + \hat{F}_{MP}\right) \tag{8}$$

Figure 5 displays the basic structure and workflow of PAM. The basic structure primarily consists of four different types of convolution, which yield the final feature map through matrix multiplication, convolution, and activation functions.

3.4. Embedding Network

We use a network structure similar to Dense Connection Architecture (DCA) as the backbone of the embedding network. There are several reasons why such architecture is suitable for DIH4RSID: (1) DCA reduces the problem of gradient disappearance by means of cross-layer connection so that information will not be lost during transmission. This is important for DIH4RSID because remote sensing images typically have large sizes and complex backgrounds that require the network to transmit and utilize feature information efficiently. (2) DCA can better utilize and enhance features in the transmission process by including the features of all previous layers in the input of each layer, which is also very beneficial for DIH4RSID because the features of RSI are often complex, and the network needs to be able to extract and transmit these features effectively. (3) DCA combines features of previous layers to form a more detailed description and discrimination of features. This enables the network to use the feature information more effectively, thus improving the hiding effect and extraction quality of DIH4RSID. (4) Reducing the number of parameters: DCA reduces the number of parameters in the model by reducing the number of connections, thus reducing the complexity and calculation cost of the model, which is also essential for DIH4RSID because RSI is usually huge and requires low model complexity and computational cost to achieve effective classification.

To maintain stealthiness, it is crucial for the stego image to bear a close resemblance to the cover image. Moreover, to ensure the integrity of the recovered image, it should be nearly identical to the RSI. Nevertheless, the process involving various convolutional and activation layers can unavoidably lead to a loss of information from input images like cover images and RSI, which is detrimental to both hiding and revealing capabilities. To mitigate this issue, introducing global and local skip connections is proposed, as shown in Figure 6. The global skip connection facilitates the direct transmission of original image data to the uppermost layer, enhancing edge and texture detail synthesis and boosting concealment and retrieval efficiency. On the other hand, the local skip connection allows for the unimpeded flow of RSI within the embedding network, ensuring that its details and semantic content are effectively incorporated into the stego. This incorporation aids the follow-up extraction network in precisely restoring the RSI. The embedding procedure is defined by a specific Equation (9).

$$Stego = EN(C, PN(RSI, \theta_{PN}), \theta_{EN}) \tag{9}$$

where EN denotes the Embedding Network, C denotes a Cover, PN denotes a Preprocessing Network, θ_{PN} denotes parameters of PN, and θ_{EN} denotes parameters of EN.

Figure 6 shows the main units of the encoder and the method of feature fusion. Due to space limitations, the complete structural diagram is not displayed. In fact, in our specific implementation, we use 16 convolution, normalization, and activation blocks.

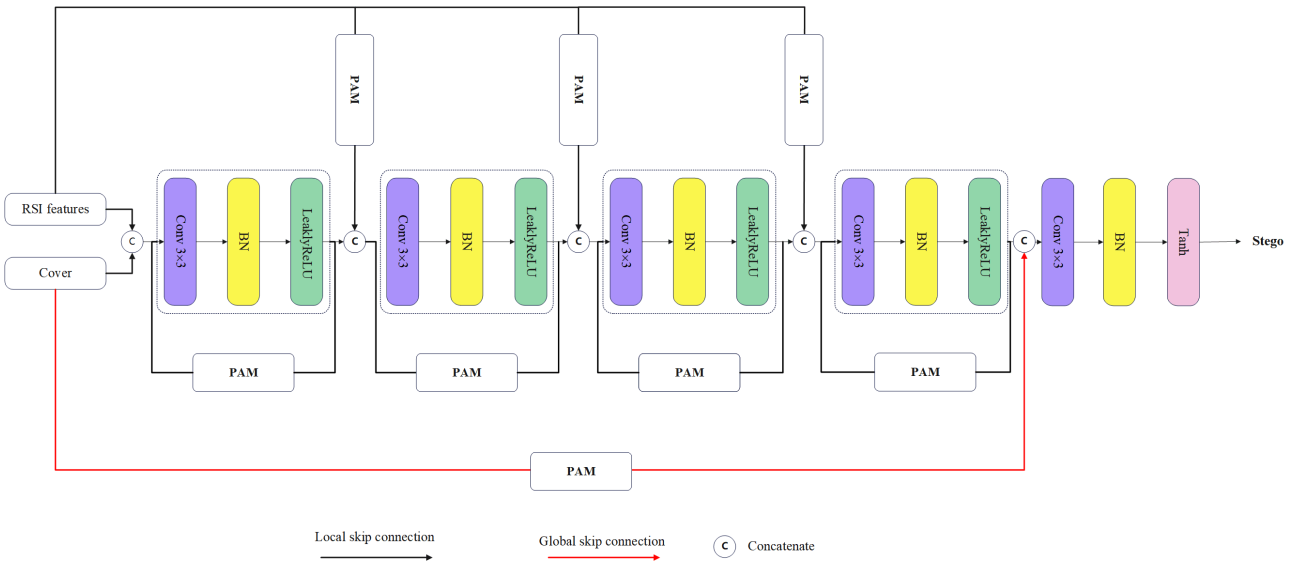

Figure 6. Embedding network architecture contains Convolution, Batch Normalization, LeaklyRelU, and PAM.

3.5. Revealing Network

RSI extraction is an inverse embedding process. The revealing network (RN) adopts a structure similar to the embedded network and still extracts depth features with dense connections, hoping to recover the RSI with a high accuracy. Unlike the embedding process, the revealing process does not consider the effect of RSI feature integration, so the modules of global skip connection and local skip connection used in the embedded network are not used in the RN. The process of revealing can be expressed by Equation (10) and is illustrated in Figure 7.

$$\hat{RSI} = RN(Stego, \theta_{RN}) \tag{10}$$

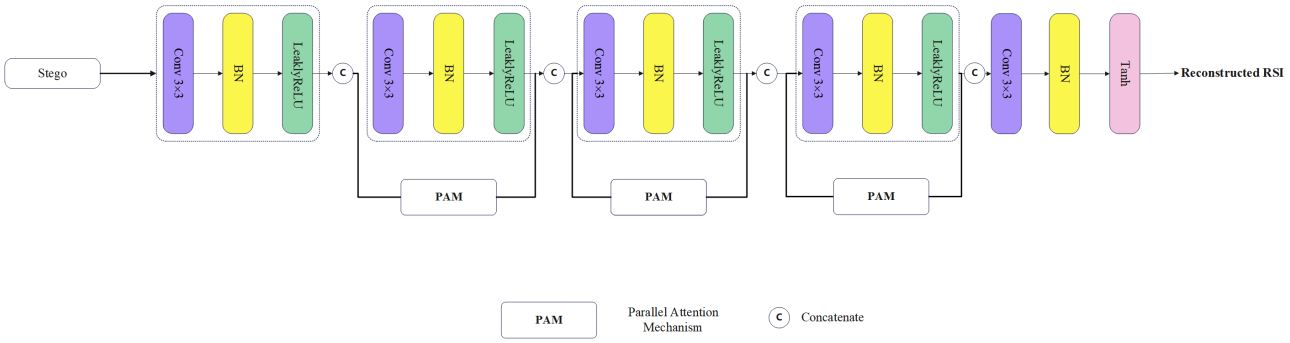

Figure 7. Revealing network add parallel attention mechanism to improve the extracted quality of a secret image.

3.6. Discriminating Network

The primary purpose of this study was to realize the safe distribution of remote sensing images, to ensure visual imperceptibility, and to ensure that there are no apparent statistical distribution anomalies. To achieve these, we add an authentication network to the entire training process, which is described later in Algorithm 1. In the selection of the discriminator network module, we mainly use the design idea of the XuNet [28] network as a reference but we make some changes. The first is to increase the number of nodes at the input side to handle the stego with three dimensions, and the second is to add an ASPP [33] pooling technology, which uses spatial pyramid pooling to obtain the context information of different scales and residual information to improve the model's ability to

perceive and distinguish stego. The details of the DN are listed in Table 1, and the process of the DN can be formulated by Equation (11).

$$\text{Probabilities of classes} = DN(Stego, \theta_{RN}) \tag{11}$$

Table 1. Overall architecture details of the revealing network.

Inputs	Modules	Kernel	Outputs
Stego ($3 \times 256 \times 256$)	HPF	$3 \times 5 \times 5$	Out1 ($3 \times 256 \times 256$)
Input1	Conv-ABS-BN-Tanh—Average	$3 \times 5 \times 5$	Out2 ($8 \times 128 \times 128$)
Input2	Conv-BN-Tanh—Average	$8 \times 5 \times 5$	Out3 ($16 \times 64 \times 64$)
Input3	Conv-BN-Tanh—Average	$16 \times 1 \times 1$	Out4 ($32 \times 32 \times 32$)
Input4	Conv-BN-Tanh—Average	$32 \times 1 \times 1$	Out5 ($64 \times 16 \times 16$)
Input5	Conv-BN-Tanh—Average	$64 \times 1 \times 1$	Out5 ($128 \times 8 \times 8$)
Input6	ASPP	$3 \times 3, 1 \times 1$	Out5 ($2560 \times 1 \times 1$)
Input7	Fully Connected	-	Out5 ($2 \times 1 \times 1$)
Input8	SoftMax	-	Probabilities of classes ($2 \times 1 \times 1$)

Algorithm 1: Training DIH4RSID. We use default values of $\lambda = 10, n_{\text{DN}} = 5$, $\alpha = 0.0001, \beta_1 = 0, \beta_2 = 0.9$

Input: Cover dataset $\mathcal{X}$, RSI dataset $\mathcal{R}$; initial PN parameters w_{PN}, initial EN parameters w_{EN}, initial DN parameters w_{DN}, initial RN parameters w_{RN}; batch size m, DN iterations per EN iteration n_{DN}; the gradient penalty coefficient λ, the number of DN iterations per generator iteration n_{DN}, the batch size m, Adam hyperparameters α, β_1, β_2., the total iterations N

Output: Trained parameters w_{PN}, w_{EN}, w_{DN}, w_{RN}

for $t \leftarrow 0$ **to** N **do**
  for $i \leftarrow 0$ **to** n_{DN} **do**
    $\mathcal{L}^{(i)} \leftarrow 0$
    for $j \leftarrow 0$ **to** m **do**
      Sample $x \in \mathcal{X}$
      Sample $r \in \mathcal{R}$
      Sample $\epsilon \in \mathcal{U}[0, 1]$
      $\overline{x} \leftarrow EN(Concat(PN(r), x))$
      $\hat{x} \leftarrow \epsilon\overline{x} + (1 - \epsilon\overline{x})$
      $\mathcal{L}^{(i)} \leftarrow \mathcal{L}_d(x, \hat{x}) + \lambda(\|\nabla_{\hat{x}} DN_{w_{DN}}(\hat{x})\|_2 - 1)^2$
    end
    $w_{DN} \leftarrow Adam\left(\nabla_{w_{DN}} \frac{1}{m}\mathcal{L}^{(i)}, w_{DN}, \alpha, \beta_1, \beta_2\right)$
  end
  Randomly select $\{x_i\}_{i=1}^{m}$ a batch from $\mathcal{X}$
  Randomly select $\{r_i\}_{i=1}^{m}$ a batch from $\mathcal{R}$
  $w_{PN} \leftarrow Adam(\nabla_{w_{PN}} \mathcal{L}_{PN,EN,RN}, w_{PN}, \alpha, \beta_1, \beta_2)$
  $w_{EN} \leftarrow Adam(\nabla_{w_{EN}} \mathcal{L}_{PN,EN,RN}, w_{EN}, \alpha, \beta_1, \beta_2)$
  $w_{RN} \leftarrow Adam(\nabla_{w_{RN}} \mathcal{L}_{PN,EN,RN}, w_{RN}, \alpha, \beta_1, \beta_2)$
end

3.7. Loss Function Design

Our framework is similar to a GAN structure, where the PN, EN, and RN sub-networks function as a cohesive unit working end-to-end in the pipeline, synchronously updating each other. Built upon the concept of mutual adversariality, DN is alternately optimized against them. The loss of the embedding and revealing processes can be denoted as $\mathcal{L}_{PN,EN,RN}$, which comprises the embedding loss $\mathcal{L}_e$, revealing loss $\mathcal{L}_r$, and adversarial loss $\mathcal{L}_{gd}$. In contrast, the loss of the discriminating process can be denoted as $\mathcal{L}_{DN}$, which

consists solely of the adversarial loss. Given that both the generated *stego* and the recovered RSI' are fundamentally RGB images, the generation loss and recovery loss are composed of pixel-wise loss, structural loss, and perceptual loss. Each loss function is a measurement between the two images, focusing on different aspects and complementing one another. Therefore, in our approach, we combine all three to enhance the embedding effect and the quality of recovery.

$$\begin{aligned}\mathcal{L}_e = MSE(Cover, Stego) + SSIM(Cover, Stego) \\ +MSE(VGG19(Cover), VGG19(Stego)).\end{aligned} \tag{12}$$

$$\mathcal{L}_r = MSE(RSI, RSI') + SSIM(RSI, RSI') + MSE(VGG19(RSI), VGG19(RSI')). \tag{13}$$

In Equations (12) and (13), MSE denotes Mean Squared Error, SSIM denotes Structural Similarity Index Measure, and VGG19 represents a deep convolutional neural network architecture. In this context, VGG19 is utilized solely to extract semantic features by employing its pre-trained model without additional training.

The DN determines the probability that the input stego belongs to the cover, and the larger the value is, the closer it is to the distribution of the cover. As a loss function, it is generally optimized by minimizing its value, so a negative sign is added before the probability value of the loss. Therefore, the adversarial loss EN and PN can be expressed in accordance with Equation (14).

$$\mathcal{L}_{gd} = -DN(Stego). \tag{14}$$

In short, training PN, EN, and RN is to optimize the loss function as expressed in Equation (15).

$$\mathcal{L}_{PN,EN,RN} = \lambda_1 \mathcal{L}_e + \lambda_2 \mathcal{L}_r + \lambda_3 \mathcal{L}_{gd}. \tag{15}$$

where $\lambda_1, \lambda_2, \lambda_3$ are weight factors that adjust the proportion of different loss functions in the total loss function. The DN strives to reduce the predicted cover score while increasing it for stegos. We optimize it by the loss function as expressed in (16):

$$\mathcal{L}_{DN} = (0 - DN(Stego))^2 + (1 - DN(Cover))^2. \tag{16}$$

3.8. Training Process

When training is complete, and the model is deployed to a real-world application scenario, the stego will be disseminated through public channels. Theoretically, a completely secure distribution system requires stegos and the cover to follow the same distribution. It is difficult to achieve the same distribution in practical applications, generally using a means to measure the distance between the two distributions; the smaller the distance, the better the hiding effect. Therefore, minimizing the distance between the two distributions becomes our optimization goal for generating the stego. Under the guidance of this idea, the original GAN proposed by Goodfellow et al. [34] can optimize the generated distribution based on the KL divergence. Subsequent researchers focused on enhancing GAN by developing appropriate network architectures and introducing novel loss functions to mitigate its numerous shortcomings. Among them, Arjovsky et al. [35] discovered that the Wasserstein distance offers benefits over both KL and JS distances, leading to the introduction of Wasserstein–GAN (WGAN) to achieve more stable training processes. In the original version of WGAN, the 1-Lipschitz constraint was imposed via weight clipping. However, this approach had several drawbacks, such as potentially leading to gradient vanishing or exploding and limiting the model's capacity. To overcome these limitations, WGAN-GP [36] was proposed. Building on the foundation of WGAN, WGAN-GP introduces a gradient penalty to more effectively enforce the 1-Lipschitz constraint, resulting in several advantages: improved training stability, elimination of problems associated with weight clipping, enhanced sample quality during training, and simplified fine-tuning. As a result, we utilize WGAN-GP to more accurately align the generator's output distribution, which similarly facilitates the achievement of a stable training regimen. For our task, the

Wasserstein-1 distance between the cover distribution P_{cover} and the stego distribution Q_{stego} is denoted by the following Equation (17).

$$W(P_{cover}, Q_{stego}) = \inf_{\gamma \in \Pi(P_{cover}, Q_{stego})} \int_{\mathcal{X} \times \mathcal{Y}} \|x - y\| \, d\gamma(x, y). \tag{17}$$

In Equation (17), $W(P_{cover}, Q_{stego})$ represents the Wasserstein-1 distance between probability distributions of Cover and Stego. $\Pi(P_{cover}, Q_{stego})$ represents the collection of all combined distributions γ that transport P_{cover} to Q_{stego}, where each γ must satisfy the marginal distributions to be consistent with P_{cover} and Q_{stego}. $\|x - y\|$ represents the Euclidean distance between two points $x \in Cover$ and $y \in Stego$ in space, and the integral calculates the expected cost of moving from x to y over all possible transport plans γ. Our training process is described later in Algorithm 1 based on the principles above. The (PN, EN, RN) and DN are trained alternately until the number of iterations reaches the maximum value, where (PN, EN, RN) jointly learn to minimize $\mathcal{L}_{PN,EN,RN}$ while DN aims to minimize$\mathcal{L}_{DN}$. Note that DN is iterated five times once (PN, EN, RN) is iterated. The generator can receive dependable gradients within this framework that consistently enhances the embedding's effectiveness.

4. Experimental Results and Analysis

In this study segment, we performed a series of comprehensive ablation experiments to methodically assess the impact and efficacy of the various design choices incorporated into our model. These experiments are critical for understanding the contribution of individual components and features to the overall performance of our system. By methodically disabling specific elements or altering configurations, we can isolate and identify the value and functionality of each discrete design decision.

4.1. Experimental Environment

We delve into the specifics of the datasets used for testing and validation and the details regarding the experimental framework structured for evaluating our model. We meticulously selected comprehensive datasets, ensuring a broad representation of various image types and complexities to train and test our system robustly. We then elaborate on the configuration of our experimental setup, which includes the hardware specifications, software environments, and the parameters set for conducting the experiments. This will clarify how the trials were performed and under what conditions, setting the stage for replicable and transparent results.

ImageNet [15] is extensively utilized as the primary data source for challenges in image recognition and classification. Meanwhile, the NWPU-RESISC45 [14] database, comprising 31,500 aerial images spanning 45 different scene categories with abundant spatial diversity and variation, serves as the repository for the RSI collection.

During our experimental setup, we utilized a workstation equipped with an NVIDIA GeForce RTX 3080 Ti GPU as the hardware. The software environment was configured with the Python 3.8 programming language and the Pytorch 1.31.0 framework, operating on the Windows 10 platform. For training purposes, we employed a corpus of 30,000 images from the ImageNet dataset as the cover images, alongside 1500 images designated for testing. Similarly, for the confidential images, we trained on 30,000 images and allocated 1500 images for testing, all sourced from the NWPU-RESISC45 dataset.

4.2. Visual Quality Test and Analysis

The primary purpose of this scheme is to hide RSIs through natural images to achieve an efficient and secure distribution. One of the most basic requirements is that the hidden image is not visually detectable by the attacker because the attacker will steal the stego or block the distribution channel if they find the image suspicious. Therefore, the first experiment we performed after the program was trained was to test the visual effects of the stego. In addition, we also tested the visual quality of the RSI extracted from the

stego, aiming to test the hiding performance of the scheme from the perspective of visual quality. Figure 8 shows the five randomly selected image pairs in our experimental results: Cover, RSI, Stego from left to right, the extracted RSI, and the residual between Cover and Stego. Intuitively, the stego generated by the scheme and the extracted RSI had sound visual effects. A subjective evaluation was not very accurate, and we made some objective evaluations, mainly using PSNR [37] and SSIM [38], two general visual quality evaluation criteria. Table 2 shows the results of the objective evaluation. It is generally believed that PSNR reaches 37 db and SSIM reaches 0.85, which means that the evaluation object has better visual quality. That is, no visual anomaly can be detected compared with the reference object.

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{18}$$

In Equation (18), MAX is the maximum possible pixel value of the two images involved in the calculation, and MSE represents the Mean Squared Error.

$$SSIM(x, y) = \frac{(2u_x u_y + C_1)}{(u_x^2 + u_y^2 + C_1)} \cdot \frac{(2\delta_{xy} + C_2)}{(\delta_x^2 + \delta_y^2 + C_2)} \tag{19}$$

In Equation (19), u_x and u_y represent the average values of images, δ_{xy} represents covariance between images, δ_x and δ_y represent the variances of images, and C_1 and C_2 are two constants which are used to prevent unstable results.

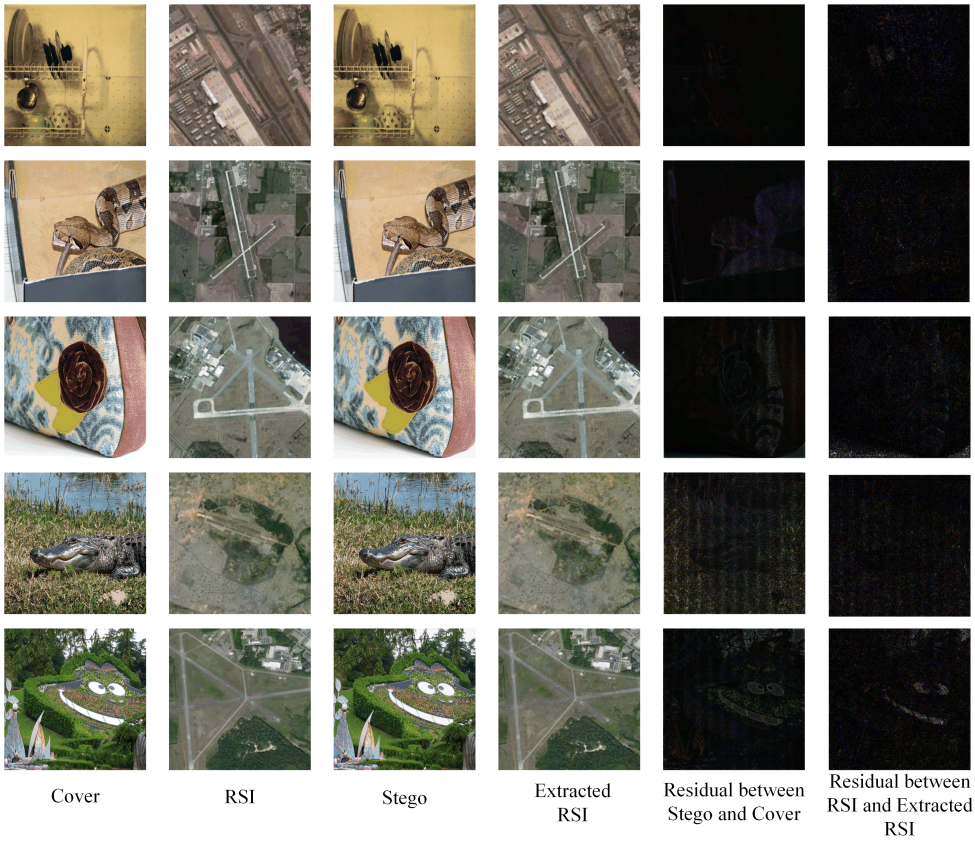

Figure 8. Visual performance display. The first column is the cover, the second column is the RSI, the third column is the stego, the fourth column is the extracted RSI, the fifth column is the residual between the cover and the stego, and the final column is the residual between the RSI and the extracted RSI.

Table 2. Objective evaluation of embedding and revealing effects.

The Test Pairs	Cover and Stego	RSI and Extracted RSI
	PSNR/SSIM *	PSNR/SSIM *
Row#1	46.8 db/0.97	38.7 db/0.86
Row#2	47.1 db/0.98	39.3 db/0.84
Row#3	46.9 db/0.96	39.2 db/0.86
Row#4	46.8 db/0.97	38.6 db/0.88
Row#5	47.2 db/0.98	38.8 db/0.86
Row#6	46.9 db/0.96	39.1 db/0.85

* PSNR and SSIM can be calculated by Equations (18) and (19), respectively.

4.3. Semantic Retention Capability Test

In image classification, three standard evaluation metrics are commonly utilized: overall accuracy, average accuracy, and the confusion matrix. Overall classification accuracy (OCA) is measured by the proportion of correctly classified samples across all classes relative to the total sample count. Average classification accuracy (ACA) calculates the mean classification accuracy for each class, independent of the class sample size. The confusion matrix, an insightful layout, dissects the classification performance, detailing each correct or mistaken prediction by class through an accumulative tabulation of tested samples.

It is important to note that in the case of the NWPU-RESISC45 dataset, each class contains an identical number of images. Consequently, the overall accuracy coincides with the average accuracy. As a result, in our study, we only employed overall accuracy as in Table 3 and the confusion matrix as in Figure 9 to gauge the effectiveness of various classification methodologies.

Table 3. Overall accuracy of three kinds of methods based on CNN and their fine-tuned variants under the training ratios of 70% and 80%.

Method Based on CNN	70% Training Ratio	80% Training Ratio
	Native/Extracted	Native/Extracted
AlexNet	$91.5 \pm 0.18/91.3 \pm 0.17$	$92.7 \pm 0.12/92.6 \pm 0.11$
VGGNet16	$90.5 \pm 0.19/90.6 \pm 0.15$	$92.6 \pm 0.20/92.6 \pm 0.19$
GoolgeLeNet	$91.8 \pm 0.13/92.1 \pm 0.12$	$92.3 \pm 0.13/92.2 \pm 0.15$
Fine-tuned AlexNet	$97.5 \pm 0.18/97.2 \pm 0.16$	$98.7 \pm 0.10/98.5 \pm 0.11$
Fine-tuned VGGNet16	$97.6 \pm 0.18/96.9 \pm 0.19$	$98.9 \pm 0.09/98.3 \pm 0.08$
Fine-tuned GoolgeLeNet	$97.3 \pm 0.18/97.5 \pm 0.17$	$98.7 \pm 0.12/98.4 \pm 0.13$

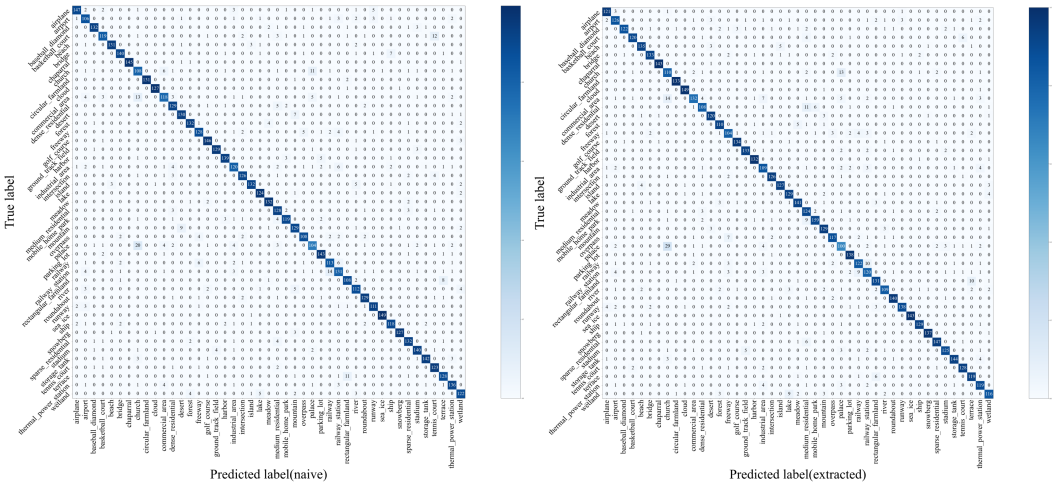

Figure 9. On the left is the confusion matrix results of scene classification for the cover; and on the right are the confusion matrix results of scene classification for the stego.

Furthermore, to ensure the dependability of the overall accuracy and confusion matrix metrics, we conducted ten iterations of the experimental process for each training–testing split. The outcomes were then presented as a mean and standard deviation, providing a robust and trustworthy statistical analysis of the classification results.

4.4. Security Test and Analysis

The security index is an essential measure of the practicality of the scheme proposed in this paper, which mainly refers to the ability of the scheme to resist detection by relevant algorithms, that is, to detect whether the stego contains an RSI. Referring to the general practice of security testing in steganography, we chose two steganographic analysis algorithms, one based on statistical analysis and the other based on deep learning. The former representative tool is stegexpose [39], and the latter representative algorithm is YeNet [40]. In order to obtain the experimental results accurately, 10,000 natural images from ImageNet were selected in the test, and the corresponding stego was obtained by inputting these 10,000 images into the trained embedded model. First, the stegexpose resistance test was carried out, and 2000 images obtained from the result were randomly selected and input into the stegexpose program. The stegexpose detection result was 0.51, similar to the random guessing result. It can be seen that the proposed algorithm does not affect the least significant bit, DCT coefficient, and noise distribution of the carrier image, and it can effectively resist the analysis attack of stegexpose. In other words, the scheme in this paper is safe when only the stegexpose test is used. In order to further verify the security of our scheme, we then carried out the experiment of resistance to YeNet, randomly selecting 8000 pairs of the above 10,000 pairs of images for training and then using the remaining 2000 pairs as a test. From the experimental results, the accuracy of YeNet reached 0.90%. However, this high level does not mean that our algorithm is unsafe because this supervised detection algorithm needs to obtain the stego generated by our algorithm and the corresponding cover for model training, which is almost impossible in many cases.

In our method, the sender and the receiver initially communicate through a secure channel to pass the trained decoder, and subsequent communications involve sending stegos through a public channel. For an attacker, reconstructing a decoder with the same effectiveness as the sender's is highly challenging, as each encoder and decoder network pair is co-trained and designed to work together. Such a design means that even if the attacker can access the transmitted stegos, without the corresponding decoder, it would be difficult to retrieve the hidden information. Nevertheless, the attacker may still have an opportunity. They might attempt to train many variants of a randomly initialized steganographic system to collect extensive statistical information about the hiding process, using this to crack the system. While such an approach might increase the likelihood of a successful attack, it also requires significant computational resources and a deep understanding of the hiding mechanism. As for this issue involving the defense and attack of neural networks, this article does not elaborate further.

4.5. Comparison

In order to verify the necessity of this study, we performed some comparative experiments to verify that the proposed algorithm is more suitable for our RSI security distribution task from different perspectives. Specifically, we selected five deep information hiding algorithms [8,11,23,24,27] that are close to the idea and method of the algorithm in this paper. They only completed the process of hiding pictures in pictures without a special design for RSI. Therefore, to make the comparison fair, we only compared some performance indicators tested jointly in these papers, including the stego's visual quality, the extracted RSI's visual quality, and the detection resistance.

In the experiment, we replicated the schemes mentioned above and tested them with the same cover and RSI. The experimental results are shown in Table 4 and Figure 10. According to the experimental results, the proposed algorithm achieved the best visual

effect and anti-steganalysis ability, which is mainly due to the PN and PAM module designed in this study according to the characteristics of RSIs and the addition of a DN network, which was further verified in a subsequent ablation experiment.

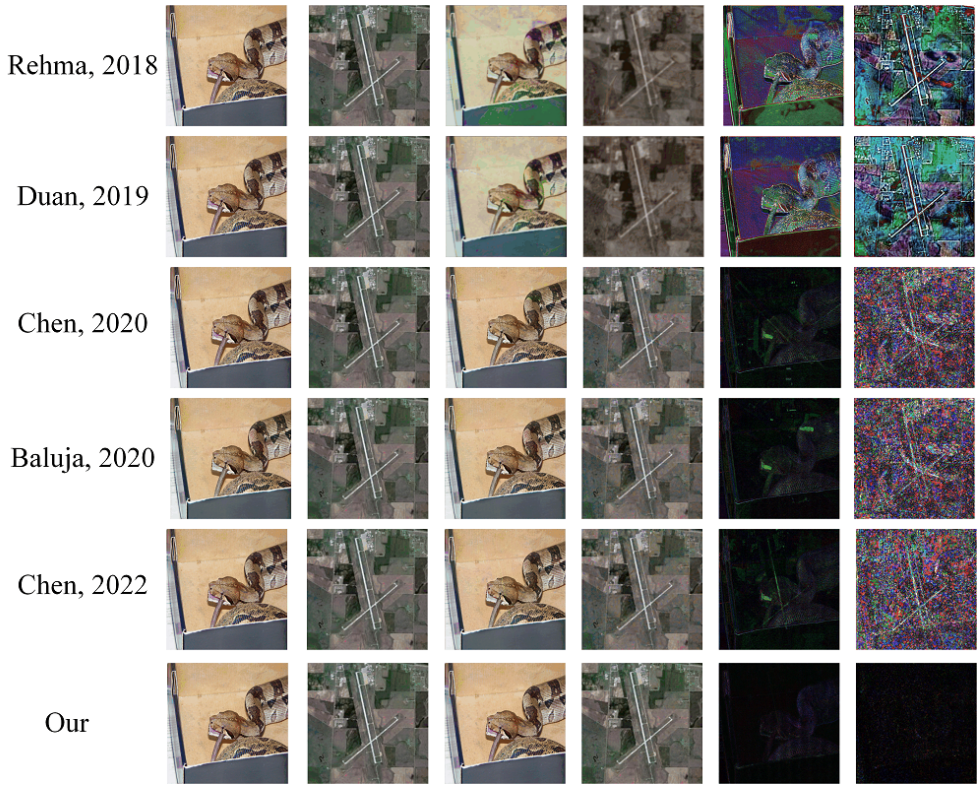

Figure 10. From top to bottom, the results of hiding and extraction on remote sensing images are from reference [8] (Rehman, 2018), reference [24] (Duan, 2019), reference [11] (Chen, 2020), reference [23] (Baluja, 2020), reference [27] (Chen, 2022), and our proposed scheme respectively.

Table 4. Performance comparison with existing algorithms.

Schemes	Cover and Stego	RSI and Extracted RSI	Acurracy of Detection	ACA
	PSNR/SSIM *	PSNR/SSIM *	Stegexpose/YeNet	
Literature [8]	34.78 db/0.92	31.5 db/0.90	0.55/0.99	0.92
Literature [24]	34.6 db/0.96	36.1 db/0.94	0.53/0.98	0.91
Literature [11]	44.1 db/0.97	39.8 db/0.98	0.58/0.98	0.93
Literature [23]	41.3 db/0.95	33.1 db/0.97	0.52/0.98	0.91
Literature [27]	44.6 db/0.97	38.6 db/0.97	0.53/0.98	0.93
Literature [12]	42.3 db/0.99	38.8 db/0.96	0.52/0.96	0.94
The proposed method	47.1 db/0.99	38.9 db/0.99	0.51/0.90	0.98

* PSNR and SSIM can be calculated by Equations (18) and (19), respectively.

4.6. Ablation Experiments

In our study, we incorporated the Position Attention Module (PAM) and the Perceptual Network (PN) into our Inception Network-based architecture to dynamically refine the distribution of RSI data. The PN within our framework serves to enhance the embedding ability of the Encoder Network (EN) by progressively approximating the visual characteristics and distribution of the cover images and to improve the restoration capability of the Revealing Network (RN) by ensuring the recovered images closely match the visual and semantic qualities of the RSI.

To assess the impact of the PAM and PN, we conducted an ablation study with variations to the original DIH4RSID configuration, producing three offshoots: (1) DIH4RSID-PAM-PN lacking both PAM and PN; (2) DIH4RSID-PAM without PAM, but including PN; (3) DIH4RSID-PN lacking PN, but incorporating PAM. We trained the main DIH4RSID network and its variants on the same dataset and evaluated their efficacy based on image quality, extraction accuracy, visual appeal, security, and classification precision. The quantitative outcomes are presented in Table 5.

Table 5. Ablation study results.

Variants	Cover and Stego	RSI and Extracted RSI	AD *	ACA
	PSNR/SSIM	PSNR/SSIM	Stegexpose/YeNet	
DIH4RSID-PAM-PN	36.8 db/0.80	29.6 db/0.80	0.53/0.96	0.89
DIH4RSID-PAM	40.9 db/0.83	30.1 db/0.82	0.55/0.92	0.94
DIH4RSID-PN	42.1 db/0.92	32.3 db/0.88	0.52/0.91	0.93
DIH4RSID	47.1 db/0.99	38.9 db/0.99	0.51/0.90	0.98

* AD denotes the accuracy of detection of Stegexpose or YeNet, and ACA denotes the average accuracy classification.

The comparison between DIH4RSID and DIH4RSID-PAM-PN demonstrates the significant enhancements PAM and PN provide across all benchmarks. These components optimize feature utilization, highlighting relevant details while suppressing extraneous ones and generating steganographic images with less perceptible noise. Additionally, including PN in DIH4RSID aids in a more accurate recovery of RSI.

A marginal decline in performance with DIH4RSID-PN suggests that PAM contributes to stable embedding by promoting the imperceptibility of steganographic content. Furthermore, the effectiveness of RSI extraction and visual quality also relies on PN, as it facilitates the concurrent training of both the EN and RN, underscoring the complementary roles of PAM and PN in our network design.

5. Discussion

In order to provide an alternative scheme for the safe distribution of remote sensing images, this paper proposes a new end-to-end network structure based on the idea of deep information hiding. The remote sensing images we want to hide are no different from ordinary natural images, so the intuitive idea is to transplant the existing image-to-hide image algorithm. Through comparative and ablation experiments, we found that it can be barely used in scenes with low-performance requirements. However, further experimental results showed that the proposed algorithm can achieve better visual quality and higher security. Through analysis, the scheme in this paper achieved better performance than the existing algorithms because of the delicate design of the network structure.

Firstly, in this study, we designed a PN according to the characteristics of RSIs. Theoretically, to capture more details of RSIs and targets of different scales, the features extracted by the PN module are used as input for subsequent embedding in the network. Secondly, a newly designed attention mechanism module, PAM, is adopted in the encoder network, similar to the Coordinate Attention Mechanism, which can realize feature enhancement and context information integration. In addition to adding PAM to the encoder, we also add local skip connections and global skip connections to achieve good information embedding and visual quality maintenance. Finally, to achieve better security, we add the discriminator module to the pipeline to achieve higher security and embedding effects.

Although this study provides a novel approach to the secure distribution of remote sensing images, the RSI images used were compressed RGB images, and the efficiency and complexity of the algorithm were not analyzed. The method presented in this paper cannot be directly applied to hyperspectral RSIs. Then, due to the complexity of the network architecture, it can be anticipated that the efficiency will not be very high. Therefore, to

extend the applicability of this algorithm and improve efficiency, future work could attempt to efficiently implement the embedding and extraction for hyperspectral images to achieve their secure distribution.

Author Contributions: Conceptualization, P.L., J.L. and D.M.; methodology, P.L. and J.L.; validation, P.L., J.L. and D.M.; formal analysis, P.L., J.L., J.X. and Q.D.; investigation, P.L.; writing—original draft preparation, P.L. and J.L.; writing—review and editing, P.L. and J.L.; experiments, P.L. and Q.D.; supervision, P.L., J.L. and D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (grant numbers 62372069, 62272478, and 6217072522).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, D.; Ren, L.; Shafiq, M.; Gu, Z. A Lightweight Privacy-Preserving System for the Security of Remote Sensing Images on IoT. *Remote Sens.* **2022**, *14*, 6371. [CrossRef]
2. Zhang, X.; Zhang, G.; Huang, X.; Poslad, S. Granular Content Distribution for IoT Remote Sensing Data Supporting Privacy Preservation. *Remote Sens.* **2022**, *14*, 5574. [CrossRef]
3. Alsubaei, F.S.; Alneil, A.A.; Mohamed, A.; Mustafa Hilal, A. Block-Scrambling-Based Encryption with Deep-Learning-Driven Remote Sensing Image Classification. *Remote Sens.* **2023**, *15*, 1022. [CrossRef]
4. Naman, S.; Bhattacharyya, S.; Saha, T. Remote sensing and advanced encryption standard using 256-Bit key. In *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 181–190.
5. He, R.; Sun, Q.; Thangasamy, P.; Chen, X.; Zhang, Y.; Wang, H.; Luo, H.; Zhou, X.D.; Zhou, M. Accelerate oxygen evolution reaction by adding chemical mediator and utilizing solar energy. *Int. J. Hydrogen Energy* **2023**, *48*, 8898–8908. [CrossRef]
6. Akhaee, M.A.; Marvasti, F. A Survey on Digital Data Hiding Schemes: Principals, Algorithms, and Applications. *ISeCure* **2013**, *5*, 5.
7. Singh, A.K. Data hiding: Current trends, innovation and potential challenges. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2020**, *16*, 1–16. [CrossRef]
8. Rehman, A.; Rahim, R.; Nadeem, M.; Hussain, S. End-to-end trained CNN encode-decoder networks for image steganography. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018; pp. 723–729.
9. Zhang, K.A.; Cuesta-Infante, A.; Xu, L.; Veeramachaneni, K. SteganoGAN: High Capacity Image Steganography with GANs. *arXiv* **2019**, arXiv:1901.03892.
10. Yu, C. Attention Based Data Hiding with Generative Adversarial Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 1120–1128.
11. Chen, F.; Xing, Q.; Liu, F. Technology of hiding and protecting the secret image based on two-channel deep hiding network. *IEEE Access* **2020**, *8*, 21966–21979. [CrossRef]
12. Chen, F.; Xing, Q.; Fan, C. Multilevel Strong Auxiliary Network for Enhancing Feature Representation to Protect Secret Images. *IEEE Trans. Ind. Inform.* **2022**, *18*, 4577–4586. [CrossRef]
13. Shi, J.; Liu, W.; Shan, H.; Li, E.; Li, X.; Zhang, L. Remote Sensing Scene Classification Based on Multibranch Fusion Attention Network. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
14. Cheng, G.; Han, J.; Lu, X. Remote sensing image scene classification: Benchmark and state of the art. *Proc. IEEE* **2017**, *105*, 1865–1883. [CrossRef]
15. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
16. Mielikainen, J. LSB matching revisited. *IEEE Signal Process. Lett.* **2006**, *13*, 285–287. [CrossRef]
17. Filler, T.; Judas, J.; Fridrich, J. Minimizing Additive Distortion in Steganography Using Syndrome-Trellis Codes. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 920–935. [CrossRef]
18. Pevnỳ, T.; Filler, T.; Bas, P. Using high-dimensional image models to perform highly undetectable steganography. In Proceedings of the Information Hiding: 12th International Conference, IH 2010, Calgary, AB, Canada, 28–30 June 2010; Revised Selected Papers 12; Springer: Berlin/Heidelberg, Germany, 2010; pp. 161–177.
19. Holub, V.; Fridrich, J. Designing steganographic distortion using directional filters. In Proceedings of the 2012 IEEE International Workshop on Information Forensics and Security (WIFS), Tenerife, Spain, 2–5 December 2012; pp. 234–239. [CrossRef]
20. Holub, V.; Fridrich, J. Digital image steganography using universal distortion. In Proceedings of the first ACM Workshop on Information Hiding and Multimedia Security, Montpellier, France, 17–19 June 2013; ACM: New York, NY, USA, 2013; pp. 59–68.

21. Li, B.; Wang, M.; Huang, J.; Li, X. A new cost function for spatial image steganography. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 4206–4210.
22. Zhu, J.; Kaplan, R.; Johnson, J.; Fei-Fei, L. Hidden: Hiding data with deep networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 657–672.
23. Baluja, S. Hiding Images within Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 1685–1697. [CrossRef]
24. Duan, X.; Jia, K.; Li, B.; Guo, D.; Zhang, E.; Qin, C. Reversible image steganography scheme based on a U-Net structure. *IEEE Access* **2019**, *7*, 9314–9323. [CrossRef]
25. Huang, J.; Luo, T.; Li, L.; Yang, G.; Xu, H.; Chang, C.C. ARWGAN: Attention-guided Robust Image Watermarking Model Based on GAN. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 5018417. [CrossRef]
26. Tan, J.; Liao, X.; Liu, J.; Cao, Y.; Jiang, H. Channel attention image steganography with generative adversarial networks. *IEEE Trans. Netw. Sci. Eng.* **2021**, *9*, 888–903. [CrossRef]
27. Chen, F.; Xing, Q.; Sun, B.; Yan, X.; Cheng, J. An Enhanced Steganography Network for Concealing and Protecting Secret Image Data. *Entropy* **2022**, *24*, 1203. [CrossRef]
28. Xu, G.; Wu, H.Z.; Shi, Y.Q. Structural Design of Convolutional Neural Networks for Steganalysis. *IEEE Signal Process. Lett.* **2016**, *23*, 708–712. [CrossRef]
29. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
30. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
31. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
32. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
33. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]
34. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 27.
35. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR.org, Sydney, Australia, 6–11 August 2017; pp. 214–223.
36. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved Training of Wasserstein GANs. 2017. Available online: https://proceedings.neurips.cc/paper_files/paper/2017/hash/892c3b1c6dccd52936e27cbd0ff683d6-Abstract.html (accessed on 5 November 2023).
37. Almohammad, A.; Ghinea, G. Stego image quality and the reliability of PSNR. In Proceedings of the International Conference on Image Processing, Hong Kong, China, 26–29 September 2010.
38. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
39. Boehm, B. Stegexpose-A tool for detecting LSB steganography. *arXiv* **2014**, arXiv:1410.6656.
40. Yedroudj, M.; Comby, F.; Chaumont, M. Yedroudj-net: An efficient CNN for spatial steganalysis. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 2092–2096.